



Modelling household finances: A Bayesian approach to a multivariate two-part model



Sarah Brown ^{a,*}, Pulak Ghosh ^b, Li Su ^c, Karl Taylor ^a

^a Department of Economics, University of Sheffield, 9 Mappin Street, Sheffield S1 4DT, UK

^b Department of Quantitative Methods and Information Systems, Indian Institute of Management, Bangalore, India

^c MRC Biostatistics Unit, Cambridge Institute of Public Health, Cambridge Biomedical Campus, Robinson Way, Cambridge, CB2 0SR, UK

ARTICLE INFO

Article history:

Received 16 December 2014

Received in revised form 17 March 2015

Accepted 19 March 2015

Available online 26 March 2015

JEL classification:

C11

G11

G21

Keywords:

Assets

Bayesian approach

Bridge distribution

Debt

Two-part model

ABSTRACT

We contribute to the empirical literature on household finances by introducing a Bayesian multivariate two-part model, which has been developed to further our understanding of household finances. Our flexible approach allows for the potential interdependence between the holding of assets and liabilities at the household level and also encompasses a two-part process to allow for differences in the influences on asset or liability holding and on the respective amounts held. Furthermore, the framework is dynamic in order to allow for persistence in household finances over time. Our findings endorse the joint modelling approach and provide evidence supporting the importance of dynamics. In addition, we find that certain independent variables exert different influences on the binary and continuous parts of the model thereby highlighting the flexibility of our framework and revealing a detailed picture of the nature of household finances.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Over the last three decades, there has been growing interest in the financial economics literature in the nature of financial portfolios at the household level. Such interest has coincided with significant changes in debt and asset 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, [Glick and Lansing \(2009\)](#) and [Brown et al. \(2013\)](#). 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](#)), savings (for example, [Browning and Lusardi, 1996](#)) or debt (for example, [Brown et al., 2005, 2008](#), and [Crook, 2001](#)).

Policy-makers have, however, commented on the importance of analysing household financial assets and liabilities together, which is at odds with the approach generally taken in the academic literature which explores specific aspects of household finances in isolation of other aspects of the household balance sheet. 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.¹ Similarly, the Monetary Policy Committee in Great Britain has acknowledged the

* Corresponding author.

¹ Remarks made by Alan Greenspan "Understanding Household Debt Obligations" at the Credit Union National Association, Government Affairs Conference, Washington, D.C. February 23, 2004.

importance of establishing whether the same households have been accumulating financial assets as well as debt over the recent years (Bank of England, Minutes of the Monetary Policy Committee, 2002 and Brown and Taylor, 2008).

One exception in the academic literature 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 since a high proportion of them hold debt yet no financial assets, i.e. they have negative net worth. Such findings highlight the importance of further research in this area. Moreover, it is apparent that, in order to predict the influences of changes in economic policy at the household level, such as changes in the interest rate, it is important to adopt a holistic approach to analysing household finances including both assets and liabilities.

In order to contribute to the literature on household finances, we analyse panel data from the U.S. Panel Study of Income Dynamics (PSID), for 1984, 1989, 1994, 1999, 2001, 2003, 2005, 2007, 2009 and 2011, with our period of analysis covering pre and post the recent financial crisis. The PSID provides detailed information at the household level as well as allowing us to track households over time. In addition to providing empirical analysis of household finances during this period, we develop a flexible empirical framework, which reveals a detailed picture of liability and asset holding at the household level and allows us to uncover interdependencies across the various aspects of household finances. We adopt a Bayesian approach, which is highly flexible in the context of complex models. Hence, given the complicated nature of household finances, it is surprising that there is a lack of Bayesian analysis in the existing literature. 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, 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).² 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 held may be characterised by different influences. In terms of evaluating the level of financial pressure faced by households, there is a significant difference, for example, between being in debt and holding high amounts of debt.

A double-hurdle model is an alternative econometric specification, which allows independent variables to have different effects on the probability of holding debt or financial assets and on the level of debt or financial assets if it is non-zero. 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 analysing financial donations). The double-hurdle model has not, however, been extended to the multivariate case. Thus, studies adopting the double-hurdle approach have been restricted to focusing on one aspect of household finances. It is clearly important to allow for different aspects of household finances: for example, as stated above, it may not be problematic in terms of levels of financial vulnerability if households holding debt simultaneously hold assets to draw upon if an adverse event arises.

In this paper, we develop a flexible Bayesian multivariate two-part model for the joint modelling of four aspects of household finances, namely unsecured debt, secured debt, non-housing financial assets and housing assets. With correlated random effects, our approach allows for the potential interdependence between household liabilities and asset holding and, hence, allows for potential complex interactions between the various components of household finances, as well as persistence over time. 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, i.e. it is important to consider debt levels in the context of asset holdings and vice versa. 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.

The results from our new framework endorse the joint modelling approach and provide evidence supporting dynamics in asset and debt holding at the household level. In addition, our results indicate that certain independent variables exert different influences on the binary and continuous parts of the model. The results are potentially interesting from a policy perspective. For example, we find evidence of strong persistence in the probability of households holding unsecured debt: if the head of household held unsecured debt in the previous period the likelihood of currently holding such debt increases by 126 percentage points. Hence, if alleviating the number of individuals in debt is a concern, policymakers may consider influencing the underlying behaviour of households or the mechanisms behind how credit is obtained, such as from a high street bank or a pay-day loan company, where the latter is likely to be more accessible but carries higher risk in terms of the rate of interest required, which may exacerbate future debt levels.

2. Empirical framework

In this section, we develop a Bayesian multivariate two-part model for the joint modelling of four aspects of household finances, namely unsecured debt, secured debt, non-housing financial assets and housing assets. 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 and

² 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.

Schafer, 2001; Tooze et al., 2002). We adopt a Bayesian estimation strategy, which has the following distinct advantages for our model. First, our Bayesian estimation procedure, with the incorporation of the recent development of the Markov Chain Monte Carlo (MCMC) method (Gelfand and Smith, 1990), is powerful and flexible in dealing with complex non-linear problems, where the classical maximum likelihood approach encounters severe computational difficulties (Lopes and Carvalho, 2013). Second, the Bayesian strategy enables us to examine the entire posterior distribution of the parameters, and avoid the dependence on asymptotic properties to assess the sampling variability of the parameter estimates.

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. The model also incorporates dynamics in the two-part process, i.e. the decision to acquire and the amount held, hence allowing for persistence over time. 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 literature to date.

In terms of the specific statistical methods proposed in this paper, our multivariate 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 household balance sheet. 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, whilst 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 model (integrated over the distribution of the random intercept) 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., unsecured debt), 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, whilst the random intercept in the Gamma GLM for the continuous part follows a Normal distribution (Lin et al. 2010; Su et al. 2015; Wang and Louis, 2003). 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 variables, i.e. secured debt; non-housing financial assets and housing assets. The interdependence between the two parts of each dependent variable as well as the interdependence between all of the dependent variables is 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 is adopted to guarantee the positive definiteness of the correlation matrix (Daniels and Pourahmadi, 2009).

2.1. Modelling the value of unsecured debt

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

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

with

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

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

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

$$\begin{aligned} \text{logit}(p_{ij}^{ud}) &= \mathbf{X}_{ij}^1 \boldsymbol{\beta}^1 + \eta_1 y_{i,j-1}^{ud} + B_i^{ud} \\ \log(s_{ij}^{ud}) &\sim N(\mu_{ij}^{ud}, \sigma_i^{ud^2}) \\ \mu_{ij}^{ud} &= \mathbf{X}_{ij}^2 \boldsymbol{\beta}^2 + \eta_2 y_{i,j-1}^{ud} + V_i^{ud} \end{aligned} \quad (1)$$

where \mathbf{X}_{ij}^1 and \mathbf{X}_{ij}^2 are the independent variable vectors with associated parameters β^1 and β^2 for the binary and continuous parts, respectively; and B_i^{1d} and V_i^{1d} are the random intercepts of the two parts of the model accounting for the dependence of the repeated observations within the household.

2.2. Modelling the value of secured debt

Let y_{ij}^{sd} be the secured debt of the i th household ($i = 1, 2, \dots, n$) in the j th wave ($j = 1, 2, \dots, m$). Let R_{ij}^{sd} be a random variable denoting whether secured debt is held where

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

with

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

Further, let $s_{ij}^{sd} = [y_{ij}^{sd} | R_{ij}^{sd} = 1]$ denote the positive secured debt of the i th household in the j th wave.

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

$$\begin{aligned} \text{logit}(p_{ij}^{sd}) &= \mathbf{X}_{ij}^3 \beta^3 + \eta_3 y_{i,j-1}^{sd} + B_i^{sd} \\ \log(s_{ij}^{sd}) &\sim N(\mu_{ij}^{sd}, \sigma_i^{sd^2}) \\ \mu_{ij}^{sd} &= \mathbf{X}_{ij}^4 \beta^4 + \eta_4 y_{i,j-1}^{sd} + V_i^{sd} \end{aligned} \tag{2}$$

where \mathbf{X}_{ij}^3 and \mathbf{X}_{ij}^4 are the independent variable vectors with associated parameters β^3 and β^4 for the binary and continuous parts, respectively; and B_i^{sd} and V_i^{sd} are the random intercepts of the two parts of the model accounting for the dependence of the repeated observations within the household.

2.3. Modelling the value of non-housing financial assets

Let y_{ij}^{fa} be the financial assets, excluding the value of housing, of the i th household ($i = 1, 2, \dots, n$) in the j th wave ($j = 1, 2, \dots, m$). Let R_{ij}^{fa} be a random variable denoting whether financial assets are held where

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

with

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

Further, let $s_{ij}^{fa} = [y_{ij}^{fa} | R_{ij}^{fa} = 1]$ denote the positive financial assets of the i th household in the j th wave.

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

$$\begin{aligned} \text{logit}(p_{ij}^{fa}) &= \mathbf{X}_{ij}^5 \beta^5 + \eta_5 y_{i,j-1}^{fa} + B_i^{fa} \\ \log(s_{ij}^{fa}) &\sim N(\mu_{ij}^{fa}, \sigma_i^{fa^2}) \\ \mu_{ij}^{fa} &= \mathbf{X}_{ij}^6 \beta^6 + \eta_6 y_{i,j-1}^{fa} + V_i^{fa} \end{aligned} \tag{3}$$

where \mathbf{X}_{ij}^5 and \mathbf{X}_{ij}^6 are the independent variable vectors with associated parameters β^5 and β^6 for the binary and continuous parts, respectively; and B_i^{fa} and V_i^{fa} are the random intercepts of the two parts of the model accounting for the dependence of the repeated observations within the household.

2.4. Modelling the value of housing assets

Let y_{ij}^{ha} be the house value of the i th household ($i = 1, 2, \dots, n$) in the j th wave ($j = 1, 2, \dots, m$). Let R_{ij}^{ha} be a random variable denoting whether housing assets are held where

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

with

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

Further, let $s_{ij}^{ha} = [y_{ij}^{ha} | R_{ij}^{ha} = 1]$ denote the value of housing assets, conditional on such assets being held, of the i th household in the j th wave.

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

$$\begin{aligned} \text{logit}(p_{ij}^{ha}) &= \mathbf{X}_{ij}^7 \boldsymbol{\beta}^7 + \eta_7 y_{i,j-1}^{ha} + B_i^{ha} \\ \log(s_{ij}^{ha}) &\sim N(\mu_{ij}^{ha}, \sigma_i^{ha^2}) \\ \mu_{ij}^{ha} &= \mathbf{X}_{ij}^8 \boldsymbol{\beta}^8 + \eta_8 y_{i,j-1}^{ha} + V_i^{ha} \end{aligned} \tag{4}$$

where \mathbf{X}_{ij}^7 and \mathbf{X}_{ij}^8 are the independent variable vectors with associated parameters $\boldsymbol{\beta}^7$ and $\boldsymbol{\beta}^8$ for the binary and continuous parts, respectively; and B_i^{ha} and V_i^{ha} are the random intercepts of the two parts of the model accounting for the dependence of the repeated observations within the household.

Note that for each of the continuous outcomes we incorporate heteroscedasticity in the error terms, i.e. $(\sigma_i^{ud})^2$, $(\sigma_i^{sd})^2$, $(\sigma_i^{fa})^2$ and $(\sigma_i^{ha})^2$ in Eqs. (1) to (4), respectively. By assuming a prior on the variance, this enables the variance to be random across households and hence household specific.

2.5. Random effects model

For each household, we have an 8-dimensional random effect vector $\mathbf{b}_i = (B_i^{ud}, V_i^{ud}, B_i^{sd}, V_i^{sd}, B_i^{fa}, V_i^{fa}, B_i^{ha}, V_i^{ha})^T$. Since the binary and continuous parts of the sub-models are highly likely to be related within the households over the follow-up waves, 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 Eqs. (1) to (4) 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 outcomes, we extend the random intercept GLMM approach in Wang and Louis (2003) to the multivariate two-part model setting.

We assume that B_i^{ud} , B_i^{sd} , B_i^{fa} and B_i^{ha} , the random intercepts in the binary parts, marginally follow the bridge distributions of Wang and Louis (2003) with densities

$$\begin{aligned} f_1(b_i^{ud} | \phi_1) &= \frac{1}{2\pi} \frac{\sin(\phi_1 \pi)}{\cosh(\phi_1 b_i^{ud}) + \cos(\phi_1 \pi)} \quad (-\infty < b_i^{ud} < \infty) \\ f_3(b_i^{sd} | \phi_3) &= \frac{1}{2\pi} \frac{\sin(\phi_3 \pi)}{\cosh(\phi_3 b_i^{sd}) + \cos(\phi_3 \pi)} \quad (-\infty < b_i^{sd} < \infty) \\ f_5(b_i^{fa} | \phi_5) &= \frac{1}{2\pi} \frac{\sin(\phi_5 \pi)}{\cosh(\phi_5 b_i^{fa}) + \cos(\phi_5 \pi)} \quad (-\infty < b_i^{fa} < \infty) \\ f_7(b_i^{ha} | \phi_7) &= \frac{1}{2\pi} \frac{\sin(\phi_7 \pi)}{\cosh(\phi_7 b_i^{ha}) + \cos(\phi_7 \pi)} \quad (-\infty < b_i^{ha} < \infty) \end{aligned}$$

with unknown parameters ϕ_1, ϕ_3, ϕ_5 and ϕ_7 (where $0 < \phi_1, \phi_3, \phi_5, \phi_7 < 1$). The bridge distribution is symmetric with mean zero and variance $\sigma_k^2 = \pi^2(\phi_k^{-2} - 1)/3$, where $k = 1, 3, 5, 7$. 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 $\mathbf{b}_i = (B_i^{ud}, V_i^{ud}, B_i^{sd}, V_i^{sd}, B_i^{fa}, V_i^{fa}, B_i^{ha}, V_i^{ha})^T$, the marginal probabilities $\text{prob}(R_{ij}^{ud} = 1)$, $\text{prob}(R_{ij}^{sd} = 1)$, $\text{prob}(R_{ij}^{fa} = 1)$ and $\text{prob}(R_{ij}^{ha} = 1)$ relate to independent

variables through the same logit link functions as for the corresponding conditional probabilities. In addition, if we specify the marginal regression structure of the binary parts as

$$\begin{aligned} \text{logit}\left\{\text{prob}\left(R_{ij}^{ud} = 1\right)\right\} &= \mathbf{X}_{ij}\boldsymbol{\theta}^1 \\ \text{logit}\left\{\text{prob}\left(R_{ij}^{sd} = 1\right)\right\} &= \mathbf{X}_{ij}\boldsymbol{\theta}^3 \\ \text{logit}\left\{\text{prob}\left(R_{ij}^{fa} = 1\right)\right\} &= \mathbf{X}_{ij}\boldsymbol{\theta}^5 \\ \text{logit}\left\{\text{prob}\left(R_{ij}^{ha} = 1\right)\right\} &= \mathbf{X}_{ij}\boldsymbol{\theta}^7 \end{aligned}$$

then the marginal independent variable effects $\boldsymbol{\theta}^k$ ($k = 1, 3, 5, 7$) are proportional to the household specific conditional independent variable effects $\boldsymbol{\beta}^k$ with $\boldsymbol{\theta}^k = \phi_k \boldsymbol{\beta}^k$. Therefore models (1) to (4) can be rewritten as

$$\text{logit}\left\{\text{prob}\left(R_{ij}^{ud} = 1 \mid B_i^{ud}\right)\right\} = \mathbf{X}_{ij}^1 \boldsymbol{\theta}^1 / \phi_1 + B_i^{ud} \tag{5}$$

$$\text{logit}\left\{\text{prob}\left(R_{ij}^{sd} = 1 \mid B_i^{sd}\right)\right\} = \mathbf{X}_{ij}^3 \boldsymbol{\theta}^3 / \phi_3 + B_i^{sd} \tag{6}$$

$$\text{logit}\left\{\text{prob}\left(R_{ij}^{fa} = 1 \mid B_i^{fa}\right)\right\} = \mathbf{X}_{ij}^5 \boldsymbol{\theta}^5 / \phi_5 + B_i^{fa} \tag{7}$$

$$\text{logit}\left\{\text{prob}\left(R_{ij}^{ha} = 1 \mid B_i^{ha}\right)\right\} = \mathbf{X}_{ij}^7 \boldsymbol{\theta}^7 / \phi_7 + B_i^{ha} \tag{8}$$

Further, $V_i^{ud}, V_i^{sd}, V_i^{fa}$ and V_i^{ha} are assumed to be marginally normally distributed with mean zero and variance, $\tau_{ud}^2, \tau_{sd}^2, \tau_{fa}^2$ and τ_{ha}^2 , respectively. Therefore, $\log(s_{ij}^{ud}), \log(s_{ij}^{sd}), \log(s_{ij}^{fa})$ and $\log(s_{ij}^{ha})$, given the vector of random effects $\mathbf{b}_i = (B_i^{ud}, V_i^{ud}, B_i^{sd}, V_i^{sd}, B_i^{fa}, V_i^{fa}, B_i^{ha}, V_i^{ha})^T$, follow GLMM with means $(\mathbf{X}_{ij}^2 \boldsymbol{\beta}^2 + V_i^{ud}), (\mathbf{X}_{ij}^4 \boldsymbol{\beta}^4 + V_i^{sd}), (\mathbf{X}_{ij}^6 \boldsymbol{\beta}^6 + V_i^{fa})$ and $(\mathbf{X}_{ij}^8 \boldsymbol{\beta}^8 + V_i^{ha})$, respectively.

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, see 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 $F_1(b_i^{ud}), F_2(v_i^{ud}), F_3(b_i^{sd}), F_4(v_i^{sd}), F_5(b_i^{fa}), F_6(v_i^{fa}), F_7(b_i^{ha})$ and $F_8(v_i^{ha})$ are the CDFs of $\mathbf{b}_i = (B_i^{ud}, V_i^{ud}, B_i^{sd}, V_i^{sd}, B_i^{fa}, V_i^{fa}, B_i^{ha}, V_i^{ha})^T$, respectively, then there exists a function C such that the joint CDF of \mathbf{b}_i is $F(b_i^{ud}, v_i^{ud}, b_i^{sd}, v_i^{sd}, b_i^{fa}, v_i^{fa}, b_i^{ha}, v_i^{ha}) = C\{F_1(b_i^{ud}), F_2(v_i^{ud}), F_3(b_i^{sd}), F_4(v_i^{sd}), F_5(b_i^{fa}), F_6(v_i^{fa}), F_7(b_i^{ha}), F_8(v_i^{ha})\}$, see Nelsen (1999) and Joe (1997).

To construct the Gaussian copula for \mathbf{b}_i , we specify a vector $\mathbf{U}_i = (U_{i1}, U_{i2}, U_{i3}, U_{i4}, U_{i5}, U_{i6}, U_{i7}, U_{i8})^T$ such that

$$\begin{bmatrix} U_{i1} \\ U_{i2} \\ U_{i3} \\ \vdots \\ U_{i8} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{18} \\ \rho_{21} & 1 & \rho_{23} & \cdots & \rho_{28} \\ \rho_{31} & \rho_{32} & 1 & \cdots & \rho_{38} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{81} & \rho_{82} & \rho_{83} & \cdots & 1 \end{bmatrix} \right) \tag{9}$$

Note that the diagonal elements of the covariance matrix $\boldsymbol{\Sigma}$ are equal to 1 so that it is also the correlation matrix. We let $\rho_{j+j+\tau} = \text{corr}(U_{ij}, U_{i,j+\tau})$, where $j = 1, 2, \dots, 8; 1 \leq \tau \leq 8$, denote the correlation between U_{ij} and $U_{i,j+\tau}$. Using the probability integral transforms (see Hoel et al. (1971)), then $B_i^{ud} = F_1^{-1}\{\Phi(U_{i1})\}$, $B_i^{sd} = F_3^{-1}\{\Phi(U_{i3})\}$, $B_i^{fa} = F_5^{-1}\{\Phi(U_{i5})\}$ and $B_i^{ha} = F_7^{-1}\{\Phi(U_{i7})\}$ have marginal CDFs of $F_1(b_i^{ud}), F_3(b_i^{sd}), F_5(b_i^{fa})$ and $F_7(b_i^{ha})$, respectively (see Lin et al. (2010), and Wang and Louis (2003)). Here $\Phi(\cdot)$ is the standard normal CDF, and $F_k^{-1}(\cdot)$ ($k = 1, 3, 5, 7$) 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 $V_i^{ud}, V_i^{sd}, V_i^{fa}$ and V_i^{ha} , we have $V_i^{ud} = \tau_{ud} U_{i2}$, $V_i^{sd} = \tau_{sd} U_{i4}$, $V_i^{fa} = \tau_{fa} U_{i6}$ and $V_i^{ha} = \tau_{ha} U_{i8}$.

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 the off-diagonal elements will be estimated separately. 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 re-parameterization of $\boldsymbol{\Sigma}$ using the notion of partial autocorrelation from time series analysis. The advantage of this approach is the 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 we will leave $\boldsymbol{\Sigma}$ completely unstructured and the inferences will be based on the correlation matrix parameters $\rho_{j+j+\tau}$ as functions of partial autocorrelations denoted by $\gamma_{j+j+\tau} = \text{corr}(U_{ij}, U_{i,j+\tau} \mid U_{it}, j < l < j + \tau)$.

3. Bayesian inference

3.1. Likelihood specification

Let $\mathbf{Y}_i^{ud} = (y_{i1}^{ud}, \dots, y_{im}^{ud})^T$, $\mathbf{Y}_i^{sd} = (y_{i1}^{sd}, \dots, y_{im}^{sd})^T$, $\mathbf{Y}_i^{fa} = (y_{i1}^{fa}, \dots, y_{im}^{fa})^T$ and $\mathbf{Y}_i^{ha} = (y_{i1}^{ha}, \dots, y_{im}^{ha})^T$. Similarly, we define $\mathbf{X}_i^k = (\mathbf{X}_{i1}^k, \dots, \mathbf{X}_{im}^k)^T$ for $k = 1, 2, \dots, 8$. Let $\boldsymbol{\Omega}_1 = (\boldsymbol{\beta}^1, \boldsymbol{\beta}^2)$, $\boldsymbol{\Omega}_2 = (\boldsymbol{\beta}^3, \boldsymbol{\beta}^4)$, $\boldsymbol{\Omega}_3 = (\boldsymbol{\beta}^5, \boldsymbol{\beta}^6)$, $\boldsymbol{\Omega}_4 = (\boldsymbol{\beta}^7, \boldsymbol{\beta}^8)$ be the vectors for the multivariate dependent variables, $\boldsymbol{\Omega}_5 = (\phi_1, \tau_{ud}, \phi_3, \tau_{sd}, \phi_5, \tau_{fa}, \phi_7, \tau_{ha})$ be the parameter vector for random effects $\mathbf{b}_i = (B_i^{ud}, V_i^{ud}, B_i^{sd}, V_i^{sd}, B_i^{fa}, V_i^{fa}, B_i^{ha}, V_i^{ha})^T$. The likelihood function for the i th household can be partitioned as

$$\begin{aligned} &L(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{\Omega}_3, \boldsymbol{\Omega}_4, \boldsymbol{\Omega}_5 | \mathbf{Y}_i^{ud}, \mathbf{Y}_i^{sd}, \mathbf{Y}_i^{fa}, \mathbf{Y}_i^{ha}, \mathbf{X}_i^1, \mathbf{X}_i^2, \mathbf{X}_i^3, \mathbf{X}_i^4, \mathbf{X}_i^5, \mathbf{X}_i^6, \mathbf{X}_i^7, \mathbf{X}_i^8; \mathbf{b}_i) \propto \\ &L(\boldsymbol{\Omega}_1 | \mathbf{Y}_i^{ud}, \mathbf{X}_i^1, \mathbf{X}_i^2, B_i^{ud}, V_i^{ud}) L(\boldsymbol{\Omega}_2 | \mathbf{Y}_i^{sd}, \mathbf{X}_i^3, \mathbf{X}_i^4, B_i^{sd}, V_i^{sd}) L(\boldsymbol{\Omega}_3 | \mathbf{Y}_i^{fa}, \mathbf{X}_i^5, \mathbf{X}_i^6, B_i^{fa}, V_i^{fa}) \\ &L(\boldsymbol{\Omega}_4 | \mathbf{Y}_i^{ha}, \mathbf{X}_i^7, \mathbf{X}_i^8, B_i^{ha}, V_i^{ha}) L(\boldsymbol{\Omega}_5 | B_i^{ud}, V_i^{ud}, B_i^{sd}, V_i^{sd}, B_i^{fa}, V_i^{fa}, B_i^{ha}, V_i^{ha}) \end{aligned} \tag{10}$$

where

$$\begin{aligned} L(\boldsymbol{\Omega}_1 | \mathbf{Y}_i^{ud}, \mathbf{X}_i^1, \mathbf{X}_i^2, B_i^{ud}, V_i^{ud}) &= \prod_{j=1}^m \left\{ 1 - \text{prob}(R_{ij}^{ud} = 1 | B_i^{ud}) \right\}^{(1-r_{ij}^{ud})} \left\{ \text{prob}(R_{ij}^{ud} = 1 | B_i^{ud}) \times \text{LN}(Y_{ij}^{ud}; \mu_{ij}^{ud}, \sigma_i^{ud^2}) \right\}^{r_{ij}^{ud}} \\ L(\boldsymbol{\Omega}_2 | \mathbf{Y}_i^{sd}, \mathbf{X}_i^3, \mathbf{X}_i^4, B_i^{sd}, V_i^{sd}) &= \prod_{j=1}^m \left\{ 1 - \text{prob}(R_{ij}^{sd} = 1 | B_i^{sd}) \right\}^{(1-r_{ij}^{sd})} \left\{ \text{prob}(R_{ij}^{sd} = 1 | B_i^{sd}) \times \text{LN}(Y_{ij}^{sd}; \mu_{ij}^{sd}, \sigma_i^{sd^2}) \right\}^{r_{ij}^{sd}} \\ L(\boldsymbol{\Omega}_3 | \mathbf{Y}_i^{fa}, \mathbf{X}_i^5, \mathbf{X}_i^6, B_i^{fa}, V_i^{fa}) &= \prod_{j=1}^m \left\{ 1 - \text{prob}(R_{ij}^{fa} = 1 | B_i^{fa}) \right\}^{(1-r_{ij}^{fa})} \left\{ \text{prob}(R_{ij}^{fa} = 1 | B_i^{fa}) \times \text{LN}(Y_{ij}^{fa}; \mu_{ij}^{fa}, \sigma_i^{fa^2}) \right\}^{r_{ij}^{fa}} \\ L(\boldsymbol{\Omega}_4 | \mathbf{Y}_i^{ha}, \mathbf{X}_i^7, \mathbf{X}_i^8, B_i^{ha}, V_i^{ha}) &= \prod_{j=1}^m \left\{ 1 - \text{prob}(R_{ij}^{ha} = 1 | B_i^{ha}) \right\}^{(1-r_{ij}^{ha})} \left\{ \text{prob}(R_{ij}^{ha} = 1 | B_i^{ha}) \times \text{LN}(Y_{ij}^{ha}; \mu_{ij}^{ha}, \sigma_i^{ha^2}) \right\}^{r_{ij}^{ha}} \end{aligned}$$

with μ_{ij}^{ud} , μ_{ij}^{sd} , μ_{ij}^{fa} and μ_{ij}^{ha} given in Eqs. (1) to (4) respectively, and

$$\begin{aligned} L(\boldsymbol{\Omega}_5 | B_i^{ud}, V_i^{ud}, B_i^{sd}, V_i^{sd}, B_i^{fa}, V_i^{fa}, B_i^{ha}, V_i^{ha}) &= C \left\{ F_1(b_i^{ud}), F_2(v_i^{ud}), F_3(b_i^{sd}), F_4(v_i^{sd}), F_5(b_i^{fa}), F_6(v_i^{fa}), F_7(b_i^{ha}), F_8(v_i^{ha}) \right\} \\ & f_1(b_i^{ud} | \phi_1) \phi\left(\frac{v_i^{ud}}{\tau_{ud}}\right) f_3(b_i^{sd} | \phi_3) \phi\left(\frac{v_i^{sd}}{\tau_{sd}}\right) f_5(b_i^{fa} | \phi_5) \phi\left(\frac{v_i^{fa}}{\tau_{fa}}\right) f_7(b_i^{ha} | \phi_7) \phi\left(\frac{v_i^{ha}}{\tau_{ha}}\right) \end{aligned} \tag{11}$$

with $f_k(\cdot)$ $k = 1, 3, 5, 7$ being the bridge density functions, $\phi(\cdot)$ being the standard normal density function and $c(\cdot)$ being the density of the copula $C(\cdot)$ from Section 2.5. This is given by

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

Here $\mathbf{q} = (q_1, \dots, q_8)$ with $(0 < q_k < 1, k = 1, \dots, 8)$, $\mathbf{u} = (u_1, \dots, u_8)^T$ a vector of normal scores $u_k = \Phi^{-1}(q_k)$, and \mathbf{I} is an 8-dimensional identity matrix.

3.2. 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 $\boldsymbol{\Omega}_1$, $\boldsymbol{\Omega}_2$, $\boldsymbol{\Omega}_3$, $\boldsymbol{\Omega}_4$, and $\boldsymbol{\Omega}_5$, are independently distributed. Because the number of independent variables is large in the joint model, they might be 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 the following:

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

where ψ is the hyper-parameter. In the regression, the use of the Laplace prior is known as the LASSO, [Efron et al. \(2004\)](#). Thus, the posterior mode estimate of $\boldsymbol{\beta}$ is the LASSO estimate.

The above LASSO shrinks all the parameters to the same degree. However, when some effects are non-null, shrinkage towards these non-null locations may be beneficial.³ Thus, we extend the above LASSO specification using a newly developed Bayesian adaptive shrinkage LASSO prior proposed by MacLehose and Dunson (2010). We assume the following prior

$$\begin{aligned}
 \beta_j | \psi_j &\sim N(\mu_j, \psi_j) \\
 \psi_j &\sim \exp(2/\lambda_j) \\
 (\mu_j, \lambda_j) &\sim \theta \delta(\mu_j|0) \times \text{Gamma}(\lambda_j|a_0, b_0) + (1-\theta)N(\mu_j|c, d) \times \text{Gamma}(\lambda_j|a_1, b_1) \\
 \theta &\sim \text{Beta}(1, 1)
 \end{aligned}
 \tag{12}$$

where $\delta(\mu_j|0)$ indicates that μ_j has a degenerate distribution with all its mass at 0. With probability θ , the coefficient β_j is shrunk towards zero as in the standard LASSO model. With probability $(1 - \theta)$, the coefficient β_j is shrunk towards non-zero mean, μ_j . The amount of shrinkage is determined by λ_j with large values resulting in greater shrinkage. We specify a_0 and b_0 to give support to large values of λ_j in order to allow for strong shrinkage of β_j towards 0, whilst specifying a_1 and b_1 to give support to smaller values of λ_j to allow less shrinkage towards non-zero values.

In the analysis shown in Sections 2.1 to 2.4, all elements of the regression coefficient vectors, $\beta^1, \beta^2, \beta^3, \beta^4, \beta^5, \beta^6, \beta^7$ and β^8 are assigned an adaptive LASSO prior. Following MacLehose and Dunson (2010), we assume, $a_0 = b_0 = 30$ and $a_1 = b_1 = 7$. For parameters in the random effects model, denote $\tau_1^2 = \pi^2(\phi_1^{-2} - 1)/3$, $\tau_3^2 = \pi^2(\phi_3^{-2} - 1)/3$, $\tau_5^2 = \pi^2(\phi_5^{-2} - 1)/3$ and $\tau_7^2 = \pi^2(\phi_7^{-2} - 1)/3$ and we use the following priors $\tau_1, \tau_{ud}, \tau_3, \tau_{sd}, \tau_5, \tau_{fa}, \tau_7$ and τ_{ha} , iid \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 $\gamma_{12}, \gamma_{13}, \gamma_{14}, \gamma_{15}, \gamma_{16}, \gamma_{17}, \gamma_{18}, \gamma_{23}, \gamma_{24}, \gamma_{25}, \gamma_{26}, \gamma_{27}, \gamma_{28}, \gamma_{34}, \gamma_{35}, \gamma_{36}, \gamma_{37}, \gamma_{38}, \gamma_{45}, \gamma_{46}, \gamma_{47}, \gamma_{48}, \gamma_{56}, \gamma_{57}, \gamma_{58}, \gamma_{67}, \gamma_{68}$ and γ_{78} . We assume a weakly informative Gamma prior distribution for error variances, $(\sigma_i^{ud})^2, (\sigma_i^{sd})^2, (\sigma_i^{fa})^2$ and $(\sigma_i^{ha})^2$.

The joint posterior distribution of the model parameters conditional on the observed data are obtained by combining the likelihood from Section 3.1 and the previously specified priors using Bayes Theorem:

$$\begin{aligned}
 &POST(\Omega_1, \Omega_2, \Omega_3, \Omega_4, \Omega_5, \mathbf{b}_i | \mathbf{Y}_i^{ud}, \mathbf{Y}_i^{sd}, \mathbf{Y}_i^{fa}, \mathbf{Y}_i^{ha}) \\
 &\propto \prod_{i=1}^n L_i(\Omega_1, \Omega_2, \Omega_3, \Omega_4, \Omega_5 | \mathbf{Y}_i^{ud}, \mathbf{Y}_i^{sd}, \mathbf{Y}_i^{fa}, \mathbf{Y}_i^{ha}, \mathbf{b}_i) f(\mathbf{b}_i) f(\beta|\psi, \theta) f(\theta) f(\tau_1) f(\tau_{ud}) \\
 &\quad \times f(\tau_3) f(\tau_{sd}) f(\tau_5) f(\tau_{fa}) f(\tau_7) f(\tau_{ha}) f(\gamma) f(\sigma_i^{ud^2}) f(\sigma_i^{sd^2}) f(\sigma_i^{fa^2}) f(\sigma_i^{ha^2}).
 \end{aligned}
 \tag{13}$$

The posterior distributions are analytically intractable. However, computation can be achieved using MCMC methods such as the Gibbs sampler (Gelfand et al., 1992). Since all the full conditional distributions are not standard, a straightforward implementation of the Gibbs sampler using standard sampling techniques may not be possible. However, sampling methods can be performed using Adaptive Rejection Sampling (ARS), metropolis Hastings and/or blocked Gibbs sampling methods (Gilks and Wild, 1992).

In this paper we have used a general program for Bayesian inference using Gibbs Sampling implemented in the WinBUGS package (version 1.4.1), Spiegelhalter et al. (1996). WinBUGS uses the Gibbs sampling algorithm to construct transition kernels for its Markov chain samplers. During compilation, WinBUGS chooses a method to draw samples for each of the full conditional distributions of the model parameters. Such sampling can be done univariately or in multivariate nodes. The sampling methods within WinBUGS include direct sampling using standard algorithms, derivative free ARS (Gilks, 1992), slice sampling (Neal, 2000) and metropolis sampling (Gelfand and Smith, 1990) and blocked Gibbs sampling. The first choice is always a standard density if it is available. This possibility arises when a full conditional is recognizable. For non-standard but log-concave full conditionals, ARS sampling is used to sample the full conditional (Gilks and Wild, 1992). WinBUGS checks if log-concavity is satisfied or not, and uses slice sampling if this condition is not met (Neal, 2000). The random walk metropolis algorithm is also used by WinBUGS for non-conjugate continuous full conditionals. The samples from the posterior distribution obtained from the MCMC allow us to achieve summary measures of the parameter estimates. Summary statistics such as posterior means and statistical significance obtained from 95% credible intervals are provided for inference.

3.3. Model selection

For comparing between alternative models, we use a selection criteria called Deviance Information Criteria (DIC) proposed by Spiegelhalter et al. (2002). This approach has been used in several previous studies involving zero-inflated data (such as Neelon et al. (2010); Montagna et al. (2012)). As with the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), the DIC also provides an assessment of model fit and a penalty for model complexity. Let θ be the set of parameters in a model, then the DIC is defined as $\bar{D}(\theta) + p_D$, where $\bar{D}(\theta) = E[D(\theta)|\mathbf{y}]$ is the posterior mean of the deviance, $D(\theta)$, and $p_D = \bar{D}(\theta) - \hat{D}(\theta) = E[D(\theta)|\mathbf{y}] - D(E[\theta|\mathbf{y}])$ is the difference in the posterior mean of the deviance and the deviance evaluated at the posterior mean of the set of parameters. The deviance is taken as negative twice the log-likelihood and is a measure of a model's relative fit, whereas p_D is a penalty for the model's complexity (Montagna et al., 2012). Whilst the AIC and the BIC are well suited for fixed effects models (since the number of parameters are easily determined), for the hierarchical random effects model the DIC is better suited as the dimension of the parameter space is less clear and depends on the degree of

³ We are grateful to an anonymous referee for pointing this out.

heterogeneity between subjects (Montagna et al., 2012). The DIC was proposed to estimate the number of effective parameters in a Bayesian hierarchical model. For finite mixture models, Celeux et al. (2006) proposed a modified DIC, termed DIC_3 , since in a mixture distribution the effective number of parameters, p_D , can be negative. The DIC_3 estimates $\hat{D}(\theta)$ using the posterior mean of the marginal likelihood and is given by $DIC_3 = 2\bar{D}(\theta) + 2 \log \left[\prod_i^n \hat{f}(y_i) \right]$, where $\hat{f}(y_i)$ is the posterior mean of the marginal likelihood contribution for subject i . A smaller DIC denotes a better model. If two models differ by more than ten, then the one with the smaller DIC is considered the best fit (Spiegelhalter et al., 2002).

We also use an alternative model selection criteria – the cross validated predictive approach of Gelfand et al. (1992), i.e. the predictive distributions conditioned on the observed data with a single data point deleted. The conditional predictive coordinate (CPO) has been widely used for model diagnostic and assessment (Chen et al., 2000; Gelfand et al., 1992) and for our model is defined as:

$$CPO_i = p\left(\mathbf{Y}_i^P, \mathbf{Y}_i^C, \mathbf{V}_i \mid \mathbf{Y}_{-i}^P, \mathbf{Y}_{-i}^C, \mathbf{V}_{-i}\right) = E_\theta \left[p\left(\mathbf{Y}_i^P, \mathbf{Y}_i^C, \mathbf{V}_i \mid \theta\right) \mid \mathbf{Y}_{-i}^P, \mathbf{Y}_{-i}^C, \mathbf{V}_{-i} \right]$$

where $p(\mathbf{Y}_i^P, \mathbf{Y}_i^C, \mathbf{V}_i \mid \mathbf{Y}_{-i}^P, \mathbf{Y}_{-i}^C, \mathbf{V}_{-i})$ is the posterior predictive density of $p(\mathbf{Y}_i^P, \mathbf{Y}_i^C, \mathbf{V}_i)$ for subject i conditional on the observed data with a single data point deleted, θ denotes all unobservables in the model under consideration, and the expectation is taken with respect to the posterior distribution of θ conditional on $p(\mathbf{Y}_{-i}^P, \mathbf{Y}_{-i}^C, \mathbf{V}_{-i})$. Gelfand et al. (1992) have proposed a harmonic mean estimator of CPO_i based on Markov chain samples from the full posterior given the entire data, y . Geisser and Eddy (1979) and Gelfand et al. (1992) proposed the pseudo-marginal likelihood $\prod p(\mathbf{Y}_i^P, \mathbf{Y}_i^C, \mathbf{V}_i \mid \mathbf{Y}_{-i}^P, \mathbf{Y}_{-i}^C, \mathbf{V}_{-i})$ or its logarithm, i.e. the Log Pseudo Marginal Likelihood (LPML):

$$LPML = \sum_{i=1}^n \log p\left(\mathbf{Y}_i^P, \mathbf{Y}_i^C, \mathbf{V}_i \mid \mathbf{Y}_{-i}^P, \mathbf{Y}_{-i}^C, \mathbf{V}_{-i}\right).$$

Higher values of the LPML denote a superior specification. In our analysis we use the DIC and LPML to select between competing models.

4. Modelling household finances

4.1. Data

Using the proposed model, we analyse data collected from the U.S. Panel Study of Income Dynamics (PSID). The 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 4800 families in 1968 to more than 7000 families by the turn of the century. Further information on the PSID is available at: <http://psidonline.isr.umich.edu>. Our data set contains information on: unsecured debt, e.g. credit card debt, (y_{ij}^{ud}); secured debt, e.g. mortgage debt, (y_{ij}^{sd}); non-housing financial assets, e.g. stocks and shares, (y_{ij}^{fa}); and housing assets (y_{ij}^{ha}). We analyse this information for a balanced panel of households measured over 10 waves, spanning over a quarter of a century, $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, 10$. To be specific, in 1984, 1989, 1994, 1999, 2001, 2003, 2005, 2007, 2009 and 2011, the head of family is asked to provide information about the household's financial assets and debt.

For debt, the head of household is asked to specify the amount remaining on the first mortgage and second mortgage which constitutes secured debt in our analysis; whilst credit card charges, student loans, medical or legal bills and other loans constitute unsecured debt. In terms of financial assets, the head of household 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. The head of household is also asked what the present value of the house or apartment is, specifically about how much would it bring if it sold was today. Hence, we have detailed information on the household balance sheet in terms of debt and 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 zeros, the logarithmic variables are recoded to zero, since there are no reported values between zero and unity in the sample.

The sample of households analysed in this study forms a balanced panel where the same heads of household are observed in each wave yielding total observations over the period of 3930.⁴ Table 1 provides summary statistics for the dependent variables. Fig. 1 presents histograms of the natural logarithm of each dependent variable (see left hand side for the sample of all households). The right hand column of Fig. 1 shows the distribution of each respective dependent variable excluding zeros. 41% (37%) of households over the period have no unsecured (secured) debt, with mean (median) values conditional on non-zeros being \$8133 (\$3125) and \$55,138 (\$43,529) for unsecured and secured debt, respectively. In terms of financial assets, 22% (24%) of households have no financial assets (housing assets), with mean (median) values conditional on possessing such assets being \$12,920 (\$3497) and \$103,586 (\$79,096) for financial assets and house value, respectively.

Fig. 2 shows how the distribution of each dependent variable has changed between 1984 and 2011 conditional on positive amounts being held of each respective dependent variable. For unsecured debt, secured debt and house value, there has been a clear shift in the distribution to the right hand side, i.e. conditional on positive amounts being held, the average amount held has

⁴ We have replicated the analysis which follows using an unbalanced panel. The results are generally very similar to those of the balanced data and so, for brevity, we do not report the analysis of the unbalanced panel. This is available on request.

Table 1
Summary statistics for dependent variables.

	% ZERO	All observations	Excluding zeros
Log unsecured debt (y_{ij}^{ud})	40.92%	4.735	8.014
Unsecured debt (\$)		\$4805	\$8133
Log secured debt (y_{ij}^{sd})	36.69%	6.674	10.542
Secured debt (\$)		\$34,906	\$55,138
Log financial assets (y_{ij}^{fa})	21.88%	6.264	8.019
Financial assets (\$)		\$10,093	\$12,920
Log house value (y_{ij}^{ha})	23.59%	8.575	11.222
House value (\$)		\$79,152	\$103,586

increased and this is driven by higher amounts being held above the mean. This is less apparent for non-housing assets, where the distribution looks similar over time: although the mean amount held has increased between 1984 and 2011 from 7.8 to 8.2 log points (\$7200 to \$13,572), the tails of the distribution are similar. What underlies the distributions over time for financial assets is that a high proportion of the sample held non-housing assets in both 1984 and 2011, at 71% and 77%, respectively. Given financial assets include checking and savings accounts, these figures are perhaps not surprising since such accounts are common being relatively low risk and highly liquid. For the other dependent variables, there are marked differences over time in the proportion of household holdings, where the corresponding figures for those holding unsecured debt, secured debt and housing in 1984 (2011) are: 56% (65%); 45% (61%); and 46% (79%), respectively.

The independent variables used in the analysis to explain both the continuous parts and binary components of household debt and financial asset holding, i.e. y_{ij}^q and p_{ij}^q , respectively (see Section 2) where $q = ud, sd, fa, ha$, follow the existing literature and consist of time invariant head of household characteristics and time varying variables. We also allow for full dynamics in both continuous and binary outcomes, by the inclusion of lagged dependent variables $y_{ij}^q - 1$ and $p_{ij}^q - 1$ (see Section 2). 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 include: 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 and full definitions of the explanatory variables are given in Table 2.

In the four binary parts of the model, p_{ij}^q , i.e. the decision to hold a particular type of debt or asset, as in Brown et al. (2013), we include a set of additional variables. Specifically, we also include the proportion of household heads employed in the financial services in the state (this follows Bertaut and Starr-McCluer, 2002), whether the head of household has been bankrupt in the past and the degree of risk tolerance of the head of household, which is increasing in risk tolerance.⁵ In terms of modelling the level of unsecured debt, y_{ij}^{ud} , and financial assets, y_{ij}^{fa} , we also condition on whether the head of household has a mortgage.

4.2. Results

Table 3 presents the findings from applying the model detailed above to the PSID data whilst Table 4 tests different model specifications specifically: model 1, the joint model with dynamics (as detailed in Sections 2 and 3 above); model 2, the joint model with no dynamics, i.e. a static version of the model in Sections 2 and 3; model 3, a joint model with dynamics but without exclusion restrictions (which are used to identify the binary outcomes); and model 4, which is characterised by independence, i.e. $\Sigma = 0$. Clearly, model 1 is the favoured specification since it has the smallest DIC and largest LPML. Hence, we can reject the null hypothesis that the preferred functional form is static. Such a finding is arguably unsurprising: we would predict the existence of state dependence in household finances with respect to asset and debt holding. For example, secured debt holding is a relatively long-term commitment made by households.⁶

In terms of the set of additional variables used to model the probability of holding unsecured debt, secured debt, financial assets and home ownership, these are jointly significant in determining the decision to hold each item on the household balance sheet (see Table 3). Moreover, the full joint dynamic model with the additional variables in the binary part (model 1) dominates the

⁵ In the 1996 PSID, a Risk Aversion Section was included containing detailed information on attitudes toward risk. This section of the PSID contains five questions related to hypothetical gambles with respect to lifetime income. Hence, it is possible to rank individuals based on an ordinal index of risk attitudes. Kimball et al. (2008, 2009) discuss some the disadvantages of this ordinal measure and alternatively assign a range of risk tolerance coefficients to each gamble response category. They argue that the imputations offer advantages over the categorical sequence of gamble responses in that the responses can be formulated into a single cardinal measure of preferences. It is this cardinal risk tolerance measure which we adopt herein.

⁶ We have also performed prior sensitivity analysis. This was based on making various choices of prior parameters by changing only one parameter at a time and keeping all other parameters constant to their default values. This follows the standard practice in the Bayesian paradigm (see, for example, Ghosh and Gonen (2008), Gelman et al. (2013), and Stroud and Johannes (2014)). The main justification behind this approach is that, with so many parameters, if the hyperprior values of all parameters are changed at the same time, it is difficult to ascertain which prior is the sensitive one if the overall results change, i.e. it is difficult to pinpoint which prior is problematic. The results are generally robust for all the choices of prior parameters that we have explored.

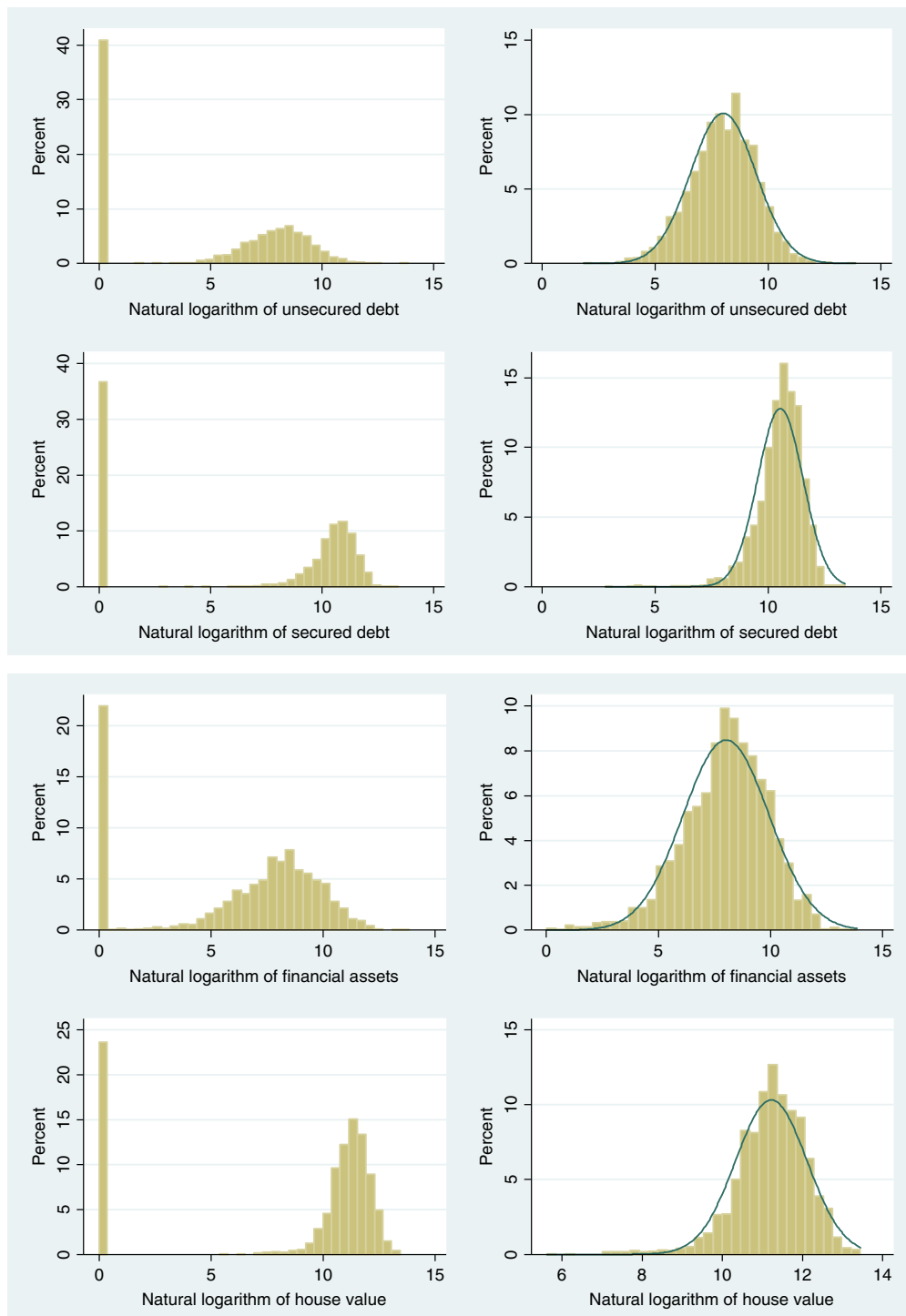


Fig. 1. Histograms of unsecured debt, secured debt, financial assets and house value.

corresponding model without such additional variables (model 3) in terms of the DIC and LPML statistics (see Table 4). Hence, the null hypothesis that the additional variables included in the binary outcomes are jointly equal to zero is rejected.

Interestingly, heads of household who have been declared bankrupt in the past have a higher (lower) probability of holding unsecured (secured) debt, as found by Brown et al. (2013). Given that debt repayments are generally financed from household income, it is apparent that if income is subject to risk (due to, for example, redundancy, unemployment, or changes in real wages), then the attitudes towards risk of the individual will potentially influence the decision to acquire debt, given the distribution of future income

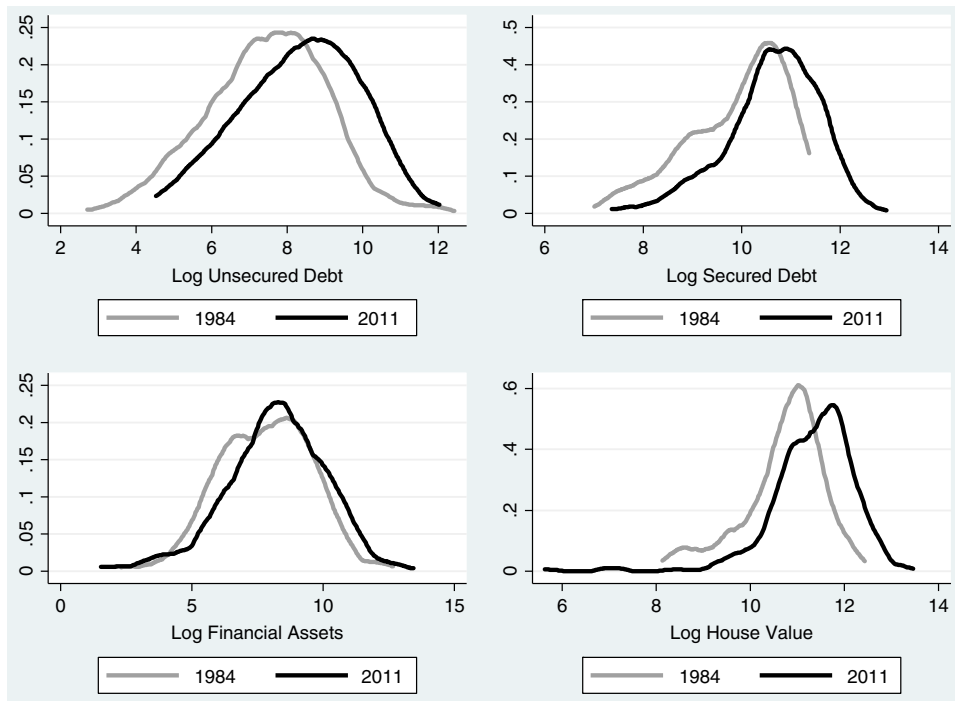


Fig. 2. Distributions of unsecured debt, secured debt, financial assets and house value over time; conditional on positive values.

and interest rates. In terms of the existing literature, [Donkers and Van Soest \(1999\)](#) find that risk averse Dutch homeowners tend to live in houses with lower mortgages, whilst [Brown et al. \(2013\)](#) report that risk aversion is negatively associated with the level of both unsecured and secured debt held. Our analysis sheds further light on the relationship between risk preference and household finances, in particular revealing that it is the decision to hold debt which is influenced by risk attitudes. To be specific, those heads of households who are more risk tolerant have a higher probability of holding unsecured and secured debt. As found by [Bertaut and Starr-McCluer \(2002\)](#), the share of household heads employed in the financial services sector by state has a positive association with the probability of owning non-housing financial assets.

In terms of the dynamics, there is clear evidence of state dependence for both the continuous and binary parts of unsecured debt and house value, whilst dynamics only appear to be important (in terms of statistical significance) for the binary parts of secured debt

Table 2
Summary statistics of explanatory variables.

Variable	Description	Mean	Standard deviation
Male	= 1 if male, 0 = female	0.4351	0.4958
Non-white	= 1 if non-white, 0 = other ethnicity	0.3257	0.4687
Married or cohabiting	= 1 if married or cohabiting, 0 = otherwise	0.8001	0.4002
Number of adults in household	Number of adults (16+) in household	2.2695	0.7995
Number of kids in household	Number of children (aged <16) in household	1.4018	1.1744
Excellent or good health	= 1 if in good/excellent health, 0 = poor/average	0.6336	0.4819
Employee	= 1 if employee, 0 = otherwise	0.8346	0.3716
Income in 0–24th percentile	= 1 if income in 0–25 percentile, 0 = otherwise	0.1295	0.3358
Income in 25–49th percentile	= 1 if income in 25–50 percentile, 0 = otherwise	0.2028	0.4021
Income in 50–75th percentile	= 1 if income in 50–75 percentile, 0 = otherwise	0.2969	0.4570
Aged 18–24	= 1 if aged 18–24, 0 = otherwise	0.0493	0.2167
Aged 25–34	= 1 if aged 25–34, 0 = otherwise	0.1692	0.3750
Aged 35–44	= 1 if aged 35–44, 0 = otherwise	0.3247	0.4683
Aged 45–54	= 1 if aged 45–54, 0 = otherwise	0.3689	0.4826
Aged 55–60	= 1 if aged 55–60, 0 = otherwise	0.0758	0.2648
Did not complete high school	= 1 if not completed high school, 0 = otherwise	0.0331	0.1789
Completed high school	= 1 if completed high school, 0 = otherwise	0.3656	0.4817
Some college	= 1 if some college, 0 = otherwise	0.2585	0.4379
Graduated	= 1 if graduated, 0 = otherwise	0.3165	0.4652
Ever bankrupt	= 1 if previously bankrupt, 0 = otherwise	0.0789	0.2696
Risk tolerance	Risk tolerance (higher values denote greater risk tolerance)	1.0678	2.1041
% in financial services by state	Proportion of employed in the financial services in the state	0.0442	0.0355
No mortgage held	= 1 if currently has mortgage, 0 = otherwise	0.3669	0.4820

Table 3

Estimated Bayesian marginal effects (posterior means) of the independent variables upon binary and continuous outcomes.

	Unsecured debt		Secured debt		Financial assets		House value	
	prob($Y > 0$)	log(Y) $Y > 0$	prob($Y > 0$)	log(Y) $Y > 0$	prob($Y > 0$)	log(Y) $Y > 0$	prob($Y > 0$)	log(Y) $Y > 0$
$\{Y > 0\}_j - 1$	1.256 *	–	2.176 *	–	1.322 *	–	0.755 *	–
$\{\log(Y) Y > 0\}_j - 1$	–	–0.474 *	–	–0.272	–	0.263	–	3.030 *
Male	–2.860 *	0.806 *	–0.734 *	–0.547	–0.142	0.738 *	0.043	0.061
Non-white	–0.047	–1.102	0.550	–0.538	–0.561	–1.202 *	–0.084	–0.492
Married or cohabiting	–1.747 *	0.610	2.898 *	2.137 *	0.633 *	0.569	0.558 *	2.050 *
Number of adults in household	1.290 *	0.131	0.144	0.275	–0.354 *	0.158	0.173 *	0.628 *
Number of kids in household	1.092 *	–0.371 *	0.451 *	0.474	–0.959 *	–0.394 *	0.272 *	0.401 *
Excellent or good health	0.088	–0.522 *	0.066 *	0.382	0.997 *	0.542 *	0.077	0.478 *
Employee	2.085 *	–0.121	0.302	0.603	–0.348	–0.088	0.281 *	0.277
Income in 0–24th percentile	3.515 *	–1.484 *	–1.712 *	–2.042	–1.302	–1.502 *	–0.760 *	–1.636 *
Income in 25–49th percentile	0.302	–1.290 *	–1.778 *	–1.576	0.152	–1.332 *	–0.833 *	–1.063 *
Income in 50–75th percentile	3.636 *	–0.224	–0.244	–0.715	0.938 *	–0.254	–0.414 *	–0.798 *
Aged 18–24	–3.490 *	2.617 *	–3.299 *	–0.135	–2.177 *	2.466 *	–1.853 *	–0.289
Aged 25–34	–3.455 *	3.071 *	–1.240	2.015 *	–0.275	2.964 *	–1.327 *	0.968
Aged 35–44	–2.541 *	2.783 *	0.167	2.817 *	0.737	2.679 *	–0.340	2.043 *
Aged 45–54	–0.866	1.947 *	–0.131	2.540 *	0.284	1.785 *	0.384	1.970 *
Aged 55–60	1.857 *	0.902	–0.250	2.608 *	–0.085	0.832	0.546	2.320 *
Did not complete high school	–0.399	0.885	–0.285	1.115	0.801 *	0.946	0.618	1.598 *
Completed high school	0.821	0.919	–1.349 *	0.212	1.203 *	0.965 *	0.196	0.989
Some college	3.713 *	1.430 *	–0.692 *	0.556	0.571	1.509 *	0.373	1.579 *
Graduated	3.761 *	1.702 *	1.812 *	0.628	0.725 *	1.820 *	–0.289	0.894
Ever bankrupt	3.319 *	–	–1.043 *	–	0.108	–	0.408	–
Risk tolerance	2.457 *	–	1.327 *	–	–0.171 *	–	0.097 *	–
% in financial services by state	–0.822	–	–2.044	–	4.436 *	–	–1.947	–
No mortgage held	0.053	–0.140	–	–	–0.085	–0.272	–	–
1989	–0.624	1.720 *	0.053	1.400 *	–0.226	1.565 *	0.260	1.539 *
1994	–1.462 *	1.345 *	0.699	1.880 *	–0.298	1.205 *	0.404	2.414 *
1999	1.934 *	1.077 *	–0.613	1.271 *	–1.212 *	0.934 *	–0.506	1.890 *
2001	–2.807 *	1.898 *	–0.402	1.005	3.143 *	1.745 *	–0.496	1.986 *
2003	0.864	1.991 *	–0.335	1.420 *	1.817 *	1.840 *	0.075	2.657 *
2005	2.353 *	2.555 *	0.057	1.563 *	1.779 *	2.389 *	–0.590	2.426 *
2007	–2.859 *	2.214 *	0.656	2.053 *	–0.628	2.079 *	–0.395	2.781 *
2009	0.979	2.696 *	0.558	2.538 *	–0.429	2.535 *	–0.372	3.059 *
2011	–5.231 *	2.796 *	0.567	2.570 *	1.263 *	2.650 *	–0.258	3.115 *
Observations	3390							

* denotes statistical significance at the 5% level.

and financial assets. In particular, focusing on unsecured debt, if the head of household held unsecured debt in the previous period then the probability of holding unsecured debt increases by 126 percentage points, *ceteris paribus*. However, the log amount of unsecured debt, conditional on holding a non-zero value, is decreasing in the amount held in the previous period. This might suggest that over time households are paying off such loans. In terms of house value, having housing equity in the previous period increases the probability of home ownership (either outright or on a mortgage) by 76 percentage points, *ceteris paribus*, and the current house value is a positive function of the value in the previous time period.

The importance of modelling both sides of the household balance sheet as a two-part process is apparent, as can be seen from Table 4, since the diagnostic statistics reject the hypothesis that $\Sigma = 0$. This can be seen by comparing the DIC and LPML from model 1 to that of model 4. Furthermore, it is clear from Table 5 that many of the variance and covariance terms in Σ are statistically significant suggesting complex interactions across the various aspects of household finances. The interrelationship between household assets and liabilities is likely to be complicated. For example, Kullmann and Siegel (2005) show that exposure to real estate risk reduces household financial asset holding, yet homeowners are more prone to invest in risky financial assets such as shares traded in the stock market. The variance–covariance matrix of the errors terms, Σ , sheds some light on this: where statistically significant, correlations in the error terms suggest that there are unobserved factors which influence the probability of jointly holding different types of liabilities and assets. In accordance with the findings of Kullmann and Siegel (2005), we find evidence of statistically significant covariance terms between non-housing financial assets, such as stocks and shares, and home ownership.

Table 4

Model selection.

Model	DIC	LPML
1. Joint model with dynamics	1011.422	–0.4676
2. Joint model with no dynamics, i.e. static	1067.297	–0.6190
3. Joint model with dynamics and exclusion restrictions	1231.103	–0.6320
4. Independence, i.e. $\Sigma = 0$	1804.912	–0.6512

Table 5
Variance–covariance matrix.

VAR (binary unsecured debt) $\sum_{1,1}$	1.7100*
COV (binary unsecured debt and log unsecured debt) $\sum_{1,2}$	2.2000*
COV (binary unsecured debt and binary secured debt) $\sum_{1,3}$	0.6740
COV (binary unsecured debt and log secured debt) $\sum_{1,4}$	−1.8300*
COV (binary unsecured debt and binary financial asset) $\sum_{1,5}$	−0.0082
COV (binary unsecured debt and log financial asset) $\sum_{1,6}$	1.6980*
COV (binary unsecured debt and binary house value) $\sum_{1,7}$	−0.3189
COV (binary unsecured debt and log house value) $\sum_{1,8}$	0.8213
VAR (log unsecured debt) $\sum_{2,2}$	2.1390*
COV (log unsecured debt and binary secured debt) $\sum_{2,3}$	−0.4711
COV (log unsecured debt and log secured debt) $\sum_{2,4}$	0.5427*
COV (log unsecured debt and binary financial asset) $\sum_{2,5}$	−1.9120*
COV (log unsecured debt and log financial asset) $\sum_{2,6}$	1.9120*
COV (log unsecured debt and binary house value) $\sum_{2,7}$	−0.5953*
COV (log unsecured debt and log house value) $\sum_{2,8}$	1.4350*
VAR (binary secured debt) $\sum_{3,3}$	4.7200*
COV (binary secured debt and log secured debt) $\sum_{3,4}$	−2.5070*
COV (binary secured debt and binary financial asset) $\sum_{3,5}$	0.3244
COV (binary secured debt and log financial asset) $\sum_{3,6}$	−0.3281
COV (binary secured debt and binary house value) $\sum_{3,7}$	1.5430*
COV (binary secured debt and log house value) $\sum_{3,8}$	−1.5620*
VAR (log secured debt) $\sum_{4,4}$	1.8610*
COV (log secured debt and binary financial asset) $\sum_{4,5}$	−0.6259
COV (log secured debt and log financial asset) $\sum_{4,6}$	0.4644
COV (log secured debt and binary house value) $\sum_{4,7}$	−0.7115*
COV (log secured debt and log house value) $\sum_{4,8}$	1.1190*
VAR (binary financial asset) $\sum_{5,5}$	2.8260*
COV (binary financial asset and log financial asset) $\sum_{5,6}$	−1.7940*
COV (binary financial asset and binary house value) $\sum_{5,7}$	0.3502
COV (binary financial asset and log house value) $\sum_{5,8}$	−1.0290*
VAR (log financial asset) $\sum_{6,6}$	1.8100*
COV (log financial asset and binary house value) $\sum_{6,7}$	−0.5206*
COV (log financial asset and log house value) $\sum_{6,8}$	1.2730*
VAR (binary house value) $\sum_{7,7}$	1.0210*
COV (binary house value and log house value) $\sum_{7,8}$	−0.9088*
VAR (log house value) $\sum_{8,8}$	1.6090*

* denotes statistical significance at the 5% level.

In addition to capturing relationships across the holding of the different types of debt and assets, the flexibility of the two-part process is also evident when comparing the influence of the explanatory variables across the binary and the continuous parts of the model, where it can be seen that some explanatory variables exert different influences across these two parts (see Table 3).

For example, it is apparent from the logistic model results that having a male head of household is inversely associated with the probability of holding unsecured debt yet exerts a statistically significant positive influence on, conditional on holding unsecured debt, the amount of unsecured debt held. As expected, having a married head of household has a very strong positive association with the probability of holding secured debt and a positive influence on the amount of secured debt held. Such findings may reflect the joint holding of debt within couples, such as a jointly held mortgage for the family home. On the opposite side of the household balance sheet, having a married head of household increases the probability of home ownership by 56 percentage points and has a positive influence on the amount of equity held where married individuals have approximately twice the amount of housing assets compared to single heads of household (conditional on home ownership).

Household composition is associated with household finances. This is particularly the case for unsecured debt, where a one standard deviation increase in the number of children in the household increases the probability of holding this type of debt by around 109 percentage points, yet conditional on holding unsecured debt, the number of children decreases the amount of debt held. The number of adults in the household, on the other hand, only influences the likelihood of holding unsecured debt.

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 suggest that a head of household in poor health holds higher levels of unsecured debt compared to those in good or excellent health. Having a head of household in good health is positively associated with the probability of holding financial assets, and, conditional on holding such assets, such individuals have around a 54 percentage points higher amount than a head of household in poor health. Such findings accord with the finding of a positive association between unsecured debt and poor health in the existing literature, in that those individuals in poor health may face financial constraints and pressures and, as such, individuals in poor health may be less likely to hold financial assets. Our findings also 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.

With respect to economic and financial factors, having a head of household in employment is positively associated with holding unsecured debt, with employees being twice as likely to hold unsecured debt compared to those not in employment. However, conditional on holding this type of debt, employment does not appear to influence the amount of unsecured debt held. Such a finding ties in with our a priori expectations in that being employed is often a prerequisite for taking out a personal loan or a credit card. The only other instance of where labour market status has a statistically significant effect in the two-part model is on the probability of home ownership where employed heads of household are 28 percentage points more likely to own their home.

The three household income quartile controls are all inversely associated with the probability of holding secured debt relative to being in the top household income quartile and statistically significant (with the exception of those above the median). This inverse association is also apparent in the continuous part of the model but does not reach levels of statistical significance. Such results reinforce the findings in the existing literature related to a positive association between income and secured debt. Worryingly, focusing on the probability of holding unsecured debt, those in the bottom income quartile are 3.5 times more likely to hold such debt compared to those households in the top part of the income distribution. Turning to the opposite side of the household balance sheet, there is a monotonic relationship between household income and the amount of financial assets (conditional on holding financial assets) and the amount of housing assets (conditional on owning a home).

Interestingly, the results suggest that the age of the head of household has a statistically significant negative influence on the probability of holding unsecured debt yet statistically significant positive effects on the amount of unsecured debt are apparent for all age groups relative to individuals aged over 60. The age effects peak for those aged 25 to 34 who have approximately three times the amount of unsecured debt than those aged over 60. Such age effects may reflect consumption smoothing over the life cycle, with individuals aged between 25 and 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 contrast to the striking association between age and unsecured debt, there are generally no effects from age on the probability of holding secured debt. The exception to this is those heads of household aged 18 to 24 who are more than three times less likely to hold mortgage debt than those aged above 60. However, age effects are apparent when focusing on the amount of secured debt held, conditional on holding such debt, where again life cycle factors would appear to matter with the level of debt culminating when aged 35 to 44.

Focusing on the role of age effects on the opposite side of the household balance sheet, age effects are not apparent in the binary part of the model for financial assets or housing assets yet all age groups have a positive influence on the amount of financial assets held relative to the aged over 60 group, with the size of the effect declining across the age groups.⁷ For example, those aged 25 to 34 have approximately three times the value of 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 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 unsecured debt relative to having below eighth grade school education. For example, a head of household who has college education and above is nearly four times more likely to hold unsecured debt than those whose highest level of education is below eighth grade. With respect to the amount of unsecured debt held, conditional on holding this type of debt, a head of household who has graduated holds 170 percentage points more unsecured debt relative to heads having below eighth grade education. Turning to secured debt, the only effect that educational attainment has is on the probability of holding mortgage debt. For example, a head of household who has graduated has a 181 percentage point higher probability of holding such debt compared to those with education below eighth grade, i.e. being nearly two and half times as likely to hold such debt. Moreover, there is no significant impact from education on the level of secured debt. Thus, the differences 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 \(2008\)](#).

Turning to financial assets and the role of educational attainment, no clear pattern emerges other than that those who have graduated have a higher probability of holding non-housing financial assets and, conditional on ownership of such assets, they hold larger amounts: specifically 182 percentage points more than those with education below eighth grade, ceteris paribus. This again concurs with the existing literature, see, for example [Hong et al. \(2004\)](#) for the U.S. and [Guiso et al. \(2008\)](#) who analyse Dutch and Italian survey data.

Controls for the year of interview are incorporated into the analysis in order to account for unobserved macroeconomic conditions that have the potential to affect all households. Interestingly, the value of each type of debt conditional on holding that type of debt has generally increased over time compared to the base year of 1984. This is especially the case post 2001 for unsecured and secured debt. This is an effect over and above inflation since monetary values are held at constant prices.

⁷ This finding for the two-part process of non housing financial assets is consistent with [Bertaut and Starr-McCluer \(2002\)](#), in that it is the level of assets which is influenced by age, but at odds with the results of [Ameriks and Zeldes \(2004\)](#) who find a significant effect of age on ownership but not the amount held.

For example, compared to a head of household in the base period, a head of household in 2011 held nearly three times as much unsecured debt (conditional on holding such debt) and over two and half times as much secured debt (again conditional on holding secured debt). However, this aggregate macroeconomic trend in terms of the accumulation of liabilities has been offset by similar magnitudes on the opposite side of the balance sheet in terms of increases in the amount of non-housing financial assets and housing assets.

To place our modelling contributions into context, we have contrasted the results from our new joint Bayesian estimator with classical alternatives, in particular the multivariate Tobit model, which has been used in the household finances literature. This is also a joint estimator but is less flexible than the approach we develop here in that the decision to acquire and the amount held of a particular aspect of household finances cannot be disentangled.⁸ Whilst it not possible to directly compare the model performance between a classical estimator and a Bayesian estimator, in [Table A1](#) in the appendix, the results of modelling the natural logarithm, $\log(Y)$, of unsecured debt, secured debt, financial assets and housing value within a joint framework are shown. Coefficients are reported throughout, which it should be noted are not directly comparable to the Bayesian marginal effects. However, we can compare the sign and statistical significance of the covariates on the dependent variables between the Bayesian and classical estimators. Clearly, a joint modelling approach is also supported within the context of a classical estimator since the null hypothesis that the correlation in the error terms is equal to zero is rejected. This suggests that a joint modelling framework yields an efficiency gain over alternative single equation estimators such as the double-hurdle model. Dynamics, i.e. state dependence, and income effects are also shown to be important, which is consistent with the Bayesian estimator although the multivariate Tobit model masks whether the effect of such covariates is operating through the censored or uncensored part of the distribution. This is an obvious advantage of our new estimator. For example, from the results shown in [Table 3](#), it is apparent that dynamic effects operate on the participation decision for each dependent variable, whilst the role of state dependence on the amounts held of each outcome is less clear.

5. Conclusion

In this paper, we have introduced a Bayesian multivariate two-part model and applied it to the modelling of the household balance sheet specifically in terms of liabilities (unsecured and secured debt) and in terms of assets (non-housing financial assets and home ownership). With correlated random effects, our approach allows for the potential interdependence between household liabilities and asset holding and, hence, allows for potential complex interactions between the various components of household finances. This is important given that policy-makers have highlighted such interdependence as being relevant for ascertaining the true financial health of households, i.e. it is important to consider debt levels in the context of asset holdings and vice versa. 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. 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. We also incorporate dynamics in the two-part process for each outcome of interest where there is evidence of persistence. This is especially the case for unsecured debt.

To understand how household assets and debt are distributed across demographic and socio-economic characteristics, our modelling framework 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 different aspects of household finances. 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 interrelated which provided the key motivation for developing the modelling framework in this regard. The framework we develop thus combines four important aspects of the modelling of household finances, which have generally been analysed separately in the existing literature: namely the two-part approach, allowing for heteroscedasticity, the incorporation of a dynamic process, and the joint modelling approach. 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 the complex nature of household financial behaviour and decision-making.

Acknowledgements

We are very grateful to the editor and two referees for their excellent comments and advice. Sarah Brown is grateful for the financial support from the Leverhulme Trust for the Major Research Fellowship 2011-007. Li Su's work was supported by the Medical Research Council [Unit Programme number U105261167]. The normal disclaimer applies.

⁸ A double-hurdle model is another alternative classical estimator, which does separate the participation decision, i.e. the binary outcome, and the monetary amount, i.e. the continuous outcome. However, the disadvantage of this approach is that a joint estimator across a set of dependent variables is not available and, moreover, given that our findings reveal the importance of interdependence across not only the two-parts of a particular dependent variable but also across the four aspects of household finances, the double-hurdle approach is arguably more restrictive as well as less efficient. However, for purposes of comparison, we have also implemented a double-hurdle model, the results from which are available on request.

Appendix A

Table A1

Dynamic multivariate Tobit model.

	Unsecured debt	Secured debt	Financial assets	House value
	log(Y)	log(Y)	log(Y)	log(Y)
{log(Y)} _j – 1	0.7012 *	0.8386 *	0.3733 *	0.5898 *
Male	0.0132	–0.4440 *	0.2331 *	–0.2076
Non-white	–0.4062 *	–1.4480 *	–1.9269 *	–1.0549 *
Married or cohabiting	0.1260	2.7019 *	0.8445 *	1.7984 *
Number of adults in household	0.3154 *	0.3603 *	–0.0990	0.2605 *
Number of kids in household	–0.0704	0.2025 *	–0.2001 *	0.0854
Excellent or good health	–0.1696	0.0755	0.2672 *	0.0648
Employee	0.0544	0.5609 *	0.0097	–0.0226
Income in 0–24th percentile	–0.9069 *	–3.8777 *	–2.8509 *	–2.8752 *
Income in 25–49th percentile	–0.5371 *	–2.5308 *	–1.4219 *	–1.7843 *
Income in 50–75th percentile	0.4429 *	–0.6484 *	–0.6560 *	–0.5714 *
Aged 18–24	1.9544	–2.3145	–1.2715	–3.9304 *
Aged 25–34	1.9035	1.2941	–1.2013	0.0456
Aged 35–44	1.3789	0.8073	–1.0006 *	0.0810
Aged 45–54	1.2306	0.5365	–0.7671	0.1175
Aged 55–60	1.0675	1.0161	–0.5314	0.5644
Did not complete high school	0.4107	–2.4468 *	–1.5035 *	–1.9157 *
Completed high school	2.3276 *	–1.2704	0.6390	–0.7065
Some college	2.8353 *	–0.5823	0.7554 *	–0.2582
Graduated	2.2285 *	–0.4874	0.9239 *	–0.2014
LR chi ² (112); p-value	6266.62; p = [0.000]			
H ₀ : ρ ₁₂ = ρ ₁₃ = ... = ρ ₃₄ = 0 chi ² (6); p-value	1954.86; p = [0.000]			
Observations	3390			

* denotes statistical significance at the 5% level. Year controls not reported for brevity. Coefficients are reported.

References

- Ameriks, J., Zeldes, S., 2004. How do household portfolio shares vary with age? Working Paper. University of Columbia
- Bertaut, C., 1998. Stockholding behaviour of U.S. households: evidence from the 1983–1989 Survey of Consumer Finances. *Rev. Econ. Stat.* 80 (2), 263–275.
- Bertaut, C., Starr-McCluer, M., 2002. Household portfolios in the US. In: Guiso, M., Haliassos, Jappelli, T. (Eds.), *Household Portfolios*. MIT Press, pp. 189–217.
- Bridges, S., Disney, R., 2010. Debt and depression. *J. Health Econ.* 29 (3), 388–403.
- Brown, S., Taylor, K., 2008. Household debt and financial assets: evidence from Germany, Great Britain and the USA. *J. R. Stat. Soc. Ser. A* 171 (3), 615–643.
- Brown, S., Garino, G., Taylor, K., Wheatley Price, S., 2005. Debt and financial expectations: an individual and household level analysis. *Econ. Inq.* 43 (1), 100–120.
- Brown, S., Garino, G., Taylor, K., 2008. Mortgages and financial expectations: a household level analysis. *South. Econ. J.* 74 (3), 857–878.
- Brown, S., Garino, G., Taylor, K., 2013. Household debt and attitudes toward risk. *Rev. Income Wealth* 59 (2), 283–304.
- Browning, M., Lusardi, A., 1996. Household saving: micro theories and micro facts. *J. Econ. Lit.* 34 (4), 1797–1855.
- Celeux, G., Forbes, F., Robert, C.P., Titterton, D.M., 2006. Deviance information criteria for missing data models. *Bayesian Anal.* 1 (4), 651–674.
- Chen, M.-H., Shao, Q., Ibrahim, J.G., 2000. *Monte Carlo Methods in Bayesian Computation*. Springer-Verlag Inc., Berlin, New York.
- Cox, P., Whitley, J., Brierley, P., 2002. Financial pressures in the UK household sector: evidence from the British Household Survey. *Bank of Engl. Q. Bull.* 410–419 (Winter).
- Crook, J., 2001. The demand for household debt in the USA: evidence from the 1995 Survey of Consumer Finance. *Appl. Financ. Econ.* 11, 83–91.
- Daniels, M.J., Pourahmadi, M., 2009. Modeling covariance matrices via partial autocorrelations. *J. Multivar. Anal.* 100, 2352–2363.
- Diggle, P., Heagerty, P., Liang, K.Y., Zeger, S., 2002. *Analysis of Longitudinal Data*. Oxford University Press, New York.
- Donkers, B., Van Soest, A., 1999. Subjective measures of household preferences and financial decisions. *J. Econ. Psychol.* 20, 613–642.
- Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., 2004. Least angle regression. *Ann. Stat.* 407–499.
- Geisser, S., Eddy, W.F., 1979. A predictive approach to model selection. *J. Am. Stat. Assoc.* 74, 153–160.
- Gelfand, A.E., Smith, M., 1990. Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* 398–409.
- Gelfand, A.E., Dey, D.K., Chang, H., 1992. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), *Model determination using predictive distributions with implementation via sampling-based methods*. Bayesian Stat. vol. 4. Oxford University Press.
- Gelman, A., Carlin, J., Stern, Dunson, H.D., Vehtari, A., Rubin, D., 2013. *Bayesian Data Analysis*. Third Edition. Chapman & Hall/CRC Texts in Statistical Science.
- Ghosh, P., Gönen, M., 2008. Bayesian modeling of multivariate average bioequivalence. *Stat. Med.* 2402–2419.
- Gilks, W.R., 1992. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), *Derivative-free adaptive rejection sampling for Gibbs sampling*. Bayesian Stat. vol. 4. Oxford University Press.
- Gilks, W.R., Wild, P., 1992. Adaptive rejection sampling for Gibbs sampling. *Appl. Stat.* 41 (2), 337–348.
- Glück, R., Lansing, K.J., 2009. U.S. Household Deleveraging and Future Consumption Growth. Federal Reserve Bank of San Francisco (FRBSF) Newsletter, Number 2009–16.
- Gropp, R., Scholz, J.K., White, M.J., 1997. Personal bankruptcy and credit supply and demand. *Q. J. Econ.* 112, 217–251.
- Guiso, L., Sapienza, P., Zingales, L., 2008. Trusting the Stock Market. *J. Financ.* 63, 2557–2600.
- Hochguertel, S., Alessie, R., Van Soest, A., 1997. Savings accounts versus stocks and bonds in a household portfolio allocation. *Scand. J. Econ.* 99 (1), 81–97.
- Hoel, P.G., Port, S.C., Stone, C.J., 1971. *Introduction to Probability Theory*. Houghton Mifflin, Boston.
- Hong, H., Kubik, J.D., Stein, J.C., 2004. Social interaction and stock-market participation. *J. Financ.* 59, 137–163.
- Jenkins, R., Bhugra, D., Bebbington, P., Brugha, T., Farrell, M., Coid, J., Fryers, T., Weich, S., Singleton, N., Meltzer, H., 2008. Debt, income and mental disorder in the general population. *Psychol. Med.* 38, 1485–1493.
- Joe, H., 1997. *Multivariate Models and Dependence Concepts*, Monographs on Statistics and Applied Probability. Chapman & Hall.
- Kimball, M.S., Sahm, C.R., Shapiro, M.D., 2008. Imputing risk tolerance from survey responses. *J. Am. Stat. Assoc.* 103, 1028–1103.

- Kimball, M.S., Sahm, C.R., Shapiro, M.D., 2009. Risk preferences in the PSID: individual imputations and family covariation. *Am. Econ. Rev. Pap. Proc.* 99, 363–368.
- Kullmann, C., Siegel, S., 2005. Real Estate and its Role in Household Portfolio Choice. EFA 2003 Annual Conference Paper No. 918. Sauder School of Business.
- Lin, L., Bandyopadhyay, D., Lipsitz, S.R., Sinha, D., 2010. Association models for clustered data with binary and continuous responses. *Biometrics* 66, 287–293.
- Lopes, H.F., Carvalho, Carlos M., 2013. Online Bayesian learning in dynamic models: an illustrative introduction to particle methods. *Hierarchical Models and Markov Chain Monte Carlo*.
- MacLehose, R.F., Dunson, D.B., 2010. Bayesian semi-parametric multiple shrinkage. *Biometrics* 66 (2), 455–462.
- Montagna, S., Tokdar, S.T., Neelon, B., Dunson, D.B., 2012. Bayesian latent factor regression for functional and longitudinal data. *Biometrics* 68, 1064–1073.
- Neal, R.M., 2000. Markov chain sampling methods for Dirichlet process mixture models. *J. Comput. Graph. Stat.* 9 (2), 249–265.
- Neelon, B.H., O'Malley, A.J., Normand, S.T., 2010. A Bayesian model for repeated measures zero-inflated count data with application to outpatient psychiatric service use. *Stat. Model.* 10, 421–439.
- Nelsen, R.B., 1999. *An Introduction to Copulas*, Lecture Notes in Statistics. Springer Verlag, New York.
- Olsen, M.K., Schafer, J.L., 2001. A two-part random-effects model for semi-continuous longitudinal data. *J. Am. Stat. Assoc.* 96, 730–745.
- Rosen, H.S., Wu, S., 2004. Portfolio choice and health status. *J. Financ. Econ.* 72, 457–484.
- Shum, P., Faig, M., 2006. What explains household stock holdings. *J. Bank. Financ.* 30 (9), 2579–2597.
- Spiegelhalter, D., Thomas, A., Best, N., Gilks, W., 1996. BUGS: Bayesian Inference using Gibbs Sampling, version 0.50. MRC Biostatistics Unit, Cambridge.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit (with discussion). *J. R. Stat. Soc. Ser. B* 64 (4), 583–639.
- Stroud, J., Johannes, M., 2014. Bayesian modeling and forecasting of 24-hour high-frequency volatility. *J. Am. Stat. Assoc.* 109, 1368–1384.
- Su, L., Tom, B.D., Farewell, V.T., 2015. A likelihood-based two-part marginal model for longitudinal semi-continuous data. *Stat. Methods Med. Res.* 24, 194–205.
- Tooze, J.A., Grunwald, G.K., Jones, R.H., 2002. Analysis of repeated measures data with clumping at zero. *Stat. Methods Med. Res.* 11 (4), 341–355.
- Wang, Z., Louis, T., 2003. Matching conditional and marginal shapes in binary mixed-effects models using a bridge distribution function. *Biometrika* 90, 765–775.
- Yen, S.T., Boxall, P.C., Adamowicz, W.L., 1997. An econometric analysis of donations for environmental conservation in Canada. *J. Agric. Resour. Econ.* 22, 246–263.