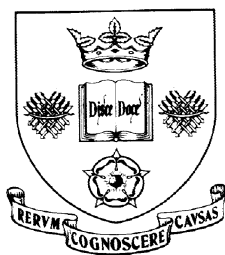


# Sheffield Economic Research Paper Series

**SERP Number: 2012009**

ISSN 1749-8368



Li Su  
Sarah Brown  
Pulak Ghosh  
Karl Taylor

## **Modelling Household Debt and Financial Assets: A Bayesian Approach to a Bivariate Two-Part Model**

**March 2012**

Department of Economics  
University of Sheffield  
9 Mappin Street  
Sheffield  
S1 4DT  
United Kingdom  
[www.shef.ac.uk/economics](http://www.shef.ac.uk/economics)

# Modelling Household Debt and Financial Assets: A Bayesian Approach to a Bivariate Two-Part Model

Li Su, Sarah Brown, Pulak Ghosh, and Karl Taylor

March 21, 2012

---

Li Su is Senior Investigator Statistician, MRC Biostatistics Unit, Cambridge, UK. Sarah Brown and Karl Taylor are Professors, Department of Economics, University of Sheffield, UK. Pulak Ghosh is Professor, Department of Quantitative Methods and Information Systems, Indian Institute of Management at Bangalore, India. The third author acknowledges the grant support from DST, Govt. of India.

## Abstract

In this paper, we contribute to the empirical literature on household finances by introducing a Bayesian bivariate two-part model. With correlated random effects, the proposed approach allows for the potential interdependence between the holding of assets and debt at the household level and also encompasses a two-part process to allow for differences in the influences of the independent variables on the decision to hold debt or assets and the influences of the independent variables on the amount of debt or assets held. Finally, we also incorporate joint modelling of household size into the framework to allow for the fact that the debt and asset information is collected at the household level and hence household size may be strongly correlated with household debt and assets. Our findings endorse our joint modelling approach and, furthermore, confirm that certain explanatory variables exert different influences on the binary and continuous parts of the model.

*Key words: Assets; Bayesian Approach; Bridge distribution; Debt; Two-Part Model*

# 1 Introduction

Over the last two decades there has been growing interest in the economics literature in the nature of financial portfolios at the household level. Such interest has coincided with significant changes in debt accumulation at the household level. Over the last decade, for example, there has initially been a considerable increase in consumer debt in the U.S. followed by a decline in household leverage, the ratio of debt to disposable income, with the onset of the recession towards the end of 2007, see Glick and Lansing (2009). Not surprisingly, increases in the level of household debt around the start of the millennium led to concern amongst policy-makers over the extent of financial vulnerability and risk at the household level.

Despite the importance for policy-making, amongst academic economists, research into the determinants of debt at the household level remains surprisingly scarce. There are, however, a small yet growing number of empirical studies on household debt. For example, Crook (2001) explores the factors that determine household debt in the U.S. over the period 1990 to 1995 and finds that income, home ownership and family size are positively associated with the level of debt. Whilst, for the U.K., Brown et al. (2005, 2008) find that financial expectations are important determinants of unsecured and secured debt at both the individual and the household level. There is a growing empirical literature exploring household financial portfolios and asset holding more generally (see, for example, Guiso et al., 2002, for a comprehensive review of this area). In general, in the existing literature, economists have focused on specific aspects of the financial portfolio including the demand for risky financial assets such as stocks and shares (for example, Bertaut, 1998, Hochguertel et al., 1997 and Shum and Faig, 2006) or savings (for example, Browning and Lusardi, 1996).

Thus, existing studies have generally focused on one aspect of household finances such as the holding of particular types of risky financial assets or household liabilities. Policy-makers have, however, commented on the importance of analysing both household financial assets and liabilities. In particular, Alan Greenspan, the former

Chairman of the U.S. Federal Reserve Board, has argued that unless one simultaneously considers financial assets along with liabilities it is hard to ascertain the true burden of debt.<sup>1</sup> One exception is Cox et al. (2002), who explore financial pressure across households in Great Britain, and find that households with the highest absolute levels of debt also tend to have the highest income and net wealth, implying that these households may be relatively well disposed towards coping with adverse financial shocks. On the other hand, the findings of Brown and Taylor (2008), who jointly model household debt and assets, suggest that the youngest households and those households who are in the lowest income quartile are the most vulnerable to changes in their financial circumstances being characterized by a high proportion of households with debt yet no financial assets, i.e. negative net worth.

Many of the statistical models used in the existing literature treat the level of household debt or assets as censored variables since they cannot have negative values. Consequently, a Tobit approach has been commonly used to allow for this truncation (see, for example, Bertaut and Starr-McCluer, 2002 and Brown et al, 2005, and, 2008). In studies, where a joint modelling approach has been adopted, a bivariate Tobit model has been used allowing for the possibility of inter-dependent decision-making with respect to financial assets and liabilities (see, for example, Brown and Taylor, 2008, where the findings endorse the joint modelling approach indicating interdependence between the holding of assets and debt).<sup>2</sup> One problem with the Tobit approach, however, lies in the possibility that the decision to hold debt or financial assets and the decision regarding the level of debt or financial assets to hold may be characterized by different influences. A double-hurdle model is an alternative econometric specification, which allows independent variables to have different effects on the probability of holding

---

<sup>1</sup>Remarks made by Alan Greenspan “Understanding Household Debt Obligations” at the Credit Union National Association, Governmental Affairs Conference, Washington, D.C. February 23, 2004.

<sup>2</sup>Where studies have explored the holding of particular financial assets, a probit or logit approach has been adopted given the discrete nature of the dependent variable. For example, Bertaut and Starr-McCluer (2002) use a multivariate probit approach to investigate household decisions relating to holding different financial assets.

debt or financial assets ('the binary part') and on the level of debt or financial assets if it is nonzero ('the continuous part'). Such an approach allows for a two-stage decision-making process: for example, a household decides whether to hold a particular asset and, conditional on the decision to hold a particular asset, the household then decides how much of that asset to hold, where there is potential correlation between the two decision-making processes (see, for example, Yen et al., 1997, in the context of analyzing financial donations).

In this paper, we contribute to the existing literature on modelling household finances by introducing a Bayesian bivariate two-part model. By two-part model, we refer to data generated from a response which is a mixture of true zeros and continuously distributed positive values (Olsen et al. 2001; Tooze et al. 2002) . With correlated random effects, the proposed approach allows for the potential interdependence between the holding of assets and debt at the household level and also encompasses a two-part process to allow for differences in the influences of the independent variables on the decision to hold debt or assets and the influences of the independent variables on the amount of debt or financial assets held. In addition to the novelty of introducing a Bayesian approach to exploring the influences on household finances, combining the joint modelling approach with the two-part approach brings together two important aspects of household financial decision-making which have been explored separately in the economics literature to date. Finally, we also incorporate joint modelling of household size into the statistical framework to allow for the fact that the debt and asset information is collected at the household level and, hence, household size may be strongly correlated with household debt and asset accumulation.

In terms of the specific statistical methods proposed in this paper, our bivariate two-part model has the advantage of offering straightforward interpretations of the effects of the independent variables, both conditionally (i.e., given the random effects) and marginally (i.e., after integrating over the random effects), when modelling the binary parts of total assets and debt. In the standard generalized linear mixed model

(GLMM) for binary dependent variables, the marginal probabilities integrated over Normal random effects in general no longer follow a generalized linear model (GLM) if a non-linear link function (e.g., logit link) is adopted (Diggle et al. 2002). In this case, the GLMM with Normal random effects can only provide subject-specific effects of independent variables conditional on random effects, while the population-averaged effects of independent variables on marginal probabilities might be of interest for the study in question. Therefore, in practice it is desirable that both the population-averaged and subject-specific effects of independent variables are readily available when drawing study conclusions. For this purpose, instead of using the usual Normal distribution, we use the bridge distribution introduced in Wang and Louis (2003) for the random intercepts in the binary parts of the dependent variables. The bridge distribution allows both the marginal (integrated over the distribution of the random intercept) model and the conditional model (conditional on the random effects) for the binary parts of the dependent variables to follow a logistic regression model, with regression coefficients proportional to each other.

Specifically, in our two-part model for a single dependent variable (e.g., total assets), we use a random intercept logistic model for modelling the binary part of the dependent variable and a random intercept Gamma GLMM with a log link for modelling the continuous part of the dependent variable. As mentioned earlier, the random intercept in the conditional logistic model for the binary part follows a bridge distribution, while the random intercept in the Gamma GLM for the continuous part follows a Normal distribution (Wang and Louis, 2003; Lin et al. 2010; Su et al. 2011). The marginal effects of the independent variables are proportional to the conditional effects of the independent variables with closed forms in both parts of the model because the marginal expectations in both parts preserve the logit and log links after integration over the random effects.

The same two-part model is specified for the other dependent variable, i.e., total debt. Further, we jointly model household size as another dependent variable using a

Poisson GLMM with a log link and a Normal random intercept. We model household size due to its potential role in influencing debt and asset holding and in particular due to the nature of the survey question which elicits this information at the household level. The interdependence between the two-parts of total assets and debt as well as the interdependence between all three dependent variables are taken into account by allowing the random effects to be correlated. Further, a multivariate density using a Gaussian copula model is assumed for the random effects, which is parameterised by the correlation matrix of the Gaussian copula and marginal variances of the random effects. Due to the multivariate nature of our data, the correlation matrix of the Gaussian copula is left as unstructured, where the new partial autocorrelation approach (Daniels and Pourahmadi, 2009) is adopted to guarantee the positive definiteness of the correlation matrix.

In Section 2, we describe the data, which is drawn from the U.S. Panel Study of Income Dynamics (PSID). Section 3 formally describes our joint model, including the bivariate two-part model for longitudinal semicontinuous data. We also introduce the bridge density for the correlated random effects. We describe the Bayesian methods for inference in Section 4 and analytical results are presented in Section 5. We conclude the article with discussion of our findings in Section 6.

## 2 Motivating the Data

The Panel Study of Income Dynamics (PSID) is an ongoing panel study of households conducted at the Institute for Social Research, University of Michigan since 1968. The sample size has grown from 4,800 families in 1968 to more than 7,000 families by the turn of the century. Further information on the PSID is available at: <http://psidonline.isr.umich.edu>.

In 1984, 1989, 1994, 1999, 2001, 2003, 2005 and 2007, the head of family is asked to provide information about the household's financial assets and debt. For debt, the head of family is asked to specify the amount remaining on the first mortgage, second



mortgage, credit card charges, student loans, medical or legal bills, or other loans. Our measure of total debt (i.e. secured and unsecured) is the summation of all of these different types of debt in each year. In terms of financial assets, in each year the head of family is asked to specify the value of shares of stock in publicly held corporations, mutual funds, investment trusts, money in current (i.e. checking) or savings accounts, money market funds, certificates of deposit, and government savings bonds and treasury bills. In order to obtain a measure of total financial assets at the household level, we aggregate over all of these different types of financial assets. All monetary variables are given in 1984 constant prices. As the distributions of debt and assets are highly skewed, following Gropp et al. (1997), we specify logarithmic dependent variables. For households reporting zero debt or zero assets, the logarithmic variables are recoded to zero, since there are no reported values between zero and unity in the sample.

[Figure 1 about here.]

The sample of households analysed in this study forms a balanced panel where the same 1,957 heads of household are observed in each year yielding total observations over the period of 15,656 heads of household who are aged between 18 and 65. Table 1 provides summary statistics for the dependent variables. Figure 1 presents histograms of the natural logarithms of total debt and total financial assets over the period. Panel A shows the natural logarithm of total debt where just under 23% of households over the period have no secured or unsecured debt. Panel B shows the distribution of total debt conditional on holding debt, where the mean (median) is \$41,760 (\$27,633). Similarly, in Panel C the natural logarithm of financial assets is shown where around 23% of households hold no assets. Panel D reveals that conditional upon holding assets the mean (median) over the period is \$32,709 (\$5000). Finally, 65% of the sample hold both assets and debt.

Figure 2 shows how the distribution of the natural logarithm of total debt has changed over time: for households who hold debt, there has clearly been a shift in the

distribution of debt from 1984 to 2007. Similarly, Figure 2 also shows the distribution of the natural logarithm of financial assets for those heads of household with positive amounts of financial assets, over the time period. Interestingly, in contrast to the distribution of debt, the distribution of financial assets has remained relatively stable over this time period.

[Figure 2 about here.]

The independent variables used in the analysis to explain household debt and financial asset holding follow the existing literature and consist of time invariant head of household characteristics and time varying independent variables. Time invariant variables are binary controls for: gender and ethnicity. Time varying binary controls include age, specifically whether: aged 18-24; aged 25-34; aged 35-44; aged 45-54 and aged 55-60 (where aged over 60 is the reference category). Other time varying controls are included for: marital status; whether the individual is in good or excellent health; employment status; whether the household is in the 0-25th income quartile; whether the household is in the 25-50th income quartile; and whether the household is in the 50-75th income quartile (where above the 75th quartile is the reference category). We also control for the level of highest educational attainment of the head of family, which is defined as: not completed high school but more than eighth grade; completed high school; some college education; and a college degree or above (where below eighth grade is the reference category). Summary statistics of the explanatory variables are given in Table 2 for three different samples: all households; households with positive debt; and households with positive assets. For all households, just under 80% of household heads are male; most heads are in paid employment; 36% of household heads are aged 35 to 44; and the most common level of educational attainment is completion of high school education. The sample of households holding debt and the sample of households holding assets are characterised by a lower proportion of non-white heads of family and a higher proportion of married or co-habiting heads of family. Interestingly, as compared to the sample of all households, these two sub-samples are characterised by a lower

proportion being in the lowest income quartile and a higher proportion being in the highest income quartile.

[Table 1 about here.]

[Table 2 about here.]

### 3 Joint Model

#### 3.1 Modelling total financial assets

Let  $y_{ij}^a$  be the financial assets of the  $i$ th household ( $i = 1, 2, \dots, n$ ) in the  $j$ th year ( $j = 1, 2, \dots, m$ ), where  $n$  is the total number of households and  $m$  is the total number of follow-up years. Let  $R_{ij}^a$  be a random variable denoting the amount of assets held where

$$R_{ij}^a = \begin{cases} 0, & \text{if } y_{ij}^a = 0 \\ 1, & \text{if } y_{ij}^a > 0 \end{cases}$$

with

$$\Pr(R_{ij}^a = r_{ij}^a) = \begin{cases} 1 - p_{ij}^a, & \text{if } r_{ij}^a = 0 \\ p_{ij}^a, & \text{if } r_{ij}^a = 1 \end{cases}.$$

Further, let  $s_{ij}^a \equiv [y_{ij}^a | R_{ij}^a = 1]$  denote the positive assets of the  $i$ th household in the  $j$ th year.

We model the probability  $p_{ij}^a$  ('the binary part') using a random intercept logistic model and the non-zero continuous observations  $s_{ij}^a$  ('the continuous part') using a Gamma GLMM with a log link as follows:

$$\text{logit}(p_{ij}^a) = \mathbf{X}_{ij}^1 \boldsymbol{\beta}^1 + B_i^a, \tag{3.1}$$

$$\log(s_{ij}^a) \sim \text{Gamma}(\nu^a, \mu_{ij}^a),$$

$$\mu_{ij}^a = \nu^a / \text{E}\{\log(s_{ij}^a)\},$$

$$\log[\text{E}\{\log(s_{ij}^a)\}] = \mathbf{X}_{ij}^2 \boldsymbol{\beta}^2 + V_i^a,$$

where  $\mathbf{X}_{ij}^1$  and  $\mathbf{X}_{ij}^2$  are the independent variable vectors with associated regression coefficients  $\boldsymbol{\beta}^1$  and  $\boldsymbol{\beta}^2$  for the binary and continuous parts, respectively;  $\nu^a$  is the shape parameter of the Gamma distribution; and  $B_i^a$  and  $V_i^a$  are the random intercepts

of the two parts accounting for the dependence of the repeated observations within households.

### 3.2 Modelling total debt

Let  $y_{ij}^d$  be the total debt of the  $i$ th household ( $i = 1, 2, \dots, n$ ) in the  $j$ th year ( $j = 1, 2, \dots, m$ ), where  $n$  is the total number of households and  $m$  is the total number of follow-up years. Let  $R_{ij}^d$  be a random variable denoting the amount of debt where

$$R_{ij}^d = \begin{cases} 0, & \text{if } y_{ij}^d = 0 \\ 1, & \text{if } y_{ij}^d > 0 \end{cases}$$

with

$$\Pr(R_{ij}^d = r_{ij}^d) = \begin{cases} 1 - p_{ij}^d, & \text{if } r_{ij}^d = 0 \\ p_{ij}^d, & \text{if } r_{ij}^d = 1 \end{cases}.$$

Also, let  $s_{ij}^d \equiv [y_{ij}^d | R_{ij}^d = 1]$  denote the positive debt of the  $i$ th household in the  $j$ th year.

We model the probability  $p_{ij}^d$  using a random intercept logistic model and the non-zero continuous observations  $s_{ij}^d$  using a Gamma GLMM as follows:

$$\begin{aligned} \text{logit}(p_{ij}^d) &= \mathbf{X}_{ij}^3 \boldsymbol{\beta}^3 + B_i^d, \\ \log(s_{ij}^d) &\sim \text{Gamma}(\nu^d, \mu_{ij}^d), \\ \mu_{ij}^d &= \nu^d / \text{E}\{\log(s_{ij}^d)\}, \\ \log[\text{E}\{\log(s_{ij}^d)\}] &= \mathbf{X}_{ij}^4 \boldsymbol{\beta}^4 + V_i^d, \end{aligned} \tag{3.2}$$

where  $\mathbf{X}_{ij}^3$  and  $\mathbf{X}_{ij}^4$  are the independent variable vectors with associated regression coefficients  $\boldsymbol{\beta}^3$  and  $\boldsymbol{\beta}^4$  for the binary and continuous parts, respectively;  $\nu^d$  is the shape parameter of the Gamma distribution; and  $B_i^d$  and  $V_i^d$  are random intercepts for the two parts.

### 3.3 Modelling household size

Instead of including household size in the set of explanatory variables, we jointly model household size since the number of household members may be strongly correlated with

total debt and financial asset holding measured at the household level. We adopt this approach since the PSID, as is often the case with such household surveys, records debt and assets at the household level. Thus, the household size variable in this context is highly likely to be endogenous if included in the set of explanatory variables.

Let  $y_{ij}^s$  be the size of the  $i$ th household in the  $j$ th year, then  $y_{ij}^s$  is assumed to follow a random intercept Poisson GLMM.

$$\begin{aligned} y_{ij}^s &\sim \text{Poisson}(\lambda_{ij}), \\ \log(\lambda_{ij}) &= \mathbf{X}_{ij}^5 \boldsymbol{\beta}^5 + V_i^s, \end{aligned} \tag{3.3}$$

where  $\mathbf{X}_{ij}^5$  is the independent variable vector with associated regression coefficients  $\boldsymbol{\beta}^5$  and  $V_i^s$  is the random intercept.

### 3.4 Random effects model

For each household, we have a 5-dimensional random effects vector  $\mathbf{b}_i = (B_i^a, V_i^a, B_i^d, V_i^d, V_i^s)^\top$ . Since the binary and continuous parts of total asset and total debt holding as well as household size are highly likely to be related within the households over the follow-up years, it is necessary to allow the elements of  $\mathbf{b}_i$  to be correlated. A typical option would be to assume a multivariate Normal distribution for  $\mathbf{b}_i$ . However, the logistic models in (3.1) and (3.2) with Normal random effects can only provide the household-specific independent variable effects conditional on the random effects. In order to provide marginal effects of the independent variables in the logistic models for the binary parts of total asset and debt holding, we extend the random intercept GLMM approach in Wang and Louis (2003) to the bivariate two-part model setting.

We assume that  $B_i^a$  and  $B_i^d$ , the random intercepts in the binary parts from (3.1) and (3.2), marginally follow the bridge distributions of Wang and Louis (2003) with densities

$$\begin{aligned} f_1(b_i^a | \phi_1) &= \frac{1}{2\pi} \frac{\sin(\phi_1 \pi)}{\cosh(\phi_1 b_i^a) + \cos(\phi_1 \pi)} \quad (-\infty < b_i^a < \infty), \\ f_3(b_i^d | \phi_3) &= \frac{1}{2\pi} \frac{\sin(\phi_3 \pi)}{\cosh(\phi_3 b_i^d) + \cos(\phi_3 \pi)} \quad (-\infty < b_i^d < \infty) \end{aligned}$$

with unknown parameters  $\phi_1$  and  $\phi_3$  ( $0 < \phi_1 < 1$ ,  $0 < \phi_3 < 1$ ). The bridge distribution is symmetric with mean zero and variance  $\sigma_k^2 = \pi^2(\phi_k^{-2} - 1)/3$  ( $k = 1, 3$ ). It is slightly heavy-tailed and more concentrated than the Normal distribution with the same variance. The key characteristic of this bridge density is that, after integration over the random effects,  $\mathbf{b}_i = (B_i^a, V_i^a, B_i^d, V_i^d, V_i^s)^\top$ , the marginal probabilities  $\Pr(R_{ij}^a = 1)$  and  $\Pr(R_{ij}^d = 1)$  relate to the independent variables through the same logit link functions as in the case of the corresponding conditional probabilities. In addition, if we specify the marginal regression structure of the binary parts as

$$\begin{aligned}\text{logit}\{\Pr(R_{ij}^a = 1)\} &= \mathbf{X}_{ij}\boldsymbol{\theta}^1, \\ \text{logit}\{\Pr(R_{ij}^d = 1)\} &= \mathbf{X}_{ij}\boldsymbol{\theta}^3,\end{aligned}$$

then the marginal independent variable effects  $\boldsymbol{\theta}^k$  ( $k = 1, 3$ ) are proportional to the household-specific conditional independent variable effects  $\boldsymbol{\beta}^k$ , with  $\boldsymbol{\theta}^k = \phi_k\boldsymbol{\beta}^k$ . Therefore, we can rewrite (3.1) and (3.2) as

$$\text{logit}\{\Pr(R_{ij}^a = 1 | B_i^a)\} = \mathbf{X}_{ij}^1\boldsymbol{\theta}^1/\phi_1 + B_i^a, \quad (3.4)$$

and

$$\text{logit}\{\Pr(R_{ij}^d = 1 | B_i^d)\} = \mathbf{X}_{ij}^3\boldsymbol{\theta}^3/\phi_3 + B_i^d. \quad (3.5)$$

Further,  $V_i^a, V_i^d, V_i^s$  are assumed to be marginally Normally distributed with mean zero and variance  $\sigma_a^2, \sigma_d^2, \sigma_s^2$ , respectively. Therefore,  $\log(s_{ij}^a), \log(s_{ij}^d), y_{ij}^s$ , given the random effects,  $\mathbf{b}_i = (B_i^a, V_i^a, B_i^d, V_i^d, V_i^s)^\top$ , follow GLMM with means  $\exp(\mathbf{X}_{ij}^2\boldsymbol{\beta}^2 + V_i^a)$ ,  $\exp(\mathbf{X}_{ij}^4\boldsymbol{\beta}^4 + V_i^d)$  and  $\exp(\mathbf{X}_{ij}^5\boldsymbol{\beta}^5 + V_i^s)$ , respectively. It follows that the marginal means of  $\log(s_{ij}^a), \log(s_{ij}^d), y_{ij}^s$  integrated over  $\mathbf{b}_i$  are  $\exp(\mathbf{X}_{ij}^2\boldsymbol{\beta}^2 + \sigma_a^2/2)$ ,  $\exp(\mathbf{X}_{ij}^4\boldsymbol{\beta}^4 + \sigma_d^2/2)$  and  $\exp(\mathbf{X}_{ij}^5\boldsymbol{\beta}^5 + \sigma_s^2/2)$ , respectively (Diggle et al. 2002). Therefore, the marginal and conditional independent variable effects in the models for  $\log(s_{ij}^a), \log(s_{ij}^d), y_{ij}^s$  coincide except that the intercepts are shifted by constants.

For the purpose of characterizing the interdependence of the dependent variables and the possible dependence between the two parts, as well as assuring the desired

marginal density of each member of  $\mathbf{b}_i$ , we construct a multivariate joint distribution for the random effects using a Gaussian copula (Nelsen, 1999). A copula is a convenient way of formulating a multivariate distribution, and is specified as a function of the marginal cumulative distribution function (CDF). If  $F_1(b_i^a), F_2(v_i^a), F_3(b_i^d), F_4(v_i^d), F_5(v_i^s)$  are the CDFs of  $B_i^a, V_i^a, B_i^d, V_i^d, V_i^s$ , respectively, then there exists a function  $C$  such that the joint CDF of  $\mathbf{b}_i$  is  $F(b_i^a, v_i^a, b_i^d, v_i^d, v_i^s) = C \{F_1(b_i^a), F_2(v_i^a), F_3(b_i^d), F_4(v_i^d), F_5(v_i^s)\}$  (Nelsen (1999), Joe (1997)).

To construct the Gaussian copula for  $\mathbf{b}_i$ , we specify a vector  $\mathbf{U}_i = (U_{i1}, U_{i2}, U_{i3}, U_{i4}, U_{i5})^T$  such that

$$\begin{bmatrix} U_{i1} \\ U_{i2} \\ U_{i3} \\ U_{i4} \\ U_{i5} \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{14} & \rho_{15} \\ \rho_{21} & 1 & \rho_{23} & \rho_{24} & \rho_{25} \\ \rho_{31} & \rho_{32} & 1 & \rho_{34} & \rho_{35} \\ \rho_{41} & \rho_{42} & \rho_{43} & 1 & \rho_{45} \\ \rho_{51} & \rho_{52} & \rho_{53} & \rho_{54} & 1 \end{bmatrix} \right), \quad (3.6)$$

Note that the diagonal elements of the covariance matrix  $\boldsymbol{\Sigma}$  equal 1 so that it is also the correlation matrix. We let  $\rho_{j,j+t} = \text{Corr}(U_{ij}, U_{i,j+t})$  ( $j = 1, 2, 3, 4; 1 \leq t \leq 4$ ) denote the correlation between  $U_{ij}$  and  $U_{i,j+t}$ . Using the probability integral transforms (Hoel, Port and Stone, 1971),

$$B_i^a = F_1^{-1}\{\Phi(U_{i1})\}, \quad B_i^d = F_3^{-1}\{\Phi(U_{i3})\}$$

have marginal CDFs  $F_1(b_i^a), F_3(b_i^d)$ , respectively (Wang and Louis, 2003, Lin et al. 2010). Here  $\Phi(\cdot)$  is the standard Normal CDF, and  $F_k^{-1}(\cdot)$  ( $k = 1, 3$ ) is the inverse cumulative distribution function,

$$F_k^{-1}(x) = \frac{1}{\phi_k} \log \left[ \frac{\sin(\phi_k \pi x)}{\sin\{\phi_k \pi(1-x)\}} \right]$$

of the bridge density for  $0 < x < 1$ . For  $V_i^a, V_i^d, V_i^s$ , we have  $V_i^a = \sigma_a U_{i2}$ ,  $V_i^d = \sigma_d U_{i4}$ , and  $V_i^s = \sigma_s U_{i5}$ .

To fully parameterize the Gaussian copula, we need to specify  $\boldsymbol{\Sigma}$ . Due to the multivariate nature of our data, we choose to leave  $\boldsymbol{\Sigma}$  as unstructured and all off-diagonal elements will be separately estimated. Difficulties in modelling correlation matrices lie in the requirement of positive definiteness and constancy along the diagonals of

the matrices. Recently, Daniels and Pourahmadi (2009) proposed an unconstrained and statistically interpretable reparameterization of  $\Sigma$  using the notion of partial autocorrelation from time series analysis. The advantage of this reparameterisation is computational simplification given that the partial autocorrelations are free to vary independently in  $[-1, 1]$  and positive definiteness is guaranteed. Although the natural ordering of the random variables is usually required in this approach, this is not an issue as in our model and we will leave the  $\Sigma$  completely unstructured and the inferences will be based on the correlation matrix parameters  $\rho_{j,j+k}$  as functions of partial autocorrelations.

Denote  $\pi_{j,j+t} = \text{Corr}(U_{ij}, U_{i,j+t} | U_{il}, j < l < j+t)$  as the partial autocorrelations.

We can establish a one-to-one correspondence between  $\rho_{j,j+t}$  and  $\pi_{j,j+t}$ :

$$\begin{aligned}
\rho_{j,j+1} &= \pi_{j,j+1}, j = 1, 2, 3, 4; \\
\rho_{j,j+2} &= \rho_{j,j+1}\rho_{j+1,j+2} + \pi_{j,j+2}(1 - \rho_{j,j+1}^2)^{1/2}(1 - \rho_{j+1,j+2}^2)^{1/2}, j = 1, 2, 3; \\
\rho_{j,j+3} &= (\rho_{j,j+1} \quad \rho_{j,j+2}) \begin{pmatrix} 1 & \rho_{j+1,j+2} \\ \rho_{j+2,j+1} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_{j+3,j+1} \\ \rho_{j+3,j+2} \end{pmatrix} \\
&+ \pi_{j,j+3} \left\{ (\rho_{j,j+1} \quad \rho_{j,j+2}) \begin{pmatrix} 1 & \rho_{j+1,j+2} \\ \rho_{j+2,j+1} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_{j,j+1} \\ \rho_{j,j+2} \end{pmatrix} \right\}^{1/2} \\
&\left\{ (\rho_{j+3,j+1} \quad \rho_{j+3,j+2}) \begin{pmatrix} 1 & \rho_{j+1,j+2} \\ \rho_{j+2,j+1} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_{j+3,j+1} \\ \rho_{j+3,j+2} \end{pmatrix} \right\}^{1/2}, j = 1, 2; \\
\rho_{j,j+4} &= (\rho_{j,j+1} \quad \rho_{j,j+2} \quad \rho_{j,j+3}) \begin{pmatrix} 1 & \rho_{j+1,j+2} & \rho_{j+1,j+3} \\ \rho_{j+2,j+1} & 1 & \rho_{j+2,j+3} \\ \rho_{j+3,j+1} & \rho_{j+3,j+2} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_{j+4,j+1} \\ \rho_{j+4,j+2} \\ \rho_{j+4,j+3} \end{pmatrix} \\
&+ \pi_{j,j+3} \left\{ \begin{pmatrix} \rho_{j,j+1} \\ \rho_{j,j+2} \\ \rho_{j,j+3} \end{pmatrix}^T \begin{pmatrix} 1 & \rho_{j+1,j+2} & \rho_{j+1,j+3} \\ \rho_{j+2,j+1} & 1 & \rho_{j+2,j+3} \\ \rho_{j+3,j+1} & \rho_{j+3,j+2} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_{j,j+1} \\ \rho_{j,j+2} \\ \rho_{j,j+3} \end{pmatrix} \right\}^{1/2} \\
&\left\{ \begin{pmatrix} \rho_{j+4,j+1} \\ \rho_{j+4,j+2} \\ \rho_{j+4,j+3} \end{pmatrix}^T \begin{pmatrix} 1 & \rho_{j+1,j+2} & \rho_{j+1,j+3} \\ \rho_{j+2,j+1} & 1 & \rho_{j+2,j+3} \\ \rho_{j+3,j+1} & \rho_{j+3,j+2} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_{j+4,j+1} \\ \rho_{j+4,j+2} \\ \rho_{j+4,j+3} \end{pmatrix} \right\}^{1/2}, j = 1.
\end{aligned}$$



## 4 Bayesian Inference

### 4.1 Likelihood specification

Let  $\mathbf{Y}_i^a = (y_{i1}^a, \dots, y_{im}^a)^\top$ ,  $\mathbf{Y}_i^d = (y_{i1}^d, \dots, y_{im}^d)^\top$  and  $\mathbf{Y}_i^s = (y_{i1}^s, \dots, y_{im}^s)^\top$ . Similarly, we define  $\mathbf{X}_i^k = (\mathbf{X}_{i1}^k, \dots, \mathbf{X}_{im}^k)^\top$  for  $k = 1, 2, 3, 4, 5$ . Let  $\boldsymbol{\Omega}_1 = (\boldsymbol{\beta}^1, \boldsymbol{\beta}^2, \nu^a)$ ,  $\boldsymbol{\Omega}_2 = (\boldsymbol{\beta}^3, \boldsymbol{\beta}^4, \nu^d)$ ,  $\boldsymbol{\Omega}_3 = \boldsymbol{\beta}^5$  be the parameter vectors for the multivariate dependent variables, and  $\boldsymbol{\Omega}_4 = (\phi_1, \sigma_a, \phi_3, \sigma_d, \sigma_s, \pi_{12}, \pi_{13}, \pi_{14}, \pi_{15}, \pi_{23}, \pi_{24}, \pi_{25}, \pi_{34}, \pi_{35}, \pi_{45})$  be the parameter vector for the random effects  $\mathbf{b}_i = (B_i^a, V_i^a, B_i^d, V_i^d, V_i^s)^\top$ .

The likelihood function from the  $i$ th household can be partitioned as

$$\begin{aligned} & L(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{\Omega}_3, \boldsymbol{\Omega}_4 \mid \mathbf{Y}_i^a, \mathbf{Y}_i^d, \mathbf{Y}_i^s, \mathbf{X}_i^1, \mathbf{X}_i^2, \mathbf{X}_i^3, \mathbf{X}_i^4, \mathbf{X}_i^5, \mathbf{b}_i) \\ & \propto L(\boldsymbol{\Omega}_1 \mid \mathbf{Y}_i^a, \mathbf{X}_i^1, \mathbf{X}_i^2, B_i^a, V_i^a) L(\boldsymbol{\Omega}_2 \mid \mathbf{Y}_i^d, \mathbf{X}_i^3, \mathbf{X}_i^4, B_i^d, V_i^d) \\ & \quad L(\boldsymbol{\Omega}_3 \mid \mathbf{Y}_i^s, \mathbf{X}_i^5, V_i^s) L(\boldsymbol{\Omega}_4 \mid B_i^a, V_i^a, B_i^d, V_i^d, V_i^s), \end{aligned} \quad (4.1)$$

where

$$\begin{aligned} & L(\boldsymbol{\Omega}_1 \mid \mathbf{Y}_i^a, \mathbf{X}_i^1, \mathbf{X}_i^2, B_i^a, V_i^a) \\ & = \prod_{j=1}^m \{1 - \Pr(R_{ij}^a = 1 \mid B_i^a)\}^{(1-r_{ij}^a)} \{\Pr(R_{ij}^a = 1 \mid B_i^a)\}^{r_{ij}^a} \\ & \quad \times [(\mu_{ij}^a)^{\nu^a} \{\log(s_{ij}^a)\}^{\nu^a-1} \exp\{-\mu_{ij}^a \log(s_{ij}^a)\} / \Gamma(\nu^a)]^{r_{ij}^a} \end{aligned}$$

with  $\mu_{ij}^a$  given in (3.1),

$$\begin{aligned} & L(\boldsymbol{\Omega}_2 \mid \mathbf{Y}_i^d, \mathbf{X}_i^3, \mathbf{X}_i^4, B_i^d, V_i^d) \\ & = \prod_{j=1}^m \{1 - \Pr(R_{ij}^d = 1 \mid B_i^d)\}^{(1-r_{ij}^d)} \{\Pr(R_{ij}^d = 1 \mid B_i^d)\}^{r_{ij}^d} \\ & \quad \times [(\mu_{ij}^d)^{\nu^d} \{\log(s_{ij}^d)\}^{\nu^d-1} \exp\{-\mu_{ij}^d \log(s_{ij}^d)\} / \Gamma(\nu^d)]^{r_{ij}^d} \end{aligned}$$

with  $\mu_{ij}^d$  given in (3.2),

$$L(\boldsymbol{\Omega}_3 \mid \mathbf{Y}_i^s, \mathbf{X}_i^5, V_i^s) = \prod_{j=1}^m (\lambda_{ij}^s)^{y_{ij}^s} \exp(-\lambda_{ij}^s) / y_{ij}^s!$$

with  $\lambda_{ij}^s$  defined in (3.3), and

$$\begin{aligned} & L(\boldsymbol{\Omega}_4 \mid B_i^a, V_i^a, B_i^d, V_i^d, V_i^s) \\ & = c \left\{ F_1(b_i^a), F_2(v_i^a), F_3(b_i^d), F_4(v_i^d), F_5(v_i^s) \right\} f_1(b_i^a \mid \phi_1) \phi \left( \frac{v_i^a}{\sigma_a} \right) f_3(b_i^d \mid \phi_3) \phi \left( \frac{v_i^d}{\sigma_d} \right) \phi \left( \frac{v_i^s}{\sigma_s} \right) \end{aligned}$$

with  $f_1(\cdot)$  and  $f_3(\cdot)$  being the bridge density functions,  $\phi(\cdot)$  being the standard normal density function and  $c(\cdot)$  being the density of the copula  $C(\cdot)$  in Section 3.4 given by

$$c(\mathbf{q} \mid \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ \frac{1}{2} \mathbf{u}^T (\mathbf{I} - \boldsymbol{\Sigma}^{-1}) \mathbf{u} \right\}.$$

Here  $\mathbf{q} = (q_1, \dots, q_5)$  ( $0 < q_k < 1$ ,  $k = 1, \dots, 5$ ),  $\mathbf{u} = (u_1, \dots, u_5)^T$  is a vector of normal scores  $u_k = \Phi^{-1}(q_k)$ , and  $\mathbf{I}$  is the 5-dimensional identity matrix.

## 4.2 Prior specification and posterior inference

To complete the Bayesian specification of the model, priors need to be assigned for all unknown parameters. We assume that the elements of  $\boldsymbol{\Omega}_1$ ,  $\boldsymbol{\Omega}_2$ ,  $\boldsymbol{\Omega}_3$ ,  $\boldsymbol{\Omega}_4$  are independently distributed. Because the numbers of independent variables are large in the joint model, quite a few of them are expected to have weak effects on the dependent variables. In order to incorporate this prior knowledge into our analysis, we set up a prior distribution such that each regression coefficient has a high probability of being near zero but a large effect is still possible.

A commonly used prior in this scenario is the Laplace prior or double exponential prior to obtain shrinkage estimates. The Laplace prior for a  $p \times 1$  regression coefficient vector  $\boldsymbol{\beta}$  is given by

$$f(\boldsymbol{\beta}; \psi) = \prod_{j=1}^p \frac{\psi}{2} \exp(-\psi |\beta_j|)$$

where  $\psi$  is the hyperparameter. In regression, the use of the Laplace prior is known as the LASSO (Efron et al., 2004). Thus, the posterior mode estimate of the coefficients  $\boldsymbol{\beta}$  is the LASSO estimate. A probabilistic re-interpretation was given by Figueiredo (2003) by rewriting the Laplace density in a hierarchical manner:

$$\beta_j \mid \tau_j \sim N(0, 1/\tau_j); \quad \tau_j \mid \psi \sim \text{Gamma}(1, \psi/2).$$

In the analysis reported in Section 5, all elements of regression coefficient vectors,  $\boldsymbol{\beta}^1$ ,  $\boldsymbol{\beta}^2$ ,  $\boldsymbol{\beta}^3$ ,  $\boldsymbol{\beta}^4$ ,  $\boldsymbol{\beta}^5$ , are assigned a Laplace prior with hyperparameter  $\psi = 1$ . For shape parameters in the Gamma distributions, we use exponential priors

$\nu^a \sim \exp(1)$ ,  $\nu^d \sim \exp(1)$ . For the parameters in the random effects model, denote  $\sigma_1^2 = \pi^2(\phi_1^{-2} - 1)/3$  and  $\sigma_3^2 = \pi^2(\phi_3^{-2} - 1)/3$  and we use the following priors  $\sigma_1, \sigma_a, \sigma_3, \sigma_d, \sigma_s \stackrel{i.i.d}{\sim} Uniform(0, 10)$ . Finally, independent uniform priors on  $[-1, 1]$  (or  $Beta(2, 1)$  priors transformed to  $[-1, 1]$ ) are chosen for the partial autocorrelations  $\pi_{12}, \pi_{13}, \pi_{14}, \pi_{15}, \pi_{23}, \pi_{24}, \pi_{25}, \pi_{34}, \pi_{35}, \pi_{45}$ .

The joint posterior distributions of the model parameters conditional on the observed data are obtained by combining the likelihood in (4.1) and the previously specified priors. For computation, we use Markov Chain Monte Carlo (MCMC) methods to sample from the posterior distributions. The model fitting procedure is implemented in the WinBUGS package (version 1.4.1); summary statistics such as posterior means and 95% credible intervals are provided for inference.

## 5 Data Analysis

### 5.1 Model Specification

Using the proposed model, we analyse data collected from the U.S. Panel Study of Income Dynamics (PSID). The data set contains information on total financial assets ( $y_{ij}^a$ ), total debt ( $y_{ij}^d$ ) and household size ( $y_{ij}^s$ ) for 1957 households measured over 8 years,  $i = 1, 2, \dots, 1957$ ;  $j = 1, 2, \dots, 8$ . The baseline independent variables and time varying independent variables included in  $\mathbf{X}_{ij}^1, \mathbf{X}_{ij}^2, \mathbf{X}_{ij}^3, \mathbf{X}_{ij}^4, \mathbf{X}_{ij}^5$  are  $Male_i, nonwhite_i, year_{ij}, marr_{ij}, nkids_{ij}, ghealth_{ij}, emp_{ij}, inc-025_{ij}, inc-2550_{ij}, inc-5075_{ij}, age-1824_{ij}, age-2534_{ij}, age-3544_{ij}, age-4554_{ij}, age-5560_{ij}, age-60_{ij}, ed1_{ij}, ed2_{ij}, ed3_{ij}, ed4_{ij}$ . These variables are described in Section 2 and also in Table 2.

We use the prior specification described in Section 4 and 2 MCMC chains, with diverse initial values, are run for obtaining the posterior samples of the parameters. Convergence is assessed by history plots and Gelman-Rubin statistics provided by the WinBUGS package. Pooled samples of 10000 are used for the inference.

## 5.2 Analytical Results

Tables 3 to 6 present the findings from applying the model detailed above to the PSID data. The importance of modelling debt at the household level as a two part process is apparent when comparing the influence of the explanatory variables across the binary part of the model and the continuous part of the model, where it can be seen that some explanatory variables exert different influences across these two parts (see Table 4). For example, it is apparent from the logit results that having a male head of household is inversely associated with the probability of holding debt yet does not exert a statistically significant influence on, conditional on holding debt, the amount of debt held. In contrast, having a non-white head of household has a very strong inverse association with the probability of holding debt as well as, conditional on holding debt, an inverse association with the amount of debt held. Although this inverse association has been found in the existing literature, see, for example, Brown and Taylor (2008), the Tobit specification typically employed does not distinguish between the effects on debt holding and the effects on the amount of debt held. Our findings indicate that being non-white is inversely associated with both of these aspects of debt holding. As expected, having a married head of household has a very strong positive influence on holding debt and a positive - albeit smaller in magnitude - influence on the amount of debt held. Specifically, a married head of household has an 81 percentage point higher probability of holding debt than an unmarried head,<sup>3</sup> and holds around 8.5 per cent more debt.<sup>4</sup> Such findings may reflect the joint holding of debt within couples, such as a jointly held mortgage for the family home. In contrast, the number of children in the household does not exert a statistically significant influence on either the binary part or the continuous part of the debt model.

---

<sup>3</sup>For a probabilistic outcome, it is possible to find the percentage impact of a binary independent variable by multiplying the coefficient through by 100 percentage points. For example, to interpret the influence of marriage on the probability of being in debt:  $\beta_{\text{married}}^3 \times 100\% = 0.8085 \times 100\% = 81$ .

<sup>4</sup>In the tobit equations, because the dependent variable is specified on a logarithmic scale, the influence of a binary independent variable can be found by multiplying the coefficient through by 100 percentage points. For example, to interpret the effect of marriage on the amount of debt held:  $\beta_{\text{married}}^4 \times 100\% = 0.0852 \times 100\% = 8.5$ .

Interestingly, the results suggest that the age of the head of household does not have a statistically significant influence on the probability of holding debt yet statistically significant positive effects on the amount of debt are apparent for the aged 25 to 34, 35 to 44 and 45 to 54 groups relative to individuals aged over 60. The age effects peak for those aged 35 to 44 who have approximately 4.7 per cent more debt than those aged over 60. Such age effects may reflect consumption smoothing over the life cycle, with individuals aged between 25 to 54 being engaged in activities such as marriage, bringing-up children or house buying at various stages of the life cycle, when consumption may exceed income for a variety of such reasons. As individuals become older, debt levels typically fall as loans are repaid and/or as income increases, which is in accordance with the signs of the estimated coefficients. In addition, it is interesting to note that the sizes of the estimated coefficients increase from the aged 18 to 24 group to the aged 34 to 44 group, then start to decrease, with the smallest effect being apparent for the aged 55 to 60 group who only have 1.5 per cent higher debt than those aged over 60. This pattern in the size of the estimated coefficients ties in with life cycle effects as discussed above.

With respect to the educational attainment of the head of household, the two highest levels of educational attainment, namely having some college education and college education and above, are both positively related to the probability of holding debt relative to having below eighth grade school education. For example, a head of household who has college education and above is 47 percentage points more likely to hold debt than those whose highest level of education is below eighth grade. With respect to the amount of debt held, conditional on holding debt, with the exception of the dummy variable indicating some college education, the three remaining levels of the head of household's educational attainment are all positively associated with the continuous part of the debt distribution relative to having below eighth grade education. Those who are educated to college education and above have 3 per cent higher debt than the reference category. Thus, the difference in the findings related to

educational attainment across the binary and the continuous parts of the framework highlight the importance of applying the two part modelling approach and may explain the mixed results relating to the relationship between education and debt reported in the existing literature, see, for example, Brown and Taylor, 2008.

With respect to the influence of health, the existing literature (see, for example, Bridges and Disney, 2010, and Jenkins et al. 2008) generally supports a positive association between being in poor health and debt, although the direction of causality remains an unresolved issue here. Such a relationship may reflect an individual's inability to work whilst in poor health or may reflect direct costs associated with being in poor health such as additional transport costs or costs associated with medical care. Our results accord with the existing literature suggesting that having a head of household in good health is inversely associated with holding debt. We also find that the head of household's health does not exert a statistically significant influence on the amount of debt held. Our findings thus contribute to our understanding of the relationship between health and debt revealing that it is the holding of debt per se which is influenced by health status rather than the amount of debt held.

With respect to economic and financial factors, having a head of household in employment is positively associated with holding debt, with employees being 24 percentage points more likely to hold debt than those not in employment. However, conditional on holding debt, employment does not appear to influence the amount of debt held. Such a finding ties in with our a priori expectations in that being employed is generally a pre-requisite for taking out a loan or a credit card. The three household income quartile controls are all inversely associated with the probability of holding debt relative to being in the top household income quartile. This inverse association is also apparent in the continuous part of the model. Such results re-inforce the findings in the existing literature related to a positive association between income and debt and tie in with intuition in that the amount of credit given is generally dependent on the level of income. Our findings additionally reveal that this positive association relates

to both the probability of holding debt and, conditional on holding debt, the amount of debt held. It is apparent therefore that such economic and financial factors play an important role in the holding of debt at the household level.

Turning to asset holding, it is also apparent that some of the explanatory variables exert different influences across the two parts of the model (see Table 3). For example, having a male head of household does not appear to influence the probability of holding financial assets yet does exert a positive influence on the continuous part of the model. Interestingly, having a non-white head of household has a very large inverse effect on the probability of holding financial assets. The opposite effect is apparent in the continuous part of the model, with having a non-white head of household being positively associated with the amount of assets held. Once again, having a married head of household has positive influences in both parts of the model, which may reflect the holding of joint financial assets between husband and wife such as a joint savings account. Interestingly, in contrast to household debt, the number of children in the household does have an influence on asset holding, exerting a positive effect in the continuous part of the model. Age effects are not apparent in the binary part of the model yet all age groups have a positive influence on the amount of assets held relative to the aged over 60 group, with the size of the effect declining across the age groups. For example, those aged 25 to 34 have approximately 11 per cent lower financial assets compared to those aged above 60. Such findings, which tie in with the concave relationship found between age and the holding of stocks and shares in the existing literature, see for example Shum and Faig, 2006, may once again be capturing life cycle effects and may, for example, reflect dis-saving associated with retirement as older individuals move out of the labour market and liquidate financial assets in order to supplement their pension income.

With respect to education, the results relating to the binary part of the model are again somewhat mixed with the not completed high school but more than eighth grade and the completed high school categories being characterised by statistically

significant negative estimated coefficients and the college degree or above category being characterised by a significant positive coefficient. The estimated coefficients for the continuous part of the model are generally characterized by negative coefficients. Inconclusive effects for the role of education in asset holding in the U.S. are also found by Brown and Taylor (2008) within a bivariate Tobit framework. The role of education within this context is clearly complicated in that educational attainment is generally associated with a higher probability of employment and higher income yet in this framework both employment and income are also controlled for, as discussed below. Hence, any effect captured by educational attainment arguably does not reflect such labour market factors.

Having a head of household in good health is positively associated with the probability of holding financial assets, at around 25 additional percentage points as compared to someone not in good health, but does not influence the amount of assets held. Such findings accord with the finding of a positive association between debt and poor health in the existing literature in that individuals in poor health may face financial constraints and pressures due to the reasons outlined above in the context of debt holding and, as such, individuals in poor health may be less likely to hold financial assets. Our findings tie in with those of Rosen and Wu (2004), who, using data from the U.S. Health and Retirement Survey, find that being in poor health is inversely associated with the probability of holding a range of financial assets including bonds and risky assets such as stocks and shares.

The influence of economic and financial factors is once again apparent with having an employed head of household being positively associated with the probability of holding financial assets and the continuous part of the model, i.e., the level of assets. The effect of employment on the probability of holding financial assets is of a similar order of magnitude as that found for debt at approximately 24 percentage points. The household income quartiles are all inversely associated with both the binary and continuous parts of the model indicating that being in the highest household income



quartile is positively associated with both the probability of holding assets as well as, conditional on holding assets, the amount of assets held.

Finally, we also model household size given its potential role in influencing debt and asset holding at the household level and in particular due to the nature of the PSID question which specifically relates to asset and debt holding at the household level (see Table 5). It is apparent that household size in this context may be strongly correlated with reported debt and assets. This is confirmed by Brown and Taylor (2008), who using the PSID, find a strong positive effect of household size when jointly modelling household debt and assets. We comment only briefly on the estimated coefficients in this part of the framework given that the focus of our contribution lies in furthering our understanding of household finances, whilst in this context importantly allowing for household size as an additional methodological contribution to the empirical literature on household finances. As expected, higher levels of household income are positively associated with household size as is the head of household's age. It is also apparent that other socio-demographic characteristics of the head of household such as gender and ethnicity play an important role here.

In terms of correlations in the unobservable effects across equations, i.e. the estimated  $\rho_{ij}$ 's, these are all statistically significant (see Table 6). Positive correlations are found to exist between the error terms of the equation for the probability of holding debt and the equation for the amount of debt held. This is also the case for financial assets, although the correlation is not as large in this case. The correlation in the error terms between the probability of holding debt and the probability of holding financial assets, on the other hand, is negative. In contrast, the correlation between the error terms between the amount of debt and the amount of assets held is positive. These findings indicate inter-dependence across the different parts of the estimated model and endorse our joint modelling approach.

[Table 3 about here.]

[Table 4 about here.]

[Table 5 about here.]

[Table 6 about here.]

### 5.3 Negative Net Worth

Our approach highlights the importance of jointly modelling household debt and assets in order to explore the overall financial position of the household and to ascertain what factors determine each side of the household's balance sheet. Similarly, Barwell et al (2006) argue that net worth (the difference between assets and liabilities) reflects the overall state of a household's balance sheet, with financial pressure at the household level being indicated by negative net worth, i.e. being in a situation where financial liabilities outweigh financial assets. In order to use our modelling framework to explore the issue of negative net worth at the household level, we use our model to predict household assets and debt allowing us to analyse the prevalence of negative worth across key individual and household characteristics, where our predictions, in contrast to the existing literature, are based on a two part modelling strategy which importantly allows for the interdependence between asset and debt accumulation at the household level.

In our sample of households, 16% of households are predicted to be in negative net worth. Furthermore, a clear pattern emerges across the income quartiles with the prevalence of negative net worth declining monotonically with household income, with 31% of households in the lowest household income quartile predicted to be in negative net worth, as compared to 17% in the second quartile, 11% in the third quartile and only 8% in the fourth, i.e. the highest income, quartile. With respect to the employment status of the head of household, our model predicts 15% of households with an employed head to be in negative net worth as compared to 25% of households with an unemployed head. Similar disparities are apparent relating to health status, with 15% of households with a head reporting good health predicted to be in negative net worth as compared to 34% of households where the head reports that they are not

in good health. Furthermore, out of those households with a head reported not to be in good health, 31% are predicted to have no financial assets as compared to 11% amongst households with a head of household in good health, suggesting that those in poor health are potentially particularly financially vulnerable. These findings once again highlight the important role that health plays in household finances. Interestingly, across the head of household age categories, the prevalence of net worth is found to be relatively stable at around 16%, with the exception of the oldest category, those aged 60 and above, where only 8% of households are predicted to have negative net worth. In terms of year effects, we find that the proportion of households predicted to be in negative net worth increases over the time period with, for example, 8.7% of households predicted to be in negative net worth in 1989 compared to 17% in 2007. It should be noted that our predictions of the proportion of net worth are slightly higher than those based on the U.S. Survey of Consumer Finances (see, for example, Kennickell, 2003, who predicts 7.3% in 1989), which may reflect the oversampling of wealthy families in the U.S. Survey of Consumer Finances. Such changes in the proportion of households experiencing negative net worth over time may reflect changes in the prevailing economic and financial climate. Overall, our analysis of negative worth once again endorses the importance of economic and financial factors, such as income and employment status, in influencing finances at the household level as well as the key role played by health status.

## 6 Discussion

In this paper, we have contributed to the empirical literature on household finances by introducing a Bayesian two-part bivariate model. With correlated random effects, our approach allows for the potential interdependence between household debt and asset holding. This is important given that policy-makers have highlighted such interdependence as being relevant for ascertaining the true financial health of or burden faced by households. In addition, our approach incorporates a two part process which allows for

differences in the effects of the explanatory variables on the decision to acquire assets or debt and on the amount of assets or debt held. Finally, given that the information related to debt and assets is recorded at the household level, we allow for the potential endogeneity of household size by specifying household size as a third dependent variable within this joint modelling framework.

Our findings endorse the modelling of household debt and assets as a two part process since some explanatory variables exert different influences across the binary and the continuous parts of the model. This provides interesting information related to the influence of particular variables, such as health, ethnicity and age, which have attracted interest in the existing literature in this area. With respect to health, for example, the findings suggest that having a head of household in good health is inversely associated with holding debt but does not influence the amount of debt held. Similarly, having a head of household in good health is positively associated with asset holding but is not associated with the amount of assets held. In contrast, variables such as ethnicity appear to influence both the binary and the continuous parts of the model. For example, having a non-white head of household has an inverse effect on household debt in terms of both the probability of holding such debt and, conditional on holding debt, the amount of debt held. However, with respect to household assets, having a non-white head exerts a very large negative influence on the probability of holding financial assets yet a positive influence on the continuous part of the model. The head of household's age, on the other hand, does not influence the holding of debt yet, conditional on holding debt, does influence the amount of debt held, with the estimated effects being in accordance with consumption smoothing over the life cycle. A similar pattern of findings is evident for the relationship between age and financial asset holding. In sum, such results indicate how our flexible modelling strategy reveals detailed information relating to debt and asset holding at the household level. In addition, our analysis of the proportion of households predicted to have negative worth once again endorses the importance of economic and financial factors, such as income and employment status,

in influencing the financial situation of households as well as further highlighting the important role played by health status.

For policy-makers to understand how household assets and debt are distributed across demographic and socio-economic characteristics, our modelling framework thus provides a detailed picture of finances at the household level, as well as importantly allowing for the joint modelling approach, as endorsed by the correlations in the unobserved effects across the equations. The findings thus suggest interdependence across the different parts of the model, which confirms our a priori prediction that these aspects of household finances are inter-related which provided the motivation for developing the modelling framework in this regard. The framework we develop thus combines two important aspects of the modelling of household finances, namely the two-part approach and the joint modelling approach, which have been analysed separately in the existing literature via, for example, double hurdle models and bivariate Tobit models, respectively. It is apparent that, in order to accurately ascertain the extent to which households are financially vulnerable or subject to financial stress or pressure, developing such econometric approaches is important in order to further our understanding of this complex aspect of household behaviour and decision-making.

## References

- 1 Barwell, R., May, O. and Pezzini, S. “The Distribution of Assets, Income and Liabilities across UK Households.” *Bank of England Quarterly Bulletin*, Spring, 35–44 (2006).
- 2 Bertaut, C. “Stockholding Behaviour of U.S. Households: Evidence from the 1983-1989 Survey of Consumer Finances.” *Review of Economics and Statistics*, 80(2):263–275 (1998).
- 3 Bertaut, C. and Starr-McCluer, M. “Household Portfolios in the US.” In *Household Portfolios* (ed. L. Guiso, M. Haliassos and T. Jappelli), 189–217, MIT Press (2002).
- 4 Bridges, S. and Disney, R. “Debt and Depression.” *Journal of Health Economics*, 29(3):388–403 (2010).
- 5 Browning, M. and Lusardi, A. “Household Saving: Micro Theories and Micro Facts.” *Journal of Economic Literature*, 34(4):1797–1855 (1996).
- 6 Brown, S., Garino, G., Taylor, K., and Wheatley Price, S. “Debt and Financial Expectations: An Individual and Household Level Analysis.” *Economic Inquiry*, 43(1):100–120 (2005).
- 7 Brown, S., Garino, G., and Taylor, K. “Mortgages and Financial Expectations: A Household Level Analysis.” *Southern Economic Journal*, 74(3):857–878 (2008).
- 8 Brown, S. and Taylor, K. “Household Debt and Financial Assets: Evidence from Germany, Great Britain and the USA.” *Journal of the Royal Statistical Society, Series A*, 171(3):615–43 (2008).
- 9 Cox, P., Whitley, J., and Brierley, P. “Financial Pressures in the UK Household Sector: Evidence from the British Household Survey.” *Bank of England Quarterly Bulletin*, Winter:410–19 (2002).
- 10 Crook, J. “The Demand for Household Debt in the USA: Evidence from the 1995 Survey of Consumer Finance.” *Applied Financial Economics*, 11:83–91 (2001).
- 11 Daniels, M.J., Pourahmadi, M. “Modeling covariance matrices via partial auto-correlations.” *Journal of Multivariate Analysis*, 100: 2352–2363 (2009).
- 12 Diggle, P., Heagerty, P., Liang, K.-Y., and Zeger, S. *Analysis of Longitudinal Data*. New York: Oxford University Press (2002).
- 13 Efron, B., Hastie, T., Johnstone, I. & Tibshirani, R. Least angle regression, *Annals of Statistics*, 407–499 (2004).
- 14 Figueiredo, M. A. T. Adaptive sparseness for supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1150–1159 (2003).
- 15 Glick, R. and Lansing, K. J. “U.S. Household Deleveraging and Future Consumption Growth.” *Federal Reserve Bank of San Francisco (FRBSF) Newsletter*, Number 2009-16, May (2009).
- 16 Gropp, R., Scholz, J.K., and White, M. J. “Personal Bankruptcy and Credit Supply and Demand.” *The Quarterly Journal of Economics*, 112:217–51 (1997).
- 17 Guiso, L., Haliassos, M. and Jappelli, T. *Household Portfolios*, MIT Press (2002).

- 18 Hochguertel, S., Alessie, R., and Van Soest, A. "Savings Accounts versus Stocks and Bonds in a Household Portfolio Allocation." *Scandinavian Journal of Economics*, 99(1):81–97 (1997).
- 19 Hoel, P. G., Port, S. C., and Stone, C. J. *Introduction to Probability Theory*, Boston: Houghton Mifflin (1971).
- 20 Jenkins R., Bhugra D., Bebbington P., Brugha T., Farrell M., Coid J., Fryers T., Weich S., Singleton N., and Meltzer H. "Debt, Income and Mental Disorder in the General Population." *Psychological Medicine*, 38:1485-1493 (2008).
- 21 Joe, H. *Multivariate Models and Dependence Concepts*, *Monographs on Statistics and Applied Probability*, Chapman & Hall (1997).
- 22 Kennickell, A. B. "A Rolling Tide: Changes in the Distribution of Wealth in the U.S., 1989-2001". *Federal Reserve Banking Working Paper*, March 2003.
- 23 Lin, L., Bandyopadhyay, D., Lipsitz, S. R., and Sinha, D. "Association Models for Clustered Data with Binary and Continuous Responses." *Biometrics*, 66:287–293 (2010).
- 24 Nelsen, R. B. *An Introduction to Copulas*, *Lecture Notes in Statistics*, Springer Verlag, New York (1999).
- 25 Olsen MK and Schafer JL. "A two-part random-effects model for semicontinuous longitudinal data." *Journal of the American Statistical Association*, 96: 730745 (2001).
- 26 Tooze JA, Grunwald GK and Jones RH. "Analysis of repeated measures data with clumping at zero." *Statistical Methods in Medical Research*, 11(4): 341355 (2002).
- 27 Rosen, H. S. and Wu, S. "Portfolio Choice and Health Status." *Journal of Financial Economics*, 72:457–484 (2004).
- 28 Shum, P. and Faig, M. "What Explains Household Stock Holdings." *Journal of Banking and Finance*, 30(9):2579-2597 (2006).
- 29 Su, L., Tom, B. D., and Farewell, V. T. "A likelihood-based two-part marginal model for longitudinal semicontinuous data." *Statistical Methods in Medical Research* (2011). DOI: 10.1177/0962280211414620.
- 30 Wang, Z. and Louis, T. "Matching conditional and marginal shapes in binary mixed-effects models using a bridge distribution function." *Biometrika*, 90:765–775 (2003).
- 31 Yen, S. T., Boxall, P. C., and Adamowicz, W. L. "An Econometric Analysis of Donations for Environmental Conservation in Canada." *Journal of Agricultural and Resource Economics*, 22:246–63 (1997).

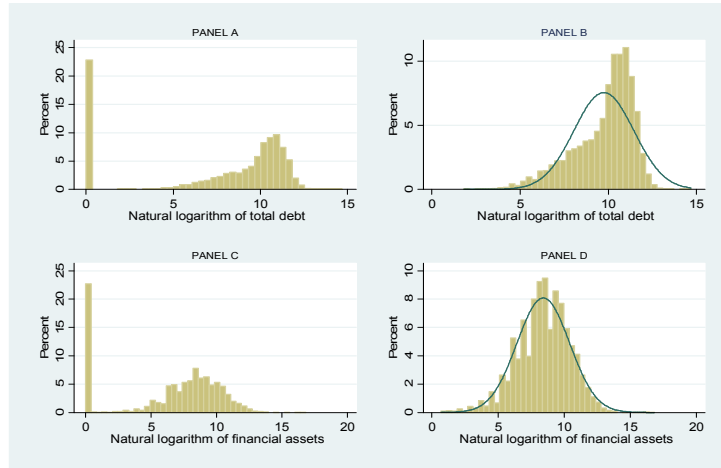


Figure 1: Histograms of Debt and financial assets



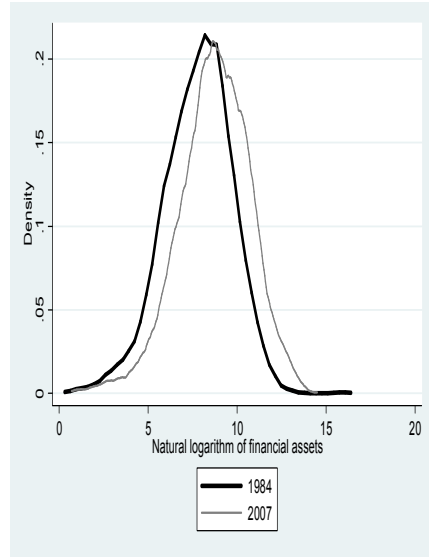
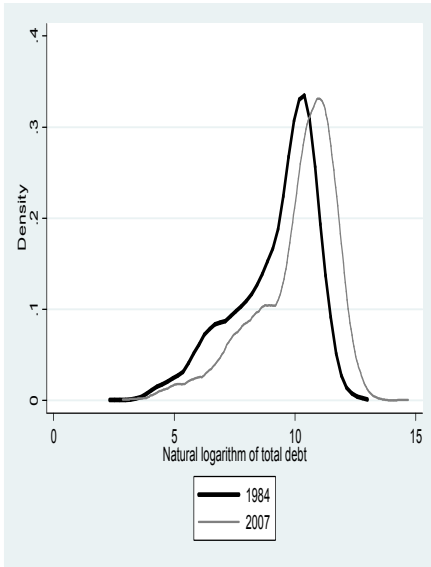


Figure 2: Distribution of Debt and financial assets

Table 1: Summary Statistics for Debt and Assets

	% Zero	Median	
		All Observations	Excluding Zero
Log Total Debt	22.85%	\$9.58	\$10.23
Total Debt		\$14,437	\$27,633
Log Total Financial Assets	22.68%	\$7.75	\$8.52
Total Financial Assets		\$2,318	\$5,000

Table 2: Summary Statistics Explanatory Variables

Variable	Description	Mean		
		All (15,656 obs)	Debt>0 (12,078 obs)	Assets >0 (12,105 obs)
male	=1 if male, 0=female	0.79	0.82	0.83
nonwhite	=1 if non white, 0=other ethnicity	0.30	0.25	0.22
marr	=1 if married or cohabiting, 0=other	0.64	0.70	0.70
hads	Number of adults in household	1.69	1.70	1.67
nkids	Number of children in household	1.02	1.02	0.99
ghealth	=1 if in good/excellent health, 0=poor/average	0.90	0.91	0.92
emp	=1 if employee, 0=otherwise	0.85	0.88	0.88
inc_025	=1 if income in 0-25 percentile, 0=other	0.25	0.18	0.18
inc_2550	=1 if income in 25-50 percentile, 0=other	0.25	0.23	0.24
inc_5075	=1 if income in 50-75 percentile, 0=other	0.25	0.27	0.27
inc_75	=1 if income in 75 percentile or above, 0=other	0.25	0.32	0.31
age_1824	=1 if aged 18-24, 0=other	0.05	0.05	0.05
age_25-34	=1 if aged 25-34, 0=other	0.21	0.19	0.19
age_3544	=1 if aged 35-44, 0=other	0.36	0.37	0.36
age_4554	=1 if aged 45-54, 0=other	0.31	0.32	0.32
age_5560	=1 if aged 55-60, 0=other	0.05	0.05	0.05
age_60	=1 if aged over 60, 0=other	0.02	0.02	0.03
ed0	=1 if completed less than eighth grade, 0=other	0.17	0.17	0.16
ed1	=1 if not completed high school, 0=other	0.09	0.08	0.07
ed2	=1 if completed high school, 0=other	0.30	0.28	0.28
ed3	=1 if some college, 0=other	0.20	0.21	0.21
ed4	=1 if graduated, 0=other	0.24	0.26	0.28

Table 3: Estimated marginal effects (posterior means and 95% credible intervals) of the independent variables on the binary and continuous parts of asset holding.

variable	Binary part				Continuous part			
	mean	2.5%	97.5%	significance	mean	2.5%	97.5%	significance
Intercept	2.504	2.060	2.970	yes	0.358	0.145	0.577	yes
time	-0.104	-0.239	0.038	no	2.092	2.052	2.137	yes
male	-0.017	-0.233	0.211	no	0.063	0.053	0.074	yes
nonwhite	-1.747	-1.920	-1.569	yes	0.037	0.012	0.056	yes
marr	0.245	0.102	0.393	yes	-0.115	-0.131	-0.099	yes
nkids	0.003	-0.043	0.050	no	0.058	0.045	0.070	yes
ghealth	0.247	0.069	0.418	yes	-0.003	-0.007	0.000	no
emp	0.236	0.091	0.378	yes	0.025	0.010	0.039	yes
inc-025	-0.975	-1.164	-0.790	yes	0.016	0.004	0.027	yes
inc-2550	-0.461	-0.638	-0.303	yes	-0.109	-0.122	-0.096	yes
inc-5075	-0.307	-0.475	-0.139	yes	-0.095	-0.106	-0.084	yes
age-1824	-0.082	-0.473	0.276	no	-0.051	-0.061	-0.041	yes
age-2534	-0.286	-0.616	0.025	no	-0.108	-0.152	-0.066	yes
age-3544	-0.079	-0.374	0.205	no	-0.073	-0.111	-0.034	yes
age-4554	0.008	-0.265	0.286	no	-0.065	-0.100	-0.029	yes
age-5560	-0.102	-0.446	0.240	no	-0.042	-0.072	-0.010	yes
ed1	-0.568	-0.782	-0.352	yes	-0.033	-0.066	0.000	no
ed2	-0.217	-0.389	-0.048	yes	-0.037	-0.057	-0.016	yes
ed3	0.093	-0.088	0.276	no	-0.034	-0.047	-0.020	yes
ed4	0.358	0.145	0.577	yes	0.011	-0.004	0.026	no

Table 4: Estimated marginal effects (posterior means and 95% credible intervals) of the independent variables on the binary and continuous parts of debt holding.

variable	Binary part				Continuous part			
	mean	2.5%	97.5%	significance	mean	2.5%	97.5%	significance
Intercept	2.340	1.872	2.821	yes	2.144	2.118	2.173	yes
time	-0.041	-0.179	0.099	no	0.091	0.082	0.099	yes
male	-0.378	-0.575	-0.186	yes	0.001	-0.014	0.016	no
nonwhite	-0.868	-1.048	-0.701	yes	-0.056	-0.067	-0.045	yes
marr	0.809	0.658	0.962	yes	0.085	0.076	0.095	yes
nkids	0.006	-0.039	0.054	no	0.001	-0.002	0.004	no
ghealth	-0.189	-0.367	-0.012	yes	-0.002	-0.014	0.009	no
emp	0.244	0.101	0.382	yes	0.005	-0.004	0.014	no
inc-025	-1.151	-1.334	-0.979	yes	-0.090	-0.100	-0.080	yes
inc-2550	-0.677	-0.842	-0.505	yes	-0.065	-0.073	-0.057	yes
inc-5075	-0.285	-0.457	-0.115	yes	-0.032	-0.039	-0.024	yes
age-1824	-0.228	-0.656	0.155	no	0.023	-0.004	0.052	no
age-2534	-0.172	-0.575	0.213	no	0.038	0.012	0.059	yes
age-3544	0.071	-0.304	0.409	no	0.047	0.024	0.067	yes
age-4554	0.111	-0.237	0.430	no	0.041	0.019	0.060	yes
age-5560	-0.262	-0.626	0.100	no	0.015	-0.010	0.038	no
ed1	-0.169	-0.397	0.049	no	-0.030	-0.045	-0.015	yes
ed2	-0.166	-0.339	0.001	no	-0.016	-0.025	-0.005	yes
ed3	0.344	0.149	0.533	yes	-0.001	-0.011	0.010	no
ed4	0.470	0.269	0.689	yes	0.030	0.020	0.040	yes

Table 5: Estimated marginal effects (posterior means and 95% credible intervals) of the independent variables on household size.

variable	mean	2.5%	97.5%	significance
Intercept	1.009	0.871	1.119	yes
time	-0.674	-0.706	-0.641	yes
male	-0.135	-0.172	-0.099	yes
nonwhite	0.129	0.100	0.158	yes
marr	0.305	0.272	0.339	yes
nkids	-0.185	-0.197	-0.174	yes
ghealth	-0.040	-0.081	0.001	no
emp	-0.046	-0.083	-0.009	yes
inc-025	-0.070	-0.110	-0.030	yes
inc-2550	-0.098	-0.134	-0.062	yes
inc-5075	-0.079	-0.113	-0.044	yes
age-1824	0.336	0.225	0.456	yes
age-2534	0.298	0.198	0.412	yes
age-3544	0.241	0.140	0.353	yes
age-4554	0.194	0.094	0.307	yes
age-5560	0.062	-0.055	0.186	no
ed1	-0.061	-0.111	-0.012	yes
ed2	-0.139	-0.179	-0.101	yes
ed3	-0.166	-0.210	-0.125	yes
ed4	-0.192	-0.234	-0.150	yes

Table 6: Variance component and correlation parameter estimates (posterior means and 95% credible intervals) for the random effects structure.

	parameter	mean	2.5%	97.5%
var-logit of lasssav	$\sigma_1^2$	1.969	1.646	2.337
corr-logit-gamma of lasssav	$\rho_{12}$	0.183	0.038	0.339
corr-logit( lasssav ) & logit(ltotdebt)	$\rho_{13}$	-0.388	-0.462	-0.310
corr-logit(lasssav) & gamma(ltotdebt)	$\rho_{14}$	-0.053	-0.172	0.061
corr-logit(lasssav) & no of household	$\rho_{15}$	-0.054	-0.238	0.157
var of gamma of lasssav	$\sigma_a^2$	0.905	0.883	0.929
corr-gamma( lasssav ) & logit(ltotdebt)	$\rho_{23}$	-0.212	-0.373	0.021
corr-gamma( lasssav ) & gamma(ltotdebt)	$\rho_{24}$	0.109	-0.128	0.506
corr-gamma(lasssav) & no of household	$\rho_{25}$	0.688	0.474	0.853
var of logit of ltotdebt	$\sigma_3^2$	1.798	1.451	2.171
corr-logit(ltotdebt) & gamma(ltotdebt)	$\rho_{34}$	0.406	0.274	0.505
corr-logit(ltotdebt) & no of household	$\rho_{35}$	-0.021	-0.236	0.212
var of gamma(ltotdebt)	$\sigma_d^2$	0.937	0.929	0.952
corr-gamma(ltotdebt) & no of household	$\rho_{45}$	0.434	0.171	0.730
var-no of household	$\sigma_s^2$	0.972	0.955	0.985