MODELLING OF SPATIAL EFFECTS IN TRANSPORT EFFICIENCY: THE ‘SPFRONTIER’ MODULE OF ‘R’ SOFTWARE

Dmitry Pavlyuk

Candidate of Economic Science, Transport and Telecommunication Institute
Lomonosova 1, Riga, LV-1019, Latvia
Phone: (+371)29958338, Email: Dmitry.Pavlyuk@tsi.lv

This research is devoted to analysis of spatial effects in modelling of transport efficiency. Four different types of spatial effects, essential for efficiency modelling, are discussed and a spatial modification of the stochastic frontier model, which includes effects of the specified types, is formulated. We also discuss a problem of maximum likelihood estimation of the SSF model and its implementation in the ‘spfrontier’ package. We suggest utilising of the SSF model and the ‘spfrontier’ package for different areas of transport modelling and compile a list of topical transport problems, where a presence of spatial effects is widely acknowledged.

Keywords: spatial effects, stochastic frontier, transport, efficiency estimation

1. Introduction

The importance of spatial effects is widely acknowledged in different areas of transport science. Spatial interactions take a place in national, regional, and less aggregated levels. For example, regional science strongly states spatial spillovers of transport infrastructure on industry, agriculture, and overall regional economic and social development (Tong et al. 2013; Cohen 2010); spatial design of public transport routes and movements are an essential part of urban spatial planning; spatial competition for traffic among transport enterprises is a feature of modern liberalised transportation market.

Spatial econometrics (Anselin 1988) provides an extensive set of tools for modelling different types of spatial effects. Spatial effects include relationships between neighbour units (e.g. spatial competition) and spatial heterogeneity (influence of area-specific factors like climate, economics and others).

Economic efficiency of units like transport enterprises and transport systems is a key indicator, frequently used in decision making processes on different levels. Stochastic frontier model (Kumbhakar & Lovell 2003) is one of the most popular tools for statistical estimation of units’ efficiency. Recently significant research efforts have been made to introduce spatial effects in stochastic frontier (Druska & Horrace 2004; Affuso 2010; Barrios & Lavado 2010).

Estimation of spatial effects is quite complicated both from theoretical and computational points of view. This complexity is a serious obstacle for inclusion of spatial effects into everyday work of economists and urban managers. A stable and convenient software tool can bridge this gap. The “spfrontier” tool is one of the first steps in this direction. The tool is implemented as a module for CRAN R, a free software environment for statistical computing and graphics. ‘spfrontier’ is developed by the author, so the main goal of this research is to introduce this tool to scientific community and investigate its facilities for modelling of efficiency in transport science and adjacent research areas.

2. Spatial stochastic frontier model

A classical linear stochastic frontier model is formulated as:

\[ Y_i = \sum_{k=1}^{K} X_{ki} \beta_k + \tilde{v}_i - \tilde{u}_i, \]

where
- \( i \) is a unit index, \( i = 1, \ldots, n, \)
- \( Y_i \) is an output of a unit \( i, \)
- \( X_{ki} \) are inputs of a unit \( i, k = 1, \ldots, K, \)
- \( \beta_k \) are coefficients, representing direct effects of inputs, \( k = 1, \ldots, K, \)
\( \tilde{v}_i \) are independent identically distributed (IID) random disturbances, 
\( \tilde{u}_i \) are IID inefficiency levels.

In the context of the stochastic frontier model, we specify a hypothesis about existence of the following four types of spatial effects:
- Endogenous spatial effects
- Exogenous spatial effects
- Spatially correlated random disturbances
- Spatially related efficiency

Endogenous spatial effects represent a relationship between an output (or, more generally, decisions) of a particular unit and outputs of its neighbours.

Exogenous spatial effects represent a relationship between an output of a particular unit and inputs (explanatory factors) of its neighbours. These effects can be explained by indirect flow of resources into neighbourhood, where a production process and output registration can be separated in space.

The third type of spatial effects, spatially correlated random disturbances, is not based on a theory, but usually is consistent with modelling features. Suppose a model, where observations are affected by an unobserved factor with a spatial nature. Some of these factors can be technically unobservable; some of them are just not available in a research sample. Spatial heterogeneity of these factors leads to spatially correlated random disturbances in a model.

The fourth type of spatial effects, spatially related efficiency, reflects a relationship between efficiency of neighbour units. This type of effects is under-researched and rarely used in applications. Researches, where spatially related efficiency is included into the model, are limited with (Druska & Horrace 2004), (Barrios & Lavado 2010), (Areal et al. 2010), (Tonini & Pede 2011), (Fusco & Vidoli 2013), among few others. A reasoning of spatially related efficiency is very similar to endogenous spatial effects and includes possible spatial impact of units.

Note that first three types of spatial effects are well known (Elhorst 2009), but spatial effects in efficiency are a relative novelty.

A complete stochastic frontier linear model with all types of spatial effects (SSF, spatial stochastic frontier) takes the form:

\[
Y_i = \rho_Y \sum_{j=1}^{n} w_{y,ij} Y_j + \sum_{k=1}^{K} X_{ki} \beta_k + \sum_{k=1}^{K} \left( \beta_k^{(s)} \sum_{j=1}^{n} w_{x,ij} X_{kj} \right) + v_i - u_i,
\]

\[
v_i = \rho_v \sum_{j=1}^{n} w_{v,ij} v_j + \tilde{v}_i,
\]

\[
u_i = \rho_u \sum_{j=1}^{n} w_{u,ij} u_j + \tilde{u}_i,
\]

where
- \( w_{y,ij} \) are spatial weights for spatial endogenous effects between units \( i \) and \( j \), \( j = 1, \ldots, n \),
- \( w_{x,ij} \) are spatial weights for spatial exogenous effects between units \( i \) and \( j \),
- \( w_{v,ij} \) are spatial weights for spatially correlated random disturbances of units \( i \) and \( j \),
- \( w_{u,ij} \) are spatial weights for spatially related efficiency of units \( i \) and \( j \),
- \( \beta_k^{(s)} \) are coefficients, representing spatial exogenous effects of inputs, \( k = 1, \ldots, K \),
- \( \rho_Y \) is a coefficient, representing spatial endogenous effects,
- \( \rho_v \) is a coefficient, representing spatially correlated random disturbances,
- \( \rho_u \) is a coefficient, representing spatially related efficiency.

Folding the model by \( i, j \), and \( k \), we formulate the model in the matrix form:

\[
Y = \rho_Y W_Y Y + X \beta + W_X X \beta^{(s)} + v - u,
\]

\[
v = \rho_v W_v v + \tilde{v},
\]

\[
u = \rho_u W_u u + \tilde{u}.
\]

A set of methods, used to estimation of a classical stochastic frontier model, includes method of moments, maximum likelihood estimator (MLE), generalised maximum entropy estimator, and Bayesian estimator. Objectives of the estimators are estimation of the production frontier parameters \( \beta \) and values of the technical inefficiency \( u_i \) for each unit in the sample. The SSF model also requires estimation of coefficients \( \rho_Y, \beta^{(s)}, \rho_v, \) and \( \rho_u \) for spatial effects of four types.
MLE is one of the most frequently utilised techniques in stochastic frontier analysis. Likelihood functions for different specifications of the classical stochastic frontier model are well known (Kumbhakar & Lovell 2003) and implemented in specialised software packages (R, Stata, LIMDEP). Such likelihood functions utilise an assumption of error term independency and represented as a product of univariate distribution densities. The independency assumption is obviously not satisfied for spatially correlated components of the SSF model’s composed error term, so a multivariate closed skew normal (CSN) distribution is used as a base for MLE (Domínguez-Molina et al. 2003). Estimation of CSN distribution parameters itself is a complicated task, which is weakly covered in literature and requires additional research. Given the probability density function for ε, the log-likelihood function can be stated as:

$$
\ln L(\beta, \beta^{(1)}, \sigma^2_v, \mu, \rho_y, \rho_v, \rho_u) =
\ln \Phi_u(0, -\mu, \Sigma_u) + \ln \Phi_u(-\Sigma_u (\Sigma_v + \Sigma_u)^{-1} (e + \mu), -\mu (\Sigma_v^{-1} + \Sigma_u^{-1})^{-1}) + \ln \varphi_u(e, -\mu, \Sigma_v + \Sigma_u),
$$

where

$$
e = Y - \rho_y W Y - X \beta - W X \beta^{(1)},
$$

$$
\Sigma_v = \sigma^2_v (I_n - \rho_v W_v)^{-1} (I_n - \rho_v W_v)^{-1},
$$

$$
\Sigma_u = \sigma^2_u, (I_n - \rho_u W_u)^{-1} (I_n - \rho_u W_u)^{-1}.
$$

3. Review of the {spfrontier} package

Implementation of the stated maximum likelihood estimator of the SSF model parameters requires a set of functions, which are well-known in theory, but computationally hard. These functions include:

- Multivariate normal probability density and distribution functions calculation is required for the likelihood function. Note that number of dimensions matches the sample size $n$ and can be very significant. Computation of multivariate normal functions is recently intensively researched (Genz 2004) and implemented in many software packages.

- Multivariate truncated normal probability density and distribution functions calculation is straightforward on the base of multivariate normal functions.

- Moments for multivariate truncated normal random variables are required for calculation of individual technical efficiency levels.

- The likelihood function also requires extensive matrix algebra. In practice, the matrixes contain a large percent of zero values (sparse), so implementation of sparse matrix algebra algorithms is helpful.

- Maximisation of the likelihood function requires implementation of optimisation algorithms (quasi-Newton BFGS, Nelder-Mead, SANN, and others).

R (CRAN 2014) is one of popular software tools, where all of the required core algorithms are implemented. The Comprehensive R Archive Network (CRAN) contains a large number of packages, implementing particular statistical tools and algorithms.

Relying on the required functions, we chose the R environment as a base for implementation of the MLE functions. The developed software package is called `spfrontier` and available in the official CRAN archive (Pavluky 2014). The main estimator of the SSF model is implemented as a function of the same name ‘spfrontier’ and encapsulates all algorithms, required for the MLE of SSF model parameters and individual inefficiency values. The function allows estimating all four types of spatial effects, specified in the SSF model. Results of the ‘spfrontier’ function include:

- vectors of parameter estimates and their standard errors,
- a Hessian matrix of the parameter estimates,
- a vector of individual efficiency estimates,
- a vector of fitted values of the dependent variables, and
- a vector of residuals.

Spatial effects, estimated with the ‘spfrontier’ package, can be qualified as an essential feature of econometric benchmarking models in different areas of transport science.

4. Potential applications of the SSF model in transport science
Transport science includes a large number of models and methods, related with benchmarking. Measuring of efficiency of transport units, nodes, routes or decisions is widely utilised in transport planning and decision making. In this paragraph we list several application areas, where spatial effects can play an important role in efficiency measurement.

4.1. Spatial interactions between transport nodes

Efficiency of transport nodes (airports, coach terminals, ports), located near to each other, can be affected by different types of spatial effects. Spatially correlated random disturbances are the most expected type of spatial effects in transport units’ benchmarking. Spatial settings, which have a significant effect on transport node productivity and efficiency, are distributed over space unevenly. Some of these effects (population density, income) can be included into a model explicitly, but a lot of them are hardly measured and rarely included into consideration (for example, population habits, weather specifics, location features). This fact leads to spatially correlated error terms of models and, as a result, to heteroscedasticity of residuals and biased estimates of efficiency.

Endogenous spatial effects also can be a feature of these models and can be explained by interaction between neighbourhood nodes. Interactions can be both positively or negatively directed. Positive endogenous spatial effects appear when traffic from one transport node continues its flow using neighbour nodes. For example, within a popular hub and spoke organisation of airport networks, traffic, attracted by a hub airport, flows to secondary airports, increasing their productivity. Negative endogenous spatial effects are also possible and can be explained by spatial competition for traffic (passengers, cargo) in a particular geographical area between transport nodes.

4.2. Spatial effects of transport infrastructure

Another application area, intensively modelled by practitioners and academic researchers, is transport infrastructure and its effects on territory enhancements. These effects can appear on the regional level, when transport infrastructure helps regional development (in terms of population, income, manufacturing, tourism, etc.), and also on less aggregated levels. For example, there are many researches, devoted to analysis of a role of public transport infrastructure in district grow and housing pricing. Estimation of decision efficiency is very important for urban planners and regional government and spatial effects should be taken into account. Endogenous and exogenous spatial effects of transport infrastructure are obviously very significant, but rarely included into efficiency modelling explicitly.

4.3. Modelling transport emission

Modelling of environmental impact of transport is another application area, where presence of spatial effects is undisputable. Different types of transport emission and noise pollution spread over space, which leads to positive endogenous spatial effects. Frequently spatial effects play a role, which is more significant than activities, executed in a particular point or region. Thus this is necessary to include spatial effects when modelling efficiency of different emission-preventing decisions and actions.

4.4. Traffic accident prediction

Traffic accident prediction models are also a popular tool of transport planners, which likely affected by spatial effects. Accidents have endogenous spatial effects, when an occurring accident leads to changes in road environment and related accidents in neighbourhood. An accident affects behaviour of drivers, who change their usual routes, depart from the rules to avoid traffic jams, and as a result frequently leads to new accidents near to the first one. Also accidents are frequently explained by area-specific reasons like weather, road conditions, speed limits, etc., which can be included into a model in form of spatially correlated random disturbances.

There are many other areas of transport modelling, where spatial effects play an important role. It includes analysis of roads and crossroads congestion, trip generation and origin-destination flows prediction, modelling of transport demand. The ‘spfrontier’ package, encapsulating different types of spatial effects in the stochastic frontier model, can be a useful tool for decision makers in these and other areas of transport researches.

5. Conclusions
This paper is devoted to analysis of spatial effects in modelling of transport efficiency. There are four different types of spatial effects, peculiar to efficiency estimation: endogenous spatial effects (dependence between neighbour units’ outputs), exogenous spatial effects (dependence between an output of a unit and inputs of its neighbours), spatially correlated random disturbances and spatially related efficiency. A spatial modification of the stochastic frontier model (SSF model), which includes all four specified spatial effects, is formulated. We also discuss a problem of maximum likelihood estimation of the SSF model and provide a derived likelihood function. Maximum likelihood estimator is implemented within the ‘spfrontier’ package.

We suggest utilising of the SSF model and the ‘spfrontier’ package for different areas of transport modelling. A number of application areas of the transport science, related with efficiency modelling, are presented and a presence of spatial effects in these areas is substantiated. The list of discussed areas includes benchmarking of transport enterprises, modelling of transport emission, predicting effects of transport infrastructure, analysis of roads congestion, trip generation prediction, modelling of transport demand, and others.

References
6. CRAN, 2014. R software, CRAN. Available at: http://cran.r-project.org/.