The DPB function package

Jack Lucchetti Claudia Pigini

April 13, 2015

Contents

1 Description **2**
1.1 List of abbreviations 2

2 List of public functions **3**

3 Contents of the model bundle **6**

1 Description

The DPB (*Dynamic Panel Binary*) function package provides the functions to estimate dynamic models for binary panel data by both fixed- and random-effects approaches. The random-effects models contained in DPB are the dynamic probit with linearised initial condition proposed in Heckman (1981) and the generalisations by Hyslop (1999) and Keane and Sauer (2009). DPB also contains the software for the estimation of the quadratic exponential model in Bartolucci and Nigro (2010). The description of the models' formulations, the illustration of the function package as well as empirical examples can be found in the preliminary draft "DPB: Dynamic Panel Binary data models in gretl". In this document, we provide a detailed list of the public functions in DPB and the description of the bundle elements.

1.1 List of abbreviations

- Models:
  - DP: Dynamic Probit model (Heckman, 1981)
  - ADP: AR(1) Dynamic Probit model (Hyslop, 1999)
  - GADP: Generalised AR(1) Dynamic Probit model (Keane and Sauer, 2009)
  - QE: Quadratic Exponential model Bartolucci and Nigro (2010)
- Algorithms:
  - GHQ: Gauss-Hermite Quadrature method (Butler and Moffitt, 1982)
  - GHK: Geweke (1989); Hajivassiliou and McFadden (1998); Keane (1994) algorithm

2 List of public functions

`DPB_setup(string mod, series depvar, list X, list Z[null])`

Return type: `bundle`
The bundle set up function is mandatory. The function arguments are:

1. `string mod`:

   `"DP"` = Dynamic Probit model (Heckman, 1981, default choice)
   `"ADP"` = AR(1) Dynamic Probit model (Hyslop, 1999)
   `"GADP"` = Generalised AR(1) Dynamic Probit model (Keane and Sauer, 2009)
   `"QE"` = Quadratic Exponential model (Bartolucci and Nigro, 2010)

2. `series y`: the binary dependent variable
3. `list X`: list of explanatory variables
4. `list Z`: list of explanatory variables for the initial condition equation in DP, ADP and GADP models (optional)

Choosing `mod = "ADP"` or `mod = "GADP"` forces the calculation of multivariate normal probabilities by GHK.

Here we:

a. subset the sample. Assuming that the dataset is already endowed with the appropriate panel structure, observations present for at least two periods are selected for the primary equation. For the DP, ADP and GADP models, observations that are present for just one period are used as well in the initial condition equation. For the QE model, matrices containing sufficient statistics by observations are built;

b. construct the data matrices and store them into the model bundle;

c. handle default settings:

   c.1 set the default algorithm for the computation of multivariate normal probabilities (GHQ or GHK); the number of quadrature point or Halton/uniform draws. For the DP model, the default choice is GHQ with 24 quadrature points, for the ADP and GADP models, the default choice is GHK with 128 draws
   c.2 set the default sequence for the GHK algorithm (Halton) for the ADP and GADP models
   c.3 set the default value for the covariance matrix estimation (Sandwich)
   c.4 set the default reporting of results: the value of the log-likelihood is printed at each iteration.

`DPB_setoption(bundle *b, string opt, scalar value)`

Return type: `scalar`

This function is for setting: method of computation of multivariate normal probabilities, number of Gauss-Hermite quadrature points or GHK draws, type of sequence for the GHK algorithm, type of variance-covariance matrix estimator, verbosity level.

The function arguments are:

- `bundle *b`: pointer to the model bundle
- `string opt, scalar value`: a string indicating which option to set and a scalar taking values accordingly:

`opt`	`value`
`"method"`	0 = GHQ, default choice for the DP model, 1 = GHK algorithm. For the ADP and GADP models, `method` is forced to 1. For the QE model a warning message is printed
`"nrep"`	number of quadrature points or GHK draws. Default is 24 for the DP model with GHQ, 128 for the DP model with GHK and for the ADP and GADP models. For the QE model a warning message is printed
`"draws"`	type sequence for the GHK algorithm 0 = Halton (default), 1 = Uniform with seed `31415927`. For the DP model with GHQ and the QE model a warning message is printed.
`"vcv"`	parameters covariance matrix 0 = Sandwich (default), 1 = Outer Product of the Gradient (OPG), 2 = Hessian
`"verbose"`	degree of output verbosity 0 = no output is printed, 1 = the log-likelihood at each iterations is printed (default), 2 = log-likelihood, parameters and gradient at each iteration are printed

The function returns a scalar equal to zero if the option has been successfully set and an error code (from 1 to 5) otherwise. Warning messages associated with the error codes are printed and related default settings are kept.

`DPB_estimate(bundle *bun, matrix *par[null]`

Return type: `void`

For the DP, ADP and GADP models, Maximum Likelihood estimation is performed calling the `gretl`-native function `BFGSmax`. For the QE model, Conditional Maximum Likelihood is performed calling the `gretl`-native function `NRmax`. After convergence is achieved, the estimated variance-covariance is computed as per the option supplied via `DPB_setoption`.

The function arguments are

- `bundle *bun`: pointer to the model bundle;
- `matrix *par[null]`: a matrix with parameters initial values (optional). If not supplied, initial values are computed by a linear probability model for models DP, ADP, and GADP. For the QE model, parameters are initialised to zero.

`DPB_printout(bundle *b)`

Return type: `void`

Prints the estimation result.

3 Contents of the model bundle

Name	Type	Purpose
draws	boolean	type of GHK draws (see DPB_setoption)
feedback	boolean	1 = report the log-lik. at each iteration; 0 = quiet maximisation
method	boolean	computation of multivariate integrals (see DPB_setoption)
AR1	integer	0 = no autocorr. (DP model), 1 = first-order autocorr. (ADP model) , 2 = generalised first-order autocorr. (GADP model)
vcvmeth	integer	covariance matrix estimation type (see DPB_setoption)
verbose	integer	output management (see DPB_setoption)
aic	scalar	Akaike Information Criterion
bic	scalar	Bayes Information Criterion
hqc	scalar	Hannan-Quinn Information Criterion
ll	scalar	log-likelihood at convergence
nk	scalar	number of primary equation parameters
nz	scalar	number of initial condition equation parameters (DP, ADP, GADP models)
nrep	scalar	number of GHQ/GHK points/draws (DP, ADP, GADP models)
npar	scalar	number of parameters
NT	scalar	number of total observations (DP, ADP, GADP models); number of observations used for likelihood contributions (QE model)
N	scalar	number of total units (DP, ADP, GADP models); number of likelihood contributions (QE model)
Tot_NT	scalar	number of total observations (QE model)
Tot_N	scalar	number of total units (QE model)
valid	series	dummy for valid consecutive observations
model	string	model type (see DPB_setup)
Xnames	string	name of explanatory variables in the primary equation
yname	string	name of the dependent variable
Znames	string	name of explanatory variables in the initial condition equation (DP, ADP and GADP models)
coeff	matrix	vector of estimated parameters
consec	matrix	number of consecutive ones in the dependent variable (QE model)
G	matrix	Score matrix by observation (QE model)
InfoMat	matrix	Hessian matrix (QE model)
inipar	matrix	vector of initial values
mX	matrix	explanatory variables (QE model)
my	matrix	dependent variable (QE model)
POS	matrix	(QE model)
sel	matrix	binary nk-vector: 0 if the corresponding column of X1 has been dropped due to collinearity, 1 otherwise
sel_aux	matrix	binary nz-vector: 0 if the corresponding column of Z0 has been dropped due to collinearity, 1 otherwise for the DP, ADP, GADP models. $nk+1$-vector referring to the variables in the last observation for the QE model.
sumy	matrix	sufficient statistics (QE model)
Ti	matrix	N rows with i-th row: i, # obs for the primary equation, # obs for the initial condition equation, total # of usable observations
vcv	matrix	estimated covariance matrix
which	matrix	(QE model)
X1	matrix	explanatory variables in the primary equation (DP, ADP and GADP models)
y1	matrix	dependent variable in the primary equation (DP, ADP, GADP models)
y0	matrix	dependent variable in the initial condition equation (DP, ADP, GADP models)
Z0	matrix	explanatory variables in the initial condition equation (DP, ADP, GADP models)

References

Bartolucci, F. and Nigro, V. (2010). A dynamic model for binary panel data with unobserved heterogeneity admitting a $\sqrt{n}$-consistent conditional estimator. *Econometrica*, 78:719–733.

Butler, J. S. and Moffitt, R. (1982). A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica*, 50(3):761–64.

Geweke, J. (1989). Bayesian inference in econometric models using monte carlo integration. *Econometrica*, 57(6):1317–1339.

Hajivassiliou, V. A. and McFadden, D. L. (1998). The method of simulated scores for the estimation of ldv models. *Econometrica*, 66(4):863–896.

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

Hyslop, D. R. (1999). State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women. *Econometrica*, 67(6):1255–1294.

Keane, M. P. (1994). A computationally practical simulation estimator for panel data. *Econometrica*, 62:95–116.

Keane, M. P. and Sauer, R. M. (2009). Classification error in dynamic discrete choice models: Implications for female labor supply behavior. *Econometrica*, 77(3):975–991.