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*...
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

\( y_{1it}^* = x'_{1it}\beta_1 + \delta_{11}y_{1i(t-1)} + \delta_{12}y_{2i(t-1)} + \eta_i + \zeta_{1it} \) (1)

\( y_{2it}^* = x'_{2it}\beta_2 + \delta_{21}y_{1i(t-1)} + \delta_{22}y_{2i(t-1)} + \eta_i + \zeta_{2it} \) (2)

with \( y_{1it} = 1(y_{1it}^* > 0) \) and \( y_{2it} = 1(y_{2it}^* > 0) \), \( x_{1it} \) and \( x_{2it} \) are \( K_1 \times 1 \) and \( K_2 \times 1 \) vectors of explanatory variables, \( \beta_1 \) and \( \beta_2 \) are vectors of coefficients, \( \eta_i = \{\eta_{1i}, \eta_{2i}\} \) are random variables representing unobserved individual heterogeneity (time-fixed), and \( \zeta_{it} = \{\zeta_{1it}, \zeta_{2it}\} \) are “idiosyncratic” shocks. We suppose \( \eta_i \) are jointly distributed with mean vector zero and covariance matrix,

\[
\Sigma_\eta = \begin{bmatrix}
\sigma_1^2 & \rho_\eta \sigma_1 \sigma_2 \\
\rho_\eta \sigma_1 \sigma_2 & \sigma_2^2
\end{bmatrix}
\]

\( \zeta_{it} \) 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 $\eta_i$

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

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

$$D(y_{1it}|y_{1it-1}, y_{2it-1}, \eta_i) = D(y_{1it}|\eta_i)$$

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 $\eta_i$ is specified (Heckman 1981),

\begin{align}
    y_{1i1}^* &= z_1' \gamma_1 + \lambda_{11} \eta_{1i} + \lambda_{12} \eta_{2i} + \xi_{1i} \\
    y_{2i1}^* &= z_2' \gamma_2 + \lambda_{21} \eta_{1i} + \lambda_{22} \eta_{2i} + \xi_{2i}
\end{align}

(3) 
(4)

with $y_{1i1} = 1(y_{1i1}^* > 0)$ and $y_{2i1} = 1(y_{2i1}^* > 0)$, $z_1$ and $z_2$ are $M_1 \times 1$ and $M_2 \times 1$ vectors of explanatory variables, and $\xi_i = \{\xi_{1i}, \xi_{2i}\}$ are jointly distributed with mean 0 and covariance $\Sigma_\xi$

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

\[
D(\eta|x, z, \zeta, \xi) = D(\eta) \quad (C1)
\]
\[
D(\zeta|x, z, \eta) = D(\zeta|\eta) \quad (C2)
\]
\[
D(\xi|x, z, \eta) = D(\xi|\eta) \quad (C3)
\]
\[
\zeta \perp \xi \mid \eta \quad (C4)
\]
\[
D(\zeta_{it}|\zeta_{is}, \eta) = D(\zeta_{it}|\eta) \quad \forall s \neq t \quad (C5)
\]
\[
D(\xi_{it}|\xi_{is}, \eta) = D(\xi_{it}|\eta) \quad \forall s \neq t \quad (C6)
\]

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 $\eta$. Finally, conditions C5-C6 rule out serial correlation for the two pairs of idiosyncratic shocks. Given that we have a Probit model we impose:

\[
\eta \sim BN(0, \Sigma_\eta); \quad \zeta|\eta \sim BN(0, \Sigma_\zeta); \quad \xi|\eta \sim BN(0, \Sigma_\xi)
\]
Estimation

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

\[
L_i = \int \int \Phi_2 \left( q_{1i0} w_{11}, q_{2i0} w_{12}, q_{1i0} q_{2i0} \rho_\xi \right) \\
\times \prod_{t=1}^{T_i} \Phi_2 \left( q_{1it} w_{21}, q_{2it} w_{22}, q_{1it} q_{2it} \rho_\zeta \right) g \left( \eta_i, \Sigma_\eta \right) d\eta_{1i} d\eta_{2i}
\]

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
- $\rho_{\eta} = 0.25$
- $\rho_{\zeta} = 0.33$
- $\rho_{\xi} = 0.25$
- $SE_{\eta_1} = \sqrt{0.30}$
- $SE_{\eta_2} = \sqrt{0.62}$
- $\eta_1$ and $\eta_2$ jointly distributed as bivariate normal
- $\xi_1$ and $\xi_2$ jointly distributed as bivariate normal
- $\zeta_1$ and $\zeta_2$ jointly distributed as bivariate normal
- $x_1, x_2, x_3, x_4, x\text{var}$ distributed as iid standard normal variates
Initial conditions

\[ y_{1\text{star}} = 0.35 + 0.5x_1 + 0.72x_2 + 0.55x_3 + 0.64\eta_1 + 0.32\eta_2 + \xi_1 + \text{if } n==1 \]
\[ y_{2\text{star}} = 0.58 + 0.98x_1 - 0.67x_2 + 0.11\eta_1 + 0.43\eta_2 + \xi_2 \text{ if } n==1 \]

by ind: replace \( y_1 = (y_{1\text{star}}>0) \) if \( n==1 \)
by ind: replace \( y_2 = (y_{2\text{star}}>0) \) 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.    | 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 | .0518609 | -0.72 | 0.475                | -.1387377 .0645605 |
| _cons    | .3648562  | .0524913 | 6.95  | 0.000                | .261975 .4677373  |
```
### Estimation

**init_y2**

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>x1</td>
<td>1.016066</td>
<td>.0522946</td>
<td>19.43</td>
<td>0.000</td>
<td>.9135701</td>
</tr>
<tr>
<td>x2</td>
<td>-.6425204</td>
<td>.0415074</td>
<td>-15.48</td>
<td>0.000</td>
<td>-.7238733</td>
</tr>
<tr>
<td>_cons</td>
<td>.602965</td>
<td>.0404014</td>
<td>14.92</td>
<td>0.000</td>
<td>.5237798</td>
</tr>
</tbody>
</table>

**y2**

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>x1</td>
<td>.262682</td>
<td>.0244236</td>
<td>10.76</td>
<td>0.000</td>
<td>.2148126</td>
</tr>
<tr>
<td>x4</td>
<td>.4210255</td>
<td>.0265955</td>
<td>15.83</td>
<td>0.000</td>
<td>.3688992</td>
</tr>
<tr>
<td>y1lag</td>
<td>-.8462671</td>
<td>.0599055</td>
<td>-14.13</td>
<td>0.000</td>
<td>-.9636798</td>
</tr>
<tr>
<td>y2lag</td>
<td>.4303569</td>
<td>.0637957</td>
<td>6.75</td>
<td>0.000</td>
<td>.3053198</td>
</tr>
<tr>
<td>xvar</td>
<td>.7336143</td>
<td>.049089</td>
<td>14.94</td>
<td>0.000</td>
<td>.6374016</td>
</tr>
<tr>
<td>y1lagxvar</td>
<td>-.4455717</td>
<td>.0657863</td>
<td>-7.72</td>
<td>0.000</td>
<td>-.5586348</td>
</tr>
<tr>
<td>y2lagxvar</td>
<td>.5443257</td>
<td>.0571247</td>
<td>9.53</td>
<td>0.000</td>
<td>.4323633</td>
</tr>
<tr>
<td>_cons</td>
<td>.7657639</td>
<td>.0650256</td>
<td>11.78</td>
<td>0.000</td>
<td>.638316</td>
</tr>
</tbody>
</table>

**lambda**

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>lambda_11</td>
<td>.602882</td>
<td>.186313</td>
<td>3.24</td>
<td>0.001</td>
<td>.2377153</td>
</tr>
<tr>
<td>lambda_12</td>
<td>.2849407</td>
<td>.0793151</td>
<td>3.59</td>
<td>0.000</td>
<td>.1294859</td>
</tr>
<tr>
<td>lambda_21</td>
<td>.0515264</td>
<td>.156512</td>
<td>0.33</td>
<td>0.742</td>
<td>-.2552316</td>
</tr>
<tr>
<td>lambda_22</td>
<td>.3900766</td>
<td>.0747893</td>
<td>5.22</td>
<td>0.000</td>
<td>.2434922</td>
</tr>
</tbody>
</table>

**SE(eta)**

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SE(eta1)</td>
<td>.5496802</td>
<td>.0618331</td>
<td>8.89</td>
<td>0.000</td>
<td>.4409193</td>
</tr>
<tr>
<td>SE(eta2)</td>
<td>.8959895</td>
<td>.0620171</td>
<td>14.45</td>
<td>0.000</td>
<td>.7823225</td>
</tr>
<tr>
<td>rho_eta</td>
<td>.2993541</td>
<td>.0909566</td>
<td>3.29</td>
<td>0.001</td>
<td>.1125119</td>
</tr>
</tbody>
</table>

**rho**

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>rho_xi</td>
<td>.3069255</td>
<td>.0561037</td>
<td>5.47</td>
<td>0.000</td>
<td>.1932879</td>
</tr>
<tr>
<td>rho_zeta</td>
<td>.354956</td>
<td>.0428158</td>
<td>8.29</td>
<td>0.000</td>
<td>.268353</td>
</tr>
</tbody>
</table>

**Likelihood ratio test for rho_eta=rho_xi=rho_zeta=0: chi2=444.90 pval = 0.000**
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 (idiosyncratic shocks) unobserved heterogeneity affecting $y_{1it}$ and $y_{2it}$ are explicitly 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 modifications 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
  - Correlated idiosyncratic shocks

- **Endogenous Treatment Effects for panel data**
  - 1 treatment dummy, 1 main response variable. Main response can be continuous or ordinal.

- **Ordinal dependent variables**
References

