Simulating Gaussian Stationary Dynamic Panel Data Models: New Features of `xtarsim`

Giovanni Bruno[1]

[1]Bocconi University

Italian SUG, 26/09/2019

- Monte Carlo analysis helps the finite-sample evaluation of estimators and tests
- provides also pedagogical benefits: allows direct experience of randomness in repeated samples
- At the base of Monte Carlo analysis there is a procedure for simulating the data from a known population: `xtarsim` (Bruno, 2005) simulates dynamic panel data in Stata
- I present new features of `xtarsim` allowing for different types of predetermined or endogenous regressors, along with possibly serially-correlated idiosyncratic errors.

Models

The data generating processes (N individuals are independently drawn, for notational simplicity the subscript indicating individuals is omitted throughout):

$$\begin{aligned} y_t &= \gamma L y_t + \beta x_t + \eta + \epsilon_t \\ x_t &= \rho L x_t + H_t \end{aligned}$$

$t = 1, ..., T$, where $|\gamma| < 1$, $|\rho| < 1$, ϵ_t and ξ_t are mutually independent i.i.d. series with $\epsilon_t|\eta \sim N\left(0, \sigma_\epsilon^2\right)$ and η is a time-invariant individual component. Depending on the form chosen for H_t, we have cases of an exogenous, endogenous or predetermined x_t.

Exogenous x

As in Kiviet (1995):

$$H_t = \omega\eta + \xi_t. \tag{1}$$

with $\xi_t|\eta \sim N\left(0, \sigma^2_\xi\right)$ independent of ϵ_t (Kiviet sets ω to zero and so does not accommodate correlated effects); I also consider the case x exogenous with MA(1) errors in the main equation, $y_t = \gamma Ly_t + \beta x_t + \eta + \epsilon_t + \delta L\epsilon_t$, as in Bowsher (2002))

Predetermined x

Two schemes (Bun and Kiviet, 2006):
Scheme 1

$$H_t = (1 - \rho L)(\alpha L \epsilon_t + \omega \eta) + \xi_t$$

Scheme 2

$$H_t = \alpha L y_t + \omega \eta + \xi_t$$

Endogenous x

Three schemes:
Scheme 1

$$H_t = (1 - \rho L)\left(\alpha\epsilon_t + \omega\eta\right) + \xi_t$$

Scheme 2

$$H_t = \alpha y_t + \omega\eta + \xi_t$$

Scheme 3

$$H_t = \omega\eta + \alpha\epsilon_t + \xi_t$$

xtarsim

Dynamic models need start-up values. I follow McLeod and Hipel (1978) and Kiviet (1995) to obtain start-up values according to the data generation process, so to avoid wasting random numbers in the estimation of start-up values and also small-sample non-stationarity problems.

xtarsim

The basic syntax of `xtarsim` is as follows

`xtarsim` *depvar indepvar ind_effect_var* `, nid(#) time(#) gamma(real) beta(real) rho(real) snratio [sigma(real) omega(real) ma1(real) seed]`

Example: exogenous x

```
. xtarsim y x fe, nid(100) time(5) gamma(.2) beta(.8) rho(.8) snratio(snr 9) seed(12345)

. describe

Contains data
  obs:           500
 vars:             5
 size:        13,000
------------------------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
------------------------------------------------------------------------------------------------
ivar            byte    %8.0g                 panel variable
tvar            byte    %8.0g                 time variable
y               double  %10.0g                dependent variable
x               double  %10.0g                regressor
fe              double  %10.0g                individual effects
------------------------------------------------------------------------------------------------
Sorted by: ivar  tvar
```

Example: exogenous x

This generates a panel of N=100 and T=5 from a model with exogenous x:

$$
\begin{aligned}
y_t &= 0.2Ly_t + 0.8x_t + \eta + \epsilon_t \\
x_t &= 0.8Lx_t + \xi_t
\end{aligned}
$$

The signal to noise is $Var\left(y_t - \frac{\eta}{1-\gamma} - \epsilon_t\right) / \sigma_\epsilon^2 = 9$, which determines the variance of ξ_t. The variance of ϵ_t is controlled by the option `sigma(real)`, set to unity by default.

Example: exogenous x with MA(1) errors

```
. xtarsim y x fe, nid(100) time(5) gamma(.2) beta(.8) rho(.8) ma1(.2) snratio(snr 9)
```

$$
\begin{aligned}
y_t &= 0.2Ly_t + 0.8x_t + \eta + \epsilon_t + 0.2L\epsilon_t \\
x_t &= 0.8Lx_t + \xi_t
\end{aligned}
$$

Example: predetermined x - scheme 1

```
. xtarsim y x fe, nid(100) time(5) gamma(.2) beta(.8) rho(.8) pred(s1 .5) snratio(snr 9)
```

$$
\begin{aligned}
y_t &= 0.2Ly_t + 0.8x_t + \eta + \epsilon_t + 0.2L\epsilon_t \\
x_t &= 0.8Lx_t + (1 - 0.8L)\, 0.5L\epsilon_t + \xi_t
\end{aligned}
$$

Example: endogenous x - scheme 1

```
. xtarsim y x fe, nid(100) time(5) gamma(.2) beta(.8) rho(.8) endog(s1 .5) snratio(snr 9)
```

$$
\begin{aligned}
y_t &= 0.2Ly_t + 0.8x_t + \eta + \epsilon_t + 0.2L\epsilon_t \\
x_t &= 0.8Lx_t + \left(1 - 0.8L\right) 0.5\epsilon_t + \xi_t
\end{aligned}
$$

Applications

Possible Monte Carlo applications of `xtarsim`,

- Evaluation of finite-sample biases and RMSEs of panel data estimators (Bruno, 2005)
- Evaluation of power and size distortion of specification tests
- Evaluation of symptoms of instrument proliferation (Roodman 2009)

Clive G. Bowsher.
On testing overidentifying restrictions in dynamic panel data models.
Economics Letters, 77:211–220, 2002.

Giovanni S. F. Bruno.
Estimation and inference in dynamic unbalanced panel data models with a small number of individuals.
The Stata Journal, 5:473–00, 2005.

M. J. G. Bun and J. F. Kiviet.
The effects of dynamic feedbacks on ls and mm estimator accuracy in panel data models.
Journal of Econometrics, 132:409–44, 2006.

J. F. Kiviet.
On bias, inconsistency and efficiency of various estimators in dynamic panel data models.
Journal of Econometrics, 68:53–78, 1995.

- A. I. McLeod and K. W. Hipel.
  Simulation procedures for box-jenkins models.
  Water Resources Research, 14:969–975, 1978.
- D. M. Roodman.
  How to do xtabond2: An introduction to difference and system gmm in stata.
  The Stata Journal, 9(1):86–136, 2009.