Bivariate dynamic probit models for panel data

Alfonso Miranda

Institute of Education, University of London

2010 Mexican Stata Users Group meeting

April 29, 2010

Two related processes...

Often the applied researcher is interested in studying two longitudinal dichotomous variables that are closely related and likely to influence each other, y_{1it} and y_{2it}; $i = \{1, \ldots N\}, t = \{1, \ldots, T_i\}$.

- Ownership of Stocks and Mutual Funds (Alessie, Hochguertel, and Van Soest, 2004)
- Spouses smoking (Clark and Etilé, 2006)
- Marital status and the decision to have children (Mosconi and Seri, 2006)
- Migration and Education (Miranda, forthcoming 2011)
- Spouses obesity (Shigeki, 2008)
- Poverty and Social Exclusion (Devicienti and Poggi, 2007)

The main interest is on the *dynamics*...

- Do past holdings of stocks affect present holdings of mutual funds? Other way round?
- Does husband's past smoking affect wife's present smoking? Other way round?
- Do father's and siblings past migration affect an individuals' chances of high school graduation today?
- Do past poverty affect today's probability of employment?

Two challenges

Problem 1

Unobserved individual heterogeneity affecting y_{1it} may be correlated with unobserved individual heterogeneity affecting y_{2it}

Problem 2

Idiosyncratic shocks affecting y_{1it} may be correlated with indiosyncratic shocks affecting y_{2it}

Dynamic equations

$$y^*_{1it} = \mathbf{x}'_{1it}\boldsymbol{\beta}_1 + \delta_{11}y_{1it-1} + \delta_{12}y_{2it-1} + \eta_{1i} + \zeta_{1it} \quad (1)$$

$$y^*_{2it} = \mathbf{x}'_{2it}\boldsymbol{\beta}_2 + \delta_{21}y_{1it-1} + \delta_{22}y_{2it-1} + \eta_{2i} + \zeta_{2it} \quad (2)$$

with $y_{1it} = 1(y^*_{1it} > 0)$ and $y_{2it} = 1(y^*_{2it} > 0)$, $\mathbf{x}_{1it}$ and $\mathbf{x}_{2it}$ are $K_1 \times 1$ and $K_2 \times 1$ vectors of explanatory variables, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are vectors of coefficients, $\boldsymbol{\eta}_i = \{\eta_{1i}, \eta_{2i}\}$ are random variables representing unobserved individual heterogeneity (time-fixed), and $\boldsymbol{\zeta}_{it} = \{\zeta_{1it}, \zeta_{2it}\}$ are "idiosyncratic" shocks. We suppose $\boldsymbol{\eta}_i$ are jointly distributed with mean vector zero and covariance matrix,

$$\Sigma_\eta = \begin{bmatrix} \sigma_1^2 & \rho_\eta\,\sigma_1\sigma_2 \\ \rho_\eta\,\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

ζ_{it} are also jointly distributed with mean vector 0 and covariance,

$$\Sigma_\zeta = \begin{bmatrix} 1 & \rho_\zeta \\ \rho_\zeta & 1 \end{bmatrix}$$

True vs *spurious* state dependence...

Take the case of y_{1it}. Correlation between y_{1it} and y_{1it-1} and y_{2it-1} can be caused because of two different reasons:

True state dependence: y_{1it-1} and y_{2it-1} are genuine shifters of the conditional distribution of y_{1it} given $\boldsymbol{\eta}_i$

$$D(y_{1it}|y_{1it-1}, y_{2it-1}, \eta) \neq D(y_{1it}|\boldsymbol{\eta}_i)$$

Spurious state dependence: y_{1it-1} and y_{2it-1} are not genuine shifters of the conditional distribution of y_{1it} given $\boldsymbol{\eta}_i$

$$D(y_{1it}|y_{1it-1}, y_{2it-1}, \boldsymbol{\eta}_i) = D(y_{1it}|\boldsymbol{\eta}_i)$$

A similar argument applies to y_{2it}.

Initial conditions

Inconsistent estimators are obtained if y_{1i1} and y_{2i1} are treated as exogenous variables in the dynamic equations (initial cond. problem). A reduced-form model for the marginal probability of y_{1i1} and y_{2i1} given $\boldsymbol{\eta}_i$ is specified (Heckman 1981),

$$
\begin{aligned}
y^*_{1i1} &= \mathbf{z'}_1\boldsymbol{\gamma}_1 + \lambda_{11}\eta_{1i} + \lambda_{12}\eta_{2i} + \xi_{1i} \qquad (3) \\
y^*_{2i1} &= \mathbf{z'}_2\boldsymbol{\gamma}_2 + \lambda_{21}\eta_{1i} + \lambda_{22}\eta_{2i} + \xi_{2i} \qquad (4)
\end{aligned}
$$

with $y_{1i1} = 1(y^*_{1i1} > 0)$ and $y_{2i1} = 1(y^*_{2i1} > 0)$, $\mathbf{z}_1$ and $\mathbf{z}_2$ are $M_1 \times 1$ and $M_2 \times 1$ vectors of explanatory variables, and $\boldsymbol{\xi}_i = \{\xi_{1i}, \xi_{2i}\}$ are jointly distributed with mean 0 and covariance Σ_ξ

$$
\Sigma_\xi = \begin{bmatrix} 1 & \rho_\xi \\ \rho_\xi & 1 \end{bmatrix}
$$

Distributional assumptions

$$D(\boldsymbol{\eta}|\mathbf{x},\mathbf{z},\boldsymbol{\zeta},\boldsymbol{\xi}) = D(\boldsymbol{\eta}) \tag{C1}$$

$$D(\boldsymbol{\zeta}|\mathbf{x},\mathbf{z},\boldsymbol{\eta}) = D(\boldsymbol{\zeta}|\boldsymbol{\eta}) \tag{C2}$$

$$D(\boldsymbol{\xi}|\mathbf{x},\mathbf{z},\boldsymbol{\eta}) = D(\boldsymbol{\xi}|\boldsymbol{\eta}) \tag{C3}$$

$$\boldsymbol{\zeta} \perp \boldsymbol{\xi} \mid \boldsymbol{\eta} \tag{C4}$$

$$D(\boldsymbol{\zeta}_{it}|\boldsymbol{\zeta}_{is},\boldsymbol{\eta}) = D(\boldsymbol{\zeta}_{it}|\boldsymbol{\eta}) \quad \forall s \neq t \tag{C5}$$

$$D(\boldsymbol{\xi}_{it}|\boldsymbol{\xi}_{is},\boldsymbol{\eta}) = D(\boldsymbol{\xi}_{it}|\boldsymbol{\eta}) \quad \forall s \neq t \tag{C6}$$

Condition C1 is the usual random effects assumption. Conditions C1-C3 ensure that all explanatory variables are exogenous. Condition C4 ensures that idiosyncratic shocks in dynamic equations and initial conditions are independent given η. Finally, conditions C5-C6 rule out serial correlation for the two pairs of idiosyncratic shocks. Given that we have a Probit model we impose:

$$\boldsymbol{\eta} \sim BN(0,\Sigma_\eta); \;\; \boldsymbol{\zeta}|\boldsymbol{\eta} \sim BN(0,\Sigma_\zeta); \;\; \boldsymbol{\xi}|\boldsymbol{\eta} \sim BN(0,\Sigma_\xi)$$

Estimation

The model is estimated by Maximum Simulated Likelihood (see, for instance, Train 2003). The contribution of the ith individual to the likelihood is,

$$
\begin{aligned}
L_i &= \int\int \Phi_2\left(q_{1i0}w_{11}, q_{2i0}w_{12}, q_{1i0}q_{2i0}\rho_{\xi}\right) \\
&\quad \times \prod_{t=1}^{T_i} \Phi_2\left(q_{1it}w_{21}, q_{2it}w_{22}, q_{1it}q_{2it}\rho_{\zeta}\right) g\left(\boldsymbol{\eta}_i, \Sigma_{\eta}\right) d\eta_{1i} d\eta_{2i}
\end{aligned}
$$

where $g(.)$ represents the bivariate normal density, $q_{1it} = 2y_{1it} - 1$, $q_{2it} = 2y_{2it} - 1$. Finally, w_{11} and w_{12} are the right-hand side of (3) and (4) excluding the idiosyncratic shocks. And w_{21} and w_{22} are defined in the same fashion using (1) and (2).

- Maximum simulated likelihood is asymptotically equivalent to ML as long as the number of draws R grows faster than $\sqrt{N}$ (Gourieroux and Monfort 1993)
- Use Halton sequences for simulation instead of uniform pseudo-random sequences
  - Better coverage of the [0,1] interval
  - Need less draws to achieve high precision
- Maximisation based on Stata's Newton-Raphson algorithm using either
  - Analytical first derivatives and numerical second derivatives (d1 method),
  - Analytical first derivatives and OPG approximation of the covariance matrix (BHHH algorithm implemented as a d2 method)
  - Really fast!!!

Let's use some simulated data...

- 2000 individuals
- 4 observations per individual
- rho_eta = 0.25
- rho_zeta = 0.33
- rho_xi = 0.25
- SEeta1 = sqrt(0.30)
- SEeta2 = sqrt(0.62)
- eta1 and eta2 jointly distributed as bivariate normal
- xi1 and x2 jointly distributed as bivariate normal
- zeta1 and zeta2 jointly distributed as bivariate normal
- x1, x2, x3, x4, xvar distributed as iid standard normal variates

Initial contions

```
y1star = 0.35 + 0.5*x1 + 0.72*x2 + 0.55*x3 + 0.64*eta1 ///
+ 0.32*eta2 + xi1 + if _n==1

y2star= 0.58 + 0.98*x1 - 0.67*x2 + 0.11*eta1 + 0.43*eta2 ///
+ xi2 if _n==1

by ind: replace y1 = (y1star>0) if _n==1
by ind: replace y2 = (y2star>0) if _n==1
```

Dynamic equations

```
#delimit ;
forval i = 2/4 {;
by ind: replace y1star = 0.42 + 0.93*x1 + 0.45*x2 - 0.64*x3 ///
+ 0.6*x4 + 0.43*y1[`i'-1] - 0.55*y2[`i'-1] + 0.21*xvar ///
+ 0.63*y1[`i'-1]*xvar + eta1 + zeta1 if _n==`i';

by ind: replace y2star = 0.65 + 0.27*x1 + 0.42*x4 ///
- 0.88*y1[`i'-1] + 0.54*y2[`i'-1] + 0.72*xvar ///
- 0.42*xvar*y1[`i'-1] + 0.5*xvar*y2[`i'-1] + eta2 ///
+ zeta2 if _n==`i';

by ind: replace y1 = (y1star>0) if _n==`i';
by ind: replace y2 = (y2star>0) if _n==`i';
};
#delimit cr
```

```
. #delimit ;
.  bprinit_v2 (y1 = x1 x2 x3 x4 y1lag y2lag xvar y1lagxvar y2lagxvar) (y2 = x1
> x4 y1lag y2lag xvar y1lagxvar y2lagxvar),
>   rep(200) id(ind) init1(x1 x2 x3) init2(x1 x2) hvec(2 1 2 100);
```

(output omitted)

```
Bivariate Dynamic RE Probit -- Maximum Simulated Likelihood
(# Halton draws = 200)

                                          Number of level 2 obs =      2000
                                          Number of level 1 obs =      8000
                                          Log likelihood        =   -7256.8
```

	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
init_y1						
x1	.5409808	.0438411	12.34	0.000	.4550538	.6269077
x2	.7443919	.0457859	16.26	0.000	.6546533	.8341306
x3	.5972203	.0420895	14.19	0.000	.5147265	.6797142
_cons	.3529803	.0381407	9.25	0.000	.2782259	.4277348
y1						
x1	.8837039	.0360177	24.54	0.000	.8131106	.9542972
x2	.4222031	.0264601	15.96	0.000	.3703423	.4740638
x3	-.6762835	.0305998	-22.10	0.000	-.736258	-.616309
x4	.6189321	.0308011	20.09	0.000	.558563	.6793011
y1lag	.4368135	.0566347	7.71	0.000	.3258116	.5478154
y2lag	-.5646897	.0610486	-9.25	0.000	-.6843427	-.4450367
xvar	.2562871	.0416498	6.15	0.000	.174655	.3379192
y1lagxvar	.5829502	.0527182	11.06	0.000	.4796244	.686276
y2lagxvar	-.0370886	.0518627	-0.72	0.475	-.1387377	.0645605
_cons	.3648562	.0524913	6.95	0.000	.261975	.4677373

```
init_y2      |
          x1 |   1.016066   .0522946    19.43   0.000     .9135701    1.118561
          x2 |  -.6425204   .0415074   -15.48   0.000    -.7238733   -.5611675
       _cons |    .602965   .0404014    14.92   0.000     .5237798    .6821502
-------------+----------------------------------------------------------------
y2           |
          x1 |    .262682   .0244236    10.76   0.000     .2148126    .3105514
          x4 |   .4210255   .0265955    15.83   0.000     .3688992    .4731518
       y1lag |  -.8462671   .0599055   -14.13   0.000    -.9636798   -.7288544
       y2lag |   .4303569   .0637957     6.75   0.000     .3053198    .5553941
        xvar |   .7336143    .049089    14.94   0.000     .6374016    .8298269
   y1lagxvar |  -.4455717   .0576863    -7.72   0.000    -.5586348   -.3325087
   y2lagxvar |   .5443257   .0571247     9.53   0.000     .4323633    .6562881
       _cons |   .7657639   .0650256    11.78   0.000      .638316    .8932118
-------------+----------------------------------------------------------------
   lambda_11 |    .602882    .186313     3.24   0.001     .2377153    .9680487
   lambda_12 |   .2849407   .0793151     3.59   0.000     .1294859    .4403954
   lambda_21 |   .0515264    .156512     0.33   0.742    -.2552316    .3582843
   lambda_22 |   .3900766   .0747893     5.22   0.000     .2434922    .5366609
-------------+----------------------------------------------------------------
    SE(eta1) |   .5496802   .0618331     8.89   0.000     .4409193    .6852691
    SE(eta2) |   .8959895   .0620171    14.45   0.000     .7823225    1.026172
     rho_eta |   .2993541   .0909566     3.29   0.001     .1125119    .4657503
-------------+----------------------------------------------------------------
      rho_xi |   .3069255   .0561037     5.47   0.000     .1932879    .4124374
    rho_zeta |    .354956   .0428158     8.29   0.000      .268353    .4358675
------------------------------------------------------------------------------
Likelihood ratio test for rho_eta=rho_xi=rho_zeta=0: chi2=444.90 pval = 0.000
------------------------------------------------------------------------------

. \#delimit cr
delimiter now cr
```

- The h() option deals with the Halton draws
  - first number sets the number of columns in the vector h
  - second and third number sets the columns that will be used for the MSL algorithm (first and second columns in this case)
  - third number sets the number of rows of vector h that will be discarded
    - number of rows of h = number of repetitions + last argument of the h() option

- Lagged dependent variables are just added as additional explanatory variables
  - Can naturally interact lagged dependent variables with other controls
  - Can add any function of the lagged explanatory variables — Will be OK as long as all the distributional assumptions are met

Discussion

Main advantage: Correlated time-fixed (individual specific) and time varying (idiosincratic shocks) unobserved heterogeneity affecting y_{1it} and y_{2it} are explicity modelled

Main disadvantage: Model is complex (4 equations). Formally identified by functional form but may suffer from *tenous identification* problems (Keane 1992)

- Need to nominate a number of *credible exclusion restrictions*. Using time varying variables to specify exclusion restrictions is, when possible, the way forward

Extensions

With minor modfifications to this model one can deal with:

- **Sample selection model for panel data** that corrects for selectivity issues due to:
  - Correlated individual specific unobserved heterogeneity
  - Correlatated idyosincratic shocks
- **Endogenous Treatment Effects for panel data**
  - 1 treatment dummy, 1 main response variable. Main response can be continous or ordinal.
- **Ordinal dependent variables**

References

- Alessie, R., Hochguertel, S., Van Soest, A., 2004. Ownership of Stocks and Mutual Funds: A Panel Data Analysis. The Review of Economics And Statistics 86, 783-796.
- Clark, AE., Etilé, F., 2006. Don't give up on me baby: Spousal correlation in smoking behaviour. Journal of Health Economics 25, 958-978.
- Devicienti, F., Poggi, A., 2007. Poverty and Social Exclusion: Two Sides of the Same Coin or Dynamically Interrelated Processes? Laboratorio R. Revelli Working Paper No. 62.
- Gourieroux, C., and Monfort, A., 1993. Simulation-based inference: A survey with special reference to panel data models. Journal of Econometrics 59, 5–33.
- Heckman, JJ., 1981. The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process. Structural Analysis of Discrete Data with Econometric Applications. MIT Press.
- Keane, M., 1992. A Note on Identification in the Multinomial Probit Model. Journal of Business & Economic Statistics 10, 193-200.
- Miranda, A., 2011. Migrant networks, migrant selection, and high school graduation in Mexico. Research in Labor Economics (in press)

- Mosconi, R., Seri, R., 2006. Non-causality in bivariate binary time series. Journal of Econometrics 132, 379–407.
- Shigeki, K., 2008. Like Husband, Like Wife: A Bivariate Dynamic Probit: Analysis of Spousal Obesities. College of Economics, Osaka Prefecture University. Manuscript.
- Train, KE., 2003. Discrete choice methods with simulation. Cambridge university press.