Fitting spatial stochastic frontier models in Stata

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

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

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

Abstract. In this article, we introduce a new command `xtsfsp` for fitting spatial stochastic frontier models in Stata. Over the last decades, stochastic frontier models have seen important theoretical progress via 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 `xtsfsp` command provides a procedure for estimating 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 present article, we introduce spatial stochastic frontier models, describing the syntax and options of the new command, and providing several examples to illustrate its usage.

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

1 Introduction

Producers might fail to optimize their productive activities, causing deviations from either maximum output or minimum cost. Economic researchers propose the concept of technical efficiency, which measures how well a producer utilizes its resources to produce goods or services. A technically efficient organization produces maximum outputs given an amount of inputs or uses a minimum amount of inputs to produce a given level of output. On the contrary, a technically inefficient organization produces 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 by employing the minimum amount of resources possible, thereby achieving potential cost savings and increased profitability.

Aigner, Lovell, and Schmidt (1977) and Meeusen and Julien (1977) introduced stochastic frontier models for the purpose of 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 inefficiency in the production process. 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 stochastic models by incorporating random effects and "true" fixed effects. Belotti and Ilardi (2018), Chen, Schmidt, and Wang (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 endogeneity in the frontier or inefficiency.

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

Over recent decades with the increasing demand for the analysis of technical efficiency, the Stata package provides official commands `frontier` and `xtfrontier` for cross-sectional and panel stochastic model estimation, respectively. Belotti, Daidone, Ilardi, and Atella (2013) developed `sfcross` and `sfpanel` commands accommodating more different distribution assumptions and allowing fixed-effect and random-effect models which consider heteroscedasticity. Karakaplan (2017) introduced the `sfkk` command for estimating endogenous stochastic frontier models. Karakaplan (2018) supplemented the `xtsfkk` command for fitting endogenous panel stochastic frontier models. Fe and Hofler (2020) provided the `sfcount` command in order to adapt fit count-data stochastic frontier models. Lian, Liu, and Parmeter (2023) developed the `sftt` command for fitting two-tier stochastic frontier models. Kumbhakar, Wang, and Horncastle (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 `xtsfsp`, a new command for fitting spatial stochastic frontier models in the style of Orea and Álvarez (2019) and Galli (2022). The proposed `xtsfsp` command not only permits more accurate inefficiency scores (see e.g., Orea, Álvarez, and Jamasb 2018) but also examines relevant economic issues that a non-spatial stochastic frontier model tends to overlook. For instance, in microdata applications, the new command can be used to test whether the production/cost function can be viewed as a purely deterministic (engineering) process where the firm controls all the inputs (see e.g., Druska and Horrace 2004). A distinctive feature of the `xtsfsp` command is that it allows the estimation of a stochastic frontier model which incorporates cross-sectional correlation in the inefficiency term. Said specification proves useful in applications where some firms benefit from the best practices implemented in adjacent firms due to, for instance, agglomeration economies, knowledge spillovers, technological diffusion or R&D spillovers. This would especially be the case if (local) firms belonging to communitarian networks (e.g. cooperatives) or common technicians (consultants) grant advice to all local firms. In practice, the proposed `xtsfsp` command can be helpful in trying to capture a kind of behavioral correlation. For instance, on those occasions when firms tend to "keep an eye" on the decisions of other peer firms in an attempt to overcome the limitations caused by the lack of information or when they simply wish to emulate each other. It is finally germane to mention that the `xtsfsp` command also allows capturing those cross-sectional effects that might be caused by non-spatial factors (e.g., the regulation environment) if we define appropriately the so-called weight (W) matrix. A proper definition of the W matrix might, for instance, allow us to examine the existence of knowledge spillovers from supplier and user firms.

Despite most applications of the SFA models use firm-level data, these models can also be implemented using macro-level data (e.g., data of countries, regions or industries). In this case, the inefficiency term of our frontier model captures production losses stemming from the inefficient allocation of resources across firms operating within an industry or economy (Restuccia and Rogerson 2013), or from differences in technology or in technical inefficiency across firms (Orea, Álvarez, and Servén 2023). As advocated by Straub (2011), this empirical strategy also allows infrastructure provision to have both a direct effect on regions or countries production as a standard input, and an indirect effect through the inefficiency term as a productivity externality. Therefore, the SFA approach provides useful information for policy makers when it is implemented using macro-data.

We advocate using the proposed `xtsfsp` command to estimate SFA models using macro-level data because the abundant evidence of important feedback processes between neighboring or non-distant regions justify the use of SAR and Durbin frontier functions in such applications. The spatial weight matrix specification commonly adopted in regional economics is based on geographical distance. However, as aforementioned, the weight matrix can be defined using a non-spatial criterion. In this sense, 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 weight matrix should be constructed using the economic distance between industries within/across national economies. In this case, the proposed `xtsfsp` command can be used to estimate spatial SAR and Durbin frontier functions in order to examine the diffusion of knowledge and technology among the participants in the international production network. When using microdata, it also makes sense to estimate a stochastic frontier model with cross-sectional correlation in the inefficiency term if we change the interpretation of the estimated correlation. In these applications, the spatial correlation in the inefficiency term likely captures barriers and distortions to the efficient allocation of resources across firms that are common to several regions, such as regulation, labor market trends or common institutions (see e.g., Orea, Álvarez, and Servén 2023).

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 `xtsfsp`; Section 4 present simulated data examples to illustrate the usage of the command; and Section 5 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 [1]; $\tilde{v}_{it}$ 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 $W_i^{vt} = (W_{i1}^{vt}, ..., W_{iN}^{vt})$, the spatial parameters of lagged independent variables (i.e., $\tilde{v}_{.t}$ and $\tilde{u}_{.t}$), 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 $N(0, \sigma_{v,it}^2)$ and $u_{it} = h(Z_{it}'\delta)u_t^*$. $h(Z_{it}'\delta) = \sqrt{exp(Z_{it}'\delta)}$ 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 $N^+(0, 1)$. Different from the original setting in Orea and Álvarez (2019), we assume $\sigma_{v,it}^2 = exp(D_{it}'\eta)$ to account for idiosyncratic error variance which depends on a vector of variables D_{it}[2]. Furthermore, we enforce the variance of u_t^* to be equal to one, thus allowing the term Z_{it} to include a constant. Using matrix notation, we can rewrite Eqs.(2) and (3) as

[1] The real output in the production function falls short of the potential output due to inefficiency, indicated by $s = -1$ and an output loss represented by $-u_{it}$. Conversely, in the cost function, the real cost exceeds the potential cost due to inefficiency, denoted by $s = 1$, with u_{it} signifying the additional cost.

[2] If $D_{it} = 1$, the error term is conditional homoscedasticity.

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

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

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 [3]. Referring to Wang and Ho (2010), we can obtain the following log-likelihood function for each period t:

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

where $\Pi_t = (I_N - \rho W^{yt})^{-1} diag(\sigma^2_{v,.t})[(I_N - \rho W^{yt})^{-1}]'$, $diag(\sigma^2_{v,.t})$ represents a diagonal matrix with $\sigma^2_{v,.t} = (\sigma^2_{v,1t}, ..., \sigma^2_{v,Nt})'$ as the diagonal elements; $\tilde{\varepsilon}_{.t} = (\tilde{\varepsilon}_{1t}, ..., \tilde{\varepsilon}_{Nt})'$,$\tilde{\varepsilon}_{it} = s(Y_{it} - X'_{it}\beta)$, and

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

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

Galli (2022) further incorporated the spatial lags of the dependent variable and the input variables into the 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 [4], associated with the spatial lagged dependent ($Y_{.t}$) and independent variables ($X_{.t}$), respectively; $Y_{.t} = (Y_{1t}, ..., Y_{Nt})'$; $X_{.t} = (X_{1t}, ..., X_{Nt})'$. This model gives rise to the following log-likelihood function for each period t:

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

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

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

[4] 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

maximize the overall log-likelihood function to obtain consistent estimates of the parameters in the above models. Specifically, we use the Stata `ml model` procedure with the `method-d0` evaluator to program the `xtsfsp` command. Following Gude, Álvarez, and Orea (2018), we parameterize ρ, γ, and τ as Eq.(11) to ensure the standard regularity condition for the spatial autoregressive models.

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

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 `xtsfsp` command.

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

Note: 0 indicates the parameter is restricted to be zero such that the corresponding spatial component is removed.

3 The xtsfsp command

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

3.1 Syntax

Estimation syntax

`xtsfsp` *depvar* [*indepvars*] `, uhet(`*varlist*[`,noconstant`]`)`
[`vhet(`*varlist*[`,noconstant`]`)` `cost` `noconstant` `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(#)

te(*newvar*) genwxvars delmissing constraints(*constraints*)]

Version syntax

xtsfsp , version

Replay syntax

xtsfsp [, level(#)]

3.2 Options

uhet(*varlist* [,noconstant]) specifies explanatory variables for scaling function depending on a linear combination of *varlist*. Use nonconstant to suppresses constant term.

vhet(*varlist* [,noconstant]) specifies explanatory variables for idiosyncratic error variance function depending on a linear combination of *varlist*. Use nonconstant to suppresses constant term.

noconstant suppresses 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 the dependent variable. The expression is wy(W_1 [$W_2 ... W_T$] [,*mata array*]). By default, the weight matrices are spmatrix objects created by Stata official command spmatrix. 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 spatial During terms. The expression is the same as wy(*wyspec*).

wu(*wuspec*) specifies spatial weight matrix for spatial spillover of inefficiency . The expression is the same as wy(*wyspec*).

wv(*wvspec*) specifies spatial weight matrix for spatial dependence of random error. 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 variables for spatial Durbin terms.

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).

`te(`*newvarname*`)` specifies a new variable name to store the estimates of technical efficiency.

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

`delmissing` allows estimation when missing values are present by removing the corresponding units from spatial matrix.

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

3.3 Dependency of xtsfsp

`xtsfsp` depends on the `moremata` package contributed by Jann (2005). If not already installed, you can install it by typing `ssc install moremata`.

4 Examples

The `xtsfsp` command described above offers great flexibility, allowing for various specifications of spatial dependence. Specifically, it enables the specification of different combinations of spatial components, with the option to have different and time-varying spatial weight matrices. Furthermore, it allows for the specification of conditional heteroscedasticity of random errors. In this section, we present four examples that demonstrate the usage of the `xtsfsp` command. To run the following examples, some Community-contributed packages `tictoc` (Kabátek 2022), `translog` (Du 2017), `graph2tex` (Statistical Consulting Group 2017) need to be installed in advance [5].

4.1 yxuv-SAR model with time-invariant spatial weight matrices

Referring to Galli (2022), we first consider the yxuv-SAR model specified by the following data-generating process (DGP 1) with $i = 1, ..., 300$ and $t = 1, .., 20$,

$$y_{it} = 0.3W_i y_{.t} + 2x_{it} + 0.5W_i x_{.t} + \tilde{v}_{it} - \tilde{u}_{it} \tag{12}$$

where $\tilde{v}_{it}$ and $\tilde{u}_{it}$ are defined as in Eqs.(2) and (3) with $\gamma = 0.3$, $\tau = 0.3$, $Z_{it} = (z_{it}, 1)'$,$\delta = (2, ln(0.2))'$, $D_{it} = 1$ and $\eta = ln(0.2)$. 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 into the following syntax.

```
. use xtsfsp_ex1.dta
. xtset id t
Panel variable: id (strongly balanced)
 Time variable: t, 1 to 20
```

[5] We run the codes with Stata 18 MP (16 Cores) in a Mac mini (M2 Chip, 16G RAM).

```
         Delta: 1 unit
. * importing spatial weight matrix from xtsfsp_wmat1.mmat
. mata mata matuse xtsfsp_w1.mmat,replace
(loading w1[300,300])
. * fitting the model
. tic
Timer is turned on!
. xtsfsp y x, uhet(z) wu(w1,mata) wy(w1,mata) wv(w1,mata) ///
>              wx(w1,mata) wxvars(x) te(te) nolog
Spatial frontier model(yxuv-SAR)                     Number of obs =      6,000
                                                     Wald chi2(2)  = 113733.00
Log likelihood = -3953.6034                          Prob > chi2   =     0.0000

------------------------------------------------------------------------------
           y | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
frontier     |
           x |   2.006736   .0082953   241.91   0.000     1.990477    2.022994
         W_x |   .5457545   .0709898     7.69   0.000     .4066171    .6848919
       _cons |   1.008833   .0460347    21.91   0.000     .9186065    1.099059
-------------+----------------------------------------------------------------
    /lnsigv2 |  -1.566886   .0184019   -85.15   0.000    -1.602953   -1.530819
-------------+----------------------------------------------------------------
uhet         |
           z |   .9858524   .0385881    25.55   0.000      .910221    1.061484
       _cons |  -1.533232   .3361919    -4.56   0.000    -2.192156   -.8743081
-------------+----------------------------------------------------------------
Wy           |
       _cons |     .57773    .062202     9.29   0.000     .4558163    .6996436
-------------+----------------------------------------------------------------
Wv           |
       _cons |   .6066382   .0743834     8.16   0.000     .4608494     .752427
-------------+----------------------------------------------------------------
Wu           |
       _cons |   .6584973   .0938095     7.02   0.000      .474634    .8423606
-------------+----------------------------------------------------------------
         rho |   .2810617   .0286408     9.81   0.000     .2240201    .3361839
       gamma |   .2943177    .033966     8.67   0.000     .2264087    .3593786
         tau |   .3178137    .042162     7.54   0.000     .2329366    .3978845
------------------------------------------------------------------------------
Note: Wy:_cons, Wv:_cons and Wu:_cons are the transformed parameters;
      rho, gamma and tau are their origin metrics in spatial components, respectiv
> ely.
      W_(x) represent Spatial Durbin terms W(x)
. toc
Elapsed time: 252.347 sec
             =4 min & 12.347 sec
```

The output shows that the command fits six equations with the `ml model`. The frontier equation has three explanatory variables x_{it}, $W_i x_{.t}$ and constant. The scaling function uhet() has two explanatory variables Z_{it} and constant. The equation /lnsigv2 is constructed for the variance parameter σ_v^2 which is 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 $W_i x_{.t}$ in the frontier equation (represented by W_x) such that we do not need to fit a separate equation. The bottom of the table reports the transformed parameters in the original metric.

4.2 xuv-SAR model with different spatial weight matrices

We consider a restricted model xuv-SAR with different spatial weight matrices, one of which is time-varying, whilst the others are time-constant. The model is described as DGP 2:

$$y_{it} = 1 + 2x_{it} + 0.5W_i^{xt}x_{it} + \tilde{v}_{it} + \tilde{u}_{it}, i = 1, .., 300; t = 1, .., 10 \tag{13}$$

where the other parameters are set the same as in DGP 1 except for $W_i^{ut} = W_i^u$, $W_i^{vt} = W_i^v$, $\delta = (4, ln(0.2))'$, and W_i^{xt} is time-varying. Different from DGP 1, which sets the production function frontier, DGP 2 specifies a cost function. The estimation of the model is shown as follows.

```
. * importing spatial weight matrices from xtsfsp_w2.mmat
. mata mata matuse xtsfsp_w2.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])
. local w w1 w2 w3 w4 w5 w6 w7 w8 w9 w10
. use xtsfsp_ex2.dta
. xtset id t
Panel variable: id (strongly balanced)
 Time variable: t, 1 to 10
         Delta: 1 unit
. * initial values for estimated parameters
. mat b=(2,0.5,1,-1.5,4,-1.5,0.6,0.6)
. * fitting the model
. tic
Timer is turned on!
. xtsfsp y x, cost uhet(z) wu(w2,mata) wv(w1,mata) wxvars(x) ///
>             wx(`w´,mata) init(b) genwxvars nolog
Spatial frontier model(xuv-SAR)                        Number of obs =     3,000
                                                       Wald chi2(2)  = 57441.23
Log likelihood = -1911.966                             Prob > chi2   =   0.0000
```

y	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
frontier						
x	1.995755	.0083855	238.00	0.000	1.97932	2.01219
W_x	.5069274	.0221354	22.90	0.000	.4635428	.5503121
_cons	.9917396	.0127299	77.91	0.000	.9667894	1.01669
/lnsigv2	-1.614693	.0260288	-62.03	0.000	-1.665708	-1.563677
uhet						
z	3.999532	.002155	1855.90	0.000	3.995308	4.003755
_cons	-1.699077	.4472546	-3.80	0.000	-2.57568	-.8224746
Wv						
_cons	.5877882	.0584159	10.06	0.000	.4732951	.7022812
Wu						
_cons	.6214134	.0008532	728.34	0.000	.6197412	.6230857

```
         tau |   .3010498   .0003879   776.12   0.000     .3002893    .3018098
       gamma |   .2856864   .0268209    10.65   0.000     .2323035     .337353
------------------------------------------------------------------------------
Note: Wv:_cons and Wu:_cons are the transformed parameters;
      gamma and tau are their origin metrics in spatial components, respectively.
      W_(x) represent Spatial Durbin terms W(x)
. toc
Elapsed time: 143.107 sec
             =2 min & 23.107 sec
```

In the second example, we use a `cost` option to specify the type of frontier. The matrix `b` is utilized as the initial value for maximum likelihood estimation. The likelihood function of spatial stochastic frontier models is intricate and typically challenging when trying to obtain optimal global solutions. Therefore, having good initial values would be beneficial for fitting spatial stochastic models. In order to acquire initial values for spatially-correlated parameters, practitioners can start by fitting non-spatial stochastic models using the `frontier` and `sfpanel` commands. This way, the initial values of the parameters involved in the frontier and the scaling function are obtained. Subsequently, the `mlplot` option can be employed to search for better initial values.

To demonstrate the usage of the *delmissing* option, two observations of y_{it} are replaced with missing values, and the aforementioned codes are re-run, resulting in the error message "*missing values found. use delmissing to remove the units from the spmatrix*". The inclusion of the *delmissing* option addresses this issue, and the generated variable __e_sample__ records the regression sample. The execution time of this example is relatively long, exceeding 17 minutes. This can be attributed to the presence of missing values, which leads to data imbalance and necessitates the computation of time-varying spatial weight matrices due to changing dimensions over time. Equations (4)-(8) involve solving the inverses of three $N \times N (N \in \{299, 300\})$ matrices, and considering the time-varying nature of these matrices, their computation needs to be performed T times in each step for optimizing likelihood functions. Consequently, this particular example takes nearly eight times longer than the previous one.

```
. * replace some observations of y to be missing
. replace y=. if _n==1 | _n==100
(2 real changes made, 2 to missing)
. local w w1 w2 w3 w4 w5 w6 w7 w8 w9 w10
. * estimation is aborted
. cap noi xtsfsp y x, cost uhet(z) wu(w2,mata) wv(w1,mata) ///
>                     wxvars(x) wx(`w´,mata) init(b) nolog
missing values found. use delmissing to remove the units from the spmatrix
invalid syntax

  . * re-estimation with delmissing option
. local w w1 w2 w3 w4 w5 w6 w7 w8 w9 w10
. tic
Timer is turned on!
. xtsfsp y x, cost uhet(z) wu(w2,mata) wv(w1,mata) wxvars(x) ///
>              wx(`w´,mata) init(b) delmissing nolog
missing values found. The corresponding units are deleted from the spmatrix
Spatial frontier model(xuv-SAR)                       Number of obs =    2,998
```

```
                                                      Wald chi2(2)  = 57427.27
Log likelihood = -1909.3843                           Prob > chi2   =   0.0000

------------------------------------------------------------------------------
           y | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
frontier     |
           x |   1.996019   .0083865   238.00   0.000     1.979582    2.012456
         W_x |   .5065641   .0221317    22.89   0.000     .4631868    .5499415
       _cons |   .9914628   .0127155    77.97   0.000     .9665409    1.016385
-------------+----------------------------------------------------------------
    /lnsigv2 |  -1.615516   .0260354   -62.05   0.000    -1.666545   -1.564488
-------------+----------------------------------------------------------------
uhet         |
           z |   3.999476   .0021537  1857.00   0.000     3.995254    4.003697
       _cons |   -1.69889   .4472536    -3.80   0.000    -2.575491   -.8222893
-------------+----------------------------------------------------------------
Wv           |
       _cons |   .5864895   .0584601    10.03   0.000     .4719098    .7010692
-------------+----------------------------------------------------------------
Wu           |
       _cons |   .6214135   .0008526   728.87   0.000     .6197425    .6230845
-------------+----------------------------------------------------------------
         tau |   .3010498   .0003876   776.69   0.000     .3002899    .3018093
       gamma |     .28509   .0268512    10.62   0.000     .2316482    .3368159
------------------------------------------------------------------------------
Note: Wv:_cons and Wu:_cons are the transformed parameters;
      gamma and tau are their origin metrics in spatial components, respectively.
      W_(x) represent Spatial Durbin terms W(x)
      Missing values found
      The regression sample recorded by variable __e_sample__

. toc
Elapsed time: 1068.042 sec
             =17 min & 48.042 sec
```

4.3 uv-SAR model with conditional heteroscedasticity of random errors

We set up a restricted model uv-SAR with time-varying spatial weight matrices and conditional heteroscedasticity of random errors. The DGP 3 is described as

$$\begin{aligned} y_{it} &= 1 + 2x_{it} + \tilde{v}_{it} - \tilde{u}_{it}, i = 1,..,300; t = 1,..,10 \\ \sigma^2_{v,it} &= exp(1 + d_{it}) \\ h(Z'_{it}\delta) &= \sqrt{exp(1 + z_{it})} \end{aligned} \tag{14}$$

where the other parameters are set the same as in DGP 1 except for $W_i^{ut} = W_i^{vt} = W_i^t$. The following syntax estimates the model alongside the results. The computation burden in this example exceeds 11 minutes, primarily attributed to the time-varying configuration of the spatial weight matrices.

```
. * importing spatial weight matrices from xtsfsp_w3.mmat
. mata mata matuse xtsfsp_w3,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])
. local w w1 w2 w3 w4 w5 w6 w7 w8 w9 w10
. use xtsfsp_ex3.dta
. xtset id t
Panel variable: id (strongly balanced)
 Time variable: t, 1 to 10
         Delta: 1 unit
. tic
Timer is turned on!
. xtsfsp y x, wu(`w´,mata) wv(`w´,mata) uhet(z) vhet(d) nolog
Spatial frontier model(uv-SAR)                         Number of obs =   3,000
                                                       Wald chi2(1)  = 7472.38
Log likelihood = -5803.0516                            Prob > chi2   =  0.0000
```

y	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
frontier						
x	1.988108	.0229991	86.44	0.000	1.943031	2.033185
_cons	.8997916	.0737763	12.20	0.000	.7551927	1.044391
lnsigv2						
d	.9897965	.0259185	38.19	0.000	.9389972	1.040596
_cons	.9950927	.0259519	38.34	0.000	.944228	1.045957
uhet						
z	1.019351	.0458701	22.22	0.000	.9294473	1.109255
_cons	1.122574	.4629594	2.42	0.015	.2151904	2.029958
Wv						
_cons	.6171649	.0431991	14.29	0.000	.5324961	.7018337
Wu						
_cons	.6471479	.0703691	9.20	0.000	.509227	.7850688
tau	.299117	.0196647	15.21	0.000	.2601042	.3371547
gamma	.3127036	.0317402	9.85	0.000	.2492256	.3735057

```
Note: Wv:_cons and Wu:_cons are the transformed parameters
      gamma and tau are their origin metrics in spatial components, respectively.
. toc
Elapsed time: 669.586 sec
             =11 min & 9.586 sec
```

4.4 Example with real data

We use the real data to exemplify how to conduct the empirical studies with the models and the command described above.The raw data on the provinces of China, including GDP (denoted as Y), investment, labor force (denoted as L), the ratio of government expenditure to GDP (denoted as fiscal), the ratio of FDI to GDP (denoted as fdi), and trade as a percentage of GDP (denoted as $trade$), is collected from the CSMAR database. All nominal variables are adjusted to constant prices in 1997. The capital stocks for the provinces are estimated using the perpetual inventory method. The production function is approximated by the translog function, and the inefficiency term's scaling function is assumed to be determined by the ratio of government expenditure to GDP, the ratio of FDI to GDP, and trade as a percentage of GDP.

```
. * Translate shapefile to Stata format
. cap spshape2dta province
.
. use province
. drop if _ID == 26 | _ID>31
(4 observations deleted)
. spset
      Sp dataset: province.dta
Linked shapefile: province_shp.dta
            Data: Cross sectional
 Spatial-unit ID: _ID
     Coordinates: _CX, _CY (planar)
. * Create spatial contiguity matrix
. spmatrix create contiguity w_con, normalize(none)
  weighting matrix in w_con contains 1 island
. * Obtain spatial matrix as Mata matrix wm from w_con
. spmatrix matafromsp wm id = w_con
. * Match the iland (_ID = 21) with the nearest province(_ID =19)
. mata: wm[19,21]=1
. mata: wm[21,19]=1
. * Create spmatrix w_con from Mata matrix wm and rwo-normalized the matrix
. spmatrix spfrommata w_con= wm id, normalize(row) replace
. * Obtain the new spatial matrix as Mata matrix wm from w_con
. spmatrix matafromsp wm id = w_con
.
. use chnempirical.dta,clear
. * Generate varables for the translog function
. qui translog Y K L , time(year) norm
. global x  lnK lnL _t lnK_lnL _t_lnK _t_lnL _t_2 lnK_2 lnL_2
. global z fiscal trade fdi
. * Fit the model with frontier command
. frontier lnY $x,uhet($z) nolog
Stoc. frontier normal/half-normal model              Number of obs =       630
                                                     Wald chi2(9)  = 33273.49
Log likelihood = 307.21728                           Prob > chi2   =    0.0000
```

lnY	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
lnY						
lnK	.7939407	.020148	39.41	0.000	.7544513	.83343
lnL	.2423919	.0152754	15.87	0.000	.2124527	.2723312
_t	-.0159651	.0029995	-5.32	0.000	-.021844	-.0100862
lnK_lnL	-.0239117	.0463983	-0.52	0.606	-.1148507	.0670274
_t_lnK	.0385162	.0100006	3.85	0.000	.0189153	.058117
_t_lnL	-.0212483	.0070556	-3.01	0.003	-.035077	-.0074197
_t_2	-.0064186	.0009174	-7.00	0.000	-.0082166	-.0046206
lnK_2	-.0151869	.0327529	-0.46	0.643	-.0793815	.0490077
lnL_2	-.0604274	.031823	-1.90	0.058	-.1227993	.0019444

```
       _cons |   .3105853   .0138557    22.42   0.000     .2834287    .3377419
-------------+----------------------------------------------------------------
lnsig2v      |
       _cons |  -4.491706   .0912372   -49.23   0.000    -4.670528   -4.312885
-------------+----------------------------------------------------------------
lnsig2u      |
      fiscal |   .0724787   .0116298     6.23   0.000     .0496848    .0952726
       trade |  -.0790757   .0165127    -4.79   0.000      -.11144   -.0467115
         fdi |  -.0262741    .006819    -3.85   0.000    -.0396391   -.0129091
       _cons |  -2.964112   .3476526    -8.53   0.000    -3.645499   -2.282726
-------------+----------------------------------------------------------------
     sigma_v |   .1058372   .0048281                      .0967849    .1157361
------------------------------------------------------------------------------

. * Predict the inefficiency term and efficiency scores
. predict double uhat, u

. generate double te0 = exp(-uhat)

. * Store the estimated parameters
. mat b0=e(b)

.
. * Fit the model with xtsfsp command
. xtset _ID year

Panel variable: _ID (strongly balanced)
 Time variable: year, 1997 to 2017
         Delta: 1 unit

. mat b1 = b0,0.6,0.6,0.6

. xtsfsp lnY $x, uhet($z) wy(wm,mata) wu(wm,mata) wv(wm,mata) ///
>                init(b1) te(tesp1) nolog

Spatial frontier model(yuv-SAR)                        Number of obs =      630
                                                       Wald chi2(9)  = 16704.73
Log likelihood = 328.51183                             Prob > chi2   =   0.0000

------------------------------------------------------------------------------
         lnY | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
frontier     |
         lnK |   .6720836   .0259441    25.91   0.000     .6212341    .7229331
         lnL |   .3390601   .0185192    18.31   0.000      .302763    .3753571
          _t |   .0141493   .0051647     2.74   0.006     .0040267    .0242719
     lnK_lnL |   .0075504    .051805     0.15   0.884    -.0939855    .1090864
      _t_lnK |   .0425662   .0115137     3.70   0.000     .0199998    .0651326
      _t_lnL |  -.0034571   .0079979    -0.43   0.666    -.0191327    .0122186
        _t_2 |  -.0069112   .0013048    -5.30   0.000    -.0094686   -.0043539
       lnK_2 |  -.0875718   .0344022    -2.55   0.011     -.154999   -.0201446
       lnL_2 |  -.0388746   .0377404    -1.03   0.303    -.1128443    .0350952
       _cons |   .4968023    .043107    11.52   0.000     .4123142    .5812903
-------------+----------------------------------------------------------------
     /lnsigv2 |  -4.025468   .0581939   -69.17   0.000    -4.139526    -3.91141
-------------+----------------------------------------------------------------
uhet         |
      fiscal |   .0098788   .0056888     1.74   0.082     -.001271    .0210287
       trade |  -.0911438   .0213384    -4.27   0.000    -.1329663   -.0493212
         fdi |  -.0185831   .0100212    -1.85   0.064    -.0382243    .0010582
       _cons |  -1.963808   .3894655    -5.04   0.000    -2.727146   -1.200469
-------------+----------------------------------------------------------------
Wy           |
       _cons |  -.0467557   .0362548    -1.29   0.197    -.1178138    .0243023
-------------+----------------------------------------------------------------
Wv           |
```

```
       _cons |  -.0671769   .1600686    -0.42   0.675    -.3809057    .2465518
-------------+----------------------------------------------------------------
Wu           |
       _cons |   1.280422   .2367245     5.41   0.000       .81645    1.744393
-------------+----------------------------------------------------------------
         rho |  -.0233713   .0181157    -1.29   0.197     -.058833    .0121493
       gamma |  -.0335725   .0799361    -0.42   0.674    -.1881642     .122643
         tau |   .5649866   .0805643     7.01   0.000     .3869258    .7024182
------------------------------------------------------------------------------
Note: Wy:_cons, Wv:_cons and Wu:_cons are the transformed parameters;
      rho, gamma and tau are their origin metrics in spatial components, respectiv
> ely.

. scalar loglikehood1 =  e(ll)

. mat b1 = b0,0.6

. xtsfsp lnY $x, uhet($z)  wu(wm,mata)  init(b1) te(tesp2)  nolog

Spatial frontier model:u-SAR                        Number of obs =      630
                                                    Wald chi2(9)  = 16174.91
Log likelihood = 327.60816                          Prob > chi2   =   0.0000
------------------------------------------------------------------------------
         lnY | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
frontier     |
         lnK |    .672366   .0264662    25.40   0.000     .6204932    .7242388
         lnL |   .3392661   .0184417    18.40   0.000     .3031209    .3754113
          _t |   .0114801    .004803     2.39   0.017     .0020664    .0208937
     lnK_lnL |   .0128618   .0467126     0.28   0.783    -.0786932    .1044169
      _t_lnK |   .0422561   .0108054     3.91   0.000     .0210779    .0634344
      _t_lnL |  -.0027543   .0074782    -0.37   0.713    -.0174113    .0119027
        _t_2 |  -.0066348   .0012629    -5.25   0.000      -.00911   -.0041595
       lnK_2 |  -.0915034   .0326073    -2.81   0.005    -.1554126   -.0275942
       lnL_2 |  -.0395255   .0328292    -1.20   0.229    -.1038696    .0248185
       _cons |   .4644181   .0315665    14.71   0.000     .4025489    .5262874
-------------+----------------------------------------------------------------
    /lnsigv2 |  -4.013496   .0575493   -69.74   0.000     -4.12629   -3.900701
-------------+----------------------------------------------------------------
uhet         |
      fiscal |   .0097757      .0055     1.78   0.076     -.001004    .0205555
       trade |  -.0869182   .0204711    -4.25   0.000    -.1270408   -.0467956
         fdi |  -.0133135   .0090702    -1.47   0.142    -.0310908    .0044638
       _cons |  -1.986602   .3822161    -5.20   0.000    -2.735732   -1.237472
-------------+----------------------------------------------------------------
Wu           |
       _cons |    1.07423   .1922999     5.59   0.000     .6973292    1.451131
-------------+----------------------------------------------------------------
         tau |    .490752   .0729815     6.72   0.000     .3351572    .6202829
------------------------------------------------------------------------------
Note: Wu:_cons is the transformed parameters;
      tau is the origin metric in the spatial components.

. scalar loglikehood2 =  e(ll)

. local lrtest = -2*(loglikehood2-loglikehood1)

. local pvalue = 1- chi2(2,`lrtest´)

. display "Likelihood-ratio test: LR chi2(2) = `lrtest´, Prob > chi2 = `pvalue´"
Likelihood-ratio test: LR chi2(2) = 1.807357699424301, Prob > chi2 = .405076698937
> 9054

.
. * Plot the density of estimates of technical efficiency from different models
. twoway (kdensity te0, color(black) lpattern(solid)) ///
```

```
>        (kdensity tesp1,color(red) lpattern(dash))   ///
>            (kdensity tesp2,color(blue) lpattern(longdash)),  ///
>         legend(pos(10) ring(0) label(1 Non-spatial Stoc. Frontier) ///
>         label(2 Spatial Stoc. Frontier:yuv) label(3 Spatial Stoc. Frontier:u)) /
> //
>         xtitle("Technical Efficiency") ytitle("Density")
. graph2tex, epsfile("./fig1") ///
>            caption(Distribution of efficency scores) label(fig1)
% exported graph to ./fig1.eps
% We can see in Figure \ref{fig:fig1} that
\begin{figure}[h]
\begin{centering}
  \includegraphics[height=3in]{./fig1}
  \caption{Distribution of efficency scores}
  \label{fig:fig1}
\end{centering}
\end{figure}
.
```

In order to construct the spatial weight matrix, we use the geographic data (province.shp and province.dbf). The Stata command **shshape2dta** is used to convert the data. Units with substantial amounts of missing values are dropped. Based on the generated province.dta, the spatial contiguity matrix is constructed using Stata official spmatrix procedures [6]. As _ID 21 represents an island, we define its contiguity unit as the nearest one. Consequently, we extract the spatial weight matrix as wm in the Mata environment from the spmatrix object w_con and assign the elements wm[19,21] and wm[21,19] a value of one. Furthermore, the spmatrix routine is used to standardize the matrix. The spmatrix object w_con can be used for the **xtsfsp** command. Alternatively, the new w_con can be placed into the Mata environment as a matrix wm.

We initially fit a non-spatial stochastic frontier model using the Stata official command **frontier**. We predict the efficiency scores and store them in the new variable te0. The parameter estimates are extracted to serve as initial values in the spatial stochastic frontier models. We then consider three sources of spatial dependence, known as the yuv-SAR model. The estimated results show that the spatial correlation coefficients ρ and γ are statistically insignificant at the 10% level. Therefore, a restricted model with spatial dependence in the inefficient term is considered. Since the restricted model is nested within the previous model, a likelihood ratio test is performed for model selection. The results indicate that the null hypothesis, which states that the restricted model cannot be rejected, is supported. The significance of τ suggests positive spatial spillovers of technical efficiency. The empirical results reveal that trade increases technical efficiency, while government intervention, as proxied by the ratio of government expenditure to GDP, decreases technical efficiency in Chinese provinces. Figure 1 illustrates the distribution of efficiency scores estimated by different models. The non-spatial stochastic frontier model significantly overestimates technical efficiency. The spatial stochastic models yuv-SAR and u-SAR provide similar estimates of technical efficiency.

[6] Alternatively, the community-contributed command **spwmatrix** by Jeanty (2010) can be used to create a spatial weight matrix in Stata

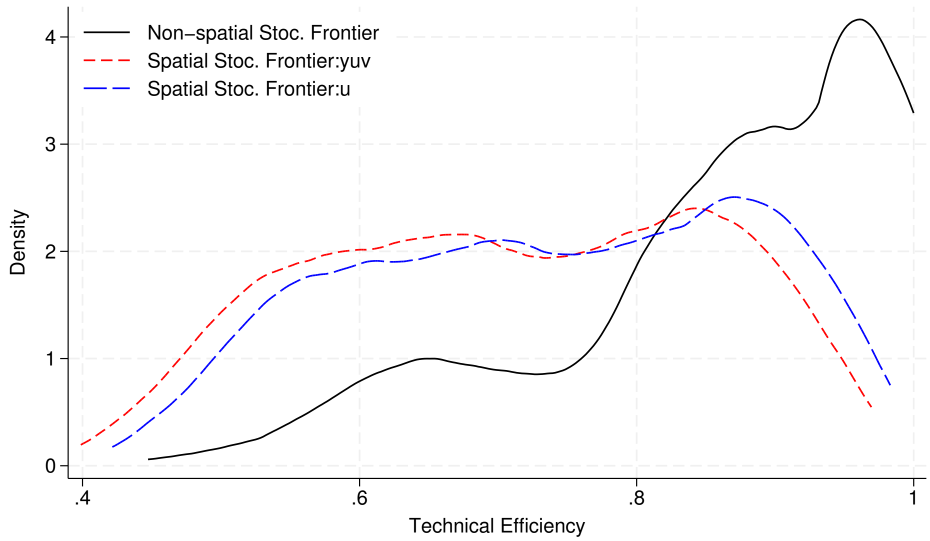

Figure 1: Distribution of efficiency scores

5 Conclusion

Geospatial units are not isolated or separated but interconnected. For instance, economic trade, social activities, and cultural exchange between different regions mutually influence each other. This spatial interdependence poses a challenge to traditional econometric methods, which generally assume cross-sectional independence. Spatial econometrics has been developed to address spatial correlation. This article presents a community-contributed command that facilitates the fitting of spatial stochastic frontier models, accounting for different sources of spatial dependence. We hope that this developed command can provide convenience to practitioners and reduce the complexity of model applications, thereby promoting robust empirical research.

However, there are certain limitations that should be acknowledged. Firstly, despite the flexibility of spatial stochastic frontier models and the introduced command, they rely on full parameterization of the spatial structure, frontier function, and distribution of random errors and inefficiency terms, following the spirit of stochastic frontier models and spatial econometrics. Model misspecification can be a potential concern, and theoretical works exploring relaxed parametric assumptions would be valuable. Secondly, as Ayouba (2023) has recently pointed out, the spatial parametric frontier models could greatly benefit from putting identification issues and economic theory at the center of the estimation process, using e.g., the fast-growing literature on peer effects in networks. Thirdly, these models require prior information on the spatial weight matrices. Different choices of spatial weight matrices may lead to varying results. Although an internalization of W in spatial stochastic frontier models remains an open issue, Ayouba (2023) states that Bayesian and model averaging approaches could mitigate the uncertainty arising from W. A related issue is the potencial endogeneity of the spatial weight matrix. Therefore, allowing W to be endogenous is another interesting direction of research. Finally, the numerical computation for maximum likelihood estimation (MLE) of spatial stochastic frontier models is highly complex. When dealing with large spatial weight matrices, especially when considering time-varying spatial weight matrices, the estimation process becomes computationally intensive as the matrices need to be repeatedly inverted. Additionally, the case of time-varying spatial weight matrices with large dimensions may prove memory intensive.

6 Acknowledgments

Kerui Du thanks the financial support of the National Natural Science Foundation of China (grant number 72074184). Luis Orea and Inmaculada Álvarez also thanks the financial support of the Ministry for Science and Technology (grant number PID2020-113076GB-I00). 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 `xtsfsp` command. We greatly appreciate Stephen P. Jenkins (the editor) and an anonymous referee for the helpful comments and suggestions that have led to an improved version of this article and the econometric package.

7 References

Aigner, D., C. Lovell, and P. Schmidt. 1977. Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics* 6: 21–37.

Ayouba, K. 2023. Spatial dependence in production frontier models. *Journal of Productivity Analysis* 60: 21–36.

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: 325–332.

Belotti, F., S. Daidone, G. Ilardi, and V. Atella. 2013. Stochastic frontier analysis using Stata. *Stata Journal* 13: 719–758.

Belotti, F., and G. Ilardi. 2018. Consistent inference in fixed-effects stochastic frontier models. *Journal of Econometrics* 202: 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: 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.

Du, K. 2017. TRANSLOG: Stata module to create new variables for a translog function.

Fe, E., and R. Hofler. 2020. sfcount: Command for count-data stochastic frontiers and underreported and overreported counts. *Stata Journal* 20: 532–547.

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: 289–300.

Greene, W. 2005. Fixed and random effects in stochastic frontier models. *Journal of Productivity Analysis* 23: 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: 155–173.

Jann, B. 2005. MOREMATA: Stata module (Mata) to provide various functions. Statistical Software Components, Boston College Department of Economics.

Jeanty, P. W. 2010. SPWMATRIX: Stata module to generate, import, and export spatial weights. Statistical Software Components, Boston College Department of Economics.

Kabátek, J. 2022. TICTOC: Stata module to provide simplified timer commands. Statistical Software Components, Boston College Department of Economics.

Karakaplan, M. 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.

Karakaplan, M., and L. Kutlu. 2017. Handling endogeneity in stochastic frontier analysis. *Economics Bulletin* 37: 889–901.

Karakaplan, M. U. 2017. Fitting endogenous stochastic frontier models in Stata. *Stata Journal* 17: 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: 389–399.

Lian, Y., C. Liu, and C. F. Parmeter. 2023. Two-tier stochastic frontier analysis using Stata. *Stata Journal* 23: 197–229.

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* forthcoming. URL https://doi.org/10.1111/roiw.12639.

Meeusen, W., and v. D. B. Julien. 1977. Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review* 18: 435–444.

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: 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: 93–116.

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

Restuccia, D., and R. Rogerson. 2013. Misallocation and productivity. *Review of Economic Dynamics* 16: 1–10.

Statistical Consulting Group. 2017. GRAPH2TEX: Convert Stata graph to .eps with LATEX code. UCLA: Statistical Consulting Group. URL https://stats.idre.ucla.edu/stat/stata/ado/analysis.

Straub, S. 2011. Infrastructure and development: A critical appraisal of the macro-level literature. *Journal of Development Studies* 47: 683–708.

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: 406–419.

Wang, H.-J., and C.-W. Ho. 2010. Estimating fixed-effect panel stochastic frontier models by model transformation. *Journal of Econometrics* 157: 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 full 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 and spatial economics.

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