Fitting spatial stochastic frontier models in Stata

Kerui Du
School of Management
Xiamen University
Xiamen, China
kerrydu@xmu.edu.cn

Luis Orea
Department of Economics
School of Economics and Business
University of Oviedo
Oviedo, Spain
lorea@uniovi.es

Inmaculada C. Álvarez
Department of Economics
Universidad Autónoma de Madrid
Madrid, Spain
inmaculada.alvarez@uam.es

Abstract. In this article, we introduce a new command spxtsfa for fitting spatial stochastic frontier models in Stata. Over the last decades, An important theoretical progress of stochastic frontier models is the incorporation of various types of spatial components. Models with the ability to account for spatial dependence and spillovers have been developed for efficiency and productivity analysis, drawing extensive attention from industry and academia. Due to the unavailability of the statistical packages, the empirical applications of the new stochastic frontier models appear to be lagging. The spxtsfa command provides a routine for estimating the spatial stochastic frontier models in the style of Orea and Álvarez (2019) and Galli (2022), enabling users to handle different sources of spatial dependence. In the presented article, we introduce the spatial stochastic frontier models, describe the syntax and options of the new command, and provide several examples to illustrate its usage.

Keywords: stochastic frontier models, SFA, spatial dependence, technical efficiency, spillovers

1 Introduction

Producers might fail in optimizing their production activities, causing deviation from the maximum output or the minimum cost. Economic researchers proposed the concept of technical efficiency, which measures how well a producer is utilizing its resources to produce goods or services. A technically efficient organization makes the maximum outputs given the amount of inputs or uses the minimum amount of inputs to produce a given level of output. On the contrary, technically inefficient organization produce fewer outputs given the same inputs or uses more inputs than necessary to produce the same output. Technical efficiency is important because it allows organizations or economies to achieve their goals with the least amount of resources possible, which can lead to cost savings and increased profitability.

Aigner et al. (1977) and Meeusen and van Den Broeck (1977) introduced stochastic frontier models for evaluating technical efficiency. The essential concept behind these models is to divide the observed output of a production process into two components, namely the "frontier" output, signifying the maximum feasible output, given the inputs utilized in the production process, and the "residual" output, denoting the production process's inefficiency. Following these initial works, stochastic frontier models gained extensive use as a tool for scrutinizing productivity and efficiency.

Methodologically,econometricians have expanded the horizons of stochastic frontier models in various directions. To name a few, Battese and Coelli (1995) incorporated the determinants of inefficiency. Wang (2003) developed the stochastic frontier model with scaling properties to capture the shape of the distribution of inefficiency. Greene (2005) extended the stochastic models with the random effects and the "true" fixed effects. Belotti and Ilardi (2018), Chen et al. (2014), and Wang and Ho (2010) circumvented the 'incidental parameters problem" in the fixed effects stochastic frontier model through model transformation. Karakaplan and Kutlu (2017) developed an endogenous stochastic frontier model to control for the endogeneity in the frontier or inefficiency.

In recent years, stochastic frontier models have undergone further extension to account for spatial dependence and spatial spillover effects. Glass et al. (2016) constructed a spatial Durbin stochastic model considering both global and local spatial dependence. Kutlu et al. (2020) proposed a spatial stochastic frontier model with endogenous frontier and environmental variables. Glass et al. (2016) and Kutlu et al. (2020) combine the concepts of spatial econometrics and stochastic frontier analysis by including the spatial lag of the dependent variable. On the other hand, Orea and Álvarez (2019) developed a new stochastic frontier model with spatial correlation in both noise and inefficiency terms. Galli (2022) integrated the two different modeling ideas to specify four different sources of spatial dependence fully.

Extending the stochastic frontier models to account for spatial dependence and spillover effects not only allows getting more accurate inefficiency scores (Orea et al. 2018) but also examining relevant economic issues that a non-spatial stochastic frontier model tends to overlook. In microdata applications, the addition of a spatial or cross-sectional correlation in the frontier can be used to test whether the firm controls all the inputs and hence the production/cost function can be viewed as a purely deterministic (engineering) process (Druska and Horrace 2004). A model with cross-sectional correlation in the inefficiency term is useful where some firms benefit from best practices implemented in adjacent firms due to, for instance, agglomeration economies, knowledge spillovers, technology diffusion or R&D spillovers. This could especially be the case if (local) firms belong to communitarian networks (e.g. cooperatives) or common technicians (consultants) are advising all local firms.

As Orea and Álvarez (2019) point out, the spatial stochastic frontier model can be implemented using macro-level data, e.g. data of countries, regions or industries. They state that abundant evidence of important feedback processes between neighboring or non-distant regions might justify the use of spatial autoregressive (SAR) and Durbin frontier functions. The spatial weight matrix specification commonly adopted in regional economics is based on geographical distance. Liu and Sickles (2023) state that the mode of production in the world economy is characterized by the division of global value chains (GVCs) and, hence, the spatial weights matrix should be constructed using the economic distance between industries within/across national economies. In this case, the spatial SAR and Durbin frontier functions allow the examination of the diffusion of knowledge and technology among the participants in the international production network. It is also makes sense to estimate a stochastic frontier model with cross-sectional correlation in the inefficiency term using macro-level data. In this case, the existence of spatially-correlated inefficiency terms likely captures barriers and distortions to the efficient allocation of resources across firms within a region (industry) that are common to several regions (industries), such as regulation, labor market trends or common institutions (Orea et al. 2023).

With the increasing demand in the last decades to analyze technical efficiency, Stata provides official commands frontier and xtfrontier for cross-sectional and panel stochastic model estimation, respectively. Belotti et al. (2013) developed sfcross and sfpanel commands accommodating more different distribution assumptions and allowing fixed-effect and random-effect models with the consideration of heteroscedasticity. Karakaplan (2017) introduced the sfkk command for estimating endogenous stochastic frontier models. Mustafa Ugur Karakaplan (2018) supplemented the xtsfkk command for fitting the endogenous panel stochastic frontier model. Kumbhakar et al. (2015) provides a practitioner's guide to stochastic frontier analysis with a suite of Stata commands (including sfmodel, sfpan, sf_fixeff, and sfprim). In this article, we introduce spxtsfa, a new command for fitting spatial stochastic frontier models in the style of Orea and Álvarez (2019) and Galli (2022). The proposed spxtsfa command not only allows capturing spatial spillovers but also other cross-sectional effects that might be caused by non-spatial factors (e.g., the regulation environment). In practice, the spxtsfa command can be useful to capture a kind of behavioral correlation, for instance when firms tend to "keep an eye" on the decisions of other peer firms trying to overcome the limitations caused by the lack of information or they simply emulate each other.

The remainder of this article unfolds as follows: Section 2 provides a brief description of the models in Orea and Álvarez (2019) and Galli (2022); Section 3 explain the syntax and options of spxtsfa; Section 4 and 5 present simulated data examples to illustrate the usage of the command; and section 6 concludes the article.

2 The model

In this section, we briefly describe the spatial stochastic frontier models developed by Orea and Álvarez (2019) and Galli (2022). The exposition here is only introductory. Please refer to the cited papers for more technical details.

Based on the transposed version of Wang and Ho (2010) model, Orea and

Álvarez (2019) proposed a spatial stochastic frontier model which accommodates spatially-correlated inefficiency and noise terms. The model is formulated as in Eqs.(1)-(3), for $i = 1, ..., N$ and $t = 1, .., T$:

$$Y_{it} = X_{it}'\beta + \tilde{v}_{it} - s\tilde{u}_{it} \tag{1}$$

$$\tilde{v}_{it} = v_{it} + \gamma W_i^{vt}\tilde{v}_{.t} \tag{2}$$

$$\tilde{u}_{it} = u_{it} + \tau W_i^{ut}\tilde{u}_{.t} \tag{3}$$

Eq.(1) describes the stochastic frontier function where Y_{it} is the dependent variable and X_{it} is a $k \times 1$ vector of variables shaping the frontier; $s = 1$ for the production function and $s = -1$ for the cost function; $\tilde{v}_{it}$ and $\tilde{u}_{it}$ represent idiosyncratic noise and inefficiency, respectively. In Eqs.(2) and (3), $W_i^{vt} = (W_{i1}^{vt}, ..., W_{iN}^{vt})$ and $W_i^{vt} = (W_{i1}^{vt}, ..., W_{iN}^{vt})$ are two known $1 \times N$ cross-sectional weight vectors depicting the structure of the cross-sectional relationship for idiosyncratic noise and inefficiency terms, respectively; $\tilde{v}_{.t} = (\tilde{v}_{1t}, ..., \tilde{v}_{Nt})'$ and $\tilde{u}_{.t} = (\tilde{u}_{1t}, ..., \tilde{u}_{Nt})'$; v_{it} is a random variable following the distribution $N(0, \sigma_v^2)$ and $u_{it} = h(Z_{it}'\delta)u_t^*$. $h(Z_{it}'\delta)$ is the scaling function where Z_{it} is a $l \times 1$ vector of variables affecting individuals' inefficiency and u_t^* is a non-negative random variable following the distribution $N^+(0, \sigma_u^2)$. Using matrix notation, we can rewrite Eqs.(2) and (3) as

$$\tilde{v}_{.t} = (I_N - \gamma W^{vt})^{-1}v_{.t} \tag{4}$$

$$\tilde{u}_{.t} = (I_N - \tau W^{ut})^{-1}h(Z_{.t}\delta)u_t^* = \tilde{h}_{.t}u_t^* \tag{5}$$

where $Z_{.t} = (Z_{1t}, ..., Z_{Nt})'$;$\tilde{h}_{.t} = (I_N - \tau W^{ut})^{-1}h(Z_{.t}\delta)$.

The above model captures the spatial correlation of the random error and inefficiency terms with the spatial autoregressive (SAR) process [1]. Referring to Wang and Ho (2010), we can obtain the following log-likelihood function for each period t:

$$\begin{aligned} \ln L_t = & -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|\Pi| - \frac{1}{2}\tilde{\varepsilon}_{.t}\Pi^{-1}\tilde{\varepsilon}_{.t} \\ & + \frac{1}{2}\left(\frac{\mu_*^2}{\sigma_*^2}\right) + \ln\left[\sigma_*\Phi\left(\frac{\mu_*}{\sigma_*}\right)\right] - \ln\left(\frac{1}{2}\sigma_u\right) \end{aligned} \tag{6}$$

where $\Pi = \sigma_v^2(I_N - \rho W^{yt})^{-1}[(I_N - \rho W^{yt})^{-1}]'$; $\tilde{\varepsilon}_{.t} = (\tilde{\varepsilon}_{1t}, ..., \tilde{\varepsilon}_{Nt})'$, $\tilde{\varepsilon}_{it} = s(Y_{it} - X_{it}'\beta)$, and

$$\mu_* = \frac{-\tilde{\varepsilon}'_{.t}\Pi^{-1}\tilde{h}_{.t}}{\tilde{h}'_{.t}\Pi^{-1}\tilde{h}_{.t} + 1/\sigma_u^2} \tag{7}$$

[1] Orea and Álvarez (2019) also considered a specification of the spatial moving average process.

$$\sigma_*^2 = \frac{1}{\tilde{h}'_{.t}\Pi^{-1}\tilde{h}_{.t} + 1/\sigma_u^2} \tag{8}$$

Galli (2022) further incorporated the spatial lags of the dependent variable and the input variables into Orea and Álvarez (2019) model, which additionally measures global and local spatial spillovers affecting the frontier function. The model is expressed as

$$Y_{it} = \rho W_i^{yt} Y_{.t} + X'_{it}\beta + W_i^{xt} X_{.t}\theta + \tilde{v}_{it} + s\tilde{u}_{it} \tag{9}$$

where $W_i^{yt} = (W_{i1}^{yt}, ..., W_{iN}^{yt})$ and $W_i^{xt} = (W_{i1}^{xt}, ..., W_{iN}^{xt})$ are two known $1 \times N$ cross-sectional weight vectors [2]; $Y_{.t} = (Y_{1t}, ..., Y_{Nt})'$; $X_{.t} = (X_{1t}, ..., X_{Nt})'$. This model gives rise to the following log-likelihood function for each period t:

$$\begin{aligned} \ln L_t =& ln|I_N - \rho W^{yt}| - \frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|\Pi| - \frac{1}{2}\tilde{\varepsilon}_{.t}\Pi^{-1}\tilde{\varepsilon}_{.t} \\ &+ \frac{1}{2}\left(\frac{\mu_*^2}{\sigma_*^2}\right) + \ln\left[\sigma_*\Phi\left(\frac{\mu_*}{\sigma_*}\right)\right] - \ln\left(\frac{1}{2}\sigma_u\right) \end{aligned} \tag{10}$$

where $\tilde{\varepsilon}_{.t} = (\tilde{\varepsilon}_{1t}, ..., \tilde{\varepsilon}_{Nt})', \tilde{\varepsilon}_{it} = s(Y_{it} - X'_{it}\beta - \rho W_i^{yt} Y_{.t} - W_i^{xt} X_{.t}\theta)$.

Summing the time-specific log-likelihood functions over all periods yields the overall likelihood function for the whole sample, i.e., $lnL = \sum_{t=1}^{T} lnL_t$. Then, numerically maximize the overall log-likelihood function to obtain consistent estimates of the parameters in the above models. Specifically, we use Stata `ml model` routine with the `method-d0` evaluator to program the `spxtsfa` command. Following Gude et al. (2018), we parameterize ρ, γ, and τ as Eq.(11) to ensure the standard regularity condition for the spatial autoregressive models.

$$\begin{aligned} &\eta = \left(\frac{1}{r_{\min}}\right)(1-p) + \left(\frac{1}{r_{\max}}\right)p \\ &0 \leq p = \frac{\exp(\delta_0)}{1 + \exp(\delta_0)} \leq 1 \end{aligned} \tag{11}$$

where η stands for one of ρ, γ, and τ; $r_{\min}$ and $r_{\max}$ are respectively the minimum and maximum eigenvalues of the corresponding spatial weight matrix.

In summary, Galli (2022) provided a fully comprehensive specification of four different types of spatial dependence: global spillovers of dependent variable Y_{it}, local spillovers of input variables X_{it}, cross-sectional correlation of idiosyncratic noise v_{it} and inefficiency u_{it}. We term this full model "$yxuv - SAR$." Some restrictions can be imposed on the specific parameters to generate the following models (summarized in Table 1), which can be estimated by the `spxtsfa` command.

[2] We index W_i^{yt}, W_i^{xt}, W_i^{ut}, and W_i^{vt} with superscript yt, xt, ut, and vt, respectively. This indicates the spatial weight matrix can be time-varying and different across various spatial components

Table 1: Specific models with restricted parameters

	yuv	*xuv*	*yv*	*yu*	*y*	*xuv*	*xv*	*xu*	*uv*	*u*	*v*
ρ		0				0	0	0	0	0	0
θ	0		0	0	0				0	0	0
γ				0	0			0		0	
τ			0		0		0				0

3 The spxtsfa command

`spxtsfa` estimates spatial stochastic frontier models in the style of Orea and Álvarez (2019) and Galli (2022).

3.1 Syntax

Estimation syntax

`spxtsfa` *depvar* [*indepvars*] *uhet(varlist)* [`noconstant cost wy(`*wyspec*`)`
`wx(`*wxspec*`) wu(`*wuspec*`) wv(`*wvspec*`) normalize(`*norm_method*`)`
`wxvars(`*varlist*`) initial(`*matname*`) mlmodel(`*model_options*`)`
`mlsearch(`*search_options*`) mlplot mlmax(`*maximize_options*`) nolog`
`mldisplay(`*display_options*`) level(`#`) lndetmc(`*numlist*`)`
`te(`*newvarname*`) genwxvars constraints(`*constraints*`)`]

Version syntax

`spxtsfa version`

Replay syntax

`spxtsfa` [`, level(`#`)`]

3.2 Options

`uhet(`*varlist*`)` specifies explanatory variables for technical inefficiency variance function depending on a linear combination of *varlist*. It is required.

`noconstant` suppresses suppress constant term.

`cost` specifies the frontier as a cost function. By default, the production function is assumed.

`wy(`*wyspec*`)` specifies the spatial weight matrix for lagged dependent variable. The expression is wy(W_1 [$W_2 ... W_T$] [,*mata array*]). By default, the weight matrices are `Sp` objects. mata declares weight matrices are mata matrices. If one weight matrix is specified, it assumes a time-constant weight matrix. For time-varying cases, T weight matrices should be specified in time order. Alternatively, using array to declare weight matrices are stored in an array. If only one matrix is stored in the specified array, the time-constant weight matrix is assumed. Otherwise, the keys of the array specify time information, and the values store time-specific weight matrices.

`wx(`*wxspec*`)` specifies the spatial weight matrix for lagged independent variable. The expression is the same as `wy(`*wyspec*`)`.

`wu(`*wuspec*`)` specifies the spatial weight matrix for lagged independent variable. The expression is the same as `wy(`*wyspec*`)`.

`wv(`*wvspec*`)` specifies the spatial weight matrix for lagged independent variable. The expression is the same as `wy(`*wyspec*`)`.

`normalize(`*norm_method*`)` specifies one of the four available normalization techniques: row, col, minmax, and spectral.

`wxvars(`*varlist*`)` specifies spatially lagged independent variables.

`initial(`*matname*`)` specifies the initial values of the estimated parameters with matrix *matname*.

`mlmodel(`*model_options*`)` specifies the `ml model` options.

`mlsearch(`*search_options*`)` specifies the `ml search` options.

`mlplot` specifies using `ml plot` to search better initial values of spatial dependence parameters.

`mlmax(`*maximize_options*`)` specifies the `ml maximize` options.

`nolog` suppresses the display of the criterion function iteration log.

`mldisplay(`*display_options*`)` specifies the `ml display` options.

`level(`#`)` sets confidence level; default is level(95).

`lndetmc(`*numlist*`)` uses the trick of Barry and Kelley Pace (1999) to solve the inverse of $(I_N - \rho W)$. The order of *numlist* is iterations, maxorder. `lndetmc(50 100)` specifies that the number of iterations is 50 and the maximum order of moments is 100.

`genwxvars` generates the spatial Durbin terms. It is activated only when `wxvars(`*varlist*`)` is specified.

`constraints(`*constraints*`)` specifies specified linear constraints for the estimated model.

3.3 Dependency of spxtsfa

`spxtsfa` depends on the *moremata* package. If not already installed, you can install it by typing ssc install moremata.

4 Examples

In this section, we use simulated data to exemplify the use of the $spxtsfa$ command. Referring to , we first consider the $yxuv-SAR$ model specified by the following data-generating process (DGP 1) with $i = 1, ..., 300$ and $t = 1, .., 20$,

$$Y_{it} = 0.3W_iY_{.t} + 2X_{it} + 0.3W_iX_{.t} + \tilde{v}_{it} - \tilde{u}_{it} \tag{12}$$

where $\tilde{v}_{it}$ and $\tilde{u}_{it}$ are defined as in Eqs.(2) and (3) with $\gamma = 0.3$, $\tau = 0.3$, $\delta = 2$, $\sigma_u^2 = 0.2$ and $\sigma_v^2 = 0.2$. All the spatial matrices for the four spatial components are the same and time-invariant, created from a binary contiguity spatial weight matrix. We generate the exogenous variables X_{it} and Z_{it} from the standard normal distribution, respectively. With the sample generated by DGP 1, we can fit the model in the following syntax.

```
. use spxtsfa_DGP1.dta

. xtset id t

Panel variable: id (strongly balanced)
 Time variable: t, 1 to 20
         Delta: 1 unit

. * importing spatial weight matrix from spxtsfa_wmat1.mmat
. mata mata matuse spxtsfa_wmat1.mmat,replace
(loading w1[300,300])

. * fitting the model
. spxtsfa y x, uhet(z) noconstant  wy(w1,mata) wx(w1,mata) wu(w1,mata) wv(w1,mata) wxvars(x) nolog

Spatial frontier model(yxuv-SAR)                     Number of obs =     6,000
                                                     Wald chi2(2)  = 118937.24
Log likelihood = -1727.016                           Prob > chi2   =    0.0000
```

y	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
frontier						
x	1.993915	.0065251	305.58	0.000	1.981126	2.006704
W_x	.4435823	.0373189	11.89	0.000	.3704386	.516726
uhet						
z	2.000371	.0013412	1491.49	0.000	1.997742	2.002999
/lnsigma2_u	-2.098104	.3163094	-6.63	0.000	-2.718059	-1.478149
/lnsigma2_v	-1.637609	.018401	-89.00	0.000	-1.673674	-1.601544
Wy						
_cons	.6605993	.0317043	20.84	0.000	.5984599	.7227386
Wu						
_cons	.5806681	.0318346	18.24	0.000	.5182735	.6430627
Wv						
_cons	.5745429	.051903	11.07	0.000	.4728148	.676271
sigma2_u	.1226888	.0388076	3.16	0.002	.0660027	.2280593
sigma2_v	.1944444	.003578	54.34	0.000	.1875567	.2015851
rho	.3187581	.0142397	22.39	0.000	.2905787	.3463849
tau	.282414	.014646	19.28	0.000	.2534626	.3108598
gamma	.2795936	.02392	11.69	0.000	.2320763	.3257792

The output shows that the command fits seven equations with `ml model`. The frontier equation has two explanatory variables X_it and $W_iX_{.t}$. The scaling function uhet() has one explanatory variable Z_{it}. Two equations (/lnsigma2_u and /lnsigma2_v) are constructed for the variance parameters σ_u^2 and σ_v^2 which are transformed by the function $exp(\cdot)$. Three Equations (Wy, Wu, and Wv) handle the spatial dependence parameters ρ, τ, and γ, which are parameterized as Eq.(11). We directly include the spatial Durbin term $W_iX_{.t}$ in the frontier equation such that we do not need to fit a separate equation. The bottom of the table reports the transformed parameters in the original metric.

We consider the restricted model $uv-SAR$ with time-varying spatial weight matrices as the second example. The DGP 2 is described as

$$Y_{it} = 1 + 2X_{it} + \tilde{v}_{it} - \tilde{u}_{it}, i = 1, .., 300; t = 1, .., 10 \tag{13}$$

where the other parameters are set the same as the DGP 1 except for $W_i^{ut} = W_i^{vt} = W_i^t$. The following syntax estimates the model alongside the results.

```
. use spxtsfa_DGP2.dta
. xtset id t
Panel variable: id (strongly balanced)
 Time variable: t, 1 to 10
         Delta: 1 unit
. * importing spatial weight matrices from spxtsfa_wmat2.mmat
. mata mata matuse spxtsfa_wmat2.mmat,replace
(loading w1[300,300], w10[300,300], w2[300,300], w3[300,300], w4[300,300],
 w5[300,300], w6[300,300], w7[300,300],  w8[300,300], w9[300,300])
. * fitting the model
. local w w1 w2 w3 w4 w5 w6 w7 w8 w9 w10
. spxtsfa y x, uhet(z) wu(`w´,mata) wv(`w´,mata) te(efficiency) nolog
Spatial frontier model(uv-SAR)                         Number of obs =    3,000
                                                       Wald chi2(1)  = 43686.91
Log likelihood = -1336.482                             Prob > chi2   =   0.0000
```

y	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
frontier						
x	2.015288	.0096419	209.01	0.000	1.99639	2.034186
_cons	.9415143	.0160786	58.56	0.000	.9100008	.9730278
uhet						
z	2.000242	.0020671	967.66	0.000	1.99619	2.004293
/lnsigma2_u	-2.006684	.4473506	-4.49	0.000	-2.883475	-1.129893
/lnsigma2_v	-1.300024	.0260099	-49.98	0.000	-1.351002	-1.249045
Wu						
_cons	.582383	.0031549	184.59	0.000	.5761995	.5885666
Wv						
_cons	.5374655	.0601775	8.93	0.000	.4195198	.6554113
sigma2_u	.1344337	.060139	2.24	0.025	.05594	.3230678
sigma2_v	.2725253	.0070883	38.45	0.000	.2589806	.2867784
tau	.2832028	.0014508	195.21	0.000	.2803569	.2860438
gamma	.262419	.0280135	9.37	0.000	.206716	.3164261

In the second example, we use option `te(efficiency)` to store the estimated efficiency score in a new variable `efficiency`. Finally, we consider another restricted model $xuv - SAR$ with different spatial weight matrices, one of which is time-varying, and the others are time-constant. The model is described as DGP 3:

$$Y_{it} = 1 + 2X_{it} + 0.5W_i^{xt} + \tilde{v}_{it} + \tilde{u}_{it}, i = 1, .., 300; t = 1, .., 10 \tag{14}$$

where the other parameters are set the same as the DGP 1 except for $W_i^{ut} = W_i^u$ and $W_i^{vt} = W_i^v$. Different from DGP 1 and DGP 2, which set the production function frontier, DGP 3 specifies a cost function. The estimation of the model is shown as follows.

```
. use spxtsfa_DGP3.dta

. xtset id t

Panel variable: id (strongly balanced)
 Time variable: t, 1 to 10
         Delta: 1 unit

. * importing spatial weight matrices from spxtsfa_wmat2.mmat
. mata mata matuse spxtsfa_wmat2.mmat,replace
(loading w1[300,300], w10[300,300], w2[300,300], w3[300,300], w4[300,300],
w5[300,300], w6[300,300], w7[300,300], w8[300,300], w9[300,300])

. * fitting the model
. local w w1 w2 w3 w4 w5 w6 w7 w8 w9 w10

. mat b = (1,1,1,1,-1,-1,0.5,0.5)

. spxtsfa y x, uhet(z) wu(w2,mata) wv(w1,mata) wxvars(x) ///
>              wx(`w´,mata) cost init(b) genwxvars nolog

Spatial frontier model(xuv-SAR)                       Number of obs =    3,000
                                                      Wald chi2(2)  = 57430.99
Log likelihood = -872.06794                           Prob > chi2   =   0.0000
```

y	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
frontier						
x	1.995734	.0083857	237.99	0.000	1.979298	2.012169
W_x	.5065867	.0221169	22.90	0.000	.4632384	.549935
_cons	.9916378	.0126993	78.09	0.000	.9667476	1.016528
uhet						
z	1.99976	.0010759	1858.74	0.000	1.997652	2.001869
/lnsigma2_u	-1.699017	.4472788	-3.80	0.000	-2.575668	-.822367
/lnsigma2_v	-1.615448	.0260424	-62.03	0.000	-1.66649	-1.564406
Wu						
_cons	.6214083	.0008566	725.44	0.000	.6197294	.6230872
Wv						
_cons	.6001869	.0595911	10.07	0.000	.4833905	.7169833
sigma2_u	.1828631	.0817908	2.24	0.025	.076103	.4393904
sigma2_v	.1988017	.0051773	38.40	0.000	.188909	.2092123
tau	.3010474	.0003894	773.04	0.000	.300284	.3018105
gamma	.291369	.0272628	10.69	0.000	.2370726	.3438504

In the third example, we use `cost` option to specify the type of frontier. The matrix `b` is used as the initial value for the maximum likelihood estimation. The likelihood function of spatial stochastic frontier models is complicated, and generally difficult to obtain the optimal global solutions. Thus, good initial values would be helpful for fitting spatial stochastic models. Practitioners might fit the non-spatial stochastic models using `fronteir` and `sfpanel` commands to obtain the initial values of the parameters involved in the frontier and the scaling function and then use the `mlplot` option to search initial values for spatially-correlated parameters.

5 Conclusion

Geospatial units are not isolated or separated but connected. For example, the economic trade, social activities, and cultural exchange between different regions affect each other. Such spatial interdependence challenges the traditional econometric methods, which generally assume cross-sectional independence. Spatial econometrics is developed to handle spatial correlation. Recently, researchers combined stochastic frontier models with spatial econometrics to account for various types of spatial effects in the field of efficiency and productivity analysis (Galli 2022; Orea and Álvarez 2019). This article presented a community-contributed command for fitting spatial stochastic frontier models with different sources of spatial dependence. We hope the developed command can provide some convenience to practitioners and reduce the difficulty of model applications, thereby promoting sound empirical research.

6 Acknowledgments

Kerui Du thanks the financial support of the National Natural Science Foundation of China (72074184). We are grateful to Federica Galli for his Matlab codes, Federico Belotti, Silvio Daidone, Giuseppe Ilardi and Vincenzo Atella for the sfcross/sfpanel package, Mustafa U. Karakaplan for the sfkk package, and Jan Ditzen, William Grieser and Morad Zekhnini for the nwxtregress package which inspired our design of the `spxtsfa` command.

7 References

Aigner, D., C. Lovell, and P. Schmidt. 1977. Formulation and Estimation of Stochastic Frontier Production Function Models. *Journal of Econometrics* 6(1): 21–37.

Barry, R. P., and R. Kelley Pace. 1999. Monte Carlo Estimates of the Log Determinant of Large Sparse Matrices. *Linear Algebra and its Applications* 289(1): 41–54.

Battese, G. E., and T. J. Coelli. 1995. A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data. *Empirical Economics* 20(2): 325–332.

Belotti, F., S. Daidone, G. Ilardi, and V. Atella. 2013. Stochastic Frontier Analysis Using Stata. *The Stata Journal: Promoting communications on statistics and Stata* 13(4): 719–758.

Belotti, F., and G. Ilardi. 2018. Consistent Inference in Fixed-Effects Stochastic Frontier Models. *Journal of Econometrics* 202(2): 161–177.

Chen, Y.-Y., P. Schmidt, and H.-J. Wang. 2014. Consistent Estimation of the Fixed Effects Stochastic Frontier Model. *Journal of Econometrics* 181(2): 65–76.

Druska, V., and W. C. Horrace. 2004. Generalized moments estimation for spatial panel data: Indonesian rice farming. *American Journal of Agricultural Economics* 185–198.

Galli, F. 2022. A Spatial Stochastic Frontier Model Including Both Frontier and Error-Based Spatial Cross-Sectional Dependence. *Spatial Economic Analysis* 1–20.

Glass, A. J., K. Kenjegalieva, and R. C. Sickles. 2016. A Spatial Autoregressive Stochastic Frontier Model for Panel Data with Asymmetric Efficiency Spillovers. *Journal of Econometrics* 190(2): 289–300.

Greene, W. 2005. Fixed and Random Effects in Stochastic Frontier Models. *Journal of Productivity Analysis* 23(1): 7–32.

Gude, A., I. Álvarez, and L. Orea. 2018. Heterogeneous Spillovers among Spanish Provinces: A Generalized Spatial Stochastic Frontier Model. *Journal of Productivity Analysis* 50(3): 155–173.

Karakaplan, M., and L. Kutlu. 2017. Handling Endogeneity in Stochastic Frontier Analysis. *Economics Bulletin* 37(2): 889–901.

Karakaplan, M. U. 2017. Fitting Endogenous Stochastic Frontier Models in Stata. *The Stata Journal: Promoting communications on statistics and Stata* 17(1): 39–55.

Kumbhakar, S. C., H. Wang, and A. P. Horncastle. 2015. *A practitioner's guide to stochastic frontier analysis using Stata*. Cambridge University Press.

Kutlu, L., K. C. Tran, and M. G. Tsionas. 2020. A Spatial Stochastic Frontier Model with Endogenous Frontier and Environmental Variables. *European Journal of Operational Research* 286(1): 389–399.

Liu, W., and R. C. Sickles. 2023. Industry-Specific Productivity and Spillovers Through Input–Output Linkages: Evidence from Asia-Pacific Value Chain. *Review of Income and Wealth* .

Meeusen, W., and J. van Den Broeck. 1977. Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. *International Economic Review* 18(2): 435.

Mustafa Ugur Karakaplan. 2018. XTSFKK: Stata Module to Estimate Endogenous Panel Stochastic Frontier Models in the Style of Karakaplan and Kutlu. *Boston College Department of Economics* Statistical Software Components S458445.

Orea, L., and I. C. Álvarez. 2019. A New Stochastic Frontier Model with Cross-Sectional Effects in Both Noise and Inefficiency Terms. *Journal of Econometrics* 213(2): 556–577.

Orea, L., I. C. Álvarez, and T. Jamasb. 2018. A spatial stochastic frontier model with omitted variables: Electricity distribution in Norway. *The Energy Journal* 39(3).

Orea, L., I. C. Álvarez, and L. Servén. 2023. The structural and productivity effects of infrastructure provision in developed and developing countries. *in Parmeter, C. and Wang, H.J. (Eds.). Essays in Honor of Subal Kumbhakar (Advances in Econometrics), Emerald Publishing Limited* .

Wang, H.-J. 2003. A Stochastic Frontier Analysis of Financing Constraints on Investment: The Case of Financial Liberalization in Taiwan. *Journal of Business & Economic Statistics* 21(3): 406–419.

Wang, H.-J., and C.-W. Ho. 2010. Estimating Fixed-Effect Panel Stochastic Frontier Models by Model Transformation. *Journal of Econometrics* 157(2): 286–296.

About the authors

Kerui Du is an associate professor at the School of Management, Xiamen University. His primary research interests include applied econometrics, energy and environmental economics.

Luis Orea is a professor at the School of Economics and Business, University of Oviedo. His primary research interests include Efficiency and productivity analysis, econometric modelling, agricultural economics, energy economics, regulation and competition, spatial economics.

Inmaculada C. Álvarez is a professor at the Department of Economics, Universidad Autónoma de Madrid. Her primary research interest include infrastructures, efficiency and productivity, economic growth and development, spatial economics and quantitative methods.