XSMLE - A Command to Estimate Spatial Panel Models in Stata

Federico Belotti⋆,
Gordon Hughes∘, Andrea Piano Mortari⋆

⋆ CEIS, University of Rome Tor Vergata
∘ School of Economics, University of Edinburg.

2013 German Stata Users Group Meeting
Potsdam, June 7 2013

Outline

1. Motivation

2. The `xsmle` command
   - Options common to all spatial models
   - The weighting matrix
   - Models' specific options
   - Postestimation command after `xsmle`

3. Examples
   - No missing data
   - Postestimation
   - Testing
   - Missing data
   - Testing with missing data

A general specification for Spatial Panel models:

$$\begin{aligned} y_{it} &= \alpha + \tau y_{it-1} + \rho \sum_{j=1}^{n} w_{ij} y_{jt} + \sum_{k=1}^{K} x_{itk}\beta_{\mathbf{k}} + \sum_{k=1}^{K}\sum_{j=1}^{n} w_{ij} x_{jtk}\theta_{\mathbf{k}} + \\ &+ \mu_i + \gamma_t + \nu_{it} \qquad (1) \\ \nu_{it} &= \lambda \sum_{j=1}^{n} m_{ij}\nu_{it} + \epsilon_{it} \qquad i = 1, ..., n \qquad t = 1, ..., T \qquad (2) \end{aligned}$$

Static Models ($\tau = 0$) and Dynamic Models ($\tau \neq 0$, Yu et al. (2008))

- if $\theta = 0$ and $\rightarrow$ Spatial Autoregressive Model with Auto Regressive disturbances (SAC)
- if $\lambda = 0 \rightarrow$ Spatial Durbin Model (SDM)
- if $\lambda = 0$ and $\theta = 0 \rightarrow$ Spatial Autoregressive Model (SAR)
- if $\rho = 0$ and $\theta = 0 \rightarrow$ Spatial Error Model (SEM)
- if $\rho = 0$, $\theta = 0$, and $\mu_i = \phi \sum_{j=1}^{n} w_{ij}\mu_i + \eta_i \rightarrow$ Generalised Spatial Panel Random Effects model (GSPRE)

A number of spatial-related routines have been written by users and available through SSC. A non-comprehensive list includes:

Data management and visualization

- `shp2dta` by K. Crow
- `spmat` by D.M. Drukker *et al*
- `spwmatrix` by P.W. Jeanty
- `spmap` by M. Pisati
- `geocode3` by S. Bernhard

Cross sectional data

- `spreg`: SAR, SEM, SAC *via* ML or GS2SLS by D. M. Drukker *et al*
- `spivreg`: SAC *via* GS2SLS by D. M. Drukker *et al*
- `spmlreg`: SAR, SEM, SDM, SAC *via* ML by P.W. Jeanty
- `spatreg`: SAR, SEM *via* ML by M. Pisati
- `spautoreg`: SAR, SEM, SDM, SAC *via* ML or GS2SLS by E.A. Shehata

Panel data

- `spreg*xt` suite SAR, SEM, SDM, SAC *via* LS, GLS, GMM or GS2SLS by E.A. Shehata (Lee, 2002)

DGP - 250 replications

$$y_{it} = \rho \sum_{j=1}^{n} w_{ij} y_{jt} + 0.3x_{1it} + 0.7x_{2it} + \mu_i + \gamma_t + \epsilon_{it} \quad n = 1, \ldots, 188 \quad t = 1, \ldots, 5$$

where the nuisance parameters μ_i $(i = 1, \ldots, n)$ are drawn from an iid standard Gaussian random variable. To allow for dependence between the unit-specific effects and the regressors, we generate the latter as follows

$$x_{kit} = 0.4\mu_i + (1 - 0.4^2)^{1/2} z_{kit},$$

where $k = 1, 2$ and the z_{kit} is an iid standard Gaussian random variable.

	$\rho = 0.3$		$\rho = 0.5$		$\rho = 0.7$	
	bias	MSE	bias	MSE	bias	MSE
xsmle	-0.0013	0.0020	-0.0016	0.0014	-0.0016	0.0007
spregfext	0.1473	0.0255	0.1972	0.0408	0.1859	0.0352
xtivreg2	0.0174	0.0091	0.0153	0.0063	0.0112	0.0033

`xsmle` fits (balanced) Spatial Panel data models *via* maximum likelihood (ML)

Requirements:

- (At least) Stata Version 10
- The $n \times n$ matrix of spatial weights. `xsmle` will deal with the longitudinal dimension automatically
- Data must be `tsset` or `xtset`

The basic `xsmle` syntax is the following

`xsmle` *depvar* [*indepvars*] [*if*] [*in*] [*weight*] [`, options`]

- The default model is the random-effects SAR model
- Only `aweight` are allowed but the declared `weight` variable must be constant within each panel unit
- The `mi` prefix is allowed
- Factor variables are allowed

Options common to all spatial models

- `model(`*name*`)` specifies the spatial model to be estimated. May be *sar* for the Spatial-AutoRegressive model, *sdm* for the Spatial Durbin Model, *sem* for the Spatial-Error Model, *sac* for the Spatial-Autoregressive with Spatially Autocorrelated Errors Model, *gspre* for the Generalised Spatial Random Effects Model.
- `re` use the random effects estimator; the default. This option cannot be specified when `model(`*sac*`)`.
- `fe` use the fixed effects estimator. This option cannot be specified when `model(`*gspre*`)`.
- `type(`*type_options* [, leeyu]`)` specifies fixed-effects type; only for `fe` estimators. May be `ind` for spatial fixed effects, `time` for time fixed effects or `both` for both spatial and time fixed effects. Suboption `leeyu` allows to transform the data according to Lee and Yu (2010) approach and can be used only when `type(ind)`.

- noconstant suppresses the constant term in the model. Only for re estimators.
- noeffects suppresses the computation of direct, indirect and total effects.
- nsim(#) sets the number of simulations for the LeSage and Pace (2009) procedure to compute the standard errors of the direct, indirect and total effects.
- constraints(*constraints_list*) applies specified linear constraints.
- from(*init_specs*) specifies initial values for the coefficients.
- level(#) sets confidence level for confidence intervals; default is level(95).
- postscore save observation-by-observation scores in the estimation results list.
- posthessian save the Hessian corresponding to the full set of coefficients in the estimation results list.
- hausman performs the Hausman test.

Variance estimation

This section describes the arguments of the `vce(`*vcetype*`)` option.

- *oim* observed information matrix.
- *opg* outer product of the gradient vectors.
- *robust* clustered sandwich estimator where *clustvar* is the *panelvar*.
- *cluster clustvar* clustered sandwich estimator.
- *dkraay(#)* Driscoll-Kraay robust estimator. Where # is the maximum lag used in the calculation.

In xsmle the spatial weighting matrix can be

- a Stata matrix
- a spmat object

In both cases the matrix can be standardized or not.

e.g.

- a Stata matrix can be created using matrix define, imported from Mata using st_matrix("*string scalar name*", *real matrix*) or imported from GIS softwares like GeoDa using

  spwmatrix gal using *path_to_gal_file*, wname(*name_of_the_matrix*)

- spmat objects are created by spmat

  spmat import *name_of_the_object* using *path_to_file*

SAR model

- wmatrix(*name*) specifies the weight matrix for the spatial-autoregressive term.
- dlag includes (time) lagged dependent variable in the model.

SDM model

- wmatrix(*name*) specifies the weight matrix for the spatial-autoregressive term.
- dmatrix(*name*) specifies the weight matrix for the spatially lagged regressors; default is to use the matrix specified in wmat(*name*).
- durbin(*dvarlist*) specifies the regressors that have to be spatially lagged; default is to lag all independent variables specified in *varlist*.
- dlag includes (time) lagged dependent variable in the model.

SEM model

- ematrix(*name*) specifies the weight matrix for the spatial-autocorrelated error term.

SAC model

- wmatrix(*name*) specifies the weight matrix for the spatial-autoregressive term.
- ematrix(*name*) specifies the weight matrix for the spatial-autocorrelated error term.

GSPRE model

- wmatrix(*name*) specifies the weight matrix for the spatial-autocorrelated random-effects.
- ematrix(*name*) specifies the weight matrix for the spatial-autocorrelated error term.
- error(#) defines the structure of the model. # is equal to 1 when $\lambda \neq \phi \neq 0$, # is equal to 2 when $\lambda = 0$, # is equal to 3 when $\phi = 0$, # is equal to 4 when $\lambda = \phi$.

Postestimation command allows to post-estimate spatial fixed or random effects. The methods implemented in this command are the panel data extensions of those available in Drukker, Prucha, and Raciborski (2011)

predict [*type*] *newvar* [*if*] [*in*] [, statistic]

where statistic includes:

- rform the default, calculates predicted values from the reduced-form equation: $y_{it} = (I_n - \rho W)^{-1}(x_{it}\beta + \alpha_i)$
- limited predicted values based on the limited information set. This option is available only when model(sac).
- naive predicted values based on the observed values of $y_{it} = \rho W y_{it} + x_{it}\beta + \alpha_i$
- xb calculates the linear prediction including the fixed or random effect $x_{it}\beta + \alpha_i$.
- a estimates α_i, the fixed or random-effect. In the case of fixed-effects models, this statistic is allowed only when type(*ind*)

DGP - Fixed effects SDM

$$
\begin{aligned}
y_{it} &= 0.3\sum_{j=1}^{n} w_{ij} y_{jt} + 0.5x_{1it} - 0.3x_{2it} - 0.2x_{3it} + 0.3\sum_{j=1}^{n} w_{ij} x_{1it} + \\
+ &\quad 0.6\sum_{j=1}^{n} w_{ij} x_{2it} + 0.9\sum_{j=1}^{n} w_{ij} x_{3it} + \mu_i + \gamma_t + \epsilon_{it} \quad n = 1, ..., 188 \;\; t = 1, ..., 5
\end{aligned}
$$

where the nuisance parameters μ_i $(i = 1, \ldots, n)$ are drawn from an iid standard Gaussian random variable. To allow for dependence between the unit-specific effects and the regressors, we generate the latter as follows

$$
x_{kit} = 0.4\mu_i + (1 - 0.4^2)^{1/2} z_{kit},
$$

where $k = 1, 2, 3$, z_{1it} is standard Gaussian, z_{2it} is $N(0, 1.5^2)$ and z_{3it} is $N(0, 2^2)$.

```
.. *** load a dta dataset containing the spatial contiguity matrix
. use ASL_contiguity_mat_ns.dta, clear

. *** get an spmat objects from dta
. spmat dta W W*, replace

. *** Summarize the spmat obj
. spmat summarize W, links

Summary of spatial-weighting object W
--------------------------------------
       Matrix |          Description
--------------+-----------------------
   Dimensions |            188 x 188
    Stored as |            188 x 188
 Links        |
        total |                  906
          min |                    1
         mean |             4.819149
          max |                   13
--------------------------------------
```

```
. ** Fixed-effects Durbin model (correctly specified, row normalized W)
. xsmle y x1 x2 x3, wmat(W) model(sdm) fe type(ind) nsim(500) nolog
Warning: All regressors will be spatially lagged


SDM with spatial fixed-effects                      Number of obs =       940

Group variable: id                                Number of groups =       188
Time variable: t                                      Panel length =         5

R-sq:    within  = 0.5727
         between = 0.3663
         overall = 0.4554

Mean of fixed-effects = -0.0137

Log-likelihood = -1230.7734
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95\% Conf. Interval]
-------------+----------------------------------------------------------------
Main         |
          x1 |   .5186041   .0364303    14.24   0.000     .4472019    .5900062
          x2 |  -.2946314   .0236541   -12.46   0.000    -.3409925   -.2482702
          x3 |  -.1923373   .0192912    -9.97   0.000    -.2301474   -.1545272
-------------+----------------------------------------------------------------
Wx           |
          x1 |   .3772047    .075502     5.00   0.000     .2292235    .5251859
          x2 |   .5765484   .0449332    12.83   0.000     .4884809    .6646159
          x3 |   .8692021   .0372769    23.32   0.000     .7961408    .9422634
-------------+----------------------------------------------------------------
Spatial      |
         rho |   .2519025   .0374278     6.73   0.000     .1785454    .3252596
-------------+----------------------------------------------------------------
Variance     |
    sigma2_e |   .7915998   .0366863    21.58   0.000     .7196959    .8635037
-------------+----------------------------------------------------------------
[CONTINUES]
```

In a spatial setting, the effect of an explanatory variable change in a particular unit affects not only that unit but also its neighbors (LeSage and Pace, 2009).

$$\left[\frac{\partial Y}{\partial x_{nk}}\right] = (I - \rho W)^{-1} \begin{bmatrix} \beta_k & w_{12}\theta_k & \cdot & w_{1n}\theta_k \\ w_{21}\theta_k & \beta_k & \cdot & w_{2n}\theta_k \\ \cdot & \cdot & \cdot & \cdot \\ w_{n1}\theta_k & w_{n2}\theta_k & \cdot & \beta_k \end{bmatrix}$$

If we have only 2 units and 1 regressor:

- SAR and SAC $\rightarrow (I - \rho W)^{-1} \begin{bmatrix} \beta_1 & 0 \\ 0 & \beta_1 \end{bmatrix}$
- SEM $\rightarrow \begin{bmatrix} \beta_1 & 0 \\ 0 & \beta_1 \end{bmatrix}$
- SDM $\rightarrow (I - \rho W)^{-1} \begin{bmatrix} \beta_1 & w_{12}\theta_1 \\ w_{21}\theta_1 & \beta_1 \end{bmatrix}$

```
[CONTINUES]
-------------+----------------------------------------------------------------
Direct       |
          x1 |   .5481382   .0362326    15.13   0.000     .4771237    .6191527
          x2 |  -.2642811   .0231199   -11.43   0.000    -.3095953   -.2189669
          x3 |  -.1422518   .0176968    -8.04   0.000    -.1769369   -.1075668
-------------+----------------------------------------------------------------
Indirect     |
          x1 |   .6480929    .090572     7.16   0.000      .470575    .8256108
          x2 |   .6450951   .0599307    10.76   0.000     .5276331    .7625571
          x3 |   1.050599    .058257    18.03   0.000     .9364176    1.164781
-------------+----------------------------------------------------------------
Total        |
          x1 |   1.196231   .1038425    11.52   0.000     .9927034    1.399759
          x2 |    .380814   .0677252     5.62   0.000     .2480751     .513553
          x3 |   .9083474   .0660288    13.76   0.000     .7789334    1.037761
------------------------------------------------------------------------------
. estimates store sdm_fe
```

```
 ** Fixed-effects Durbin model (correctly specified, row normalized W)
 . xsmle y x1 x2 x3, wmat(W) model(sdm) re type(ind) nsim(500) nolog noeff
Warning: Option type(ind) will be ignored
Warning: All regressors will be spatially lagged


SDM with random-effects                               Number of obs =       940

Group variable: id                                 Number of groups =       188
Time variable: t                                       Panel length =         5

R-sq:    within  = 0.5666
         between = 0.4543
         overall = 0.4936

Log-likelihood = -1513.7006
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95\% Conf. Interval]
-------------+----------------------------------------------------------------
Main         |
          x1 |   .6230976   .0408605    15.25   0.000     .5430126    .7031826
          x2 |  -.2439834   .0264129    -9.24   0.000    -.2957518    -.192215
          x3 |  -.1688081   .0211584    -7.98   0.000    -.2102778   -.1273385
       _cons |  -.0169191   .0811545    -0.21   0.835    -.1759791    .1421409
-------------+----------------------------------------------------------------
Wx           |
          x1 |   .3706183   .0824133     4.50   0.000     .2090911    .5321454
          x2 |    .557779   .0493092    11.31   0.000     .4611347    .6544234
          x3 |   .8845199   .0411496    21.50   0.000     .8038681    .9651717
-------------+----------------------------------------------------------------
Spatial      |
         rho |   .2472432   .0376366     6.57   0.000     .1734769    .3210096
-------------+----------------------------------------------------------------
Variance     |
   lgt_theta |  -.3920581   .1040247    -3.77   0.000    -.5959428   -.1881735
     sigma_e |   1.005536   .0528831    19.01   0.000     .9018867    1.109185
------------------------------------------------------------------------------
```

```
hausman sdm_fe sdm_re, eq(1:1 2:2 3:3)

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     sdm_fe       sdm_re        Difference          S.E.
-------------+----------------------------------------------------------------
comp1        |
          x1 |    .5186041     .6230976       -.1044935               .
          x2 |   -.2946314    -.2439834        -.050648               .
          x3 |   -.1923373    -.1688081       -.0235292               .
-------------+----------------------------------------------------------------
comp2        |
          x1 |    .3772047     .3706183        .0065864               .
          x2 |    .5765484      .557779        .0187694               .
          x3 |    .8692021     .8845199       -.0153178               .
-------------+----------------------------------------------------------------
comp3        |
         rho |    .2519025     .2472432        .0046593               .
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from xsmle
           B = inconsistent under Ha, efficient under Ho; obtained from xsmle

    Test:  Ho:  difference in coefficients not systematic
                  chi2(7) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =   -47.08    chi2<0 ==> model fitted on these
                                        data fails to meet the asymptotic
                                        assumptions of the Hausman test;
                                        see suest for a generalized test
```

```
. ** Fixed-effects Durbin model (correctly specified, row normalized W)
. xsmle y x1 x2 x3, wmat(W) model(sdm) fe type(ind) hausman noeff nolog
Warning: All regressors will be spatially lagged
... estimating random-effects model to perform Hausman test
SDM with spatial fixed-effects                     Number of obs =       940

Group variable: id                               Number of groups =       188
Time variable: t                                     Panel length =         5

R-sq:    within  = 0.5727
         between = 0.3663
         overall = 0.4554

Mean of fixed-effects = -0.0137

Log-likelihood = -1230.7734
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95\% Conf. Interval]
-------------+----------------------------------------------------------------
Main         |
          x1 |   .5186041   .0364303    14.24   0.000     .4472019    .5900062
          x2 |  -.2946314   .0236541   -12.46   0.000    -.3409925   -.2482702
          x3 |  -.1923373   .0192912    -9.97   0.000    -.2301474   -.1545272
-------------+----------------------------------------------------------------
Wx           |
          x1 |   .3772047    .075502     5.00   0.000     .2292235    .5251859
          x2 |   .5765484   .0449332    12.83   0.000     .4884809    .6646159
          x3 |   .8692021   .0372769    23.32   0.000     .7961408    .9422634
-------------+----------------------------------------------------------------
Spatial      |
         rho |   .2519025   .0374278     6.73   0.000     .1785454    .3252596
-------------+----------------------------------------------------------------
Variance     |
    sigma2_e |   .7915998   .0366863    21.58   0.000     .7196959    .8635037
------------------------------------------------------------------------------
Ho: difference in coeffs not systematic chi2(7) = 89.58    Prob>=chi2 = 0.0000
------------------------------------------------------------------------------
```

```
. estat ic

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
           . |    940           .   -1230.773      8     2477.547    2516.314
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note

. estimates store sdm
```

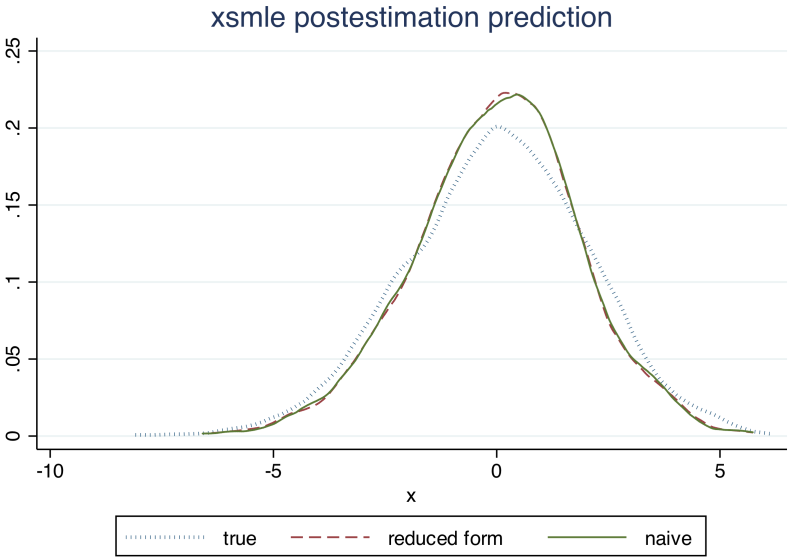
```
. ********* Postestimation
. predict yhat, rform

. predict yhat1, naive

. predict alphahat, a

. sum alpha alphahat

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       alpha |       940     .037577    1.058726  -2.261747   3.343453
    alphahat |       940    -.013692    1.182919  -2.688471   4.028156
```


xsmle postestimation prediction
0
.05
.1
.15
.2
.25
-10
-5
0
5
x
true
reduced form
naive

Using matrix notation the SDM ($\lambda = 0$) may be derived from a SEM model

$$\left\{ \begin{array}{rcr} \mathbf{y} & = & X\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{u} & = & \lambda W\mathbf{u} + \boldsymbol{\epsilon} \end{array} \right.$$

hence

$$\begin{array}{rcl} \mathbf{u}(1 - \lambda W) & = & \boldsymbol{\epsilon} \\ \mathbf{y}(1 - \lambda W) & = & X\boldsymbol{\beta}(1 - \lambda W) + \boldsymbol{\epsilon} \\ \mathbf{y} & = & \lambda W\mathbf{y} + X\boldsymbol{\beta} - \lambda WX\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ \mathbf{y} & = & \lambda W\mathbf{y} + X\boldsymbol{\beta} + \theta WX + \boldsymbol{\epsilon} \end{array}$$

and test the following constraints

1. $\theta = 0$ and $\lambda \neq 0 \Rightarrow$ the model is a SAR
2. $\theta = -\beta\lambda \Rightarrow$ the model is a SDM.

```
 ** Test for SAR
. test [Wx]x1 = [Wx]x2 = [Wx]x3 = 0

 ( 1)  [Wx]x1 - [Wx]x2 = 0
 ( 2)  [Wx]x1 - [Wx]x3 = 0
 ( 3)  [Wx]x1 = 0

           chi2(  3) =  740.80
         Prob > chi2 =    0.0000

.
. ** Test for SEM
. testnl ([Wx]x1 = -[Spatial]rho*[Main]x1) ([Wx]x2 = -[Spatial]rho*[Main]x2) ([
> Wx]x3 = -[Spatial]rho*[Main]x3)

  (1)  [Wx]x1 = -[Spatial]rho*[Main]x1
  (2)  [Wx]x2 = -[Spatial]rho*[Main]x2
  (3)  [Wx]x3 = -[Spatial]rho*[Main]x3

               chi2(3) =      545.31
           Prob > chi2 =        0.0000
```

```
 ** Test for SAC
. xsmle y x1 x2 x3, wmat(W) emat(W) model(sac) fe type(ind) noeff nolog

SAC with spatial fixed-effects                      Number of obs =       940

Group variable: id                                Number of groups =       188
Time variable: t                                      Panel length =         5

R-sq:    within  = 0.2652
         between = 0.0011
         overall = 0.0912

Mean of fixed-effects = -0.0117

Log-likelihood = -1386.0860
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95\% Conf. Interval]
-------------+----------------------------------------------------------------
Main         |
          x1 |   .3212791   .0341734     9.40   0.000     .2543005    .3882577
          x2 |  -.3135993   .0232111   -13.51   0.000    -.3590923   -.2681064
          x3 |  -.2997975   .0178884   -16.76   0.000     -.334858   -.2647369
-------------+----------------------------------------------------------------
Spatial      |
         rho |  -.6676721   .0542468   -12.31   0.000    -.7739939   -.5613504
      lambda |   .8426981   .0209346    40.25   0.000      .801667    .8837293
-------------+----------------------------------------------------------------
Variance     |
    sigma2_e |   1.001782   .0440957    22.72   0.000     .9153562    1.088208
------------------------------------------------------------------------------
```

```
. estat ic

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        940         .  -1386.086       6    2784.172   2813.247
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note

.
```

DGP - Fixed effects SDM with missing values

$$
\begin{aligned}
y_{it} &= 0.5 \sum_{j=1}^{n} w_{ij} y_{jt} - 1.5 x_{1it} - 0.7 x_{2it} - 0.3 x_{3it} - 0.9 x_{4it} + \\
&+ 0.75 \sum_{j=1}^{n} w_{ij} x_{1it} + 0.35 \sum_{j=1}^{n} w_{ij} x_{2it} + .15 \sum_{j=1}^{n} w_{ij} x_{3it} + \\
&+ .45 \sum_{j=1}^{n} w_{ij} x_{4it} + \mu_i + \gamma_t + \epsilon_{it} \quad n = 1, \ldots, 188 \quad t = 1, \ldots, 5
\end{aligned}
$$

where the nuisance parameters μ_i $(i = 1, \ldots, n)$ are drawn from an iid standard Gaussian random variable. To allow for dependence between the unit-specific effects and the regressors, we generate the latter as follows

$$x_{kit} = 0.4\mu_i + (1 - 0.4^2)^{1/2} z_{kit},$$

where $k = 1, 2, 3, 4$, z_{1it} and z_{3it} are $N(0, 1.5^2)$ and z_{2it} and z_{4it} are standard Gaussian. 5% missing values are randomly assigned to x_{1it} and x_{3it} to generate xm_{1it} and xm_{3it}.

```
. sum y x1 x1m x2 x3 x3m x4

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
           y |       940    .0392334    2.649594  -7.947968   7.694182
          x1 |       940    .0447793    1.456754  -5.412737    4.76092
         x1m |       901     .039974    1.452938  -5.412737    4.76092
          x2 |       940    .0209358     1.03554  -3.141913     3.0158
          x3 |       940   -.0236917    1.426842   -4.23553   4.351037
-------------+--------------------------------------------------------
         x3m |       884   -.0292849    1.433476   -4.23553   4.351037
          x4 |       940   -.0297022    1.020829  -3.721222   3.210468

.
. ** Fixed-effects SDM model (correctly specified)
. xsmle y x1 x2 x3 x4, wmat(W) model(sdm) fe type(ind) nolog noeff
Warning: All regressors will be spatially lagged

[OUTPUT OMITTED]

. estimates store sdm_nomissing
```

```
mi set wide
. mi register imputed x1m x3m
. mi impute mvn x1m x3m = x2 x4, add(50) rseed(12345)

Performing EM optimization:
note: 7 observations omitted from EM estimation because of all imputation
      variables missing
  observed log likelihood = -1518.1423 at iteration 5

Performing MCMC data augmentation ...

Multivariate imputation                       Imputations =        50
Multivariate normal regression                      added =        50
Imputed: m=1 through m=50                         updated =         0

Prior: uniform                                 Iterations =      5000
                                                  burn-in =       100
                                                  between =       100

               |              Observations per m
               |----------------------------------------------
      Variable |   complete   incomplete   imputed |     total
---------------+-----------------------------------+----------
           x1m |        901           39        39 |       940
           x3m |        884           56        56 |       940
--------------------------------------------------------------
(complete + incomplete = total; imputed is the minimum across m
 of the number of filled in observations.)
```

```
. ***** SDM estimates using multiple imputed data
. mi estimate (coeff1: [Wx]x1m + [Spatial]rho*[Main]x1m) ///
>                          (coeff2: [Wx]x2 + [Spatial]rho*[Main]x2)   ///
>                          (coeff3: [Wx]x3m + [Spatial]rho*[Main]x3m) ///
>                          (coeff4: [Wx]x4 + [Spatial]rho*[Main]x4),  ///
>                          dots post saving(sdm_imputed, replace):    ///
>                          xsmle y x1m x2 x3m x4, wmat(W) model(sdm)  ///
>                          fe type(ind) nolog noeff

Imputations (50):
  .........10.........20.........30.........40.........50 done

Multiple-imputation estimates                     Imputations     =         50
SDM with spatial fixed-effects                    Number of obs   =        940
                                                  Average RVI     =     0.3120
DF adjustment:   Large sample                     DF:      min    =     168.38
                                                           avg    =    1679.48
                                                           max    =    4655.10
Model F test:       Equal FMI                     F(  10, 8545.3) =     332.30
Within VCE type:          OIM                     Prob > F        =     0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95\% Conf. Interval]
-------------+----------------------------------------------------------------
Main         |
         x1m |  -1.430412   .0337574   -42.37   0.000    -1.496734    -1.36409
          x2 |  -.7335296   .0466659   -15.72   0.000     -.825132   -.6419272
         x3m |   -.274662   .0313065    -8.77   0.000    -.3361047   -.2132194
          x4 |  -.9359276    .044795   -20.89   0.000    -1.023766   -.8480894
-------------+----------------------------------------------------------------
Wx           |
         x1m |   .5188195   .0864699     6.00   0.000     .3491775    .6884615
          x2 |   .1321088   .0904967     1.46   0.144    -.0453216    .3095391
         x3m |   .0546988   .0643793     0.85   0.396    -.0716161    .1810137
          x4 |   .3360411   .1018539     3.30   0.001     .1363592     .535723
-------------+----------------------------------------------------------------
```

```
-------------+----------------------------------------------------------------
Spatial      |
         rho |   .3869484   .0429191     9.02   0.000     .3027466    .4711502
-------------+----------------------------------------------------------------
Variance     |
     sigma2_e |   1.068042   .0737025    14.49   0.000     .9225424    1.213542
------------------------------------------------------------------------------

Transformations                                   Average RVI     =     0.1161
DF adjustment:   Large sample                     DF:     min     =    1606.62
                                                          avg     =    7184.40
Within VCE type:          OIM                             max     =   12514.96

       coeff1: [Wx]x1m + [Spatial]rho*[Main]x1m
       coeff2: [Wx]x2 + [Spatial]rho*[Main]x2
       coeff3: [Wx]x3m + [Spatial]rho*[Main]x3m
       coeff4: [Wx]x4 + [Spatial]rho*[Main]x4

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95\% Conf. Interval]
-------------+----------------------------------------------------------------
      coeff1 |  -.0347924   .0575259    -0.60   0.545     -.147552    .0779673
      coeff2 |  -.1516852   .0829808    -1.83   0.068     -.314351    .0109806
      coeff3 |  -.0516019   .0636324    -0.81   0.418    -.1764131    .0732093
      coeff4 |  -.0261852   .0942367    -0.28   0.781     -.210917    .1585465
------------------------------------------------------------------------------

. estimates store sdm_imputed
```

```
. ** Test for SAR
. mi test [Wx]x1m [Wx]x2 [Wx]x3m [Wx]x4
note: assuming equal fractions of missing information

 ( 1)  [Wx]x1m = 0
 ( 2)  [Wx]x2 = 0
 ( 3)  [Wx]x3m = 0
 ( 4)  [Wx]x4 = 0

       F(  4,8383.2) =   10.09
            Prob > F =    0.0000

. ** Test for SEM
. mi testtr coeff1 coeff2 coeff3 coeff4
note: assuming equal fractions of missing information

       coeff1: [Wx]x1m + [Spatial]rho*[Main]x1m
       coeff2: [Wx]x2 + [Spatial]rho*[Main]x2
       coeff3: [Wx]x3m + [Spatial]rho*[Main]x3m
       coeff4: [Wx]x4 + [Spatial]rho*[Main]x4

 ( 1)  coeff1 = 0
 ( 2)  coeff2 = 0
 ( 3)  coeff3 = 0
 ( 4)  coeff4 = 0

       F(  4,17424.6) =    1.13
            Prob > F =    0.3395
```

```
. estout sdm_nomissing sdm_imputed, c(b se) ren(x1m x1 x3m x3)
---------------------------------------
             sdm_nomiss~g  sdm_imputed
                     b/se         b/se
---------------------------------------
Main
x1               -1.49692    -1.430412
                  .023471     .0337574
x2              -.7440764    -.7335296
                 .0340518     .0466659
x3               -.297554     -.274662
                 .0233121     .0313065
x4              -.9451319    -.9359276
                 .0349799      .044795
---------------------------------------
Wx
x1               .6416344     .5188195
                  .070266     .0864699
x2               .1982546     .1321088
                 .0729333     .0904967
x3                .057889     .0546988
                 .0494267     .0643793
x4               .4091564     .3360411
                 .0829535     .1018539
---------------------------------------
Spatial
rho              .4530939     .3869484
                 .0356911     .0429191
---------------------------------------
Variance
sigma2_e         .7564478     1.068042
                 .0355401     .0737025
---------------------------------------
```

Lee, L. F. (2002). Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. *Econometric Theory*, 18(02):252–277.

Lee, L. F. and Yu, J. (2010). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics*, 154(2):165–185.

LeSage, J. P. and Pace, R. K. (2009). *Introduction to Spatial Econometrics*. Taylor & Francis.

Yu, J., de Jong, R., and Lee, L.-F. (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and t are large. *Journal of Econometrics*, 146:118–134.