ESTIMATION OF DYNAMIC LINEAR MODELS IN SHORT PANELS WITH ORDINAL OBSERVATION

Stephen Pudney

THE INSTITUTE FOR FISCAL STUDIES
DEPARTMENT OF ECONOMICS, UCL
cemmap working paper CWP05/05

Estimation of dynamic linear models in short panels with ordinal observation

Stephen Pudney

Centre for Microdata Methods and Practice
Institute for Fiscal Studies

and

Institute for Social and Economic Research
University of Essex

September 2004, revised May 2005

Abstract: We develop a simulated ML method for short-panel estimation of one or more dynamic linear equations, where the dependent variables are only partially observed through ordinal scales. We argue that this latent autoregression (LAR) model is often more appropriate than the usual state-dependence (SD) probit model for attitudinal and interval variables. We propose a score test for assisting in the treatment of initial conditions and a new simulation approach to calculate the required partial derivative matrices. An illustrative application to a model of households' perceptions of their financial well-being demonstrates the superior fit of the LAR model.

Keywords: Dynamic panel data models, ordinal variables, simulated maximum likelihood, GHK simulator, BHPS.

JEL classifications: C23, C25, C33, C35, D84

Address for correspondence: Steve Pudney, ISER, University of Essex, Wivenhoe Park, Colchester CO4 3SQ. E-mail: spudney@essex.ac.uk.

I am grateful to Sarah Brown and Karl Taylor for allowing me to use their BHPS dataset and to them and to Manuel Arellano, Olympia Bover, Stephen Jenkins, Cheti Nicoletti and participants at the ISER JESS seminar and CIDE conference, Venice 2005, for helpful comments. This work was supported by the Economic and Social research Council through project "Fertility and poverty in developing countries" (award no. RES000230462) and the UK Longitudinal Studies Centre.

1 Introduction

In discrete data modelling there is an important distinction to be made between inherent and observational discreteness. *Inherent discreteness* refers to a case where the variables of interest are naturally discrete. For example, an individual is either employed or not employed; she has a university degree or not; she is married or not. *Observational discreteness* arises when the variables of interest are naturally continuous, but the survey instrument used to observe them imposes discreteness via a pre-specified ordinal scale of allowable responses. This applies to a wide range of attitudinal questions, which ask respondents to record their perceptions or beliefs on a Likert scale. Examples of econometric analysis of attitudinal variables have proliferated in recent years, with the development of the economic literature on happiness and satisfaction (see Van Praag and Ferrer-i-Carbonell, 2004, for a recent survey). There has been important work on econometric methodology in this area, particularly the choice between random and fixed effects modelling in panels (Ferrer-i-Carbonell and Frijters, 2004), but there has so far been little discussion of the dynamics of perceptions or of the most appropriate type of dynamic model to use.

Observational discreteness does not only arise with attitudinal data. It may also occur in survey questions about more 'objective' entities like income, when respondents are required to place themselves within one of a number of given income ranges. The discreteness here is an essentially artificial consequence of questionnaire design. For example, business surveys often ask about expectations of future sales or investment intentions. The respondent's expected value for sales or investment conditional on his information set is a continuous variable but the survey questions typically ask for a response graded as "up", "down" or "no change".

Most of the econometric literature dealing with discrete models for longitudinal data assumes inherent discreteness. The pioneering work of Heckman (1978, 1981a,b) centred on binary response models of the form:

$$\begin{aligned} y_{it}^* &= \alpha y_{it-1} + \boldsymbol{\beta}'\mathbf{x}_{it} + u_i + \varepsilon_{it} \\ y_{it} &= 1(y_{it}^* > 0) \end{aligned} \tag{1}$$

where 1(.) is the indicator function, $\mathbf{x}_{it}$ is a vector of strictly exogenous covariates, u_i is an unobserved individual effect uncorrelated with $\mathbf{x}_{it}$ and ε_{it} is a random residual uncorrelated across individuals and time. We refer to (1) as the state dependence (SD) model. It was developed primarily for applications in labour economics, where discreteness is inherent in the problem and where past outcomes of y_{it} represent state dependence. In these applications, the latent variable y^*_{it} is essentially an artificial construct and there is no reason why y^*_{it-1} should appear in the model (1).

However, attitudes, expectations and incomes are not inherently discrete and the use of models like (1), although common in the applied literature, is questionable.[1] If the discrete nature of y_{it} is only an artificial construct imposed by the questionnaire designer, then behaviour centres on the continuous variable y^*_{it}, rather than the observed indicator y_{it}. In these cases, y^*_{it-1} rather than y_{it-1}, should carry the dynamic feedback if the dynamic equation is to be a description of behaviour.

The paper has four main objectives. Firstly, (above and in section 2) we make the case for using dynamics in y^*_{it-1}, rather than y_{it-1}, in applications where the discreteness is observational rather than inherent; and then explore the interpretation of the model and its dynamic implications. The second objective, which is the subject of sections 3-4, is to consider identification and propose a practical method of estimation. The third aim is to set up a procedure for dealing with the initial conditions problem, using new specification tests which are proposed in section 5. Fourthly, we propose a new simulation method of estimating the cross-derivative matrices required for these tests; this is described in the appendix. Section 6 of the paper presents an illustrative application to a panel data model of individuals' financial expectations and section 7 concludes.

2 The model

2.1 The statistical structure

We work with a behavioural model specified in terms of the 'natural' continuous variables as follows:

$$y^*_{it} = \alpha \ y^*_{it-1} + \boldsymbol{\beta}'\mathbf{x}_{it} + u_i + \varepsilon_{it} \tag{2}$$

We refer to this as the Latent Autoregression (LAR) model. The vector $\mathbf{x}_{it}$ is assumed strictly exogenous and individuals are sampled independently from the underlying population. We make the standard assumption of Gaussian random effects so that the unobservables u_i and ε_{it} satisfy the following assumptions:

$$(u_i\ ,\ \varepsilon_{it}) \perp \mathbf{X}_i \qquad (3)$$

$$u_i \perp \varepsilon_{it} \qquad (4)$$

$$\varepsilon_{it} \perp \varepsilon_{is} \qquad \text{for every } s \neq t \qquad (5)$$

$$\begin{pmatrix} u_i \\ \varepsilon_i \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{bmatrix}\right) \qquad (6)$$

where $\mathbf{X}_i = (\mathbf{x}_{i0}, ..., \mathbf{x}_{iT})$. We only observe y_{it}^* according to a grading scale and thus:

$$y_{it} = r \qquad \text{iff} \quad y_{it}^* \in [\Gamma_{r-1}, \Gamma_r), \qquad r = 1...R \qquad (7)$$

where $\Gamma_0 = -\infty$ and $\Gamma_R = \infty$. Note that the thresholds Γ_r will be observable in the case of interval censoring (such as earnings models for grouped data) or specified as unknown parameters in the ordered probit case (such as Likert responses). In the latter case, the model is normalised by omitting the intercept from $\mathbf{x}_{it}$ and setting var(ε_{it}) = 1, which is equivalent to dividing y_{it}^*, y_{it-1}^*, $\boldsymbol{\beta}$, u_i and ε_i through by σ_ε in (2). Note that α is not affected by this normalisation.

2.2 Interpretation of parameters

There are two cases to consider. In models where the discreteness arises through interval censoring of a dynamic regression (such as an earnings model applied to grouped earnings data), the grading thresholds Γ_r are observed and thus the scale of y_{it}^* is determined by the model. Consequently, the coefficients $\boldsymbol{\beta}$ have the usual regression interpretation as the instantaneous response $\partial E(y_{it}^* \mid y_{it-1}^*, \mathbf{x}_{it}, u_i) / \partial \mathbf{x}_{it}$ and the long-run response is given by $\partial E(y_i^* \mid \bar{\mathbf{x}}_i, u_i) / \partial \mathbf{x}_i = \boldsymbol{\beta}/(1-\alpha)$ as usual, where $\bar{\mathbf{x}}_i$ is the long-run static value of $\mathbf{x}_{it}$ for individual *i*. Both of these responses are independent of the values taken by **x** and *u*.

The case of unobserved grading thresholds is less simple. Here we are dealing with variables like the expected inflation rate or the strength of a subjective response such as job satisfaction or happiness. In all such cases the scale of y_{it}^* is unobserved and we estimate $\boldsymbol{\beta}/\sigma_\varepsilon$ rather than $\boldsymbol{\beta}$. Consequently, the estimated coefficients are interpretable as $\partial E\left([y_{it}^* / \sigma_\varepsilon] \mid y_{it-1}^*, \mathbf{x}_{it}, u_i\right) / \partial \mathbf{x}_{it}$. In many applications ('happiness', for example) this problem is more fundamental than a lack of identification induced by

[1] Exceptions to this general neglect of models involving latent dynamics are papers by Arellano *et. al.* (1997) and Bover and Arellano (1997). However, the context and models considered in those studies is quite different from the case considered here, as is the approach to estimation.

imperfect observation: there is a lack of natural units for 'happiness' or 'utility' which renders the scale of β inherently ambiguous. In some cases, where there are natural units of measurement for y^*, we can fix σ_ε at a reasonable hypothetical value. For example, for an analysis of survey responses to a question about expected inflation we might reasonably set σ_ε at (say) half a percentage point to allow a rough but direct interpretation of the model in terms of the natural units. Note that α is identifiable independently of σ_ε. As a consequence, we can estimate unambiguously the speed of adjustment. For example, following a shock, the proportion of disequilibrium which is eliminated within s periods is $1-\alpha^s$ and this is unaffected by normalisation.

2.3 Dynamics

The SD and LAR processes (1) and (2) imply different patterns of dynamic behaviour. Consider the following artificial example:

$$\textit{SD model:} \quad y_t^* = 0.8 y_{t-1} + x + \varepsilon_t \tag{8}$$

$$\textit{LAR model:} \quad y_t^* = 0.422 y_{t-1}^* + 0.355 + 0.770x + \varepsilon_t \tag{9}$$

where $x = 0.5$, $\varepsilon_t \sim N(0,1)$ and $y_t = \mathbf{1}(y_t^* > 0)$. The parameters of the LAR process (9) have been chosen to reproduce exactly three properties of the SD process (8): [2]

(i) $\Pr(y = 1) = 0.877$;

(ii) $\partial \Pr(y = 1 \mid x)/\partial x = 0.246$;

(iii) $\Pr(y_t \neq y_{t-1}) = 0.170$.

With the LAR parameters chosen in this way, the distributions of run lengths in states 0 and 1 are identical for the two processes. However, the relationship between successive run lengths is not. This is reflected in the autocorrelation functions (Figure 1). As we would expect, the LAR model has much higher autocorrelations than the SD model for y_t^*. For the observed y_t, the ACF decays faster for the SD than the LAR process, despite the fact that they have the same 1st-order autocorrelation by construction. Thus, an LAR model will display greater persistence than an observationally similar SD model, in this quite subtle sense.

[2] Conditions (i) and (ii) are imposed analytically to determine β_0 and β_1 for given α; Monte Carlo simulation was then used to find the value of α to equalise the 1st order autocorrelations.

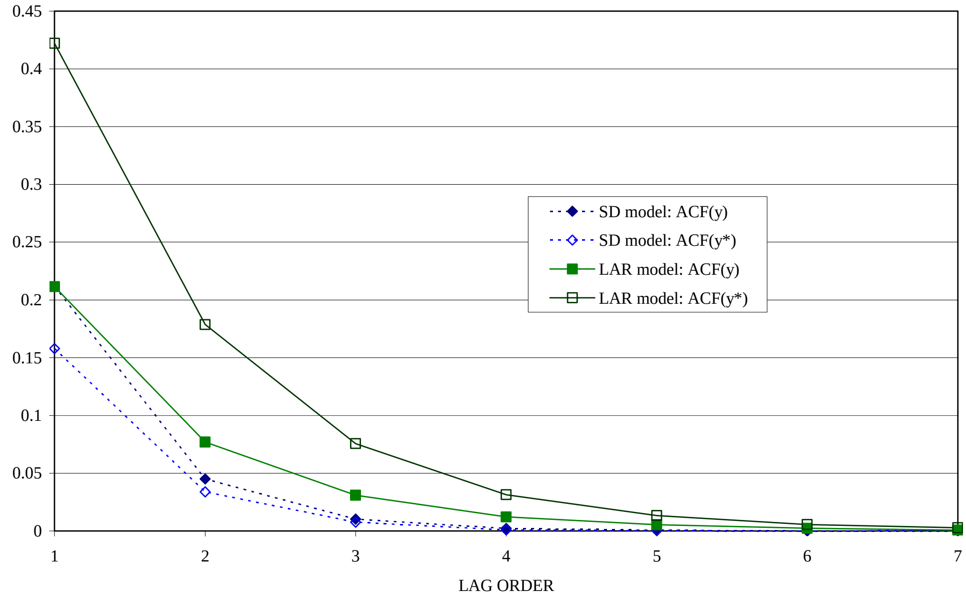

Figure 1 ACFs for the SD and LAR models

The two models also differ in terms of the implied dynamic multiplier effects of $\mathbf{x}$ on y. To illustrate this, consider again the binary case and focus on two important features: the impact on $\Pr(y_{it}=1 \mid y_{it-1}, \mathbf{X}_i, u_i)$ of switching the conditioning event from $y_{it-1} = 0$ to $y_{it-1} = 1$; and the impact of the history of $\{\mathbf{x}_{it}\}$ on the probability of a positive response, without conditioning on y_{it-1}.

For the former, the SD model is relatively simple:

$$\Pr(y_{it}=1 \mid y_{it-1}=1, \mathbf{X}_i, u_i) - \Pr(y_{it}=1 \mid y_{it-1}=0, \mathbf{X}_i, u_i) = \Phi(\alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + u_i) - \Phi(\boldsymbol{\beta}'\mathbf{x}_{it} + u_i) \quad (10)$$

where $\Phi(.)$ is the cdf of the N(0,1) distribution. For the LAR model, we have instead:

$$\Pr(y_{it}=1 \mid y_{it-1}=1, \mathbf{X}_i, u_i) - \Pr(y_{it}=1 \mid y_{it-1}=0, \mathbf{X}_i, u_i)$$

$$= \frac{\Pr(y_{it}=1, y_{it-1}=1 \mid \mathbf{X}_i, u_i)}{\Pr(y_{it-1}=1 \mid \mathbf{X}_i, u_i)} - \frac{\Pr(y_{it}=1 \mid \mathbf{X}_i, u_i) - \Pr(y_{it}=1, y_{it-1}=1 \mid \mathbf{X}_i, u_i)}{1 - \Pr(y_{it-1}=1 \mid \mathbf{X}_i, u_i)}$$

$$= \frac{\Pr(y_{it}=1, y_{it-1}=1 \mid \mathbf{X}_i, u_i) - \Pr(y_{it}=1 \mid \mathbf{X}_i, u_i)\Pr(y_{it-1}=1 \mid \mathbf{X}_i, u_i)}{\Pr(y_{it-1}=1 \mid \mathbf{X}_i, u_i)[1 - \Pr(y_{it-1}=1 \mid \mathbf{X}_i, u_i)]} \quad (11)$$

Assume the process (2) is stable and long-established. Then:

$$y_{it}^{*} = \sum_{s=0}^{\infty} \alpha^{s} \boldsymbol{\beta}'\mathbf{x}_{it-s} + \frac{u_i}{1-\alpha} + \sum_{s=0}^{\infty} \alpha^{s} \varepsilon_{it-s} \quad (12)$$

and therefore $\Pr(y_{it}=1, y_{it-1}=1 \mid \mathbf{X}_i, u_i) = \Phi^*(\mu_{it}, \mu_{it-1}; \alpha)$ and $\Pr(y_{it}=1 \mid \mathbf{X}_i, u_i) = \Phi(\mu_{it})$, where $\Phi^*(.,.;\alpha)$ is the bivariate standard normal cdf with correlation α and μ_{it} is the scaled conditional mean $(1-\alpha^2)^{1/2}[\sum_s \alpha^s \boldsymbol{\beta}'\mathbf{x}_{it-s}+u_i/(1-\alpha)]$. Thus:

$$\Pr(y_{it}=1 \mid y_{it-1}=1, \mathbf{X}_i, u_i) - \Pr(y_{it}=1 \mid y_{it-1}=0, \mathbf{X}_i, u_i) = \frac{\Phi(\mu_{it}, \mu_{it-1}; \alpha) - \Phi(\mu_{it})\Phi(\mu_{it-1})}{\Phi(\mu_{it-1})[1-\Phi(\mu_{it-1})]} \tag{13}$$

The important difference between (10) and (13) is that the former depends only on the current vector $\mathbf{x}_{it}$, whereas the latter depends on the entire history of $\mathbf{x}_{it}$.

Consider now the alternative summary measure, $\Pr(y_{it}=1 \mid \mathbf{X}_i, u_i)$. The LAR process gives a relatively simple form:

$$\Pr(y_{it}=1 \mid \mathbf{X}_i, u_i) = \Phi(\mu_{it}) \tag{14}$$

implying that the lagged marginal response decays geometrically:

$$\frac{\partial \Pr(y_{it}=1 \mid \mathbf{X}_i, u_i)}{\partial \mathbf{x}_{it-s}} = \phi(\mu_{it})\alpha^s\sqrt{1-\alpha^2}\boldsymbol{\beta} \tag{15}$$

where $\phi(.)$ is the standard normal pdf.

For the state-dependence model, we can write:

$$\begin{aligned}\Pr(y_{it}=1 \mid \mathbf{X}_i, u_i) &= \Pr(y_{it}=1 \mid y_{it-1}=0, \mathbf{X}_i, u_i)\Pr(y_{it-1}=0 \mid \mathbf{X}_i, u_i) \\ &+ \Pr(y_{it}=1 \mid y_{it-1}=1, \mathbf{X}_i, u_i)\Pr(y_{it-1}=1 \mid \mathbf{X}_i, u_i)\end{aligned} \tag{16}$$

Rearrange and write this as a recursion:

$$P_{it} = P_{it-1}\delta_{it} + \rho_{it} \tag{17}$$

where:

$P_{it} = \Pr(y_{it}=1 \mid \mathbf{X}_i, u_i)$

$\delta_{it} = \Phi(\alpha+\boldsymbol{\beta}'\mathbf{x}_{it}+u_i) - \Phi(\boldsymbol{\beta}'\mathbf{x}_{it}+u_i)$

$\rho_{it} = \Phi(\boldsymbol{\beta}'\mathbf{x}_{it}+u_i)$.

Solving back to an arbitrary period 0:

$$P_{it} = P_{i0}\prod_{j=0}^{t-1}\delta_{it-j} + \sum_{s=0}^{t-1}\rho_{it-s}\prod_{j=0}^{s-1}\delta_{it-j} \tag{18}$$

where we use the convention $\prod_{j=0}^{j=-1}\delta_{it-j} \equiv 1$. On reasonable assumptions about the **x**-process, solving back indefinitely leads to the following representation:

$$P_{it} = \sum_{s=0}^{\infty}\rho_{it-s}\prod_{j=0}^{s-1}\delta_{it-j} \tag{19}$$

Thus:

$$\frac{\partial \Pr(y_{it}=1|\mathbf{X}_i,u_i)}{\partial \mathbf{x}_{it-s}} = \frac{\partial \rho_{it-s}}{\partial \mathbf{x}_{it-s}}\prod_{j=0}^{s-1}\delta_{it-j} + \sum_{k=s+1}^{\infty}\frac{\rho_{it-k}}{\delta_{it-s}}\prod_{j=0}^{k-1}\delta_{it-j}\frac{\partial \delta_{it-s}}{\partial \mathbf{x}_{it-s}}$$

$$= \left(\prod_{j=0}^{s-1}\delta_{it-j}\right)\phi(\boldsymbol{\beta}'\mathbf{x}_{it-s}+u_i)\boldsymbol{\beta}$$

$$+\sum_{k=s+1}^{\infty}\left(\frac{\rho_{it-k}}{\delta_{it-s}}\prod_{j=0}^{k-1}\delta_{it-j}\right)\left[\phi(\alpha+\boldsymbol{\beta}'\mathbf{x}_{it-s}+u_i)-\phi(\boldsymbol{\beta}'\mathbf{x}_{it-s}+u_i)\right]\boldsymbol{\beta} \tag{20}$$

The profile of $\partial \Pr(y_{it}=1 \mid \mathbf{X}_i, u_i)/ \partial \mathbf{x}_{it-s}$ is thus considerably more complicated than the geometric decay implied by the SD model (1).

3 Estimation

3.1 Initial conditions

In the SD model, there are two alternative approaches for dealing with the random effects u. Heckman (1981b) specifies an approximation to the distribution of $y_{i0} \mid \mathbf{X}_i$, u_i, and then derives the distribution of y_{i1} ... $y_{iT} \mid y_{i0}$, $\mathbf{X}_i$, u_i using sequential conditioning. The random effects are then integrated out by numerical quadrature. The alternative approach, used by Wooldridge (2000) is to specify instead the distribution of $u_i \mid y_{i0}$, $\mathbf{X}_i$. A semi-parametric variant due to Arellano and Carrasco (2003) involves the sequence of conditional means $\lambda_{it} = E(u_i \mid y_{i0}...y_{it}, \mathbf{x}_{i0}...\mathbf{x}_{it})$, which are estimated as nuisance parameters. The latter approach has many advantages in models like (1) but is problematic in LAR models, where the lagged dependent variable is not observable and cannot be conditioned on. Conditioning on its observable counterpart complicates matters enormously. For this reason, we use the Heckman treatment of initial conditions, together with an explicit hypothesis testing procedure to control the bias induced by approximation error in the assumed distribution of $\mathbf{y}_{i0} \mid \mathbf{X}_i$, u_i.

Assume that we observe y and $\mathbf{x}$ over a period $t = 0 \ldots T$. The LAR process (2) implies the following distributed lag representation:[3]

[3] In the case where the Γ_r are not observable, we impose the normalisation $\sigma_\varepsilon = 1$ and henceforth y^*_{it}, $\boldsymbol{\beta}$ and σ_u are re-interpreted accordingly.

$$y_{it}^* = \alpha^t y_{i0}^* + \sum_{s=0}^{t-1} \alpha^s \boldsymbol{\beta}' \mathbf{x}_{it-s} + \frac{1-\alpha^t}{1-\alpha} u_i + \sum_{s=0}^{t-1} \alpha^s \varepsilon_{it-s} \tag{21}$$

This is a useful basis for estimation if either t is sufficiently large and α^t decays sufficiently rapidly with t or if we can find a good empirical approximation for y_{i0}^*.

Write this approximation to $y_{i0}^* \mid \mathbf{X}_i, u_i$ as:

$$y_{i0}^* = \boldsymbol{\delta}' \mathbf{w}_i + \gamma\, u_i + \eta_i \tag{22}$$

$$y_{i0} = r \qquad \text{iff} \quad y_{i0}^* \in [\Gamma_{r-1}^0, \Gamma_r^0), \qquad r = 1 \ldots R \tag{23}$$

where $\mathbf{w}_i$ is a vector constructed from $\mathbf{X}_i$; δ and γ are parameters and, in the ordered probit case, Γ_r^0 may differ from Γ_r. The random term η_i satisfies the following assumptions:

$$\eta_i \perp u_i \mid \mathbf{X}_i \tag{24}$$

$$\eta_i \perp \varepsilon_{it} \mid \mathbf{X}_i \qquad \text{for every } t > 0 \tag{25}$$

$$u_i \mid \mathbf{X}_i \sim \mathrm{N}(0, \sigma_\eta^2) \tag{26}$$

Note that, even in the ordered probit case, η is not normalised to have unit variance.

In principle, the vector $\mathbf{w}_i$ may contain all distinct elements of $\{\mathbf{x}_{i0}, \mathbf{X}_i\}$. However, in practice it may be found that $\mathbf{w}_i = \mathbf{x}_{i0}$ is adequate, or that limited summaries, such as $\mathbf{w}_i = \{\mathbf{x}_{i0}, T^{-1}\sum_1^T \mathbf{x}_{it}\}$, work well. This is essentially an empirical issue.

With approximation (22)-(23), equation (21) becomes:

$$y_{it}^* = \alpha^t \boldsymbol{\delta}' \mathbf{w}_i + \sum_{s=0}^{t-1} \alpha^s \boldsymbol{\beta}' \mathbf{x}_{it-s} + c_t u_i + \sum_{s=0}^{t-1} \alpha^s \varepsilon_{it-s} + \alpha^t \eta_i \tag{27}$$

where $c_t = (1 - \alpha^t)/ (1 - \alpha) + \alpha^t \gamma$.

The model now consists of equation (22) and a set of equations (27) for any collection of periods $t > 0$. In practice, the initial conditions model (22) is only an approximation and is a potential source of specification error. However, if $|\alpha| < 1$ so that $\alpha^t \to 0$ as $t \to \infty$, then the influence of the initial conditions declines as we consider later periods. There is, therefore, a case for leaving a gap (of S periods) between the initial period 0 and the subsequent periods used to estimate the LAR model. Consequently, we work with a system of (T-S+1) equations consisting of (22) and (27) for $t = S+1 \ldots T$. Data on $\{y_{i1} \ldots y_{iS}\}$ are not used. The choice of S involves a trade-off between possible misspecification bias and efficiency, since increasing S

reduces both the influence of initial conditions and the amount of data used for estimation. Increasing S also reduces the scale of the computational problem. This system is nonlinear in its parameters $\theta = \{\alpha, \beta, \delta, \gamma, \sigma_u, \sigma_\varepsilon, \sigma_\eta\}$, where $\sigma_\varepsilon = 1$ in the case of observable interval boundaries.

3.2 Identification

Consider the model with unobserved grading thresholds. Partition the covariates into a common set of time-invariant variables ζ_i and a sequence of time-varying covariates ξ_{it}, so that $\mathbf{x}_{it} = (\zeta_i, \xi_{it})$. Assume a full specification of the initial condition (9), so that $\mathbf{w}_i = (\zeta_i, \xi_{i1} \ldots \xi_{iT})$. Make the further assumption that the matrix $\text{plim}(n^{-1}\sum \mathbf{w}_i \mathbf{w}_i')$ is positive definite. An ordered probit model for y_{i0} on $\mathbf{w}_i$ will consistently estimate the normed coefficient vector δ/v_0, where $v_0^2 = \sigma_\eta^2 + \gamma^2\sigma_u^2$.

Consider equation (27), for any period, $t > 0$. Rewrite it in standardised form:

$$\left(\frac{y_{it}^*}{v_t}\right) = \left(\frac{\alpha^t v_0}{v_t}\right)\omega_i + \left(\frac{(1-\alpha^t)\boldsymbol{\beta}_\zeta}{(1-\alpha)v_t}\right)'\boldsymbol{\zeta}_i + \left(\frac{\boldsymbol{\beta}_\xi}{v_t}\right)'\boldsymbol{\xi}_{it} + \left(\frac{\alpha\boldsymbol{\beta}_\xi}{v_t}\right)'\boldsymbol{\xi}_{it-1} + \ldots + \left(\frac{\alpha^{t-1}\boldsymbol{\beta}_\xi}{v_t}\right)'\boldsymbol{\xi}_{i1}$$
$$+ \left[c_t u_i + \sum_{s=0}^{t-1}\alpha^s \varepsilon_{it-s} + \alpha^t \eta_i\right] / v_t \tag{28}$$

where $\beta' = (\beta_\zeta', \beta_\xi')$, $v_t^2 = c_t^2\sigma_u^2 + (1-\alpha^{2t})/(1-\alpha^2) + \alpha^{2t}\sigma_\eta^2$ and ω_i is the variable $\delta'\mathbf{w}_i/v_0$ which can be constructed from the coefficients of the initial conditions model (22). Rewrite (28) in simplified notation as:

$$y_{it}^* / v_t = a_t\omega_i + \mathbf{b}_t'\boldsymbol{\zeta}_i + \mathbf{d}_{0t}'\boldsymbol{\xi}_{it} + \mathbf{d}_{1t}'\boldsymbol{\xi}_{it-1} + \ldots + \mathbf{d}_{t-1,t}'\boldsymbol{\xi}_{i1} + \upsilon_{it} \tag{29}$$

Note that the covariates $(\omega_i, \zeta_i, \xi_{i1} \ldots \xi_{it})$ are (asymptotically) non-collinear. Thus, ordered probit estimation of (29) will generate consistent estimates of the scaled coefficients $(a_t, \boldsymbol{b}_t, \boldsymbol{d}_{0t}, \ldots, \boldsymbol{d}_{t-1,t})$. Identification then proceeds as follows. First, the value of α can be constructed as any element of any of the vectors of ratios $\boldsymbol{d}_{st}/\boldsymbol{d}_{s-1,t}$. If α is zero, the model becomes a static random effects ordered probit, so there is no new identification issue; we consider the case $\alpha \neq 0$ henceforth. With α known, β can be inferred up to scale as $\mathbf{g}/\|\mathbf{g}\|$ where $\boldsymbol{g} = [\boldsymbol{b}_t(1-\alpha)/(1-\alpha^t), \ \boldsymbol{d}_{0t}]$. Thus, the key behavioural parameters α and the direction of the vector β are essentially identifiable from only two waves of the panel.

The ratio, R_t, of a_t to α^t gives the value v_0/v_t, thus:

$$R_t\, v_t \;=\, v_0 \tag{30}$$

The correlation between the random errors in equations (22) and (27), which can be estimated consistently by joint estimation or from the generalised residuals, is ρ_{0t} satisfying the following:

$$\rho_{0t} v_0 \;\; v_t \;= c_t\, \gamma \;\; \sigma_u^2 + \alpha^t \sigma_\eta^2 \tag{31}$$

Equations (30) and (31) are clearly insufficient to determine the three remaining unknowns, γ, σ_u^2 and σ_η^2, so full identification requires at least two waves of data, in addition to wave 0.

Consider the 3-wave case, where we have data for t = 0, 1, 2. Calculate each of the ratios $(v_t/v_0)^2$ as α^{2t}/a_t^2. Using the definition (31), after some manipulation the quantity $\gamma\, \sigma_u^2 / v_0^2$ can be expressed as:

$$\frac{\gamma \;\, \sigma_u^2}{v_0^2} = \alpha^2 \left[\frac{\rho_{02}}{a_2} - \frac{\rho_{01}}{a_1} \right] \tag{32}$$

Note that $a_t \neq 0$ for $\alpha \neq 0$, so (32) is well-defined. Now express v_t^2 as $(A_t + \alpha^t\gamma)^2\sigma_u^2 + B_t + \alpha^{2t}\sigma_\eta^2$, where $A_t = (1-\alpha^t)/(1-\alpha)$ and $B_t = (1-\alpha^{2t})/(1-\alpha^2)$. Thus:

$$\left(\frac{v_t}{v_0}\right)^2 = A_t^2\left(\frac{\sigma_u^2}{v_0^2}\right) + 2A_t\alpha^t\left(\frac{\gamma\sigma_u^2}{v_0^2}\right) + \alpha^{2t}\left(\frac{\gamma^2\sigma_u^2 + \sigma_\eta^2}{v_0^2}\right) + B_t\left(\frac{1}{v_0^2}\right) \tag{33}$$

We know the value of $\gamma\, \sigma_u^2 / v_0^2$ from (32) and we know *a priori* that $(\gamma^2\sigma_u^2 + \sigma_\eta^2)/v_0^2$ is equal to 1. This gives the following pair of equations with known right-hand sides:

$$A_t^2\left(\frac{\sigma_u^2}{v_0^2}\right) + B_t\left(\frac{1}{v_0^2}\right) = \left(\frac{v_t}{v_0}\right)^2 - 2A_t\alpha^t\left(\frac{\gamma\sigma_u^2}{v_0^2}\right) - \alpha^{2t}, \quad t = 1, 2 \tag{34}$$

Note that the matrix $\begin{pmatrix} A_1^2 & B_1 \\ A_2^2 & B_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ (1+\alpha)^2 & (1+\alpha^2) \end{pmatrix}$ is non-singular for all $\alpha \neq 0$, so there is a unique solution for (σ_u^2/v_0^2) and $(1/v_0^2)$. From these, σ_u^2 and v_0^2 are determined. The value of γ is then given by (32) and σ_η^2 by $v_0^2 - \gamma^2\sigma_u^2$, so all parameters are identified.

3.3 SML estimation

This identification argument does not lead to an efficient estimator, since it does not impose all the restrictions on the coefficients (a_t, $\boldsymbol{b}_t$, $\boldsymbol{d}_{0t}$, ..., $\boldsymbol{d}_{t-1,t}$) in (29), nor does it exploit the relationship between the residual correlation ρ_{12} and the model parameters. Instead we use a simulated ML procedure. Let the observed outcome for y_{it} be r_{it}, implying $y^*_{it} \in [\Gamma_{r_{it}-1}, \Gamma_{r_{it}})$. The likelihood for this set of events is:

$$\Pr\left(y_{i0} = r_{i0},\ y_{iS+1} = r_{iS+1},..., y_{iT} = r_{iT} \mid \mathbf{X}_i\right) = \Pr\left(\upsilon_{i0} \in A_{i0}, \upsilon_{iS+1} \in A_{iS+1},..., \upsilon_{iT} \in A_{iT}\right) \tag{35}$$

where $\upsilon_{it} = c_t u_i + \sum_0^{t-1} \alpha^s \varepsilon_{it-s} + \alpha^t \eta_i$, $\mu_{it} = \alpha^t \boldsymbol{\delta}'\mathbf{w}_i + \sum_0^{t-1} \alpha^s \boldsymbol{\beta}'\mathbf{x}_{it-s}$ and A_{it} is the interval $[\Gamma_{r_{it}} - \mu_{it}, \Gamma_{r_{it}+1} - \mu_{it})$. The residual vector υ_i = (υ_{i0}, υ_{iS+1} … υ_{iT}) has a covariance matrix with elements:

$$\omega_{00} = \gamma^2 \sigma_u^2 + \sigma_\eta^2 \tag{36}$$

$$\omega_{0t} = \gamma\ c_t \sigma_u^2 + \sigma_\eta^2 \alpha^t, \qquad S < t \le T \tag{37}$$

$$\omega_{st} = c_s c_t \sigma_u^2 + \sum_{p=0}^{\min(s,t)-1} \alpha^{s+t-2p} \sigma_\varepsilon^2 + \alpha^{s+t} \sigma_\eta^2, \qquad S < (s,\, t) \le T \tag{38}$$

The probability (35) is a (T-S+1)-dimensional rectangle probability. Under normality, probabilities of this kind can be calculated using the GHK simulator (Hajivassiliou and Ruud, 1994), with antithetic acceleration used to improve simulation precision. We construct the following simulated log-likelihood function:

$$\ln \hat{L}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ln \hat{P}_i(\boldsymbol{\theta}) \tag{39}$$

where $\hat{P}_i(\boldsymbol{\theta})$ is the predicted probability (35) for individual i, estimated using the GHK algorithm. The simulated likelihood is maximised numerically with respect to $\boldsymbol{\theta}$.

4 The extension to higher-order and multi-equation models

Most applications of the method proposed here will be to single-equation models. However, there is no difficulty in the generalisation to a general J-dimensional system of the reduced-form equations[4]:

$$y^*_{jit} = \sum_{k=1}^{J} \alpha_{jk} y^*_{kit-1} + \boldsymbol{\beta}_j'\mathbf{x}_{it} + u_{ji} + \varepsilon_{jit}\ , \quad j = 1...J \tag{40}$$

One important way in which multi-equation systems may arise is through higher-order lags. Consider the model:

[4] Note that the case of a structural form with contemporaneous feedback can be put in the reduced form (38) in the usual way and then estimated subject to the nonlinear structural restrictions on the reduced form coefficients α_{jk} and $\boldsymbol{\beta}_j$.

$$y_{it}^* = \alpha_1 y_{it-1}^* + \alpha_2 y_{it-2}^* + \boldsymbol{\beta}'\mathbf{x}_{it} + u_i + \varepsilon_{it} \tag{41}$$

This is equivalent to the 2-equation system:

$$\begin{aligned} y_{1it}^* &= \alpha_1 y_{1it-1}^* + \alpha_2 y_{2it-1}^* + \boldsymbol{\beta}'\mathbf{x}_{it} + u_i + \varepsilon_{it} \\ y_{2it}^* &= y_{1it-1}^* \end{aligned} \tag{42}$$

System (42) is a special case of (40), with one nonstochastic equation and therefore a singular error covariance matrix. We return to this example below.

In matrix notation, the general system (40) becomes:

$$\mathbf{y}_{it}^* = \mathbf{A}\mathbf{y}_{it-1}^* + \mathbf{B}\mathbf{x}_{it} + \mathbf{u}_i + \boldsymbol{\varepsilon}_{it} \tag{43}$$

where $\mathbf{y}_{it}^* = (y_{1it}^* \dots y_{Jit}^*)'$. The coefficient matrices are $\mathbf{A} = \{\alpha_{jk}\}$ and $\mathbf{B} = (\beta_1 \dots \beta_J)'$. The approximation to the initial values distribution is generalised to:

$$\mathbf{y}_{i0}^* = \mathbf{D}\mathbf{w}_i + \mathbf{G}\mathbf{u}_i + \boldsymbol{\eta}_i \tag{44}$$

where **D** and **G** are coefficient matrices. The corresponding grading thresholds are Γ_{jr} and Γ_{jr}^0, $j = 1 \dots J$, $r = 0 \dots R_j$.

The independence assumptions (3), (4), (24) and (25) are extended to the vector case and we assume:

$$\mathbf{u}_i \mid \mathbf{X}_i \sim \mathrm{N}(\mathbf{0}, \Sigma_u) \tag{45}$$

$$\boldsymbol{\varepsilon}_{it} \mid \mathbf{X}_i \sim \mathrm{N}(\mathbf{0}, \Sigma_\varepsilon) \quad \text{for every } t \tag{46}$$

$$\boldsymbol{\eta}_i \mid \mathbf{X}_i \sim \mathrm{N}(\mathbf{0}, \Sigma_\eta) \tag{47}$$

The jth diagonal element of Σ_ε is normalised to unity if y_{jit} has unobservable thresholds.

The analogue of (27) is:

$$\mathbf{y}_{it}^* = \mathbf{A}^t\mathbf{D}\mathbf{w}_i + \sum_{s=0}^{t-1}\mathbf{A}^s\mathbf{B}\mathbf{x}_{it-s} + \mathbf{C}_t\mathbf{u}_i + \sum_{s=0}^{t-1}\mathbf{A}^s\varepsilon_{it-s} + \mathbf{A}^t\boldsymbol{\eta}_i \tag{48}$$

where $\mathbf{C}_t = (\mathbf{I} - \mathbf{A})^{-1}(\mathbf{I} - \mathbf{A}^t) + \mathbf{A}^t\mathbf{G}$. Let the observed outcome for y_{jit} be r_{jit}, implying $y_{jit}^* \in [\Gamma_{j,r_{jit}-1}, \Gamma_{j,r_{jit}})$. The likelihood for this set of events is:

$$\begin{aligned} &\Pr\left(y_{jit} = r_{jit} \text{ for } j = 1 \dots J, t \in \mathrm{T} \mid \mathbf{w}_i, \mathbf{x}_{i1}, \dots, \mathbf{x}_{iT}\right) = \\ &\qquad \Pr\left(v_{jit} \in [\Gamma_{j,r_{jit}-1} - \mu_{jit}, \Gamma_{j,r_{jit}} - \mu_{jit}) \text{ for } j = 1 \dots J, t \in \mathrm{T}\right) \end{aligned} \tag{49}$$

where T is the index set $\{0, S+1 \dots T\}$, v_{jit} and μ_{jit} are the jth elements of the vectors $\mathbf{v}_{it} = \mathbf{C}_t\mathbf{u}_i + \sum\mathbf{A}^s\varepsilon_{it-s} + \mathbf{A}^t\boldsymbol{\eta}_i$ and $\boldsymbol{\mu}_{it} = \mathbf{A}^t\mathbf{D}\mathbf{w}_i + \sum\mathbf{A}^s\mathbf{B}\mathbf{x}_{it-s}$ respectively. The covariance matrix of the residual vector $\mathbf{v}_i = (\mathbf{v}_{i0}, \mathbf{v}_{iS+1} \ldots \mathbf{v}_{iT})$ has a block structure, where blocks (0, 0), (0, t) and (s, t) are respectively:

$$\mathbf{\Omega}_{00} = \mathbf{G\Sigma}_u\mathbf{G}' + \mathbf{\Sigma}_\eta \tag{50}$$

$$\mathbf{\Omega}_{0t} = \mathbf{G\Sigma}_u\mathbf{C}_t' + \mathbf{\Sigma}_\eta(\mathbf{A}^t)' \tag{51}$$

$$\mathbf{\Omega}_{st} = \mathbf{C}_s\mathbf{\Sigma}_u\mathbf{C}_t' + \sum_{p=0}^{\min(s,t)-1}\mathbf{A}^{s-p}\mathbf{\Sigma}_\varepsilon(\mathbf{A}^{t-p})' + \mathbf{A}^s\mathbf{\Sigma}_\eta(\mathbf{A}^t)', \quad S < (s, t) \le T \tag{52}$$

The probability (49) is a $J(T\text{-}S\text{+}1)$-dimensional rectangle probability, that can again be approximated by the GHK simulator in moderately-sized systems.

In the special case where multiple equations have arisen from an original model with dynamics of order higher than 1 there are redundancies among the set of inequalities defining the probability (49). For example, in the model (40), the event $y_{1it} = r_{1it}$ implies the event $y_{2it} = r_{1it\text{-}1} \equiv r_{2it}$ with probability one. The T-S+1 redundancies of this kind halves the dimensionality of the probability (49).

5 Specification tests

How do we choose the number of panel waves, S, to skip? Considerations of estimation efficiency suggest a small value for S, while worries about misspecification bias introduced by the initial condition approximation suggests a large value. To resolve this issue, I suggest use of a test which examines the consistency of estimates based on the waves S+1 ... T with the observed outcomes in wave S. If no significant conflict is found for wave S, we then reduce the skip rate from S to S-1 waves and re-estimate to improve efficiency. This can be done sequentially until a satisfactory point on the bias-efficiency tradeoff is reached. We consider an approach based on the score vector for wave S-1.[5]

This approach allows wave-specific parameters such as time dummies. Write the log-likelihood function based on y-data for waves 0, S+1 ... T as $L(\psi, \tau_1)$, where ψ is the subvector of θ which is common to all waves and τ_1 is the vector of any further parameters identifiable from the estimation sample (typically time dummies for periods S+1 ... T-1). Let $L^*(\psi, \tau_2)$ be the log-likelihood for an estimation sample

[5] Another possibility is a Hausman parameter contrast test, comparing the parameter estimates resulting from skipping S and S-1 waves. In practice, this often encounters problems arising from non-positive-definiteness of the estimated variance matrix of the contrast vector. The Hausman test also requires two major estimation steps.

covering only waves 0 and S. The vector τ_2 will usually contain only the coefficient of a dummy for period S, which is essentially an intercept term. Now maximise $L^*(\hat{\boldsymbol{\psi}}, \boldsymbol{\tau}_2)$ with respect to the unknown wave-S parameter to give $\tilde{\boldsymbol{\tau}}_2$. This is a low-dimensional (usually scalar) optimisation and relatively easy to perform. Expanding the first-order condition for $\tilde{\boldsymbol{\tau}}_2$ about (ψ, τ_2) gives:

$$\begin{aligned} \tilde{\boldsymbol{l}}_\tau^* &= \boldsymbol{l}_\tau^* + \mathbf{L}_{\tau\psi}^* (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) + \mathbf{L}_{\tau\tau}^* (\tilde{\boldsymbol{\tau}}_2 - \boldsymbol{\tau}_2) + O_p(1) \\ &= \mathbf{0} \end{aligned} \tag{53}$$

Differentiating $L^*(\hat{\boldsymbol{\psi}}, \tilde{\boldsymbol{\tau}}_2)$ with respect to $\hat{\boldsymbol{\psi}}$ gives:

$$\tilde{\boldsymbol{l}}_\psi^* = \boldsymbol{l}_\psi^* + n^{-1}\mathbf{L}_{\psi\psi}^* (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) + n^{-1}\mathbf{L}_{\psi\tau}^* (\tilde{\boldsymbol{\tau}}_2 - \boldsymbol{\tau}_2) + O_p(1) \tag{54}$$

where: $\tilde{\boldsymbol{l}}_\tau^*$ and $\tilde{\boldsymbol{l}}_\psi^*$ are the partial derivatives of $L^*(\hat{\boldsymbol{\psi}}, \tilde{\boldsymbol{\tau}}_2)$ with respect to $\hat{\boldsymbol{\psi}}$ and $\tilde{\boldsymbol{\tau}}_2$; $\boldsymbol{l}_\tau^*$ and $\boldsymbol{l}_\psi^*$ are the derivative vectors of $L^*(\psi, \tau_2)$; and $\mathbf{L}_{\psi\psi}^*$, $\mathbf{L}_{\psi\tau}^* = \mathbf{L}_{\tau\psi}^*{}'$ and $\mathbf{L}_{\tau\tau}^*$ are the second-derivative matrices of $n^{-1}L^*$ evaluated at the true parameter values. Standard likelihood results imply:

$$\sqrt{n}\,(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = -n^{-1/2}\mathbf{H}(\boldsymbol{\theta}^+)^{-1}\boldsymbol{l}(\boldsymbol{\theta}) + o_p(1) \tag{55}$$

where $\mathbf{H}(\theta)$ is the Hessian matrix of the mean log-likelihood L/n; and $\boldsymbol{l}(\theta) = \sum \boldsymbol{l}_i(\theta)$ is the score vector. Using (53)-(55), the normed second-stage score vector for $\hat{\boldsymbol{\psi}}$ is:

$$n^{-1/2}\,\tilde{\boldsymbol{l}}_\psi^* = \left(n^{-1/2}\boldsymbol{l}_\psi^*\right) - \mathbf{L}_{\psi\tau}^*\mathbf{L}_{\tau\tau}^{*-1}\left(n^{-1/2}\boldsymbol{l}_\tau^*\right) - \left[\mathbf{L}_{\psi\psi}^* - \mathbf{L}_{\psi\tau}^*\mathbf{L}_{\tau\tau}^{*-1}\mathbf{L}_{\tau\psi}^* \mid \mathbf{0}\right]\mathbf{H}(\boldsymbol{\theta})^{-1}\left(n^{-1/2}\boldsymbol{l}(\boldsymbol{\theta})\right) + o_p(1) \tag{56}$$

Under H_0, the vectors $n^{-1/2}\boldsymbol{l}_\tau^*$, $n^{-1/2}\boldsymbol{l}_\psi^*$ and $n^{-1/2}$ $\boldsymbol{l}(\theta)$ converge in distribution to a limiting zero-mean normal distribution. Consequently, $n^{-1/2}\,\tilde{\boldsymbol{l}}_\psi^*$ converges to a normal distribution with zero mean vector and covariance matrix estimated consistently by:

$$\tilde{\mathbf{V}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\xi}_i\boldsymbol{\xi}_i{}' \tag{57}$$

where:

$$\boldsymbol{\xi}_i = \tilde{\boldsymbol{l}}_{\psi i}^* - \tilde{\mathbf{L}}_{\psi\tau}^*\tilde{\mathbf{L}}_{\tau\tau}^{*-1}\tilde{\boldsymbol{l}}_{\tau i}^* - \left[\tilde{\mathbf{L}}_{\psi\psi}^* - \tilde{\mathbf{L}}_{\psi\tau}^*\tilde{\mathbf{L}}_{\tau\tau}^{*-1}\tilde{\mathbf{L}}_{\tau\psi}^* \mid \mathbf{0}\right]\mathbf{H}(\hat{\boldsymbol{\theta}})^{-1}\boldsymbol{l}_i(\hat{\boldsymbol{\theta}}) \tag{58}$$

Here the subscript i denotes the score contribution of the ith observation and the tilde denotes derivatives evaluated at the point $(\hat{\boldsymbol{\theta}}, \tilde{\boldsymbol{\tau}}_2)$. The score test statistic is then:

$$LM = n\,\tilde{\boldsymbol{l}}_\psi^*{}'\tilde{\mathbf{V}}^{-1}\tilde{\boldsymbol{l}}_\psi^* \tag{59}$$

which has a χ^2 distribution under H_0 with degrees of freedom equal to the dimension of ψ. This test can be viewed either as a specific test for the presence of bias induced by the initial conditions approximation or more generally as a test of the specified dynamic structure relating successive waves.

The main technical difficulty with the test is the computation of the second derivative matrices $\tilde{\mathbf{L}}^*_{\psi\tau}$ and $\tilde{\mathbf{L}}^*_{\psi\psi}$. The last of these is particularly troublesome, owing to its high-dimensionality. These matrices are very complicated to calculate through analytical formulae and numerical approximations to large Hessian matrices tend to be very inaccurate. In the implementation described below, we have computed $\tilde{\mathbf{L}}^*_{\psi\psi}$ by means of a simulation algorithm (described in Appendix 1) and $\tilde{\mathbf{L}}^*_{\psi\tau}$ using recursive central difference approximations.

6 An application to individual expectations data

The British Household Panel Survey (BHPS) is the principal source of nationally-representative household- and individual-level panel data in the UK. This application is based on the first 11 waves, relating to the years 1991-2001. Each year, BHPS participants are asked a series of questions about their attitudes. Here we analyse responses to the following question, using a sample of 1,277 male household heads: "*How well would you say you yourself are managing financially these days*?" Responses have been recoded as: y = 1 "*Finding it very difficult*"; y = 2 "*Finding it quite difficult*"; y = 3 "*Just about getting by*", y = 4 "*Doing alright*"; y = 5 "*Living comfortably*". Under the LAR model, the individual's underlying assessment of his financial position at time t is a naturally continuous variable, y^*_{it}, which we assume to be generated according to the panel autoregression (2). The respondent is then assumed to translate y^*_{it} into a response to the categorical survey question according to the rule (7).

The final parameter estimates for this LAR model are given in Table 1. Computation was done using the GHK simulator, using successive passes, initially with 50 replications (with antithetic variance reduction), rising to 500 once the neighbourhood of the optimum was reached. Following convergence, a single iteration was performed with 2000 replications as a check on convergence and the optimised likelihood value.

Following initial experimentation with alternative specifications, we used a subset of the **x**-variables from wave 0 for the initial conditions model. Estimation was then done sequentially, starting with S = 8, so that it initially involved only the y-observations from waves 0, 9 and 10. The skip rate S was then reduced sequentially while the score test remained insignificant. We encountered no rejection at any stage, so our final specification uses all available waves of data. Consequently, the rectangle probabilities involved in SML estimation are 11-dimensional.

The analogous SD model is:

$$y_{it}^* = \sum_{m=1}^{4} \alpha_m d_{mit-1} + \boldsymbol{\beta}'\mathbf{x}_{it} + u_i + \varepsilon_{it} \tag{60}$$

where $d_{mit} = \mathbf{1}(y_{it}^* = m)$. Estimated parameters for the SD model are given in Appendix Table A2.1. They were computed using 48-point Gauss-Hermite quadrature. Despite the fact that the SD model has 3 more parameters than the LAR model, the latter achieves a substantially higher log-likelihood.

In the model, economic circumstances are represented by the level of household income per capita, the proportion of household income earned by the respondent himself, and a dummy for owner-occupation, together with the estimated value of the equity in the house. As expected, the level of household per capita income has a significant positive effect on perceptions of financial well-being. The magnitude of the respondent's personal contribution to household finances has a significant positive influence on his reported perceptions. There is strong evidence to support the widely-held view that homeowners' perceptions respond to rising house values. Human capital also appears to be an important element in perceived financial well-being, since there is a strong positive influence of educational attainment. Recent changes in circumstances are represented by the first differences in per capita household income, the respondent's own income and the estimated house value. None of these is statistically significant.

However, these 'objective' financial factors are not sufficient to explain the determination and evolution of perceived financial well-being. Other explanatory variables are mostly are time-invariant. The small number of time-varying covariates are included in the form of current levels and changes from the previous year.

Ethnicity is represented by dummies for the Black and Asian groups and there is evidence of a negative difference, which is statistically significant for the latter group. The effect of marital status is captured by dummies for being married/cohabiting, divorced/separated or widowed. A further dummy identifies those who have made a transition into the divorced/separated group within the last year. Other status transitions were insignificant or too few in number to permit reliable estimation. There is a significant positive influence of a marital or cohabitation relationship and of widowhood. Divorce or separation reduces perceived well-being, with a further temporary reduction in the year of separation. Labour market status is represented by dummies for employment and self-employment, with significant positive effects. The status of unemployment has a significant negative effect, with a further temporary effect in the year of transition into unemployment.

Differences in household size and structure are important. There is a significant positive coefficient for the number of household members and a nearly offsetting negative coefficient for the number of children in the household. There is no detectable impact of a new birth on financial perceptions. These effects are in addition to the per capita equivalisation used for the household income variables. The relationship between perceived well-being and age is inverse U-shaped with a peak at the 51-65 age group, reflecting a standard life-cycle pattern of asset accumulation. Health status has no significant impact. The year dummies show a strongly rising trend from wave 1 (1992) to wave 9 (2000), followed by an abrupt fall in 2001. These year effects reflect quite closely the macroeconomic trend in average earnings growth.

Dynamic adjustment is captured by the autoregressive coefficient α, which is positive and strongly significant. On this evidence, there is a significant degree of persistence in perceptions. The comparison between the LAR and SD estimates shows that the latter generates too little persistence through the inherent dynamics and compensates for this misspecification by overestimating the variance of the individual effect. In our application, the intra-person correlation, $\sigma_u^2/(1+\sigma_u^2)$, is estimated to be 0.273 for the LAR model, compared to 0.406 for the SD model. The lagged responses of y to $\mathbf{x}$ decay rather faster in the SD model. For example, consider the probability of a good or very good response ($y_t > 3$). Let $\delta(s)$ be the derivative $\partial\Pr(y_t > 3|\mathbf{X})/\partial\boldsymbol{\beta}'\mathbf{x}_{t-s}$ evaluated at the point $\mathbf{x}_t = \mathbf{x}_{t-1} = \mathbf{x}_{t-2} \ldots = \bar{\mathbf{x}}$ and consider the scaled sequence $\delta^*(s) = \delta(s)/\delta(1)$. We find $\delta^*(1) = 0.331$ and 0.279 for the LAR and SD models respectively, decaying to $\delta^*(2) = 0.110$ and 0.076 and $\delta^*(3) = 0.036$ and 0.021. These are substantial differences. For applications to data displaying greater persistence than is apparent here, the difference between SD- and LAR-estimated dynamics could be very important indeed.

Table 1 Estimates and tests (standard errors in parentheses)

Covariate	Dynamic model $\hat{\boldsymbol{\beta}}$	Dynamic model Std. err.	Initial conditions $\hat{\boldsymbol{\delta}} / \hat{\sigma}_\eta$	Initial conditions Std. err.
α	0.331	0.012		
Black	-0.224	0.290	0.153	0.602
Asian	-0.366	0.128	-0.030	0.235
In relationship	0.081	0.043	-0.069	0.100
Divorced/separated	-0.161	0.079	-0.729	0.181
Widowed	0.300	0.159	0.765	0.828
Newly divorced/separated/widowed	-0.178	0.114		
Employed	0.291	0.048	0.346	0.129
Self-employed	0.234	0.060	0.250	0.155
Unemployed	-0.340	0.076	-0.653	0.171
Newly unemployed	-0.396	0.096		
Degree or other further education	0.222	0.055	0.448	0.111
A-level	0.269	0.066	0.384	0.125
O-level / GCSE / CSE / other qualification	0.130	0.062	0.373	0.110
Household size	0.169	0.026	0.175	0.050
Number of children in household	-0.105	0.018	-0.137	0.039
Δ number of children	-0.026	0.029		
Homeowner	0.078	0.057	0.053	0.199
Annual household income per head (£$\times 10^{-3}$)	0.199	0.013	0.474	0.053
Δ household income per head (£$\times 10^{-3}$)	0.005	0.016		
Δ own income (£$\times 10^{-3}$)	-0.018	0.019		
Own income as share of household income	0.356	0.061	0.498	0.639
Value of house (£$\times 10^{-5}$)	0.211	0.062	0.245	0.268
Proportionate change in value of house	0.000	0.005		
Age 18-30	-0.065	0.052	-0.293	0.119
Age 31-40	-0.084	0.044	-0.101	0.129
Age 41-50	-0.052	0.045	-0.009	0.130
Age 51-65	0.117	0.052	-0.004	0.157
Poor health	0.014	0.067	0.016	0.122
Newly-developed ill-health	0.036	0.082		
Wave 1 dummy	-0.270	0.130		
Wave 2 dummy	-0.136	0.051		
Wave 3 dummy	-0.075	0.051		
Wave 4 dummy	-0.148	0.051		
Wave 5 dummy	0.041	0.051		
Wave 6 dummy	0.007	0.052		
Wave 7 dummy	0.018	0.051		
Wave 8 dummy	-0.081	0.052		
Wave 9 dummy	-0.111	0.052		
Γ_1	-1.542	0.168	-1.598	0.315
Γ_2	-0.620	0.166	-0.577	0.309
Γ_3	0.934	0.166	1.233	0.323
Γ_4	2.463	0.167	2.870	0.360
σ_u^2	0.375	0.027		
γ	1.005	0.081		
σ_η^2	2.666	0.193		
Score test for the S = 1 model $\chi^2(59)$	54.78 (P = 0.368)			
Log-likelihood	-14,535.50			

7 Conclusions

We have considered an alternative to the discrete state dependence (SD) model for dynamic modelling of ordinal variables from panel data. The alternative LAR model involves ordinal observation of a latent autoregression, rather than lagged feedback of the previous period's discrete outcome. It is argued that this specification is more appropriate for a range of applications involving observational, rather than inherent, discreteness. Examples include interval regressions and models of expectations, and satisfaction.

We have developed a simulated maximum likelihood estimator and an associated test procedure designed to assist in handling the initial conditions problem. As part of this procedure, a novel simulation algorithm has been implemented for computing a required numerical Hessian matrix.

The method has been applied to a simple model of individual perceptions of financial well-being, applied to UK household panel data. The LAR model provides a robust description of the evolution of financial perceptions over time, with a significant role for lagged adjustment. The LAR model fits the data considerably better than the conventional SD model and has quite different equilibrium and dynamic properties. In particular, the SD model generally displays less persistence than the LAR model, and when misused to model highly-persistent data, the estimated variance of the individual effect is biased upwards to compensate.

References

Arellano, M. and Carrasco, R. (2003). Binary choice panel data models with predetermined variables, *Journal of Econometrics* **115**, 125-157.

Arellano, M., Bover, O. and Labeaga, J. M. (1997). Autoregressive models with sample selectivity for panel data, unpublished working paper (http://rt001hfd.eresmas.net/obmajl97.pdf)

Bover, O. and Arellano, M. (1997). Estimating dynamic limited dependent variable models from panel data, *Investigaciones Economicas* **21**, 141-165.

Chow, G. S. (1960). Tests for equality between sets of coefficients in two linear regressions, *Econometrica* **28**, 591-605.

Ferrer-i-Carbonell, A. and Frijters, P. (2004) How important is methodology for the estimates of the determinants of happiness? *Economic Journal* **114**, 641-659.

Hajivassiliou, V. and Ruud, P. (1994). Classical estimation methods for LDV models using simulation, in Engle, R. F. and McFadden, D. L. (eds.) *Handbook of Econometrics* **4**, 2383-2441. Amsterdam: North-Holland.

Heckman, J. J. (1978). Simple statistical models for discrete panel data developed and applied to test the hypothesis of true state dependence against the hypothesis of spurious state dependence, *Annales de l'INSEE* **30**, 227-269.

Heckman, J. J. (1981a). Statistical models for discrete panel data, in *Structural Analysis of Discrete Data with Econometric Applications*, Manski, C. F. and McFadden, D. (eds.), 227-269. Cambridge MA: MIT Press.

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

Van Praag, B. and Ferrer-i-Carbonell, A. (2004). *Happiness Quantified. A Satisfaction Calculus Approach*. Oxford: Oxford University Press.

Wooldridge, J.M. (2000), A framework for estimating dynamic, unobserved effects panel data models with possible feedback to future explanatory variables, *Economics Letters* **68**, 245-250.

Appendix 1: A simulation approximation to $\tilde{\mathbf{L}}^*_{\psi\psi}$

Our aim is to estimate the probability limit of the partial Hessian $n^{-1}\partial^2 L^*(\hat{\boldsymbol{\psi}},\tilde{\boldsymbol{\tau}}_2)/\partial\hat{\boldsymbol{\psi}}\partial\hat{\boldsymbol{\psi}}'$. Let $\mathbf{V}$ be an asymptotically valid approximation to the covariance matrix of $\hat{\boldsymbol{\psi}}$: for example, we might use $\mathbf{V} = -n^{-1}\mathbf{H}(\hat{\boldsymbol{\theta}})^{-1}$. Since $\mathbf{V}$ reflects the variability of $\hat{\boldsymbol{\psi}}$ and is $O_p(n^{-1})$, it provides a good metric for the approximation of derivatives. Let $\mathbf{K}$ be the Choleski factor such that $\mathbf{V} = \mathbf{KK}'$. Note that $\mathbf{K} = O_p(n^{-1/2})$. Now generate a sequence of independent pseudo-random N($\mathbf{0}$, $\mathbf{I}$) vectors $\mathbf{v}^1 \ldots \mathbf{v}^M$ and construct $z^m = n^{-1}\left(L^*(\hat{\boldsymbol{\psi}} + \mu_n \mathbf{K}\mathbf{v}^m, \tilde{\boldsymbol{\tau}}_2) - L^*(\hat{\boldsymbol{\psi}},\tilde{\boldsymbol{\tau}}_2)\right)$, where μ is a steplength parameter. Note that z^m, and any covariance (with respect to the distribution of $\mathbf{v}^m$) of z^m with powers of $\mathbf{v}^m$, are $O_p(n^{-1/2}\mu)$. Expand z^m in a Taylor series:

$$z^m = \frac{\mu}{\sqrt{n}}\sum_p d_p v_p^m + \frac{\mu^2}{n}\sum_p\sum_{q>p} D_{pq} v_p^m v_q^m + \frac{\mu^2}{2n}\sum_p D_{pp}\left(v_p^m\right)^2 + \zeta^m \tag{A1}$$

where d_p and D_{pq} are elements of $\mathbf{d} = \sqrt{n}\mathbf{K}'\left(n^{-1}\partial L^*(\hat{\boldsymbol{\psi}},\tilde{\boldsymbol{\tau}}_2)/\partial\hat{\boldsymbol{\psi}}\right)$ and $\mathbf{D} = n\mathbf{K}'\left(n^{-1}\partial^2 L^*(\hat{\boldsymbol{\psi}},\tilde{\boldsymbol{\tau}}_2)/\partial\hat{\boldsymbol{\psi}}\partial\hat{\boldsymbol{\psi}}'\right)\mathbf{K}$, and ζ is a remainder term. Note that $\mathbf{d}$ and $\mathbf{D}$ are $O_p(1)$. We then have the following results:

$$E\left(z^m v_p^m\right) = \frac{\mu}{\sqrt{n}} d_p + E\left(v_p^m \zeta^m\right) \tag{A2}$$

$$E\left(z^m v_p^m v_q^m\right) = \frac{\mu^2}{n} D_{pq} + E\left(v_p^m v_q^m \zeta^m\right), \qquad p \neq q \tag{A3}$$

$$E\left(z^m (v_p^m)^2\right) = \frac{\mu^2}{2n}\left[3D_{pp} + \sum_{q\neq p} D_{qq}\right] + E\left((v_p^m)^2\zeta^m\right) \tag{A4}$$

where the expectations are taken with respect to $\mathbf{v}^m$.

If necessary, the process for generating the variates $\mathbf{v}^m$ should be appropriately truncated to ensure that $L^*(\hat{\boldsymbol{\psi}} + \mu_n \mathbf{K}\mathbf{v}^m, \tilde{\boldsymbol{\tau}}_2)$ always exists and that terms of the form $E_v\left(v_i v_j\left(\partial^3 L^*/\partial\psi_p\partial\psi_q\partial\psi_r\,|_{\psi=\psi^+(\mathbf{v})}\right)\right)$ exist for any function $\psi^+(\mathbf{v})$ bounded by $(\hat{\boldsymbol{\psi}},\ \hat{\boldsymbol{\psi}} + \mu\mathbf{K}\mathbf{v})$. In practice, this requires ensuring that simulation of $\hat{\boldsymbol{\psi}} + \mu\mathbf{K}\mathbf{v}^m$ avoids regions where the Γ_r are non-ordered. A sufficiently small value for μ (for example

0.001) will generally achieve this. Under these circumstances, the remainder terms in (A2)-(A4) are $o_p(1)$. Thus we can estimate the cross-partials consistently as follows:

$$\hat{D}_{pq} = \frac{n}{\mu^2} M^{-1} \sum\nolimits_{m=1}^{M} z^m v_p^m v_q^m \ , \qquad p \neq q \tag{A5}$$

To calculate the remaining second derivatives, we solve the simulated analogue of the system of equations (A4), to give:

$$\hat{D}_{pp} = \frac{n}{\mu^2} M^{-1} \sum_{m=1}^{M} z^m \left[{v_p^m}^2 - \left(\frac{1}{2+P} \right) \sum_{q=1}^{P} {v_q^m}^2 \right] \tag{A6}$$

Now transform **D** back to ψ-space:

$$\hat{\mathbf{L}}^*_{\psi\psi} \approx n^{-1} \mathbf{K}'^{-1} \, \mathbf{D} \, \mathbf{K} \tag{A7}$$

One advantage of this method is that replications can be made sequentially and the procedure stopped when the quantities (A5) and (A6) appear to have reached convergence. The simulated analogue of expression (A2) can be compared with a conventional numerical derivative to check on the negligibility assumption for the remainder terms. Antithetic variance reduction can be used to improve simulation precision in (A5) and (A6).

Appendix 2: SD model estimates

Table A2.1 Estimates for the SD model (standard errors in parentheses)

Covariate	Dynamic model $\hat{\boldsymbol{\beta}}$	Dynamic model Std. err.	Initial conditions $\hat{\boldsymbol{\delta}}/\hat{\sigma}_\eta$	Initial conditions Std. err.
α_1	0.286	0.080		
α_2	0.631	0.071		
α_3	1.083	0.074		
α_4	1.524	0.076		
Black	-0.254	0.327	0.186	0.603
Asian	-0.398	0.144	-0.034	0.247
In relationship	0.088	0.045	-0.085	0.100
Divorced/separated	-0.173	0.083	-0.708	0.181
Widowed	0.328	0.165	0.780	0.818
Newly divorced/separated/widowed	-0.160	0.113		
Employed	0.302	0.050	0.342	0.131
Self-employed	0.239	0.062	0.270	0.156
Unemployed	-0.387	0.077	-0.634	0.172
Newly unemployed	-0.321	0.096		
Degree or other further education	0.240	0.059	0.454	0.111
A-level	0.297	0.071	0.398	0.127
O-level / GCSE / CSE / other qualification	0.144	0.066	0.369	0.111
Household size	0.181	0.027	0.184	0.050
Number of children in household	-0.116	0.019	-0.131	0.039
Δ number of children	-0.016	0.029		
Homeowner	0.087	0.059	0.064	0.207
Annual household income per head (£×10^{-3})	0.215	0.012	0.484	0.056
Δ household income per head (£×10^{-3})	-0.010	0.016		
Δ own income (£×10^{-3})	-0.010	0.019		
Own income as share of household income	0.365	0.063	0.574	0.639
Value of house (£×10^{-5})	0.215	0.065	0.215	0.290
Δ value of house (£×10^{-5})	0.000	0.005		
Age 18-30	-0.077	0.053	-0.283	0.118
Age 31-40	-0.090	0.044	-0.092	0.128
Age 41-50	-0.052	0.045	-0.007	0.129
Age 51-65	0.124	0.053	-0.001	0.154
Poor health	0.017	0.068	0.025	0.122
Newly-developed ill-health	0.029	0.082		
Wave 1 dummy	-0.175	0.053		
Wave 2 dummy	-0.148	0.051		
Wave 3 dummy	-0.084	0.051		
Wave 4 dummy	-0.155	0.051		
Wave 5 dummy	0.031	0.051		
Wave 6 dummy	0.010	0.051		
Wave 7 dummy	0.020	0.050		
Wave 8 dummy	-0.075	0.051		
Wave 9 dummy	-0.109	0.050		
Γ_1	-0.954	0.132	-1.092	0.212
Γ_2	-0.037	0.131	-0.367	0.213
Γ_3	1.511	0.132	0.892	0.219
Γ_4	3.028	0.132	2.035	0.221
σ_u^2	0.684	0.022		
γ	0.945	0.071		
Log-likelihood	-14,560.686			