Estimation and Specification Testing of Panel Data Models with Non-Ignorable Persistent Heterogeneity, Contemporaneous and Intertemporal Simultaneity, and Regime Classification Errors

by

Vassilis Hajivassiliou*

Department of Economics and Financial Markets Group,

London School of Economics

Revised January 2011

Abstract

This paper proposes efficient estimation methods for panel data limited dependent variables (LDV) models possessing a variety of complications: non-ignorable persistent heterogeneity; contemporaneous and intertemporal endogeneity; observable and unobservable dynamics; and imperfect regime classification information. It first shows how a simple modification of estimators based on the Random Effects principle can preserve the consistency and asymptotic efficiency of the method in panel data despite non-ignorable persistent heterogeneity driven by correlations between the individual-specific component of the error term and the regressors. The approach is extremely easy to implement and allows straightforward classical and omnibus tests of the significance of such correlations that lie behind the non-ignorable persistent heterogeneity. The method applies to linear as well as nonlinear panel data models, static or dynamic. Two major extensions of the existing literature are that the method works for time-invariant as well as time-varying regressors, and that these dependencies may be non-linear functions of the regressors.

The paper then combines this modified random effects approach with two simulation-based estimation strategies to overcome *analytical* as well as computational intractabilities in a widely applicable class of nonlinear models for panel data, namely the class of LDV models with contemporaneous and intertemporal endogeneity. The effectiveness of the estimation methods in providing asymptotically efficient estimates in such cases is illustrated with three discrete-response econometric models for panel data. An important problem handled by the framework developed in this paper involves contemporaneous and intertemporal simultaneity caused by strategic interactive effects or contagion across economic agents over time.

The final contribution of the paper is the development of an algorithm that allows for the first time efficient maximum likelihood estimation (MLE) of a panel data LDV model with regime classification imperfections in the presence of Markovian state dependence.

Keywords: Limited Dependent Variable Models, Simulation-Based Estimation, Endogeneity, Correlated Random Effects, Initial Conditions in Nonlinear Dynamic Panel Data Models, Regime Misclassification Information

JEL Classifications: C51, C52, C15

*I have benefited from constructive comments by participants at seminars at the London School of Economics, at the University of Toulouse, and at the World Bank Development Research Group, as well as at the 2006 Conference on Panel Data at the University of Cambridge. A preliminary version of this paper was circulated as LSE working paper Hajivassiliou (2006)[12]. I thank in particular Alain Trognon, Hashem Pesaran, Steve Pudney, Martin Ravallion Yannis Ioannides, and Frank Windmeijer for useful comments and suggestions. The usual disclaimer applies for any remaining errors.

Address correspondence to:
E-mail: vassilis@lse.ac.uk
Address: Department of Economics
London School of Economics
London WC2A 2AE, England

1 Introduction

This paper proposes efficient estimation methods for panel data limited dependent variables (LDV) models possessing a variety of complications: non-ignorable persistent heterogeneity; contemporaneous and intertemporal endogeneity; observable and unobservable dynamics; and imperfect regime classification information. Section 2 shows how a simple modification of estimators based on the Random Effects principle can preserve the consistency and asymptotic efficiency of the method in panel data despite non-ignorable persistent heterogeneity driven by correlations between the individual-specific component of the error term and the regressors. The approach is extremely easy to implement and allows straightforward tests of the significance of such correlations that lie behind the non-ignorable persistent heterogeneity. The method applies to linear as well as nonlinear panel data models, static or dynamic. In addition, the method works for time-invariant as well as time-varying regressors, and allows for nonlinear dependencies, thus providing important extensions to the existing literature.

In Section 3 we combine this modified random effects approach with two simulation-based estimation strategies to overcome *analytical* as well as computational intractabilities in an important class of nonlinear models for panel data, namely the class of LDV models with contemporaneous and intertemporal endogeneity. The simulation-based methods are: (a) Method of Maximum Simulated Likelihood employing the Geweke-Hajivassiliou-Keane importance-sampling simulator (MSL/GHK) and (b) the Method of Simulated Scores with Gibbs resampling (MSS/GRS). Subsection 3.1 sets up the theoretical framework, while Subsections 3.2.1-3.2.3 present three illustrative applications that employ the estimation strategy developed here. The effectiveness of the estimation methods in providing asymptotically efficient estimators in such cases is illustrated with three discrete-response econometric models for panel data. Application 1 is a simultaneous system determining a binary LDV indicator and trinomial ordered LDV indicator, whereas application 2 extends the endogeneity over time. An important problem handled by the framework developed in this paper involves contemporaneous and intertemporal simultaneity caused by strategic interactive effects and contagion across economic agents over time. More specifically, application 3 illustrates how to user our framework to model strategic interactions over time across subjects in experimental settings and contagion across countries in international finance. Subsection 3.2.4 shows how our methods allow for flexible serial and contemporaneous correlations in the unobservable disturbances of our panel models.

The final contribution of the paper is presented in Section 4 where we discuss panel data LDV models with regime classification imperfections in the presence of Markovian state dependence. It develops a novel algorithm that allows for the first time efficient maximum likelihood estimation (MLE) of this class of models. Our algorithm relies on a recursive set of matrix equations.

Section 5 concludes.

2 Problem I: Non-Ignorable Persistent Heterogeneity

Consider three classic cases of panel data models with time-varying and time-invariant regressors x and z respectively:

A. Linear Static:

$$y_{it} = x_{it}'\beta + z_i'\gamma + \epsilon_{it} \tag{1}$$

B. Linear Dynamic:

$$y_{it} = \delta y_{i,t-1} + x_{it}'\beta + z_i'\gamma + \epsilon_{it} \tag{2}$$

C. Nonlinear with nonadditive errors:

$$y_{it} = h\left(x_{it}'\beta + z_i'\gamma + \epsilon_{it}\right) \tag{3}$$

where $h(\cdot)$ is a known function, allowed to be nondifferentiable and discontinuous. LDV models are clearly a special version of this. For simplicity, we assume a balanced data set indexed by $i = 1, \cdots, N$ and $t = 1, \cdots, T$. We concentrate on the common situation of large N, and small to moderately large T.[1] In each case, suppose that ϵ_{it} follows the one-factor error components structure $\epsilon_{it} = \alpha_i + \nu_{it}$, with $E(\nu_{it}|X, Z) = 0$ and α and ν independent for any i,t. We let X and Z denote the matrices of the complete sample data on the time-varying and time-invariant regressors respectively.

A usual problem in many practical cases is that α_i may be believed to be correlated with one or more of the regressors (x_{it}', z_i'). We define this problem as "Non-Ignorable Persistent Heterogeneity," which results in inconsistency of estimators based on the Random-Effects (RE) principle. This problem very frequently leads applied researchers to adopt Fixed-Effects type estimators (FE), which are not affected by such random effects-regressors correlations. These decisions are predicated on the well-known fact that such correlations normally wreak havoc to estimators that are based on the standard RE principle of accounting for the non-sphericality of the error term distribution through suitable generalized least squares (GLS) and maximum likelihood estimation (MLE) methods.

Estimators based on the FE principle either eliminate or condition upon the persistent heterogeneity term α_i and are thus consistent irrespective of any regressor-heterogeneity correlations. These estimators for (1) yield Ordinary Least Squares estimation after applying either first-differencing $(w_{it} - w_{i,t-1})$ or the within transformation $(w_{it} - \frac{1}{T}\sum_{t=1}^{T} w_{it})$, where w_{it} stands in for the dependent variable y_{it} and all the regressors x_{it}^j and z_{it}^l; for (2) they yield Instrumental Variables estimation using sufficiently older lags of the dependent variable ($y_{i,t-l}$, $l > 1$) [see Arellano and

[1] Exogenously unbalanced data sets can be readily accommodated. In case the causes of unbalancedness are endogenously determined, all models become of category C, since a valid probability model characterizing the data availability necessarily introduces a nonlinearity of type (3).

Bond (1991)[1]]; and for (3) they are in general inconsistent due to the incidental parameters problem.[2]

It is our view that abandoning RE estimation in favour of FE in such situations is premature, unnecessary, and likely to have rather unfortunate consequences. This is because well-understood shortcomings of estimators based on the FE principle include, *inter alia*: (a) FE-type methods provide no estimates in general for the time-invariant coefficients γ; (b) since N α_i parameters are implicitly or explicitly estimated, such methods suffer substantial efficiency losses as compared to methods based on the RE principle; and (c) the within and first-differencing transformations typically reduce very significantly the signal-to-noise ratio of the time-varying regressors, thus resulting in serious inconsistencies in FE-based methods. These shortcomings can be explained in an intuitive way by noting that the FE-based methods sweep away also *ignorable* heterogeneity (that is uncorrelated with regressors). Hence, they clean out "too much" and make it harder to precisely identify the effects of main interest (β).

2.1 Modified Random Effects Estimation

We show how a simple modification of estimators based on the RE principle, following ideas of Mundlak (1978)[25] and Chamberlain (1984)[5], can preserve the consistency and asymptotic efficiency of the RE methodology.

Our approach models explicitly the suspected non-ignorable persistent heterogeneity by characterizing its correlation with the regressors as:

$$E(\alpha_i|X,Z) = \mu_i = g(X,Z) \tag{4}$$

and considering specific functions $g(\cdot)$. For example for the case without time-invariant regressors z_i, Mundlak (1978)[25] proposed $\mu_i = \bar{x}_{i\cdot}'\xi$ where $\bar{x}_{i\cdot} \equiv \frac{1}{T_i}\sum_{t=1}^{T_i} x_{it}$ is the time average of the regressor vector.[3] An alternative proposal was Chamberlain (1984)[5] who instead modelled this conditional mean as $E(\alpha_i|X) = \sum_{t=1}^{T_i} r_t x_{it}$ where r_t are period-specific weights.

It is important to emphasize that, in marked contrast to the Mundlak-Chamberlain work, our framework explicitly allows for the presence of time-invariant regressors, which should be useful in many real-world applications.[4] To this end, we introduce three assumptions concerning the conditional mean function $g(\cdot)$ characterizing the

[2] In very specific cases, consistent FE estimators exist for (3), e.g., the conditional logit model of Chamberlain (1980)[4].

[3] Hajivassiliou (1985)[10] used a similar approach for deriving formal tests of regressor-heterogeneity correlations in a switching regressions framework.

[4] As Wooldridge (2005)[26] explains (see his sections 11.3.2 and 15.8.2), the Chamberlain-Mundlak setup allowed "only time-varying explanatory variables". Yet the whole focus of the approach here is to analyze the presence of time-invariant regressors and the major impact of that on interpreting the coefficients (for policy analysis etc.)

correlation between the unobserved persistent heterogeneity α_i and regressors x and z:

Assumption 1: $g(\cdot)$ is a linear function of the regressors;
Assumption 2: $g(\cdot)$ depends only on the regressor data for individual i; and
Assumption 3: $g(\cdot)$ only depends on the regressors in a time-invariant way.

Assumptions (1)-(3) are satisfied by the Mundlak error model after extending it for the presence of invariant regressors, by defining:

$$E(\alpha_i|X,Z) = g(X,Z) = \bar{x}_{i.}'\xi + z_i'\zeta \tag{5}$$

If we now write

$$\alpha_i^* \equiv \alpha_i - \bar{x}_{i.}'\xi - z_i'\zeta \tag{6}$$

this new persistent heterogeneity term has by construction conditional mean zero. We can thus substitute out α_i from (1), (2), and (3) in each of the three classic cases considered and collect terms.

Specifically, for each of the canonical models above we obtain:

A. Modified Linear Static:

$$y_{it} = x_{it}'\beta + \bar{x}_{i.}'\xi + z_i'(\gamma + \zeta) + \alpha_i^* + \nu_{it} \tag{7}$$

B. Modified Linear Dynamic:

$$y_{it} = \delta y_{i,t-1} + x_{it}'\beta + \bar{x}_{i.}'\xi + z_i'(\gamma + \zeta) + \alpha_i^* + \nu_{it} \tag{8}$$

C. Modified Nonlinear with nonadditive errors:

$$y_{it} = h\left(x_{it}'\beta + \bar{x}_{i.}'\xi + z_i'\left(\gamma + \zeta\right) + \alpha_i^* + \nu_{it}\right) \tag{9}$$

Since by construction $E(\alpha_i^*|X,Z) = 0$ and $E(\nu_{it}|X,Z) = 0$ by assumption, this approach results in modified models with well-behaved random persistent heterogeneity effects that do not pose consistency problems for GLS/MLE estimation: *the solution proposed here thus involves simply adding the time-averages of the time-varying regressors as additional regressors in the right hand side of the respective panel data model and proceeding with the RE estimator that is appropriate for each case.* Consequently, our modified RE estimators will have the usual optimality properties: for case A the optimal RE/GLS estimator corresponds to OLS of the model (7) made spherical by applying the transformation $(w_{it} - \lambda\bar{w}_{i.})$; for case B, optimal RE corresponds to full information maximum likelihood (FIML) and three stage least squares (3SLS) applied to (8) written as a cross-sectional simultaneous equations system of T equations, one per period [see Barghava and Sargan (1982)[2]]; and for case C, efficient estimation is achieved through MLE, possibly with the aid of simulation-based inference in case likelihood contributions involve high dimensional integrals. This case is the focus of Section 3 below.[5]

[5]For nonlinear *dynamic* models, the methods of Wooldridge (2005)[26] are useful for handling the initial conditions problem inherent in such models.

The RE modification presented here offers several important extensions to the existing literature: first, as already noted above, our framework extends the Mundlak-Chamberlain approach to accommodate the empirically important case of time-invariant regressors. Implications and interpretation of this extension are discussed in Subsection 2.2. The second extension, discussed in Subsection 2.3, is the development of formal tests for the presence of non-ignorable heterogeneity. The third useful extension is the following: If it is believed that the correlation function $g(X, Z)$ should allow for nonlinearities in the regressors, we can modify suitably Assumption 1 so as to expand (5) to contain polynomials in $\bar{x}_{i\cdot}$ and z_i and hence obtain the new conditional mean function:

$$E(\alpha_i|X, Z) = \sum_{l=1}^{L} \left((\bar{x}_{i\cdot})^l\right)' \xi_l + \sum_{m=1}^{M} ((z_i)^m)' \zeta_m \tag{5'}$$

This specification allows for the first L powers of $\bar{x}_{i\cdot}$ and the first M powers of z_i to characterize the nonlinear time-invariant dependency of $E(\alpha_i|X, Z)$ on the regressors.[6]

2.2 Interpreting Coefficients of Time-Invariant Regressors $\bar{x}_{i\cdot}$ and z_i

It is a direct consequence of our approach that the time-invariant regressor coefficients γ are not identifiable separately from parameter vector ζ, as can be seen from equations (7)-(9). At first glance this may appear as a limitation of the approach we propose. Upon further reflection, however, one realizes that our approach actually yields the correct marginal effects with respect to changes in regressor variables, taking into account both the direct as well as the indirect effects of such changes. To illustrate, consider a change in time-varying regressor j, say Δx_{it}^j and a change in a time-invariant regressor m, say Δz_i^m. Given that we focus on the case $E(\alpha_i|X, Z) = g(X, Z)$ where we assume specifically that $g(X, Z)$ is well modelled by $\bar{x}_{i\cdot}'\xi + z_i'\zeta$, it follows that for panel data Model A the expected marginal effect of a change Δx_{it}^j that is relevant for policy-making purposes is:[7]

$$\Delta E(y_{it}|X, Z)/\Delta x_{it}^j = \beta^j + \frac{1}{T}\xi^j$$

while for a change Δz_i^m it is:

$$\Delta E(y_{it}|X, Z)/\Delta z_i^m = \gamma^m + \zeta^m$$

[6] Alternatively, we could specify the time-average of the lth power of x_{it}. I.e., we would use: $\frac{1}{T}\sum_{t=1}^{T}(x_{it})^l$ instead of $(\bar{x}_{i\cdot})^l$.

[7] This formula needs to be adjusted accordingly in case the change in x^j is assumed to persist for longer than one period.

Our method provides estimates of both marginal effects as derived here, since it yields separately parameter vectors β and ξ as well as the combined vector $\gamma + \zeta$. Similar logic gives also the marginal effects for cases B and C, *mutatis mutandis*.[8]

2.3 Testing for Non-Ignorable Persistent Heterogeneity

Our approach enables also straightforward testing of the significance of correlation between the regressors and the persistent heterogeneity term (i.e., the individual-specific component of the error term), which would render it non-ignorable: under the maintained hypothesis of this paper, a classical test (by employing any of the traditional methods of Lagrange Multiplier, Likelihood Ratio, or Wald) of the time-averages $\bar{x}_{i.}$ when entered as additional regressors, provides a formal test as to whether the conditional mean function $E(\alpha_i|X, Z)$ indeed depends on the X regressors. To the extent that the conditional mean model is only an approximation, such significance tests should be viewed as omnibus specification tests of the presence of important Regressor-Heterogeneity correlations that are modelled less precisely.

Finally, specification tests in the Wu-Hausman mould can be constructed by comparing alternative estimators of the β parameters. In particular consider the traditional FE estimator $\hat{\beta}_{FE}$ that is consistent irrespective of Heterogeneity-Regressor correlations; the traditional $\hat{\beta}_{RE}$ estimator that is consistent and efficient under the assumption of no Regressor-Heterogeneity correlations $E(\alpha_i|X, Z) = 0$; and the modified RE $\hat{\beta}_{MRE}$ estimator here that is consistent and efficient under the correlation model $E(\alpha_i|X, Z) = \bar{x}'_{i.}\xi + z'_i\zeta$. Constructing Wu-Hausman quadratic forms based on pairing $\hat{\beta}_{MRE}$ with $\hat{\beta}_{FE}$ on one hand and with $\hat{\beta}_{RE}$ on the other yields straightforward specification tests in this context.

3 Problem II: LDV Panel Models with Contemporaneous and Intertemporal Simultaneity

It is now shown that the approach developed above can be readily applied to general additive and non-additive nonlinear panel data models, which may be static or dynamic, through the introduction of Simulation-Based inference. For an introduction to these methods, see inter alia Hajivassiliou (1993)[11]]. For the dynamic case, the framework here extends the Barghava and Sargan (1982)[2] approach to *nonlinear* dynamic models.

[8]Note that under certain scenarios (e.g., Hausman and Taylor (1981)[17]) it may be possible to extend the FE approach to recover estimates of the time-invariant parameters γ. That would allow one to identify separately the indirect effect vector ζ from the combined estimate generated by our modified RE method. In general, whether one desires the combined *direct plus indirect* $\gamma + \zeta$ or the two parameters separately will depend on the specific policy analysis one has in mind.

To focus on the most challenging case, namely nonlinear models with non-additive errors, we consider the leading case for the non-additive error nonlinear model of equation (3): this is the LDV model for panel data with T_i periods of observation on individual unit $i = 1, \cdots, N$. This model is defined by a $G_i \cdot T_i \times 1$ vector of limited dependent variables y_i induced by an $M_i \cdot T_i \times 1$ vector of latent variables y_i^* observed through the *partial observability rule*:

$$y_i = \tau(y_i^*).$$

The limited dependent vectors y_i are independently drawn across i and the $G_i \times 1$ y_{it} and $M_i \times 1$ y_{it}^* vectors are stacked in the obvious way to form the $G_i \cdot T_i \times 1$ y_i and $M_i \cdot T_i \times 1$ y_i^* vectors:

$$y_i \equiv \begin{pmatrix} y_{i1} \\ \vdots \\ y_{it} \\ \vdots \\ y_{iT_i} \end{pmatrix} \quad \text{and} \quad y_i^* \equiv \begin{pmatrix} y_{i1}^* \\ \vdots \\ y_{it}^* \\ \vdots \\ y_{iT_i}^* \end{pmatrix}$$

Particularly useful LDV models $y_i = \tau(y_i^*)$ correspond to a set of linear inequalities on y_i^* defined by lower and upper limit vectors a_i and b_i respectively, with:

$$y_i = \tau(y_i^*) \quad \text{such that} \quad \{y_i^* | a(y_i) < y_i^* < b(y_i)\}. \tag{10}$$

It should be noted that the function characterizing the latent vector y_i^* may depend on, in addition to exogenous regressors, the limited vector y and the latent vector $y^*$of other economic agents and from different points in time.

3.1 Estimation by Simulation: Maximum Simulated Likelihood (MSL) and Simulated Scores (MSS)

It is well known that maximum simulated likelihood in conjunction with the Geweke-Hajivassiliou-Keane simulator (MSL/GHK) and the method of simulated scores based on Gibbs resampling (MSS/GRS) overcome the well-known computation intractabilities of the multiperiod (panel) limited-dependent-variable models. See *inter alia* Börsch-Supan and Hajivassiliou (1993)[3], Hajivassiliou, McFadden, and Ruud (1996)[16], Hajivassiliou and McFadden (1998)[15].

In this paper we stress an additional feature of the MSL/GHK and MSS/GRS methods that is less well known and understood, namely that it overcomes *analytical intractabilities* associated with LDV models (for both panel and cross-sectional data) with complicated error correlations and endogeneity.

Let us first define the MSL method: Let the log-likelihood function for the unknown parameter vector θ given the sample of observations $(y_i, i = 1, \ldots, N)$ be

$$\ell_N(\theta) \equiv \sum_{i=1}^{N} [\log f(\theta; y_i)] \tag{11}$$

and let $\tilde{f}(\theta; y, \omega)$ be a simulator that is: (1) unbiased so that $f(\theta; y) = \mathrm{E}_\omega[\tilde{f}(\theta; y, \omega)|y]$ where ω is a simulated vector of R random variates, and (2) a continuous function of θ and ω. The maximum simulated likelihood estimator is

$$\hat{\theta}_{MSL} \equiv \arg\max_{\theta} \tilde{\ell}_N(\theta) \tag{12}$$

where

$$\tilde{\ell}_N(\theta) \equiv \sum_{n=1}^{N} \log \tilde{f}(\theta; y_n, \omega_n) \tag{13}$$

for some given simulation sequence $\{\omega_n\}$.

When $\tilde{f}(\cdot)$ is generated according to the GHK method, which is based on the importance sampling principle, $\tilde{f}$ satisfies the unbiasedness and continuity requirements of the MSL definition.

We next turn to MSS estimation: define the score of observation i by $s(\theta; y_i) = \nabla_\theta \log f(\theta; y_i)$ where ∇_θ is the first derivative operator with respect to θ. Adding up over all observations, we have

$$s_N(\theta) = \sum_{i=1}^{N} s(\theta; y_i) = \sum_{i=1}^{N} \nabla_\theta \log f(\theta; y_i) \tag{14}$$

Let $\tilde{s}(\theta; y, \omega, r_G)$ be a simulator based on r_G Gibbs resamplings that is: (1) asymptotically unbiased as $r_G \longrightarrow \infty$ so that $\tilde{s}(\theta; y, \omega, r_G) \longrightarrow \mathrm{E}_\omega[\tilde{s}(\theta; y, \omega)|y]$ where ω is a simulated vector of r_G random variates, and (2) a continuous function of θ and ω. The simulated scores estimator $\hat{\theta}_{MSS}$ is then the argument that solves the vector equation:

$$\tilde{s}_N(\theta) = 0 \tag{15}$$

where

$$\tilde{s}_N(\theta) \equiv \sum_{i=1}^{N} \tilde{s}(\theta; y_i, \omega_i, r_G) \tag{16}$$

For detailed description and analysis of the GHK and Gibbs simulators, the reader is referred to Hajivassiliou and McFadden (1998)[15]. It is proved there that the MSL/GHK estimator will be consistent, asymptotically normal, and fully efficient provided that R, the number of simulations employed per individual observation i, rises without bound at least as fast as $\sqrt{N}$. It is also proved there that the MSS/GRS estimator will be consistent, asymptotically normal, and fully efficient asymptotically if in addition r_G rises without bound at least as fast as $\ln N$.

Therefore, what remains for us to establish in the following section is that LDV models for panel data with all the complications discussed in the outset of this paper, possess likelihood contribution and score functions that can be written as sets of linear inequalities of the form (10). Specifically we need to show that these models correspond to:

$$\{Z_i | a_i < Z_i < b_i\} \tag{17}$$

with conditioning probability

$$\Pr(a_i < Z_i < b_i) \tag{18}$$

where the $M_i \times 1$ latent vector is distributed $Z_i \sim N(\mu_{Z_i}, \Sigma_{Z_i})$.The optimality properties of the MSL/GHK and MSS/GRS estimators for these models will then follow directly.[9] Consequently, we will be able to illustrate that our framework is applicable to a very general class of nonlinear, non-additive panel data models with complicated dynamics.[10]

3.2 Three Illustrative Applications with Contemporaneous and Intertemporal Endogeneities

The key difficulty with the models of Applications 1 and 2 is the presence of contemporaneous endogeneity between discrete LDV indicators at a given point in time, as well as spreading over time. Furthermore, Application 3 introduces the additional important problem of endogenous LDV factors because of strategic and social interactions. Without the estimation strategies introduced in this paper, researchers were stumped as to how to derive analytically and then compute efficiently the likelihood contributions and scores for these types of models.

---

[9]In terms of implementing these methods, one can rely on the modular procedures for GHK and GRS that return the simulated probability, $\tilde{P}_{GHK}$, and the simulated score, $\tilde{s}_{GRS}$, as a function of the following arguments:
m=dimension of multivariate normal vector Z;
mu=EZ;
w=V(Z);
wi=w^{-1};
c=Cholesky factor of w;
vectors a and b, defining the restriction region $a < Z < b$;
r=number of replications;
u=a $m \times r$ matrix of i.i.d. uniform [0,1] variates.
These procedures are publicly available at:
`http://econ.lse.ac.uk/staff/vassilis/pub/simulation/`
Versions are available in three alternative programming languages: C, Fortran, and Gauss.

[10]This generality is in marked contrast to the existing literature, which develops specialized, ad hoc methods to handle highly specific models (e.g., see Wooldridge (2005)[26].

3.2.1 Application 1: Simultaneous Determination of a Binary LDV Indicator and a Trinomial Ordered LDV Indicator

In this application, the MSL/GHK and MSS/GRS estimators provide asymptotically efficient simulation-based estimation of the Liquidity and Employment Constraint Indicator model of Hajivassiliou and Ioannides (2007)[14]. Traditional approaches deemed this model to be computationally as well as analytically intractable due to the contemporaneous as well as across-periods endogeneity due to dynamics in the limited dependent variables.

Define two latent dependent variables y^*_{1it} and y^*_{2it} and for simplicity drop the it subscripts:

$$y_1 = \begin{cases} 1 & if \quad y^*_1 > 0 \quad \text{(liquidity constraint binding),} \\ 0 & if \quad y^*_1 \leq 0 \quad \text{liquidity constraint not binding.} \end{cases} \tag{19}$$

$$y_2 = \begin{cases} -1 & if \quad y^*_2 \leq \lambda^- \quad \text{overemployed} \\ 0 & if \quad \lambda^- \leq y^*_2 < \lambda^+ \quad \text{voluntarily employed} \\ +1 & if \quad \lambda^+ \leq y^*_2 \quad \text{under-/unemployed.} \end{cases} \tag{20}$$

$$y^*_1 = \mathbf{1}(y^*_2 < \lambda^-)\gamma_{11} + \mathbf{1}(\lambda^- < y^*_2 < \lambda^+)\gamma_{12} + x'_1\beta_1 + \epsilon_1 \tag{21}$$

$$y^*_2 = \mathbf{1}(y^*_1 > 0)\delta + x_2\beta_2 + \epsilon_2 \tag{22}$$

where $\mathbf{1}(event)$ is the usual indicator function defined by $\mathbf{1}(event) = \begin{cases} 1 & \text{if } event \text{ is true} \\ 0 & \text{if } event \text{ is false} \end{cases}$.
Since (y_1, y_2) lie in $\{0,1\} \times \{-1,0,1\}$, the 6 possible configurations may be enumerated as follows:

y_1	y_2	y^*_1	y^*_2
0	-1	$\gamma_{11} + x_1\beta_1 + \epsilon_1 < 0,$	$x_2\beta_2 + \epsilon_2 < \lambda^-$
0	0	$x_1\beta_1 + \epsilon_1 < 0,$	$\lambda^- < x_2\beta_2 + \epsilon_2 < \lambda^+$
0	+1	$\gamma_{12} + x_1\beta_1 + \epsilon_1 < 0,$	$\lambda^+ < x_2\beta_2 + \epsilon_2$
1	-1	$\gamma_{11} + x_1\beta_1 + \epsilon_1 > 0,$	$\delta + x_2\beta_2 + \epsilon_2 < \lambda^-$
1	0	$x_1\beta_1 + \epsilon_1 > 0,$	$\lambda^- < \delta + x_2\beta_2 + \epsilon_2 < \lambda^+$
1	+1	$\gamma_{12} + x_1\beta_1 + \epsilon_1 > 0,$	$\lambda^+ < \delta + x_2\beta_2 + \epsilon_2$

In terms of the unobservables as in the GHK simulator implementation described above, the probability of a (y_1, y_2) observed pair is equivalent to the probability:

$$a \equiv \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} < \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix} < \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \equiv b \tag{23}$$

where $(\epsilon_1, \epsilon_2)' \sim N(0, \Sigma_\epsilon)$, and a and b are given by:

y_1	y_2	a_1	a_2	b_1	b_2
0	-1	$-\infty$	$-\infty$	$-(\gamma_{11}+x_1\beta_1)$	$\lambda^- - x_2\beta_2$
0	0	$-\infty$	$\lambda^- - x_2\beta_2$	$-x_1\beta_1$	$\lambda^+ - x_2\beta_2$
0	+1	$-\infty$	$\lambda^+ - x_2\beta_2$	$-(\gamma_{12}+x_1\beta_1)$	$+\infty$
1	-1	$-(\gamma_{11}+x_1\beta_1)$	$-\infty$	$+\infty$	$\lambda^- - \delta - x_2\beta_2$
1	0	$-x_1\beta_1$	$\lambda^- - \delta - x_2\beta_2$	$+\infty$	$\lambda^+ - \delta - x_2\beta_2$
1	+1	$-(\gamma_{12}+x_1\beta_1)$	$\lambda^+ - \delta - x_2\beta_2$	$+\infty$	$+\infty$

The variance-covariance matrix captures the contemporaneous correlation between ϵ_1 and ϵ_2. Given the binary nature of y_1, σ_{11} is normalized to 1. Section 3.2.4 below explains how our estimations take full account of this contemporaneous correlation *as well as* flexible forms of serial correlation. Section 2 above showed how to allow the random error components to be correlated with the regressors.

3.2.2 Application 2: Simultaneous Determination of Two Binary Indicators with Observable Dynamic Endogeneity

In our second illustrative application, the MSL/GHK and MSS/GRS estimators provide asymptotically efficient simulation-based estimation of the Currency and Banking Crises model of External Financing of Falcetti and Tudela (2007)[8]. Traditional approaches deemed this model to be computationally as well as analytically even more intractable than the one discussed in the previous application, because of the complicated dynamics across multiple periods involving the endogenous limited dependent variables.

Define two latent dependent variables y^*_{1it} and y^*_{2it} and two binary limited dependent variables y_{1it} and y_{2it} as follows:

$$y_{1it} = \begin{cases} 1 \quad if \quad y^*_{1it} \equiv x'_{1it}\beta_1 + \mathbf{1}\left(\sum_{s=1}^{L} y_{2i,t-s} > 0\right)\cdot\gamma + \epsilon_{1it} > 0,, \\ 0 \quad if \qquad otherwise. \end{cases} \tag{24}$$

$$y_{2it} = \begin{cases} 1 \quad if \quad y^*_{2it} \equiv x'_{2it}\beta_2 + \epsilon_{2it} > 0, \\ 0 \quad if \qquad otherwise. \end{cases} \tag{25}$$

where the distributed lag on the RHS of (24) is over L periods. For concreteness, in the illustration here we use $L = 4$, which is a natural choice if the data are quarterly. Consider the probability expression for $t \geq 5$:

$$Prob(y_{1it}, y_{2it}, \cdots, y_{1iT_i}, y_{2iT_i} | X_{1i}, X_{2i}, y_{1i,t-1}, \cdots, y_{1i,t-4}, y_{2i,t-1}, \cdots, y_{2i,t-4}, \theta) \tag{26}$$

We define $X_{1i} \equiv [x_{1i1}, x_{1i2}, \cdots, x_{1it}, x_{2it}, \cdots, x_{1iT_i}, x_{2iT_i}]$. For a typical observation it:

y_{1it}=1	$y^*_{1it} > 0$	$\epsilon_{1it} + x'_{1it}\beta_1 + \mathbf{1}\left(\sum_{s=1}^{4} y_{2i,t-s} > 0\right)\cdot\gamma > 0$
y_{1it}=0	$y^*_{1it} < 0$	$\epsilon_{1it} + x'_{1it}\beta_1 + \mathbf{1}\left(\sum_{s=1}^{4} y_{2i,t-s} > 0\right)\cdot\gamma < 0$
B_{it}=1	$y^*_{2it} > 0$	$\epsilon_{2it} + x'_{2it}\beta_2 > 0$
B_{it}=0	$y^*_{2it} < 0$	$\epsilon_{2it} + x'_{2it}\beta_2 < 0$

Therefore:

$$Prob(y_{1it}, y_{2it}|X_{1i}, X_{2i}, y_{1i,t-1}, \cdots, y_{1i,t-4}, y_{2i,t-1}, \cdots, y_{2i,t-4}, \theta) = \quad (27)$$

$$Prob\left((1-2y_{1it})\left[\epsilon_{1it} + x'_{1it}\beta_1 + \mathbf{1}\left(\sum_{s=1}^{4} y_{2i,t-s} > 0\right)\cdot\gamma\right] < 0, (1-2y_{2it})\left[\epsilon_{2it} + x'_{2it}\beta_2\right] < 0\right) \quad (28)$$

In terms of the canonical GHK and GRS formulations:

$$\begin{pmatrix} a_{1it} \\ a_{2it} \end{pmatrix} < \begin{pmatrix} \epsilon_{1it} \\ \epsilon_{2it} \end{pmatrix} < \begin{matrix} b_{1it} \\ b_{2it} \end{matrix} \quad (29)$$

we obtain the configuration:

y_{1it}	y_{2it}	a_{1it}	a_{2it}	b_{1it}	b_{2it}
0	0	$-\infty$	$-[x'_{it}\beta + Hy_{2it}\gamma]$	$-\infty$	$-x'_{2it}\beta_2$
0	1	$-\infty$	$-[x'_{it}\beta + Hy_{2it}\gamma]$	$-x'_{2it}\beta_2$	∞
1	0	$-[x'_{1it}\beta_1 + Hy_{2it}\gamma]$	∞	$-\infty$	$-x'_{2it}\beta_2$
1	1	$-[x'_{1it}\beta_1 + Hy_{2it}\gamma]$	∞	$-x'_{2it}\beta_2$	∞

where $Hy_{2it} \equiv \mathbf{1}\left(\sum_{s=1}^{4} y_{2i,t-s} > 0\right)$.

As already mentioned, Section 3.2.4 below explains how our estimations take full account of this contemporaneous correlation *as well as* flexible forms of serial correlation, and Section 2 above showed how to allow the random error components to be correlated with the regressors.

3.2.3 Application 3: Strategic Interaction Effects across Economic Agents

We now consider general dynamic LDV models with strategic interactive effects, which can arise because of game-theoretic considerations in laboratory experimental settings or because of macroeconomic contagion in panels of countries. Liu et al (2008)[23] provide an example of the first type. Here we will show how to cast these models in the linear inequality framework (10), thus making our MSL/GHK and MSS/GRS simulation-based approaches directly applicable. Consequently, our approach eliminates the need for the ad hoc specialized methodology developed by Liu et al. (2008)[23].

For individual agent i, consider the latent dependent variables for periods 1 to t. We assume $i = 1, \cdots, N$. Let u_{it} denote the time-varying component of the corresponding error (assumed to be *i.i.d.* over i and t in the simplest version) and let the heterogeneity component α_i be i.i.d. over i. Assuming that u_{it} enters in an additive form, the latent values are given by:

$$\begin{aligned}
y_{it}^* &= h_{it}(y_{i,t-1}^*, y_{i,t-2}^*, \cdots, y_{i0}^*, y_{1,t-1}, y_{2,t-1}, \cdots, y_{N,t-1}, y_{1,t-2}, y_{2,t-2}, \cdots, y_{N,t-2}, \\
&\quad \cdots, y_{11}, y_{21}, \cdots, y_{N1}, y_{10}, y_{20}, \cdots, y_{N0}, X_t, \alpha_i) + u_{it} \\
&\quad \vdots \\
y_{i2}^* &= h_{i2}(y_{i1}^*, y_{i0}^*, y_{11}, y_{21}, \cdots, y_{N1}, y_{10}, y_{20}, \cdots, y_{N0}, X_2, \alpha_i) + u_{i2} \\
y_{i1}^* &= h_{i1}(y_{i0}^*, y_{10}, y_{20}, \cdots, y_{N0}, X_1, \alpha_i) + u_{i1}
\end{aligned} \tag{30}$$

while the observed limited value y_{it} follows a binary threshold crossing specification $y_{it} \equiv \begin{cases} 1 & \text{if } y_{it}^* > 0 \\ 0 & \text{otherwise} \end{cases}$. The $h_{it}(\cdot)$ functions are assumed to follow a Polya scheme that makes them linear in their arguments.[11]

This specification makes the latent value for i in period t depend on: (a) all lagged values of same individual agent back to the initial time 0; (b) the observed binary choices y of all individuals (including own) from all the previous periods for $t-1$, $t-2$, $\cdots$, 2, 1, and 0; (c) the exogenous regressor values, X_t, for all individuals in that period; and (d) the heterogeneity effect α_i. In a game-theoretic experimental setting as in Liu et al (2008)[23], feature (b) represents strategic interactions across agents, while in a macroeconomic panel model of countries, where y_{it}^* represents the propensity of a country i in period t to run into external finance problems, feature (b) captures the phenomenon of *crisis contagion* spreading across countries.

The full vector of latent variables is:

$$Y^* = (y_{11}^*, \cdots, y_{1t}^*, \cdots, y_{1T}^*, \cdots, y_{i1}^*, \cdots, y_{it}^*, \cdots, y_{iT}^*, \cdots, y_{N1}^*, \cdots, y_{Nt}^*, \cdots, y_{NT}^*)'$$

Conditional on the strictly exogenous regressors, this vector is stochastically driven by α_i, $i = 1, \cdots, N$ and u_{it}, $i = 1, \cdots, N$, $t = 1, \cdots, T$. Note that our methods above allow α_i to be correlated with the regressors through the modified RE framework of Section 2. Our methods also allow u_{it} to follow more complicated processes than *i.i.d.* over i and t, e.g., $ARMA(p,q)$. Since the set of equations (30) specifies that y_{it}^* depends on all the past latent variables of this individual i, as well as the past observed binary choices of every individual, the set can be summarized as:

$$BY^* = CY + DX + \epsilon$$

where ϵ contains the two error components α and u. The observed set of choices Y corresponds to linear restrictions on the elements of Y^* through the binary scheme $y_{it} \equiv \begin{cases} 1 & \text{if } y_{it}^* > 0 \\ 0 & \text{otherwise} \end{cases}$ and hence the model is equivalent to:

$$a(Y, X, \theta) < Y^* < b(Y, X, \theta)$$

[11] The initial conditions $y_{10}, y_{20}, \cdots, y_{N0}$ do not pose particular modelling problems in this setting, since it is reasonable to assume that they are exogenous in an experimental setting. Similarly, the latent initial conditions $y_{10}^*, y_{20}^*, \cdots, y_{N0}^*$ are assumed to be 0.

where a and b are vectors of lower and upper bounds similar to the ones specified by Liu et al. (2008)[23]. These bounds depend on Y, X, and θ, where θ is the parameter vector to be estimated that combines B, C, D and the parameters characterizing the variance-covariance structure of ϵ. Consequently, the MSL/GHK and MSS/GRS methods can be employed as a "black-box" without the need for ad hoc derivations of the likelihood function etc.

3.2.4 Treatment of Flexible Serial and Contemporaneous Correlations in the Panel Structure

In the previous three subsections, we have described how the probability of the LDV y_i can be expressed in terms of the fundamental GHK/GRS implementation through the vector of linear inequalities:

$$a_i < \epsilon_i < b_i \tag{31}$$

The suitably stacked ϵ_i will have variance-covariance matrix with structure characterized by the precise serial correlation assumptions made on the ϵ_{it}'s. In particular, one-factor random effect assumptions will imply an equicorrelated block structure on Σ_ϵ, while the more general assumption of one-factor random effects *combined with* an AR(1) process for each error implies that Σ_ϵ combines equicorrelated and Toeplitz-matrix features. In addition, in case it is believed that the model exhibits non-ignorable individual heterogeneity in the form of regressor-random effects correlations, the modified random effects approach described above in section 2 can be invoked.

Through this representation, the probability of a complete sequence of the observable LDV behaviour for individual unit i, $\Pr(y_i)$, conditional on regressors and parameters, corresponds to:

$$Prob(a_i < \epsilon_i < b_i)$$

Consequently, our approach incorporates fully:

1. the contemporaneous correlations in vector ϵ_{it};

2. the full variance-covariance structure in ϵ_i, e.g., a one-factor plus ARMA(p,q) serial correlations in ϵ_i;

3. the interdependencies and spillovers among the LDVs due to contemporaneous, intertemporal, or strategic/social interaction factors; and

4. non-ignorable individual heterogeneity in the form of regressor-random effects correlations.

It is important to note that most features of our modelling approach summarized by properties 1.-4. are thus *testable*, since they correspond to contemporaneous and intertemporal restrictions on model parameters.[12]

One other important issue with our modelling approach that is not addressed in this paper is the identification issue of *coherency*. The interested reader is referred to Hajivassiliou (2010)[13] which develops some novel methods for establishing the coherency conditions of the models we discussed above.[13]

4 Problem III: Imperfections in Regime Classification in LDV Switching Models

The final problem we discuss are regime classification imperfections for classic panel data LDV models and present an algorithm for overcoming certain intractabilities of such models. These problems appear quite prevalent in empirical work that models LDV response in a panel data context.

For concreteness, we consider the general switching-regression model with two states, indexed by $s = 1, 2$. We consider the sequence of T observations on individual i in a panel data set, and drop the individual index i for convenience:

$$y^*_{st} = h_s(X_s\delta_s) + \epsilon_{st} \quad s = 1, 2; \quad t = 1, ..., T \tag{32}$$

$$y^*_{3t} = h_3(Z_t\zeta) + \epsilon_{3t} \tag{33}$$

$$y^*_{4t} = y^*_{3t} + \eta_t. \tag{34}$$

Here, y^*_{1t}, y^*_{2t}, y^*_{3t}, and y^*_{4t} are latent variables, unobservable by the econometrician; X_1, X_2, X_3, and Z are matrices of explanatory (exogenous) variables; and

[12] This is in great contrast to the Arellano-Bond (1991)[1] and Honoré-Kyriazidou (2000)[19] approaches, where first-differencing plus IV-type of estimation is used to estimate panel data models with observable dynamics that are linear and non-linear LDV respectively. In the A/B and H/K approaches the variance-covariance error structure is typically *necessary for identification* hence cannot be tested.

[13] Another remaining modelling issue for our panel LDV with observable dynamics is the likely endogeneity of the initial conditions in such models. The approximate solution we propose here considers the marginal LDV model for the initial condition and estimates it while allowing for flexible correlations with the future periods. This is the nonlinear analogue of the solution proposed by Barghava and Sargan (1982)[2] for the linear dynamic model and uses the best nonlinear regression for the latent variable of the initial condition by using *all* data for *all* periods available to the econometrician, which of course was not available to the decision-maker at the time t. This approach implies a new error term (u_{i1}) for the approximate initial condition equation that is different from the other periods' structural equations errors (ϵ_{it}). As Heckman (1981b)[18] explains, in general the error u_{i1} does not have the same distribution as the ϵs (assumed here to be Gaussian), nor is it likely that such a stable representation of the initial condition will exist. Such approximations are shown by Heckman's Monte-Carlo evidence not to be too critical when working with panel data with a moderately large time dimension (about 8 or higher). This gives confidence in the quality of the approximate solution described here in case relatively large number of time-periods are available for each individual in the panel. The leading alternative approach to the problem of initial conditions in dynamic panel data LDV models is that of Wooldridge (2005)[26].

$(\epsilon_{1t}, \epsilon_{2t}, \epsilon_{3t}, \eta_t)'$ is multivariate normally distributed, *i.i.d.* over time, with zero-mean. The functions $h_i(\cdot)$, $i = 1, 2, 3$, are known to the econometrician up to the vectors of parameters δ_1, δ_2 and ζ, which will be estimated.

The econometrician observes the (endogenous) variable Y_t, which is generated as follows:

$$Y_t = \begin{cases} y_{1t}^* & iff \quad y_{3t}^* \geq 0 \\ y_{2t}^* & iff \quad y_{3t}^* < 0 \end{cases} \tag{35}$$

In standard terminology, the two equations (32), $s = 1, 2$, are termed the "switched" equations and (33) the"switching" equation. Using the indicator function introduced above, we define the dummy variables $I_t \equiv \mathbf{1}\ (y_{3t}^* \geq 0)$ and $D_t \equiv \mathbf{1}\ (y_{4t}^* \geq 0)$. The econometrician observes D_t but not I_t. As long as $\sigma_\eta^2 > 0$, D_t is an imperfect measurement of I_t. In this sense, η_t can be thought of as errors in the coding of the regime information.

In its general form without measurement errors in regime classification, the switching-regression model was used by Lee (1978)[21] to study union/nonunion wage determination.[14] As Lee and Porter (1984)[22] explain, using inaccurate regime classification information in ML estimation leads to inconsistency. Moreover, Goldfeld and Quandt (1975)[9] show that if perfect information is not used, ML estimation is seriously inefficient.

Lee and Porter (1984)[22] allowed for a *constant* probability that observations were misclassified into the two regimes; their only explanatory variable in the switching equation, Z, was a constant. But assuming a constant probability of misclassification is inappropriate if one expects the probability of misclassification to vary over time, and especially so if one has exogenous information represented by Z_t, which, as theory suggests, should affect switching.

We model the misclassification probability as a monotonic function of the (unobservable) propensity of the industry to be in a particular regime measured by the latent variable y_{2t}^*. For example, in the disequilibrium version of the switching model (Fair and Jaffee (1972)[7], it seems plausible to assume that the probability of misclassification is smaller the larger the level of excess demand in the system. We demonstrate shortly that the coding error equation (34) incorporates this property into the model.

The contribution of an (independent) observation t to the likelihood function of the switching-regression model with coding error can be derived as follows. First observe that

$$\begin{array}{llllll} for \quad D_t = 1: & (y_{4t}^* \geq 0) & if \quad y_{3t}^* \geq 0, & \eta_t \geq -y_{3t}^* & Y_t = y_{1t}^* & (I_t = 1) \\ & & if \quad y_{3t}^* < 0, & \eta_t \geq -y_{3t}^* & Y_t = y_{2t}^* & (I_t = 2) \\ for \quad D_t = 2: & (y_{4t}^* < 0) & if \quad y_{3t}^* < 0, & \eta_t < -y_{3t}^* & Y_t = y_{2t}^* & (I_t = 2) \\ & & if \quad y_{3t}^* \geq 0, & \eta_t < -y_{3t}^* & Y_t = y_{1t}^* & (I_t = 1) \end{array} \tag{36}$$

[14] Fair and Jaffee (1972)[7], *inter alia*, used the model to analyze markets in disequilibrium.

Let us use the notation $p_{d|it} \equiv prob(D_t = d|I_t = i)$, $p_{dit} \equiv prob(D_t = d, I_t = i)$, $p_{dt} = prob(D_t = d)$, $\pi_{it} = prob(I_t = i)$, and $f_{it} = pdf(y^*_{i\,t})$, where d and i take values 1 or 2. For simplicity assume that ϵ_{1t} and ϵ_{2t} are independent of ϵ_{3t} and η_t.

Note that the $p_{d|i}$'s involve bivariate integrals of the form

$$p_{d|i} = \int\int_{S_{DI}} f(\epsilon_3, \theta) d\epsilon_3 d\theta \;/\; \int_{S_I} f(\epsilon_3) d\epsilon_3, \tag{37}$$

where $\theta \equiv \epsilon_2 - \eta$, and the regions of integration (as described in (36)) are the sets: $S_{DI} = \{\epsilon_2 \underset{I}{\overset{>}{<}} - Z\zeta, \eta \underset{D}{\overset{>}{<}} - (Z\zeta + \epsilon_2)\}$ and $S_I = \{\epsilon_2 \underset{I}{\overset{>}{<}} - Z\zeta\}$,
where $\underset{I}{\overset{>}{<}} \equiv \{\geq \text{ if } I = 1, < \text{ if } I = 0\}$ and $\underset{D}{\overset{>}{<}} \equiv \{\geq \text{ if } D = 1, < \text{ if } D = 0\}$.
The customary distributional assumption of normality is imposed.

The coding error model with the likelihood function defined by using (36)–(37) possesses the desired property that the misclassification probability is highest at the borderline case when a regime switch appears most likely, and falls monotonically as the exogenous classifying information becomes stronger. To see this, first note that the probabilities of misclassification are:

$$\begin{array}{ll} (D = 1|I = 2): & p_{1|2} = \Pr(\eta_t \geq -y^*_{3t}|y^*_{3t} < 0) \\ (D = 2|I = 1): & p_{2|1} = \Pr(\eta_t < -y^*_{3t}|y^*_{3t} \geq 0) \end{array} \tag{38}$$

Figures 1-3 in the Appendix present probability plots for the misclassification case of $D = 1$ and $I = 2$ as a function of the exogenous part of the switching equation, $Z\zeta$.[15] Various values of the standard deviation of the coding error η are considered. As can be seen from Figure 1, the conditional probability of misclassification, $p_{1|2}$, is monotonic in $Z\zeta$ in the desired direction, rising when the signal $Z\zeta$ tends to suggest the wrong regime more strongly. For example, when the true state of the system is $I = 2$, higher values of $Z\zeta$ are further at odds with the truth, hence $\text{Prob}(D = 1|I = 2)$ rises. As the standard deviation of the coding error η rises, the signal becomes less informative; in the limit, when $\sigma_\eta \to \infty$, the misclassification probabilities $(Prob(D = d|I = i), d \neq i,)$ approach 0.5 . Hence, we confirm that the switching model with coding error introduced here possesses the desired property that the misclassification probability falls as the tendency to lie in a particular regime rises. In Figure 2 we see that the joint probability of misclassification p_{12} has a unique mode at the least informative value of the signal, $Z\zeta = 0$, since in such a case it is most difficult to correctly classify the particular period.

An important caveat is that the coding-error switching-regression model allows only a limited degree of systematic misclassification. For example, despite the presence of the coding errors, the only change in the discrete part of the model, (38), is in the variance of the latent variable y^*_{3t}, which is, of course, unidentified. This is illustrated in Figure 3. Hence, one can obtain consistent estimates for ζ up to scale despite

[15] The corresponding plots for the case with $D = 3$ and $I = 1$ are exact mirror images with respect to $Z\zeta = 0$ of those in Figures 1-3 and are not given separately.

such misclassification.[16] This, however, does not imply that the presence of the coding error is unimportant, because ML estimation of the complete discrete/continuous switching-regression model would still yield inconsistent results if the measurement errors were neglected.[17]

4.1 A Markovian Switching Model with Imperfect Classification

Aiming for greater realism, we now introduce Markovian elements to the switching model introduced above. Because of the *i.i.d.* assumptions across i and t on the error vector $(\epsilon_{1t}, \epsilon_{2t}, \epsilon_{3t}, \eta_t)'$, the models of the previous section exhibit a Bernoulli switching structure, conditional on the exogenous variables. This is characterized by a transition matrix

$$\begin{array}{cccc} & I_t = 1 & I_t = 0 & \\ I_{t-1} = 1 & \tau_t & 1 - \tau_t & \textit{Bernoulli} \\ I_{t-1} = 1 & \tau_t & 1 - \tau_t & \end{array} \quad (40)$$

In (40) the transition probabilities τ's depend on time only through the exogenous variables, but not on the past state variable. Next we introduce a model that allows the switching process to exhibit Markov dependence over time.

If I_t is a Markov process, then it has the transition structure

[16] The importance of this restrictive feature of my measurement errors model will be investigated in future work.

[17] A natural extension of the model with imperfect regime classification allows M multiple indicators $D_1, ..., D_M$ of regime classification. This is the nonlinear analogue of the classic MIMIC model of Joreskog and Goldberger (1975)[20]. We then obtain 2^{M+1} categories with respect to $D_1, ..., D_M$, and I. For example, in the case of two imperfect classification indicators, I define $R \equiv Z\zeta$ and give the eight possibilities in that case:

D_1	D_2	I	$\epsilon_2 - \eta_1$	$\epsilon_2 - \eta_2$	ϵ_2	
1	1	1	$\leq R$	$\leq R$	$\leq R$	
1	1	0	$\leq R$	$\leq R$	$> R$	
1	0	1	$\leq R$	$> R$	$\leq R$	
1	0	0	$\leq R$	$> R$	$> R$	(39)
0	1	1	$> R$	$\leq R$	$\leq R$	
0	1	0	$> R$	$\leq R$	$> R$	
0	0	1	$> R$	$> R$	$\leq R$	
0	0	0	$> R$	$> R$	$> R$	

The likelihood contributions will in general involve $(M+1)$–fold integrals, which can be calculated by numerical methods for M up to 2 or 3. This modelling approach, like the coding-error model with a single indicator, (32)–(33), also has the desirable property that the misclassification probabilities vary over the sample period depending on the true probability of switching. The higher dimension integrals implied by multiple imperfect indicators can be accommodated by simulation estimation methods. See Hajivassiliou (1993)[11] for a discussion.

$$
\begin{array}{lccc}
 & I_t = 1 & I_t = 0 & \\
I_{t-1} = 1 & \tau_{11t} & 1 - \tau_{11t} & Markov \\
I_{t-1} = 1 & \tau_{10t} & 1 - \tau_{10t} &
\end{array} \tag{41}
$$

where $\tau_{ijt} = Prob(I_t = i | I_{t-1} = j)$.[18] Specifically, to introduce a Markov structure of order 1, we modify the switching equation so that the true propensity to switch, y^*_{2t}, depends on the lagged state I_{t-1}, i.e.,

$$
I_t = \begin{cases} 1 & \text{if } Z_t\zeta + \rho I_{t-1} + \epsilon_{2t} \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{42}
$$

With perfect classification information, this structure is straightforward to estimate since[19]

$$
p[\mathbf{Y}, \mathbf{I}, I_0 | \mathbf{X}] \equiv p[Y_1, \cdots, Y_T, I_1, \cdots, I_T, I_0 | X_1, \cdots, X_T] \tag{43}
$$

$$
= p[Y_T, I_T | I_{T-1}, X_T] \cdot p[Y_{T-1}, I_{T-1} | I_{T-2}, X_{T-1}] \cdots p[Y_2, I_2 | I_1, X_2] \cdot p[Y_1, I_1 | I_0, X_1] \cdot p[I_0].
$$

The likelihood function for process (43), however, becomes extremely intractable in the presence of imperfect regime-classification information because it then requires the evaluation of 2^T terms. The reason is as follows. We can readily show that

$$
p[D_t, I_t | I_{t-1}] = \tag{44}
$$

$$
I_t p[D_t, I_t = 1 | I_{t-1}] + (1 - I_t) p[D_t, I_t = 0 | I_{t-1}]
$$

$$
p[Y_t, D_t | I_{t-1}] = \tag{45}
$$

$$
D_t[f_1 I_t p[1,1|I_{t-1}] + f_0(1 - I_t) p[1,0|I_{t-1}]] + (1 - D_t)[f_1 I_t p[0,1|I_{t-1}] + f_0(1 - I_t) p[0,0|I_{t-1}]],
$$

where I_t is determined by (42). But the econometrician only observes D_t, given by

$$
D_t = \begin{cases} 1 & \text{if } Z_t\zeta + \rho I_{t-1} + \epsilon_{3t} + \eta_t \geq 0 \\ 0 & \text{otherwise.} \end{cases}
$$

Since I_{t-1} is unobserved by the econometrician for all t, the likelihood function is

$$
p[\mathbf{Y}, \mathbf{D} | \mathbf{X}] = \sum_{I_T} \sum_{I_{T-1}} \cdots \sum_{I_2} \sum_{I_1} \sum_{I_0} p[Y_T, D_t, I_T | I_{T-1}] \cdots p[Y_1, d_1, I_1 | I_0] \cdot p[I_0]. \tag{46}
$$

Because each pair of consecutive terms involves I_{t-1} , the likelihood $p[\mathbf{Y}, \mathbf{D} | \mathbf{X}]$ will in general require the evaluation of 2^T terms, a patently intractable task when T is of the order of 20-30 as it is frequently the case with several panel data sets. To solve this problem, we show in Appendix I how extending ideas in Cosslett and

[18] One expects positive serial persistence, in the sense of $\tau_{11t} > \tau_{10t}$.

[19] Note that it is not crucial how one treats $p[I_0]$, since this term has asymptotically vanishing influence. This is in contrast to the longitudinal data set case.

Lee (1985)[6] and Moran (1986)[24], a recursion relation can be derived that makes evaluation of (46) feasible.

Note again that the approach here differs fundamentally from earlier approaches in that the probability of misclassification is not constant but varies monotonically with the magnitude of $Z_t\zeta + \rho I_{t-1}$. A priori, this is a realistic feature. Given the dependence over time described in (42), one should expect the probability of misclassification to vary over time; it should be highest close to the boundary points when a switch occurs. These properties are exhibited by the conditional probability expressions above.[20]

5 Conclusions

This paper proposed efficient estimation methods for panel data LDV models possessing a variety of complications: non-ignorable persistent heterogeneity; contemporaneous and intertemporal endogeneity; observable and unobservable dynamics; and imperfect regime classification information. We first showed how a simple modification of estimators based on the Random Effects principle can preserve the consistency and asymptotic efficiency of the method in panel data despite non-ignorable persistent heterogeneity driven by correlations between the individual-specific component of the error term and the regressors. The approach is extremely easy to implement and allows straightforward tests of the significance of such correlations that lie behind the non-ignorable persistent heterogeneity. The method applies to linear as well as nonlinear panel data models, static or dynamic. In addition, the method works for time-invariant as well as time-varying regressors, and allows for the heterogeneity components to depend nonlinearly on regressors. These two features extend the existing literature in important dimensions. A particular focus of the approach was to analyze the presence of time-invariant regressors and to provide an interpretation of the coefficients of such regressors, which should prove especially useful for policy analysis and many real world applications.

We then combined this modified random effects approach with two simulation-based estimation strategies to overcome *analytical* as well as computational intractabilities in a widely applicable class of nonlinear models for panel data, namely the class of LDV models with contemporaneous and intertemporal endogeneity. We showed that the approach can be readily applied to general additive and non-additive nonlinear panel data models, which may be static or dynamic. For the dynamic case, the framework of this paper extends the Barghava and Sargan (1982)[2] approach to nonlinear dynamic nonlinear models. The simulation-based methods we employed were maximum simulated likelihood employing the GHK importance-sampling simu-

[20] There is a cost, however, in terms of computational complexity because the conditional probability expressions $p[D_t, I_t|I_{t-1}]$ now involve bivariate normal integrals (and in general $(M+1)$–fold integrals when M imperfect regime indicator variables are available).

lator and the method of simulated scores with Gibbs resampling. The effectiveness of the estimation methods in providing asymptotically efficient estimators in such cases was illustrated with three discrete-response econometric models for panel data: a simultaneous system determining a binary LDV indicator and a trinomial ordered LDV indicator; a model with simultaneous determination of two binary indicators with observable dynamic endogeneity; and a model with an important type of contemporaneous and intertemporal simultaneity due to strategic interactive effects over time across economic agents or financial crisis contagion across countries.

The final contribution of the paper was to discuss panel data LDV models with regime classification imperfections in the presence of Markovian state dependence, and to develop a novel algorithm that allows for the first time efficient maximum likelihood estimation of this class of models.

6 Appendix I: A Matrix Recursion for Markovian Switching with Imperfect Classification

Our aim is to facilitate evaluation of the likelihood function (46). The difficulty in evaluating it directly is that each pair of consecutive terms involves I_{t-1}; hence, each likelihood evaluation will require calculating 2^T terms, which is a computationally prohibitive task.[21]

Define the set of available endogenous information at time t by S_t, i.e., $S_t \equiv (y_1, D_1, y_2, D_2, ..., y_t, D_t)$. Further define $Q_t(I_t) \equiv p[S_t, I_t]$. Since we can always write

$$Q_t(I_t) = p[S_{t-1}, y_t, D_t, I_t] = \sum_{I_{t-1}} p[S_{t-1}, I_{t-1}, y_t, D_t, I_t], \tag{47}$$

it follows that

$$Q_t(I_t) = \sum_{I_{t-1}} p[y_t, D_t, I_t | I_{t-1}, S_{t-1}].p[I_{t-1}, S_{t-1}] = \sum_{I_{t-1}} p[y_t, D_t, I_t | I_{t-1}] \cdot Q_{t-1}(I_{t-1}), \tag{48}$$

where we have used the Markov structure $p[y_t, D_t, I_t | I_{t-1}, S_{t-1}] = p[y_t, D_t, I_t | I_{t-1}]$ and the definition $Q_{t-1}(I_{t-1}) \equiv p[I_{t-1}, S_{t-1}]$. But calculation of (48) only requires information up to t, as the following matrix equation shows:

$$\begin{pmatrix} Q_t(2) \\ Q_t(1) \end{pmatrix} = \tag{49}$$

$$\begin{pmatrix} p[y_t, I_t = 2 | I_{t-1} = 2] & p[y_t, I_t = 2 | I_{t-1} = 1] \\ p[y_t, I_t = 1 | I_{t-1} = 2] & p[y_t, I_t = 1 | I_{t-1} = 1] \end{pmatrix} \cdot \begin{pmatrix} Q_{t-1}(2) \\ Q_{t-1}(1) \end{pmatrix}$$

or,

$$\mathbf{Q}_t = \mathbf{M}_t \cdot \mathbf{Q}_{t-1}. \tag{50}$$

The likelihood (46) can thus be calculated recursively from (49) and

$$p[\mathbf{y}, \mathbf{D} | X] = \sum_{I_T} Q_T(I_T) = Q_T(0) + Q_T(1).$$

[21] A proof of identification of this imperfect regime classification switching model can be obtained from the author upon request.

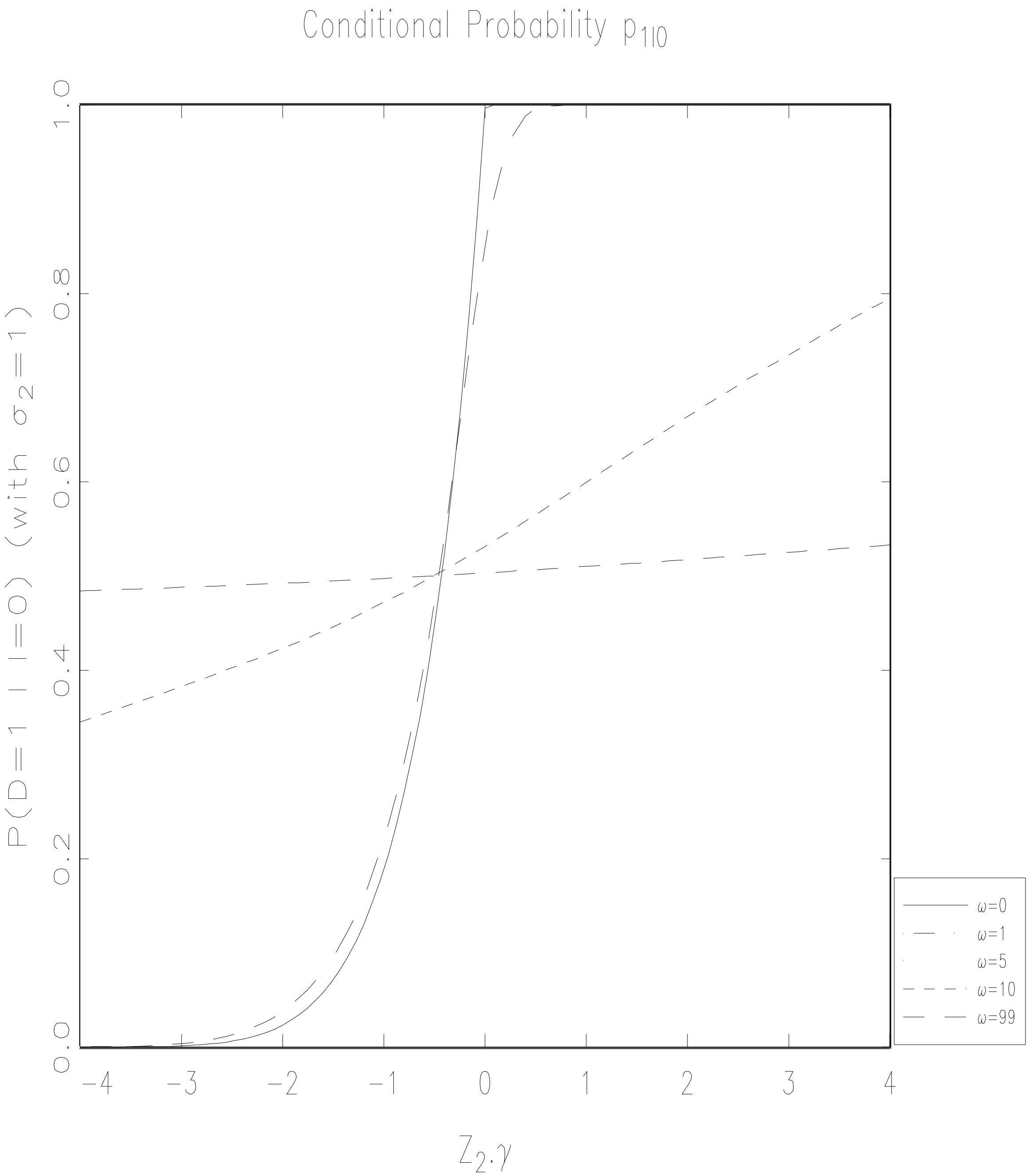

Figure 1: Conditional Probabilities

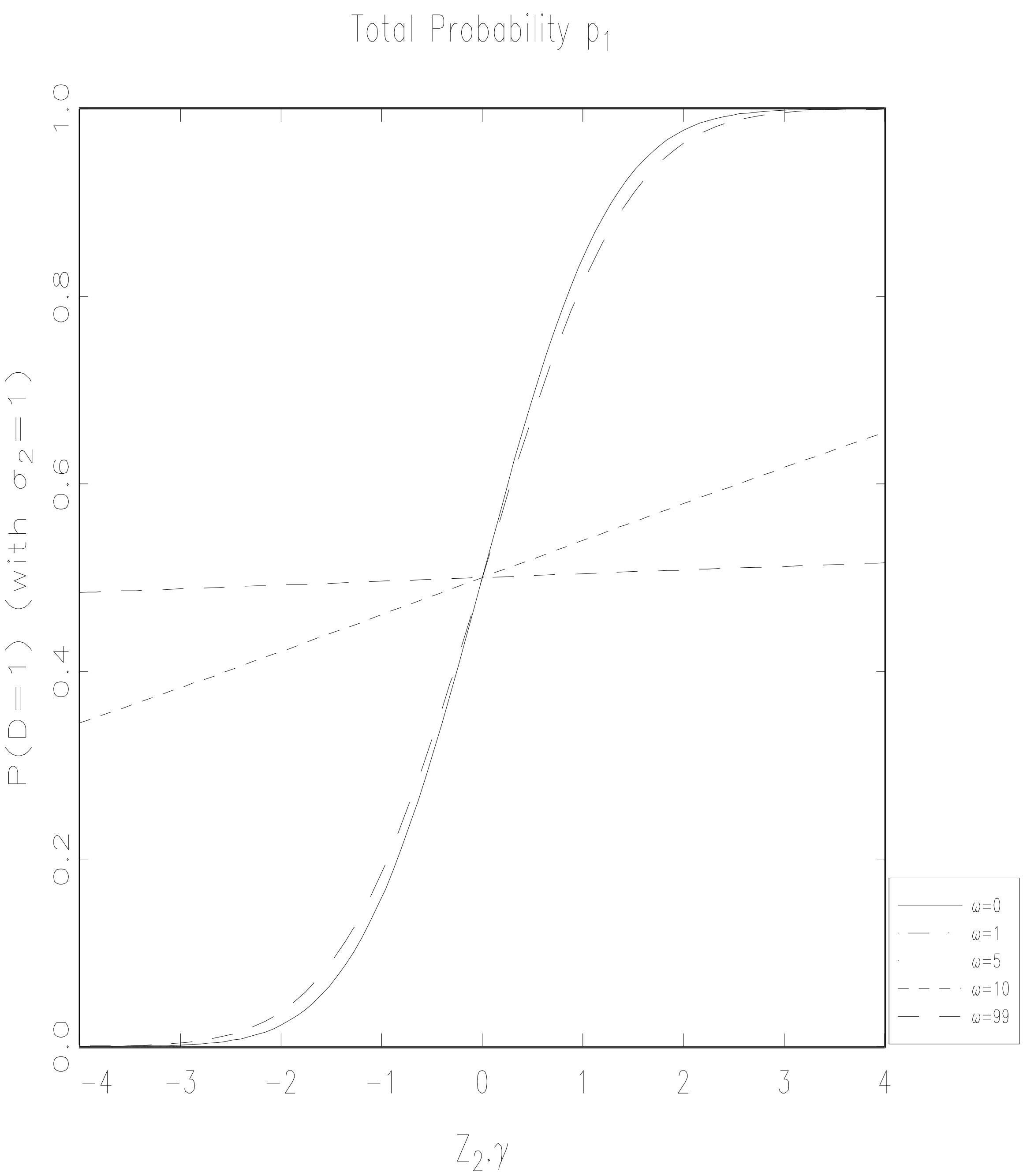

Figure 2: Total Probabilities

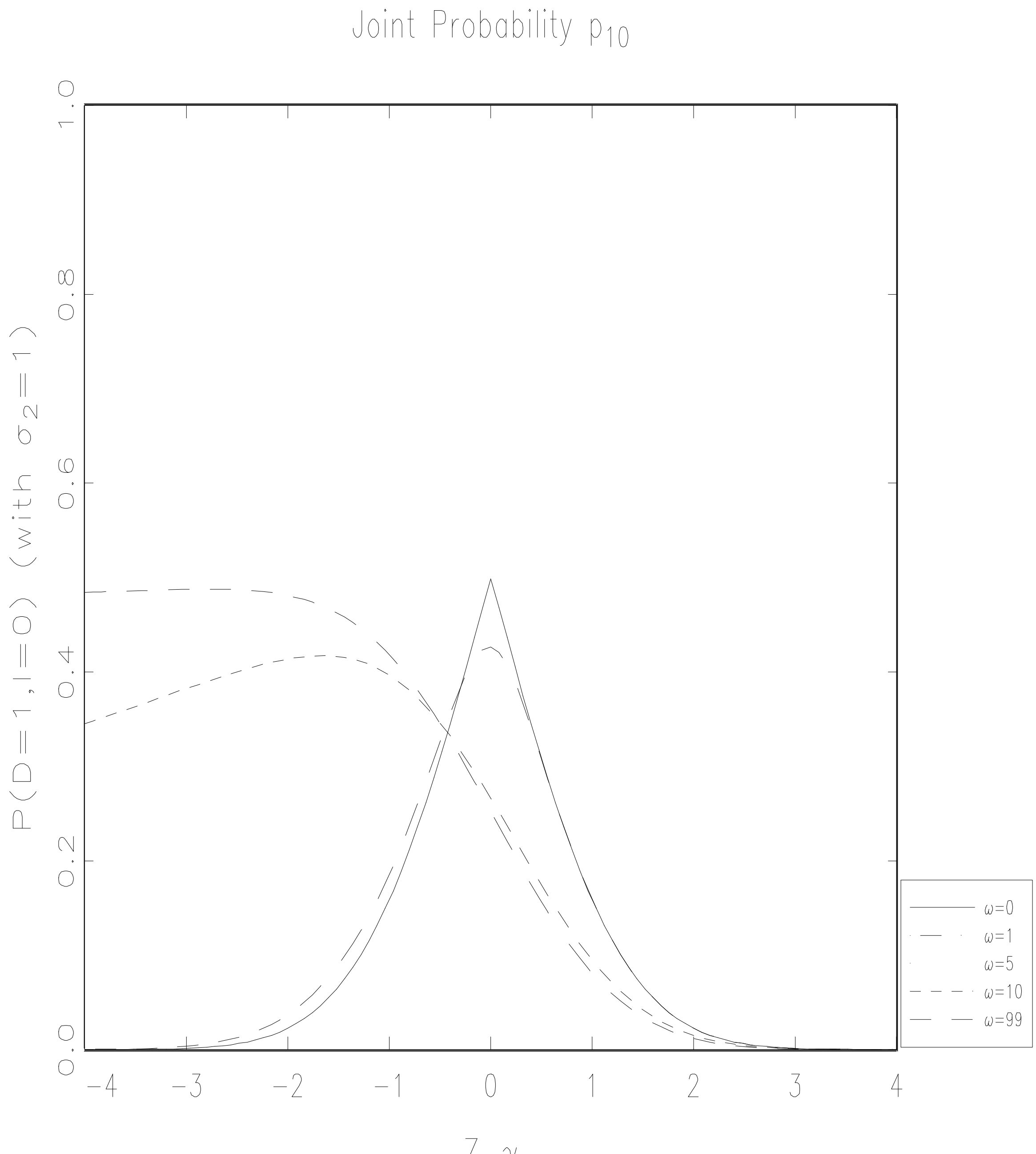

Figure 3: Joint Probabilities

References

[1] Arellano, M. and S. Bond (1991): "Test of Specification for Panel Data: Monte-Carlo Evidence and an Application to Employment," *Review of Economic Studies* 58, 277-297.

[2] Barghava, A. and J. Sargan (1982): "Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods, *Econometrica* **51**, 1635-1660.

[3] Börsch-Supan, A. and V. Hajivassiliou (1993): "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models," *Journal of Econometrics* 58(3), 347–368.

[4] Chamberlain, G. (1980): "Analysis of Covariance with Qualitative Data," *Review of Economic Studies*, **47**, 225-238.

[5] Chamberlain, G. (1984): "Panel Data", in Z. Griliches and M. D. Intriligator (eds.), *Handbook of Econometrics*, Vol. 2, Elsevier Science.

[6] Cosslett, S. and L.-F. Lee (1985): "Serial Correlation in Latent Discrete Variable Models," *Journal of Econometrics* 27, 79-97.

[7] Fair, R. and D. Jaffee (1972): "Methods of Estimation for Markets in Disequilibrium," *Econometrica* 40, 497-514.

[8] Falcetti, E. and M. Tudela (2007): "Modelling Currency Crises in Emerging Markets: A Dynamic Probit Model with Unobserved Heterogeneity and Autocorrelated Errors," *Oxford Bulletin of Economics and Statistics*, Volume 68, Number 4, August 2006 , 445-471(27).

[9] Goldfeld, S. and R. Quandt (1975): "Estimation of a Disequilibrium Model and the Value of Information," *Journal of Econometrics* 3, 325-348.

[10] Hajivassiliou, V. (1985): "Disequilibrium Modelling in Economics and Related Limited Dependent Variables Models," M.I.T. *unpublished doctoral dissertation.*

[11] Hajivassiliou, V. (1993): "Simulation Estimation Methods for Limited Dependent Variable Models." In *Handbook of Statistics*, **11** (Econometrics), G.S. Maddala, C.R. Rao and H.D. Vinod (eds.). Amsterdam: North-Holland, 519–543.

[12] Hajivassiliou, V. (2006): "A Modified Random Effects Estimator for Linear Panel Data Models with Regressor-Heterogeneity Correlations," *LSE Department of Economics working paper.*

[13] Hajivassiliou V. (2010): "Novel Approaches to Coherency Conditions in LDV Models", *LSE Department of Economics working paper.*

[14] Hajivassiliou, V., Y. Ioannides (2007): "Unemployment and Liquidity Constraints" *Journal of Applied Econometrics*, 22(3), 479-510.

[15] Hajivassiliou V., McFadden D.L. (1998): "The Method of Simulated Scores for the Estimation of LDV Models", *Econometrica*, 66(4), 863-896.

[16] Hajivassiliou, V., McFadden, D.L., and Ruud, P.A. (1996): "Simulation of Multivariate Normal Rectangle Probabilities and Derivatives: Theoretical and Computational Results", *Journal of Econometrics* 72(1&2), 85–134.

[17] Hausman, J. and W. Taylor (1981): "Panel Data and Unobservable Individual Effects," *Econometrica* **49**, 1377-1398.

[18] Heckman, J. (1981): "Dynamic Discrete Models," in Manski and McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, Massachusetts.

[19] Honoré, B. and E. Kyriazidou (2000): "Panel Data Discrete Choice Models with Lagged Dependent Variables", *Econometrica* 68, 839-874.

[20] Joreskog, K.G. and Goldberger, A.S. (1975): "Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable". *Journal of the American Statistical Association* 70, 631-639.

[21] Lee, L.-F. (1978): "Unionism and Wages: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables," *International Economic Review* 19, 415-433.

[22] Lee, L.-F. and R. Porter (1984): "Switching Regression Models with Imperfect Sample Separation Information — with an Application on Cartel Stability," *Econometrica* 52, 391-418.

[23] Liu, X., Lee, L.-F. and Kagel, J. (2008): "Learning from Peers in Signalling Game Experiments" Department of Economics,University of Colorado at Boulder working paper.

[24] Moran, P.A.P. (1986): "Orthant Probabilities and Gaussian Markov Processes," *Journal of Applied Probability*, 413-417.

[25] Mundlak, Y. (1978): "On the Pooling of Time Series and Cross Section Data," *Econometrica* **46**, 69-85.

[26] Wooldridge (2005): "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity", *Journal of Applied Econometrics* 20, 39-54.