TRANSPORTA UN SAKARU INSTITŪTS

DMITRY PAVLYUK

EIROPAS LIDOSTU EFEKTIVITĀTES PĒTĪJUMS, PAMATOJOTIES UZ TELPISKO STOHASTISKĀS ROBEŽAS ANALĪZI

PROMOCIJAS DARBS
Izvirzīts inženierzinātņu doktora zinātniskā grāda iegūšanai

Zinātnes nozare „Transports un satiksme”
Apakšnozare „Telemātika un loģistika”

Zinātniskais konsultants:
Dr.sc.ing., profesors
Aleksandrs Andronovs

RĪGA - 2015

UDK 519.2:656
P-34

Transporta un sakaru institūts

Pavlyuk D.

P-34 Eiropas lidostu efektivitātes pētījums, pamatojoties uz telpisko stohastiskās robežas analīzi. Promocijas darbs. Rīga: Transporta un sakaru institūts, 2015. 156 lpp.

TRANSPORT AND TELECOMMUNICATION INSTITUTE

DMITRY PAVLYUK

STUDY OF EUROPEAN AIRPORTS' EFFICIENCY ON THE BASIS OF SPATIAL STOCHASTIC FRONTIER ANALYSIS

DOCTORAL THESIS
to obtain the scientific degree Doctor of Science in Engineering

Scientific area "Transport and Communications"
Scientific subarea "Telematics and Logistics"

Scientific consultant:
Dr.sc.ing., professor
Alexander Andronov

RIGA - 2015

ИНСТИТУТ ТРАНСПОРТА И СВЯЗИ

ПАВЛЮК ДМИТРИЙ ВЯЧЕСЛАВОВИЧ

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ АЭРОПОРТОВ ЕВРОПЫ НА ОСНОВЕ ПРОСТРАНСТВЕННОГО СТОХАСТИЧЕСКОГО ГРАНИЧНОГО АНАЛИЗА

ДИССЕРТАЦИОННАЯ РАБОТА
на соискание степени доктора инженерных наук

Научная область «Транспорт и сообщение»
Научная подобласть «Телематика и логистика»

Научный консультант:
Dr.sc.ing., профессор
Андронов А.М.

РИГА - 2015

Dedicated to my wife and daughters

ANOTĀCIJA

Dmitrija Pavļuka (Dmitry Pavlyuk) promocijas darbs „Eiropas lidostu efektivitātes pētījums, pamatojoties uz telpisko stohastiskās robežas analīzi”. Zinātniskais konsultants Dr.sc.ing., profesors Aleksandrs Andronovs.

Pētījuma pamatmērķis ir efektivitātes novērtēšanas metodoloģijas izstrāde, ņemot vērā telpisko efektu ietekmi, un šīs metodoloģijas pielietošanu Eiropas lidostu efektivitātes praktiskajā analīzē.

Darbā sniegts esošo lidostu efektivitātes pētījumu apskats un aplūkoti mūsdienu efektivitātes novērtēšanas metožu pielietojumi lidostu analīzē, ņemot vērā telpisko efektu esamību.

Autors piedāvā stohastiskās ražošanas iespēju robežas telpisko modeli, kas ietver dažādus telpisko efektu tipus. Darbā sniegts piedāvātā modeļa kopskats, kā arī daži tā praktiski nozīmīgie īpaši gadījumi.

Disertācija ietver piedāvātā ražošanas iespēju robežas telpiskā modeļa autora piedāvātās koeficientu novērtēšanas metodes detalizētu aprakstu, kas balstīta uz maksimālo ticamības principu. Metode balstīta uz sadalījuma likuma iegūto modeļa salikto nejaušo locekli, kas ir daudzdimensiju noslēgtā sašķiebta normālā sadalījuma īpašs gadījums.

Novērtēšanas procedūras realizētas kā pakotne *spfrontier* programmatūrai R. Modulis pieejams publiski oficiālajā arhīvā CRAN. Izstrādāto procedūru validācija tika īstenota, pamatojoties uz statistisko eksperimentu sēriju un reālām datu kopām.

Darbā veikts praktisks telpisko efektu pētījums četrās datu kopās: Eiropas lidostu apvienotā izlase un atsevišķas Spānijas, Lielbritānijas un Grieķijas lidostu izlases. Pētījums ietver telpiskās autokorelācijas statistisko testēšanu starp lidostu privātās veiktspējas rādītājiem, kā arī telpisko efektu esamības analīzi, novērtējot piedāvātā stohastiskās ražošanas iespēju robežas telpiskā modeļa dažādas specifikācijas.

Pētījuma pamatrezultāti ir prezentēti 8 starptautiskajās un pētnieciskajās konferencēs un atspoguļoti 15 publikācijās.

Promocijas darbs ir uzrakstīts angļu valodā, sastāv no ievada, 4 nodaļām un nobeiguma, iekļauj 23 attēlus, 27 tabulas un 18 pielikumus, 156 lappuses. Izmantotās literatūras sarakstā ir 271 avoti.

ABSTRACT

The thesis of Dmitry Pavlyuk “Study of European airports’ efficiency on the basis of spatial stochastic frontier analysis”. Scientific consultant Dr.sc.ing., professor Alexander Andronov.

This research is devoted to incorporation of spatial effects into an efficiency estimation methodology and its empirical application to the European airport industry.

The thesis contains a critical review of existing airport benchmarking researches. Modern methodologies of efficiency analysis are discussed and classified, and a wide range of their applications to the airport industry are reviewed. The review is focused on revealing spatial effects in the airport industry, notably spatial heterogeneity and spatial dependence.

The spatial stochastic frontier (SSF) model, incorporating spatial effects, is proposed by the author. The SSF model is stated in a reasonably general form and a number of practically effective private cases of the SSF model are also discussed.

The thesis contains a detailed description of a derived maximum likelihood estimator for the SSF model parameters. The author obtains a distribution law of a composed error term of the SSF model as a private case of the closed multivariate skew normal distribution. A likelihood function for the SSF model’s error term is specified and a related estimator is constructed. Also formulas for estimation of individual inefficiency values are provided in the thesis.

The estimator for the SSF model parameters is implemented as a package for CRAN R software and called *spfrontier*. The package is accepted and published in the official CRAN archive. The derived estimator and the developed package are validated using designed statistical simulation studies and real-world examples.

Empirical analysis of spatial effects in four data sets of European airports is executed. The data sets consist of jointed European airports, and separately Spanish, UK, and Greek airports. The analysis consists of testing of spatial autocorrelation between airports’ partial factor productivity indicators and estimating of alternative specifications of the SSF model. Detailed conclusions on each data set are presented in the thesis.

Main results of the thesis are presented at 8 international scientific and research conferences and reflected in 15 scientific publications.

The thesis consists of introduction, 4 chapters and conclusions. It includes 156 pages, 23 figures, 27 tables in the main body, 18 appendixes and 271 publication titles in the bibliography.

АННОТАЦИЯ

Промоционная работа Павлюка Дмитрия Вячеславовича «Исследование эффективности аэропортов Европы на основе пространственного стохастического граничного анализа». Научный консультант Dr.sc.ing., профессор А.М. Андронов.

Основной целью исследования является разработка методологии оценивания эффективности с учетом влияния пространственных эффектов и применение данной методологии к практическому анализу эффективности аэропортов Европы.

В работе представлен обзор современных методов оценивания эффективности, учитывающих наличие пространственных эффектов, и существующих исследований эффективности работы аэропортов. Автором предложена пространственная стохастическая граничная модель, включающая различные типы пространственных эффектов. В работе представлена общая формализация предложенной модели, а также несколько практически важных частных случаев.

Диссертация содержит подробное описание предложенного автором метода оценивания коэффициентов пространственной модели производственной границы, основанного на принципе максимального правдоподобия. Основу метода составляет полученный закон распределения составного случайного члена модели, являющийся частным случаем многомерного замкнутого скошенного нормального распределения.

Процедуры оценивания реализованы в виде программного пакета *spfrontier* для среды R, доступного в официальном публичном архиве CRAN. Валидация разработанных процедур осуществлялась на основе серии статистических экспериментов и реальных наборов данных.

В работе проведено практическое исследование пространственных эффектов в 4 наборах данных: объединенная выборка аэропортов Европы и отдельные выборки аэропортов Испании, Великобритании и Греции. Исследование включает в себя статистическое тестирование пространственной автокорреляции между показателями частной производительности аэропортов, а также анализ наличия пространственных эффектов путем оценивания различных спецификаций предложенной пространственной модели стохастической производственной границы.

Основные результаты исследования представлены на 8 международных научно-исследовательских конференциях и отражены в 15 научных публикациях.

Диссертация состоит из введения, 4 глав и заключения. Она включает в себя 156 страниц, 23 иллюстрации, 27 таблиц в основном тексте работы, 18 приложений и 271 название публикаций в библиографии.

CONTENT

ABBREVIATIONS

AENA	*Spanish Airports and Air Navigation (Aeropuertos Españoles y Navegación Aérea)*
AMI	*Amazon Machine Image*
APM	*Air Passenger Movements*
AR	*Autoregression*
ARMA	*Autoregression and Moving Average*
ATM	*Air Transport Movements*
ATRS	*Air Transport Research Society*
BFGS	*Broyden–Fletcher–Goldfarb–Shanno*
CES	*Constant Elasticity of Substitution*
CIESIN	*Centre for International Earth Science Information Network*
CRAN	*Comprehensive R Archive Network*
CSN	*Closed Skew-Normal distribution*
DAFIF	*Digital Aeronautical Flight Information File*
DEA	*Data Envelopment Analysis*
DFA	*Distribution-Free Approach*
DGP	*Data Generating Process*
EBITDA	*Earnings Before Interest, Taxes, Depreciation and Amortization*
EEA	*European Economic Area*
EM	*Expectation-Maximization*
EU	*European Union*
FDH	*Free Disposal Hull*
FTE	*Full-Time Equivalent*
GARS	*German Aviation Research Society*
GDP	*Gross Domestic Product*
GME	*Generalised Maximum Entropy*
IID	*Independent Identically Distributed*
MA	*Moving Average*
MLE	*Maximum Likelihood Estimator*
MOM	*Method of Moments*
MVN	*Multivariate Normal distribution*
MVTN	*Multivariate Truncated Normal distribution*
NUTS	*Nomenclature of Territorial Units for Statistics*
OECD	*Organisation for Economic Co-operation and Development*

OLS	*Ordinary Least Squares*
PAX	*Passengers*
PFP	*Partial Factor Productivity*
PPS	*Production Possibility Set*
RMSD	*Root-Mean-Square Deviation*
SANN	*Simulated ANNealing*
SAR	*Spatial Autoregressive Model*
SEM	*Spatial Error Model*
SF	*Stochastic Frontier*
SFA	*Stochastic Frontier Analysis*
SSF	*Spatial Stochastic Frontier*
TFA	*Thick Frontier Approach*
TFP	*Total Factor Productivity*
TN	*Truncated Normal Distribution*
UK	*United Kingdom*
US	*United States*
VIF	*Variance Inflation Factor*
WLU	*Work Load Unit*

LIST OF ILLUSTRATIONS IN THE BODY OF THESIS

LIST OF TABLES IN THE BODY OF THESIS

INTRODUCTION

Relevance of the problem and motivation of the research

The legislative liberalisation process of the European[1] air transportation market was completed in 1997[1–3]. The growing competition in the air transport industry also concerned airport enterprises and initialised significant changes in airports' ownership and management. Airlines, operating in a competitive environment[4], gained an option to choose partner airports and therefore obtained influence possibilities. Those changes forced airports, originally considered as natural monopolies, to adapt to new, competitive market conditions. Development of high-speed rails, interregional bus transportation and, generally, transport networks also can be considered as a factor, strengthening competition between airports[5].

A competitive industry advances severe claims for enterprises' capitalisation and efficiency. Historically managed by governments, many airports were involved into a privatisation process to attract private investments and improve operational efficiency. Since 1987, when the UK government sold its seven major airports (including London Heathrow, Gatwick, and Stansted) to a private sector company, many European airports have become partly or completely private. Being under government ownership, airports' management was oriented (in an ideal case) to maximizing of social welfare at national and regional levels. After privatisation, these objectives were superseded by profit maximisation, obligatory for a commercial marketplace. Operational efficiency is one of the main profit maximisation sources, so efficiency estimation and enhancement became a subject of interest of privately managed airports[6].

Airport efficiency estimation, or benchmarking, can serve different purposes[7] and has important implications for involved stakeholders. A list of interested parties include[8]:

- airport management, which require efficiency comparison between airports to improve airport operations and enhance its standing in a competitive environment;
- airline management, interested in identifying of efficient airports for their operational activities;
- municipalities, which require efficient airports for attracting businesses and tourists into their regions; and
- policy makers, which need for benchmarking results for airport improvement programs and optimal decisions about subsidies and resource allocation.

[1] In this thesis terms "Europe" and "European" are related to 31 countries of the European Economic Area (EEA) and Switzerland

There are several well established scientific approaches to estimation of efficiency, based on indexes and frontier techniques. Nevertheless, application of these approaches to analysis of airports has its own specific complexities, frequently related with spatial effects of different types. Spatial heterogeneity and spatial dependence, two types of spatial effects, are widely acknowledged in the airport industry.

Spatial heterogeneity is based on uneven distribution of efficiency-related factors within a geographic area. These factors like climate features, economic and legislative environments, and population habits significantly affect airport productivity and must be considered during airport benchmarking.

Spatial dependence refers to interactions between neighbour airports. Mainly these interactions are explained by spatial competition for passenger and cargo traffic, served airlines, local labour forces, and other resources. Even in a legislatively competitive environment, competition between airports is limited by their geographical locations and thus obviously has a spatial nature. The problem is aggravated by an irregular pattern of airports' spatial dependence. Although number of European airports is increasing during last decades[9], there are geographical areas in Europe where a competition pressure is weak or absent completely. Frequently authorities try to compensate this lack of competition pressure by different forms of regulation[10], which also complicates airport benchmarking.

We expect that spatial effects, which affect activity of European airports, will strengthen in the nearest future. Currently there is a lack of theoretical and empirical studies of airport efficiency, where spatial effects are incorporated into a methodology. Methods of recently developed spatial econometrics[11] can be used for enhancing airport benchmarking procedures.

The degree of the theme study

This research is devoted to incorporation of spatial effects into an efficiency estimation methodology and its empirical applications to the European airport industry. We consider the current level of development both methodological and application areas and a place of this research on their junction. A corresponding diagram is presented on the Fig. 1.

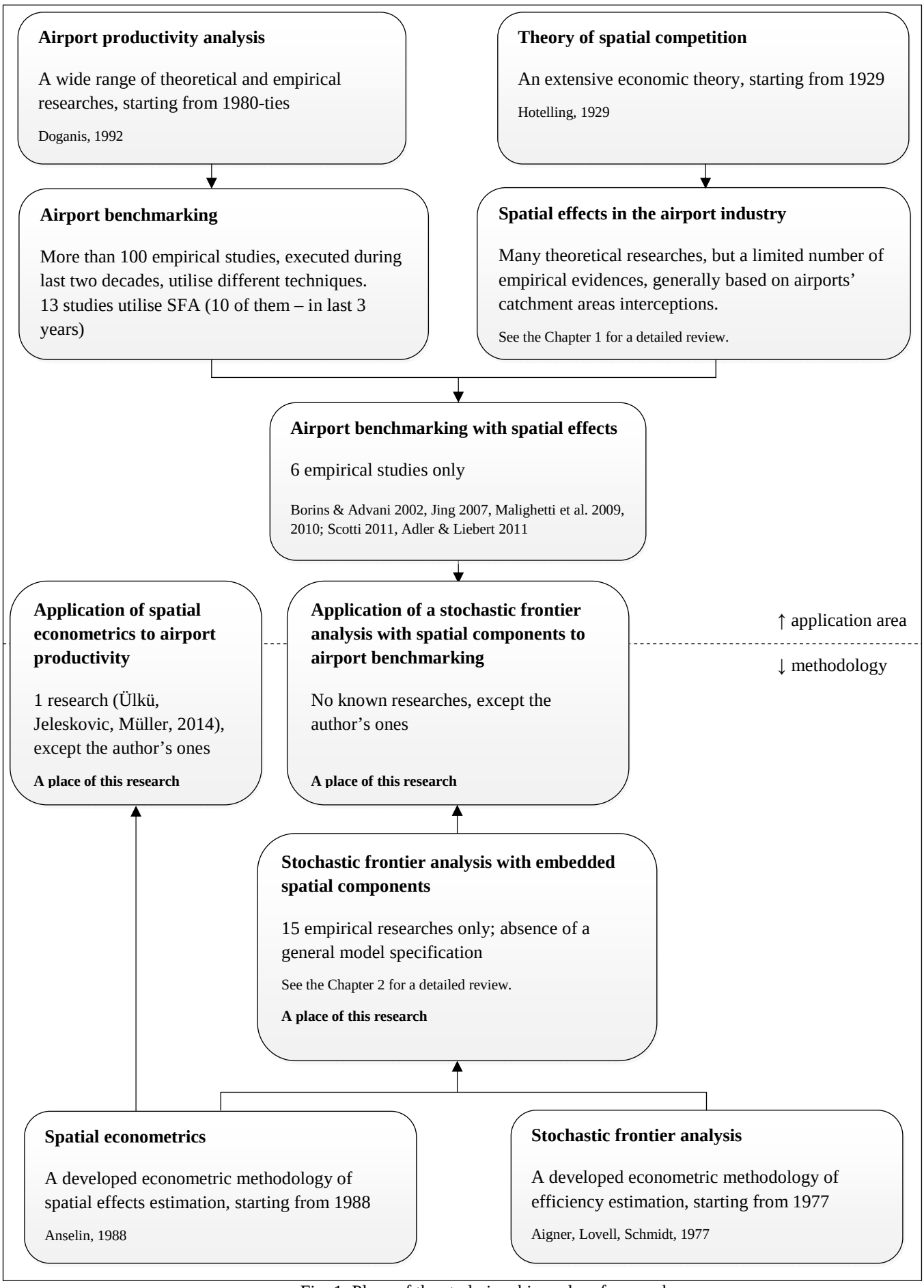


Fig. 1. Place of the study in a hierarchy of research areas

Since 1980-ties great efforts have been made in developing of performance measurements in the airport industry[12]. The growing demand for studies in this area has been stimulated by industry deregulation and led to a wide range of recently executed theoretical and empirical researches. More than a hundred of research papers, devoted to airport benchmarking, are published during last two decades. A considerable contribution was made by Graham[13–16], Gillen and Lall[17], Barros[18–22], Gitto and Mancuso[23], [24], and Liebert[25–27], among others. An extensive review of related researches is provided by Liebert and Niemeier[25]. A number of valuable reports in this area are also published: the Global Airport Performance Benchmarking Reports 2003-2011 produced by Air Transport Research Society (ATRS)[28], the Airport Performance Indicators and Review of Airport Charges reports by Jacobs Consulting, the Airport Service Quality programme by Airports Council International. Some local authorities, which control the airport sector, also provide their own benchmarking reports, e.g. Avinor AS (Norway)[29], Civil Aviation Authority (UK)[30].

Different methodologies are utilised in literature for airport performance measurement: partial factor productivity (PFP) indicators, data envelopment analysis (DEA), stochastic frontier analysis (SFA). SFA[31], a methodological base of this research, is a popular frontier-based econometric approach to efficiency estimation. It was originally presented by Aigner, Lovell, Schmidt[32], and Meeusen and van den Broeck[33] in 1977. The main advantage of SFA is a statistical approach both to frontier and unit efficiency estimation, which makes confidence intervals, significance, hypothesis testing, and other statistical procedures easily available. A list of researches, utilised the SFA approach, includes works of Pels[34], Abrate and Erbetta[35], Jing[36], Barros[19], [20], [37], Martin and Voltes[38], [39], Muller[40], Malighetti, Martini, and Scotti[41], [42].

Despite a large number of airport benchmarking studies, spatial effects are rarely included into consideration. Researches, conducted by Borins and Advani[43], Jing[36], Malighetti and Scotti[8], [41], [44], and Adler and Liebert[45], can be mentioned among a few others. Technically spatial effects can be embedded into models in different ways. Spatial heterogeneity is usually modelled using observable variables like average annual temperature or acting government subsidies. Spatial dependence between airports, in turn, is modelled by interception of airport catchment areas or airport management's subjective perception of competition. At the same time, modern methods of spatial econometrics are rarely utilised.

Spatial econometrics[11] is a set of techniques for analysis of spatial relationships. This approach deals with spatial dependence and spatial heterogeneity in regression models and is widely used in practice[46], [47]. Nevertheless, to the best of our knowledge, the only application

of these methods (except the author's ones) to analysis of airport productivity and efficiency is Ulku et al.[48].

Incorporating of spatial econometrics' principles into the stochastic frontier analysis in other application areas is also weakly covered in literature. A complete list of related studies, known to us, includes papers of Druska and Horrace[49], Fahr and Sunde[50], Barrios[51], Schettini et al.[52], Affuso[53], Lin et al.[54], [55], Areal et al.[56], Tonini and Pede[57], Mastromarco et al.[58], Glass et al.[59], and Fusco and Vidoli[60]. All presented studies consider one particular type of possible spatial effects, so a general model specification seems to be necessary.

Statement of the problem

On the base of literature analysis, we postulate the following research problems:

1. Spatial effects play an increasing role in the airport industry, but currently they are rarely included into procedures of airport benchmarking.
2. The methodology of efficiency estimation in presence of spatial effects is weakly developed.

The object and subject of the research

The object of the research is a system of European airports.

The subject of the research is statistical benchmarking of European airports subject to presence of spatial effects.

The goal and tasks of the research

The main goal of the research is to develop a methodology of statistical efficiency estimation in presence of spatial effects and apply this methodology to analysis of the European airport industry. To achieve the goal, the following principal tasks were stated and solved:

1. reviewing of existing statistical methodologies of efficiency modelling and their applications to the airport industry, paying special attention to analysis of spatial relationships;
2. proposing a new statistical model for estimation of efficiency, which explicitly includes different types of spatial effects;
3. deriving an estimator of the proposed model's parameters;
4. developing a software tool, which implements the derived estimator and related procedures;
5. testing of statistical properties of the derived estimator, using a set of statistical simulation studies;
6. testing of the proposed model on real-world data sets; and

7. analysing of the European airport industry, using existing methods of spatial statistics and the proposed statistical model.

The methodology and methods of the research

The methodological foundation of this research mainly consists of probability theory, mathematical statistics, and econometrics methods. In particular, we applied principles and techniques of spatial econometrics and stochastic frontier analysis for formulation and estimation of the new statistical model. We also applied methods of statistical simulation studies for validation and analysis of statistical properties of the developed estimator.

We used the CRAN R environment for statistical computing and as a base for the developed package. Also the study's software toolbox includes MySQL database management system as data storage; developed Java procedures for data collecting and pre-handling.

The theses, which are submitted for defence

- The proposed spatial stochastic frontier (SSF) model with explicit incorporation of spatial effects. Four types of spatial effects are incorporated into the model: endogenous spatial effects (spatial dependence), exogenous spatial effects, spatially correlated random disturbances (spatial heterogeneity), and spatially related efficiency.
- The derived estimator for the proposed SSF model. The estimator is based on maximum likelihood principles and allows estimating the SSF model parameters and unit-specific inefficiency levels. The derived estimator procedure and statistical properties of resulting estimates are tested using a developed simulation study and real-world data sets.
- The developed software package *spfrontier*, implementing the derived estimator and a set of related procedures. The package is implemented as a module for the R environment and accepted and publicly available in the official CRAN archive[61]. The package includes functions for: estimation of the SSF model parameters; estimation of unit-specific inefficiency values; numerical calculation of the estimates' Hessian matrix; testing of significance of parameter estimates; and designed simulation studies for analysis of estimates' statistical properties. Also the package is accompanied with all real-world data sets on the European airport industry, used in this research.
- The results of application of spatial statistics, including the developed SSF model, to the European airport industry. Four data sets were separately investigated: Spanish airports, UK airports, Greek airports, and joined European airports. The main goal of this empirical analysis is revealing spatial effects (or their absence) in the data sets, related to different spatial and economic settings. Using the developed SSF model, significant spatial effects are discovered and their analysis is executed and presented in this thesis.

The work approbation

The main results of the research were presented at the following conferences:

1. International Conference “Reliability and Statistics in Transportation and Communication”, Riga, Latvia, 2009, 2010, 2013, 2014.
2. III International Scientific Conference “Spatial Strategy for sustainable development”, Kuldiga, Latvia, 2011.
3. III International Youth Scientific Conference “Mathematical modelling in economics and risk management”, Saratov, Russia, 2014.
4. International Scientific Conference “Knowledge, Education and Change Management in Business and Culture”, Riga, Latvia, 2013.
5. VI International Conference „Regional Development” Entitled „Strategic Management of the Region’s Development – Perspective 2014-2020. Recommendations for Poland and Central-Eastern Europe”, Torun, Poland, 2013.

The results of the research were published in the following proceedings and journals:

1. Pavlyuk, D. (2014). Modelling of Spatial Effects in Transport Efficiency: the ‘Spfrontier’ Module of ‘R’ Software, in Proceedings of the 14th International Conference “RELIABILITY and STATISTICS in TRANSPORTATION and COMMUNICATION” (RelStat’14), Riga, Latvia, pp. 329–334.
2. Pavlyuk, D. (2014). Spatial Aspects of European Airports’ Partial Factor Productivity, Transport and Telecommunication, Vol. 15, No 1, pp. 20–26.
3. Pavlyuk, D., Gode, N. (2014). Spatial Aspects of International Migration in European Countries, in "Problems of Economic Policy of the Central and Eastern Europe Countries: Macroeconomic and Regional Aspects, A. Ignasiak-Szulc and W. Kosiedowski, Eds. Torun, Poland: Wydawnistvo Naukowe, pp. 73–92.
4. Pavlyuk, D. (2013). Distinguishing Between Spatial Heterogeneity and Inefficiency: Spatial Stochastic Frontier Analysis of European Airports, Transport and Telecommunication, Vol. 14, No 1, pp.29-38.
5. Pavlyuk, D. (2012). Airport Benchmarking and Spatial Competition: A Critical Review, Transport and Telecommunication, Vol. 13, No 2, pp. 123–137.
6. Pavlyuk, D. (2012). Maximum Likelihood Estimator for Spatial Stochastic Frontier Models, in Proceedings of the 12th International Conference “Reliability and Statistics in Transportation and Communication” (RelStat’12), Riga, Latvia, pp. 11–19.
7. Pavlyuk, D. (2011). Application of the Spatial Stochastic Frontier Model for analysis of a regional tourism sector, Transport and Telecommunication, Vol. 12, No 2, pp. 28–38.

8. Pavlyuk, D. (2011). Spatial Analysis of Regional Employment Rates in Latvia, Scientific proceedings of Riga Technical University. Ser. 14. Sustainable spatial development, Vol. 2, pp.56-62.
9. Pavlyuk, D. (2011). Efficiency of Broadband Internet Adoption In European Union Member States, in Proceedings of the 11th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'11), Riga, Latvia, pp. 19–27.
10. Pavlyuk, D. (2010). Multi-tier Spatial Stochastic Frontier Model for Competition and Cooperation of European Airports, Transport and Telecommunication, Vol. 11, No 3, pp. 57–66.
11. Pavlyuk, D. (2010). Spatial Competition and Cooperation Effects on European Airports' Efficiency, in Proceedings of the 10th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'10), Riga, Latvia, pp. 123–130.
12. Pavlyuk, D. (2010). Regional Tourism Competition in the Baltic States: a Spatial Stochastic Frontier Approach, in Proceedings of the 10th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'10), Riga, Latvia, pp. 183–191.
13. Pavlyuk, D. (2009). Spatial Competition Pressure as a Factor of European Airports' Efficiency, Transport and Telecommunication, Vol. 10, No 4, pp. 8–17.
14. Pavlyuk, D. (2009). Statistical Analysis of the Relationship between Public Transport Accessibility and Flat Prices in Riga, Transp. Telecommun., Vol. 10, No 2, pp. 26–32.
15. Pavlyuk, D. (2008). Efficiency Analysis of European Countries Railways, in Proceedings of the 8th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'08), Riga, Latvia, pp. 229–236.

The structure of the thesis

The thesis consists of the introduction, four chapters, conclusions, and 18 appendices. It contains 156 pages, 23 figures, and 27 tables. The list of references and information sources contains 271 titles.

In ***the introduction***, the relevance and motivation of the research are explained, the goal and the tasks of research are formulated, the object and subject of the research are stated, and the scientific novelty and practical value of the obtained results are presented.

The first chapter contains a critical review of existing researches of airport efficiency. Present methodologies of efficiency analysis are discussed and classified, and their applications to the airport industry are reviewed. A special attention is paid to different approaches to revealing spatial effects (spatial heterogeneity and spatial dependence). A theoretical background

of spatial interactions between airports is reviewed and existing empirical evidences of presence of spatial effects in the European airport industry are presented.

The second chapter contains an overview of basic concepts of the production theory and stochastic frontier analysis as a comprehensive tool for efficiency modelling. Mathematical formalisation is stated for a task of estimation of production possibility frontier parameters and technical efficiency. Single- and multi-output production processes and possible approaches to their econometric modelling are discussed. A special attention in the chapter is paid to known approaches to integration of spatial relationships into the stochastic frontier model.

The third chapter contains a detailed description of the SSF model, proposed by the author. Different types of spatial effects are discussed and reasoning for these spatial effects as phenomena in different branches of knowledge is presented. The SSF model specification, which explicitly includes all four types of spatial effects, is proposed. The chapter also contains a formal derivation of a maximum likelihood estimator for the SSF model parameters, including a procedure for estimation of unit-specific efficiency values. Results of the simulation study, designed for analysis of the estimator's statistical properties, are also presented. Finally, the chapter contains a description of the developed *spfrontier* package, which implements all derived methods and procedures.

The forth chapter is devoted to empirical analysis of spatial effects in four different European airports' data sets. The analysis consists of testing of spatial autocorrelation between selected PFP indicators of airports and estimating of spatial effects using alternative specifications of the SSF model. The research data sets consist of jointed European airports (359 airports, 2008-2012), Spanish airports (38 airports, 2009-2010), UK airports (48 airports), and Greek airports (42 airports, 2007). The chapter contains a description of each data set and detailed results of the conducted analysis. Separate conclusions for the data sets are presented at the end of corresponding paragraphs.

Conclusions contain summary of the executed work, description of most significant results obtained, and directions for future researches.

1. AIRPORT BENCHMARKING METHODOLOGIES AND THEIR EMPIRICAL APPLICATIONS IN SPATIAL SETTINGS

1.1. Review of airport benchmarking methodologies

1.1.1. Airport efficiency estimation

A classical definition refers economic efficiency[62] as usage of available resources (inputs) to maximise the production of goods and services (outputs). The first and one the most critical steps of efficiency estimation is definition of resources (inputs) and results (outputs) of airports. This definition is empirically complicated due to a very heterogeneous nature of the airport business, widely acknowledged in classic literature[12]. There are two most popular approaches to the airport business, which lead to different definition of airport inputs and outputs:

- airport as a commercial organisation;
- airport as an intermediary between airlines and passengers or freight being transported by air.

Analysing an airport as a commercial organisation, results of its activity can be defined in terms of economics as total revenue or profit. This definition is quite convenient in order to assess economic performance of an airport, but on a closer examination it also raises specific issues. Nowadays, activity of airport enterprises is not limited with aeronautical services, but includes parking, retailing, food and beverages, passenger access, and other services. Currently these non-aeronautical services, originally considered as complementary, play an important role in the airport business[63]. According to the ATRS reports[28], a share of non-aeronautical revenues is increasing during the last decade and for some European airports exceeds 50% (for example, for German busiest Munich and Frankfurt airports). Thereby results of various non-aeronautical activities become an important component of overall airport performance. Also nowadays majority of airports outsource some of their services to third-party organisation, so airports' total revenues become not comparable. Legal and regulatory differences between countries and regions also reinforce this problem significantly. It should be noted that comparability of research units is a critical requirement of all frontier-based approaches to estimation of efficiency, considered in this chapter.

Taking an airport as an intermediary leads to another definition of its outputs. From airlines side, the main goal of airport activity is handling their aircrafts, so the output can be specified as a number of air transport movements (ATM). From population side, an airport serves passengers and cargo, so a number of passengers served (PAX or air passenger movements, APM) and a

volume of loaded/unloaded cargo are appropriate metrics of an airport output. Frequently, cargo and passenger are grouped into work load units (WLU), which usually equals to 1 passenger or 100 kg of cargo, for simpler comparison of airport productivity. The outputs heterogeneity is also a problem in this approach. Costs and revenues vary for different types of served passengers. For example, international passengers need more space (for customs, visa checks, etc.), but at the same time spend more time in terminals and provide more revenue. Also serving of transit passengers is a quite specific operation for costs and spending. Serving cargo is also quite different as a result of specific transportation requirements and loading features.

Recently some researches also included negative outputs into airport benchmarking. These undesired outputs can have different forms like environmental emission and noise[64] or passenger delays[65].

Definition of airport resources is more classical, but also has its own specifics. Classical economics recognises three categories of resources: labour, capital, and land. Labour is usually represented as a number of employees, or in form of full-time equivalent (FTE) to make the resources comparable. Again, heterogeneity of labour resources and outsourcing make these metrics less confident, so frequently total employment costs are used instead of physical characteristics. Capital includes infrastructure objects like runways, terminals, gates, check-in desks, aircraft stands, baggage belts, vehicle parking spaces, and others. Usually infrastructure objects are measured in physical units (number, area, length, etc.) and used separately, but in some cases they are grouped into financial indicators like amortisation or capital stock. Fuel, maintenance, insurance, and other operating resources are also usually used in a financial form, called operating or soft costs. An airport location (distance to nearest cities, population in the catchment area, connections with other transport nodes) can be classified as a land resource. Aircraft noise and air pollution, usually considered as outputs, also can be reckoned to the same group of input resources.

Thus we conclude that inputs and outputs of an airport's operations are very heterogeneous, and researches usually use their own discretion for benchmarking. We summarised inputs and outputs of models, used in 96 applied studies, in the Table 1.1 (a full list of researches with inputs and outputs used can be found in the Appendix 1).

Almost all researchers use APM and ATM as an airport's outputs (75 and 74 of 96 studies respectively); majority of studies also takes loaded/unloaded cargo into account (56 studies). If financial indicators are included into the model, then revenues are usually classified to aeronautical and non-aeronautical. Other output characteristics are rarely used.

Table 1.1. Summary of inputs and outputs used in existing studies

Inputs		*Outputs*	
Indicator	Number of researches (of 96)	Indicator	Number of researches (of 96)
Employment (FTE)	48	APM	75
Terminals (area)	45	ATM	74
Operating costs	36	Cargo	56
Runways (number)	32	Non-aeronautical revenue	20
Runways (length)	17	Aeronautical revenue	19
Baggage belts (number)	16	WLU	5
Check-in desks (number)	16	Total revenue	5
Employment (costs)	15	Delays (time)	2
Aircraft stands (number)	13	Aggregated Pollution	1
Runways (area)	13	Noise Pollution	1
Gates (number)	13		
Capital (stock)	10		
Aircraft stands (area)	10		
Airport (area)	10		
Capital (costs)	9		
Gates (number)	8		
Car parking (places)	7		
Runways (capacity)	7		
Capital (investments)	5		
ATM	5		
Terminals (number)	2		
Distance to city centres	1		
Minimum connecting time	1		
Population	1		
Potential passengers (number)	1		
Opening hours	1		

A list of used resources is more diversified. More than a half of studies include labour resources in form of full-time employees. Used infrastructure resources (runways, terminal area, etc.) vary in researches, but we need to note that these indicators can be correlated, which make it unnecessary to include all of them into a model. Another popular input model component, used in 36 studies of 96, is operating costs. Surprisingly, location resources of airports are rarely included into consideration.

It should be noted that a problem of data availability becomes a significant obstacle for researches. Many European airports don’t provide disaggregated statistics on their operations, especially on financial indicators. Where statistics are available, indicators are frequently not consistent due to different accounting and classification methodologies, used in different countries.

The problem of data availability plays even more important role for spatial models, considered in this research. Many classical methods can be applied to a selected set of airports (for example, busiest ones). Spatial models require data about all neighbour airports in a research area, because only in this case identification of spatial effects becomes possible. Estimation of

spatial econometric models using the data set with missed data is very under-researched area, so a complete data set becomes critically important.

Summarising, we can note a great variety of approaches to the airport outputs and inputs definitions, and, as a result, different ways of airport benchmarking. Recently several voluminous reviews of empirical studies, related with airport efficiency estimation, were published[25], [30].

Theory of efficiency estimation provides a wide range of estimation methods with their own advantages and limitations. Scientific airport benchmarking approaches start from relatively simple linear indexes, but further include more complicated frontier-based models[66].

1.1.2. Partial factor productivity indicators

The simplest one-dimensional way of efficiency estimation is a direct ratio of a chosen airport output to a given resource used. Indicators, constructed on the base of this strategy, are called PFP indexes. Due to a great diverse of airport outputs and inputs, a range of PFP indexes is very wide. PFP indexes are not related to overall efficiency, but reflect a particular aspect of airport activity[67]:

- Labour productivity indexes: APM per employee, ATM per employee, WLU per employee.
- Infrastructure productivity indexes: APM per terminal, WLU per airport's surface square meter, ATM per runway.
- Financial performance indexes: operational costs per WLU, overall/aeronautical revenue per WLU, overall revenue to expenses ratio.
- PFP indexes for undesired outputs: delay minutes per ATM, green gas emission per ATM, etc.

PFP indexes are widely used by airport management, because their simplicity and straightforward meaning. Also calculation of PFP indexes is technically simple, and each index separately doesn't require full set of data. A PFP index provides valuable information about a particular area of interest, but by definition cannot provide a full picture of airport performance. PFP indexes don't consider differences in input/output prices and other operating environment conditions; leave factor substitution out of account[30], and so can be considered just as a good complementary research tool.

Stated weaknesses of PFP indexes led to development of methodologies, which allow calculating overall efficiency values. All methodologies can be classified on the base of their principle (averaging or comparing with frontier values) and presence of a random component (deterministic or stochastic approaches). A classification of widely used methodologies is presented in the Table 1.2.

Table 1.2. Classification of efficiency and productivity estimation methodologies

	Deterministic	*Stochastic*
Averaging	Total productivity factor (TFP)	Classical regression models
Frontier	Data envelopment analysis (DEA) Free disposal hull (FDH)	Stochastic Frontier analysis (SFA) Distribution-free approach (DFA) Thick frontier approach (TFA)

Source: own classification, based on Liebert and Niemeier [25], and Hirschhausen and Culman [66]

Methodologies, based on averaging of values, consider a relationship between weighted airport outputs and inputs. Total factor productivity(TFP) indexes use prices to weight input/output values, when regression estimates these 'weights' by minimizing a sum of squared residuals. Averaging methodologies assume that all airports in a sample operate efficiently, so the only source of deviation from the average result is a random noise. This obviously doesn't match a real situation, when a difference between outputs of two airports with similar resources can be explained not only by a random component, but also by technical or managerial efficiency. Frontier-based methodologies (like data envelopment analysis and stochastic frontier analysis) allow presence of inefficiency components by construction.

1.1.3. Parametric approaches to airport benchmarking

TFP indexes are ratios of weighted outputs to weighted inputs, where market prices are used as weights. Two most frequently used TFP indexes are Tornqvist index[68] and Caves, Christensen and Diewert index[69], which can be considered as flexible forms of classical Laspeyres or Paasche indices.

Market prices, required for calculation of TFP indexes, are rarely available and valid, which can be a reason of a limited number of TFP applications to the airport industry. The most frequently cited researches, based on TFP, are the ATRS Global Airport Performance Benchmarking Reports[28] and related analytical studies[6]. The authors constructed a variable factor productivity index and used it for productivity comparison of airports around the world. Nyshadham and Rao[70] applied TFP indexes to estimation of European airports' efficiency and compared obtained results with partial indexes. Gitto [24] applied TFP indexes as one of the tools for analysis of Italian airports efficiency. As it was described earlier, TFP indexes don't directly take airport inefficiency into account.

In 1978 Charnes, Cooper, and Rhodes[71] proposed DEA approach to estimate overall company efficiency. DEA is a frontier-based approach, based on linear programming techniques, which allows directly calculate airport inefficiency components. DEA constructs an efficiency frontier without market price values and without assumptions about a functional form of the

frontier, which makes it an easy-to-use and powerful efficiency estimation tool. A complementary Malmquist index[69], defined using distance functions for a multi-input, multi-output technology, is frequently used to analyse airport efficiency changes over time.

The DEA estimator is deterministic by construction, and this fact prevents usage of popular statistical techniques like confidence intervals and hypothesis testing and makes the DEA frontier sensitive to data problems. Moreover, the DEA estimator is biased upward[72] and inconsistent for non-convex frontiers. Simar and Wilson[72] suggested bootstrapping procedures to solve these problems and improve statistical properties of DEA estimates.

A practically important research area, which is lying outside the basic DEA model, is examination of factors, which influence airport efficiency values (like airport ownership, hub status, etc.). A typical two-stage approach, which deals with these factors, includes calculation of DEA efficiency values and their further regression on explanatory factors. DEA efficiency values are obviously limited to the [0, 1] closed interval, so regressions with a censored dependent variable are used. Simar and Wilson[73] discussed properties of two most frequently used regression models – Tobit and truncated, and suggested an alternative double bootstrapping procedure.

DEA is the most frequently used academic approach to airports benchmarking. More than a hundred scientific researches, oriented on different practical and theoretical aspects of the DEA model, were published during last two decades. Comprehensive literature reviews on this subject can be found in [19], [25], [74]; further in this paragraph we just present several DEA-based researches, published in last years.

Gillen and Lall[17] published an analysis of US airports, based on the two-stage DEA approach with a second stage Tobit regression with environmental, structural and managerial variables. This research can be considered as a pioneering one and a base for many modern DEA-based airport benchmarking researches. Another frequently cited DEA application is Sarkis' US airports performance analysis[75].

Recently published studies include several country-specific DEA application for Spanish[76], [77], Greek[78], Malaysian[79], and Latin American[80] airports. Barros et al. applied Gillen-Lall's approach to analyse airports in United States[81], Argentine[82], United Kingdom[19], [21], Italy and Portugal jointly [83], and Canada[84].

To the best of our knowledge, the most researched European countries in this aspect are Germany and Italy. German Aviation Research Society (GARS) published a set of researches ([67], [85], [86]), where the Malmqvist-DEA approach was applied to a sample of German airports. Adler and Liebert[27] complemented DEA efficiency values with second stage OLS, Tobit, and truncated regressions on ownership, regulation, and management characteristics. Ulku,

Muller, et al.[40], [74] analysed German airports applying Simar-Wilson's double bootstrapping procedure (among other research approaches).

Gitto and Mancuso published some articles[24], [87–89] with application of Simar-Wilson's double bootstrapping procedure to Italian airports. Other recent DEA applications to Italian airports performance are presented by Barros and Dieke[90] and Malighetti et al.[91].

European airports' efficiency was analysed by the University of Bergamo researchers[41], [92]. A special attention was devoted to competitive characteristics of the European airport network, which were included as a factor, influencing airport efficiency in Simar-Wilson's model. Also the DEA approach was applied to European airports by Pels et al.[34], [93].

DEA is not the only deterministic approach to efficiency estimation. The free disposal hull (FDH) method [94] is a popular extension of DEA, which relaxes DEA's assumption about a convex form of the frontier.

FDH has few applications to the airport industry. Holvad and Graham[14] applied FDH approach to analysis of European and Australian airports and discovered difference between DEA and FDH efficiency estimates for European airports.

However, since DEA and FDH are non-statistical, any deviation from the frontier is considered as inefficiency, making DEA estimates non-robust and exacting to data quality. Statistical models with a random component in specification solve this issue and allow applying standard powerful statistical techniques. Therefore statistical models (both averaging and frontier) became a more popular airport benchmarking tool during the last decade.

1.1.4. Stochastic approaches to airport benchmarking

The most popular statistical model is a classical regression, which estimates a relationship between an expected value of a dependent variable (usually output) and a set of explanatory variables (inputs). The classical regression requires a predefined functional form of this dependency. Cobb-Douglass function with a constant substitution elasticity and more flexible Translog are the two most frequently used functional forms in airport industry studies. The classical regression is based on averaging technique, so doesn't contain efficiency as a component of a model specification. In relation to airports, the classical regression represents a model of airport productivity, but not efficiency.

A pioneering airport regression analysis studies executed by Keeler[95] and Doganis and Thompson[96]. Keeler estimated the Cobb-Douglass regression between operating costs and ATM on the base of pooled panel data of US airports. Doganis and Thompson constructed Cobb-Douglass regression using WLU as an output and estimated its parameters for British airport cross-sectional data.

Later several similar studies with enhanced model specification (Translog) and estimation techniques (panel data econometrics) were published. Good literature reviews on this subject can be found in [38] and [97].

A statistical approach to frontier construction and efficiency estimation brought to development of a set of models: stochastic frontier model, thick frontier model, and distribution-free model are frequently used ones. Stochastic frontier analysis (SFA), one of the most popular approach, was presented by Aigner, Lovell, Schmidt[32], and Meeusen and van den Broeck[33] in 1977. This approach, rarely used for airports efficiency analysis before, recently became quite popular. The main strength of SFA is a statistical method both of frontier and unit efficiency estimation, which makes standard statistical tools easily available. These advantages require mandatory specification of a frontier functional form and a law of efficiency distribution. Selection of a frontier form is usually made from Cobb-Douglass and Translog functions, and rarely includes more flexible, but data-consuming forms like Fourier-Flexible. Half-normal and truncated normal distribution laws are the most frequently used options for the efficiency component. The latter (truncated) distribution allows direct inclusion of factors influencing airports efficiency into a model, and simultaneous estimation of all model parameters. In 2005 Greene[98] extended the SFA model with a cross-firm heterogeneity, which is considered as one of the most important problems in airport benchmarking. Estimation of Greene's models (called true fixed and random effects models) requires panel data, which are currently available for airport applications.

The first (to the best of our knowledge) SFA application to airport benchmarking was presented by Pels et al.[34], [99]. They applied the homogeneous Cobb-Douglass frontier model to a sample of European airports and made comparison of estimation results with DEA-based estimates. Later Oum et al.[100] applied the Translog stochastic frontier model to estimate influence of airports' ownership on its efficiency.

During last five years number of studies significantly increased. Barros et al. presented a set of heterogeneous SFA applications to European[20], Japanese[37], and UK[19] airports. Voltes[38] analysed European, American, Oceanian, and Asia-Pacific samples of airports, and later Spanish airports separately[39]. Muller, Ulku, and Zivanovic[40], within the bounds of GAP project, executed a comparison of British and German airports' performance, utilising different techniques (PFP, DEA, and SFA). The author of this thesis[101], [102] analysed efficiency of European airports using the SF model and taking spatial competition among airports into consideration. Scotti applied the homogeneous SFA model for Italian airports in his doctoral dissertation[8] and related articles[44]. Summing up SFA model applications, we can note a growing academic interest to usage of this approach to airports efficiency estimation and a lack

of studies with a heterogeneous frontier, which supposed to be a right choice for variegated environment of the airport industry.

Two other stochastic frontier methods, which are mentioned in the Table 1.2, are distribution-free and thick frontier approaches. Both methods remove restrictions of SFA related with the mandatory specification of the frontier functional form and inefficiency distribution law and make estimation more flexible, but exacting to a volume of data. These strong requirements to a data volume can be considered as one of the main reasons why there are no empirical applications of these methods to airport efficiency analysis.

Summarising this paragraph, we note that a very complicated nature of the airport benchmarking problem. The problem becomes even more complicated due to diverse nature of the airport business, allowing different approaches to definition of resources and outputs. Despite the complexity of the problem (or maybe thanks to this fact), airport benchmarking attracted a significant attention of world-wide scientific community.

1.2. Review of spatial heterogeneity in the airport industry

1.2.1. Airport heterogeneity problem

The majority of airport benchmarking methodologies are based on comparison between airports in a sample. For example, methodologies like SFA and DEA construct a surface of best performers (airports, obtained optimal results), called a frontier, and estimate an airport's efficiency by comparing its outputs with the frontier. Calculation of PFP indexes, in turn, doesn't require direct matching of airports, but these indexes are frequently used for comparison in further analysis. Effective utilisation of these approaches requires general compatibility that is homogeneity of airports. In practice, airports are highly heterogeneous.

There is an extensive background for airports heterogeneity. It can be related with airport size (large or small airports), traffic specialisation (passengers or cargo, international or cargo), ownership (public or private), social particularities, government regulations, and others.

Factors of airport heterogeneity are commonly arranged to endogenous, or controlled by airport management, and exogenous, lying beyond managerial control[26]. Endogenous heterogeneity in practice is frequently noticed as inefficiency, when exogenous is stated as a benchmarking difficulty. Discussing exogenous heterogeneity, Forsyth and Niemeier state that "a central problem of benchmarking is the heterogeneity of airports, which must be taken account" [103]. The importance of heterogeneity in airport benchmarking is widely acknowledged in literature[100], [104], [105].

For purposes of modelling, airport heterogeneity (both endogenous and exogenous) is classified to observed and unobserved. Observed heterogeneity can be represented in a model

using a set of measurable and practically available factors. For example, ownership of airports is publicly available and can be included into a model as a set of dummy variables for airports' primary owners or a set of ownership shares for more complicated ownership structures. Observed climate heterogeneity can be represented as an average temperature, average annual precipitation, annual number of days with snow cover, etc. Acting heterogeneity, which cannot be directly represented by a set of indicators, is classified as unobserved. Barros et al.[37] and Liebert[26] note the importance of unobserved heterogeneity for airport benchmarking.

1.2.2. Spatial heterogeneity of airports and its sources

In this research we focus on factors of spatial heterogeneity, related with airports' geographical positions. Spatial heterogeneity is based on uneven distribution of efficiency-related factors within a geographic area. These factors, like climate features, economic and legislative environments, and population habits, can significantly affect airport productivity and must be considered in airport benchmarking. Spatial heterogeneity can be partly represented in models by observed factors, but also latent accounting of unobserved factors is technically possible. The main premise, which allows indirect including of airport heterogeneity into a model specification, is a similarity of unobserved spatial factors' effects for neighbour airports.

There is a wide range of spatial heterogeneity sources, summarised in the list below:

1. Natural sources: spatial heterogeneity of natural conditions
    a. *Climate* exerts influence on activity of neighbour airports. Necessity of snow removal from runways and aircraft anti-icing procedures significantly change airport operations; thunderstorms and strong winds trouble airports' activity and break schedules; high temperatures leads to low air density and additional requirement for airplanes.
    b. Complicated *landscape* also significantly limits airport activity. Mountains limit aircraft landing trajectories; high altitude creates additional landing problems; mountainous area leads to higher risk of weather changes, desert airport suffered from sand storms, and so on.
2. Origin sources: spatial heterogeneity of traffic origins
    a. *Population* of airport's catchment area is the main source of outgoing traffic flows, and population density has obvious spatial patterns (see Fig. 1.1 for distribution of population density over the Europe).
    b. *Economic and social conditions* also play an important role in traffic generation. Although economic convergence in the EU is stated as a strategic development direction, a level of regional disparities is still high. Population welfare becomes

even more important in view of a growing role of non-aeronautical services in airports' income structure.

c. *Labour market*. Neighbour airports act on the same labour market and utilise local labour resources in similar ways. This factor is related with different levels of salaries, qualification and availability of labour forces.

d. *Population habits* are another factor, influencing outgoing traffic flows. Local peculiarities (like mobility, travelling directions, etc.) are still present in the European countries and can affect nearby airports' performance.

3. Destination sources: spatial heterogeneity of traffic flow attractors

   a. *Touristic places*, located near to an airport, obviously attract incoming traffic flows. Distribution of touristic attractors (seashores, health resorts, heritage objects, etc.) over the space is not even, which leads to spatial heterogeneity.

   b. *Logistic centres, ports*, and other objects of a cargo distribution network can positively affect traffic flows of all airports in the surrounding area.

   c. Similar to cargo distribution centres, a *transport infrastructure* (railway and road density, secondary airports and sea ports) is also a factor of incoming airports' traffic. A level of transport infrastructure development also differs significantly over the Europe.

   d. *Population* of airport's catchment area can also be considered as a destination attractor for cargo and visitor flows.

4. Administrative and historical sources

   a. Common *ownership* of airports. The majority of European airports were originally managed by governments, and public ownership of airports is still a widespread form. Frequently, all airports, located within a particular region (or country) are managed by the same agency. Such common ownership of neighbour airports is a major source of airport spatial heterogeneity.

   b. *Legislative environment* (including taxes, transportation laws, and air pollution limitations), affecting airport performance, is usually country-specific.

   c. *Economic regulation* of airports is another country-specific factor of heterogeneity. It will be separately discussed in the next paragraph.

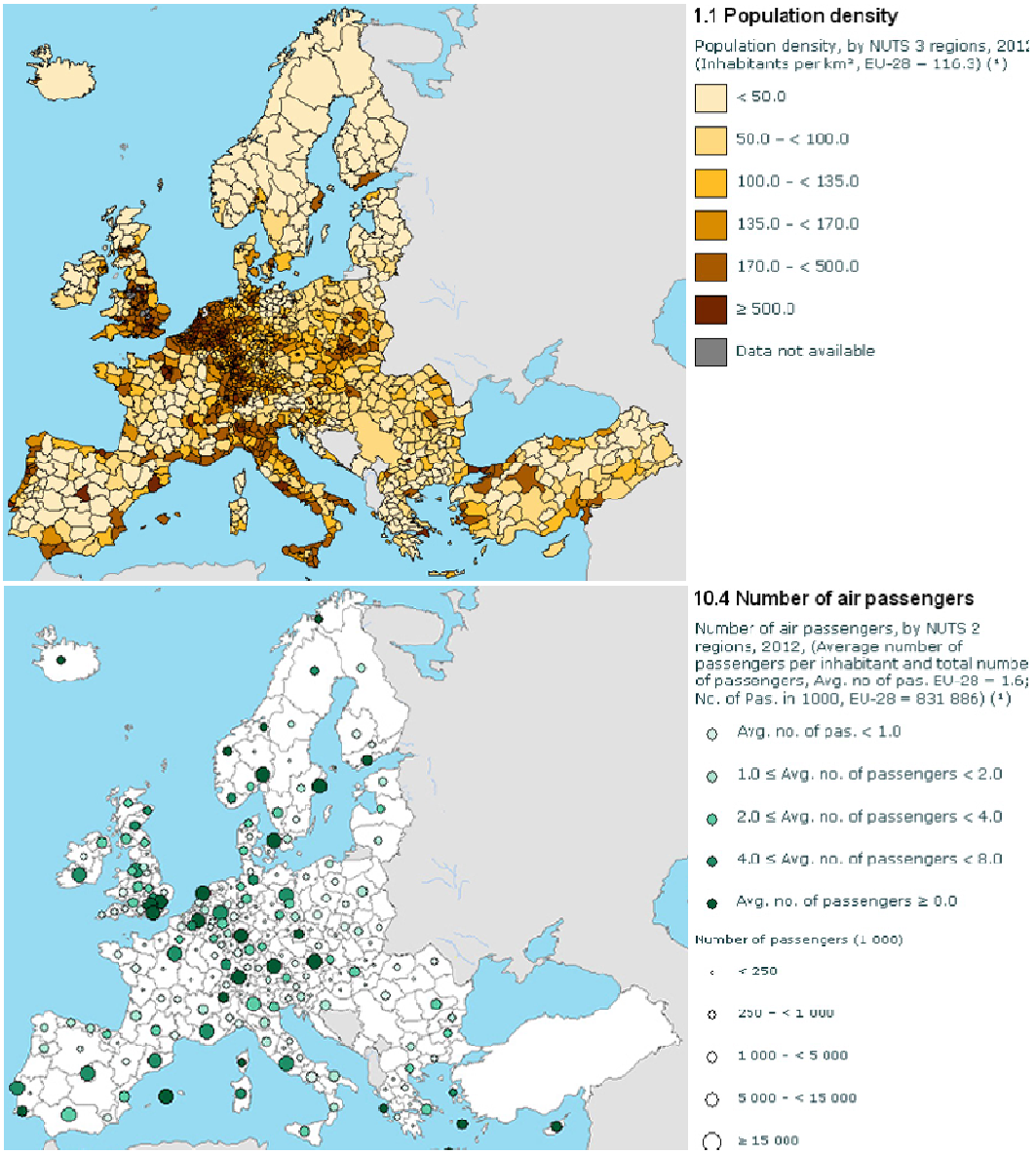


Fig. 1.1 Spatial distribution of population density and air passengers in the European countries. Source: Eurostat.

Note that mentioned factors affect both frontier and efficiency parameters, which leads to spatial heterogeneity of the frontier and spatially related inefficiencies of airports. These two consequences are modelled separately in this research.

Generally, a wide range of spatial factors create a very heterogeneous structure of the airport industry. Taking spatial heterogeneity (both observed and unobserved) into account for modelling can be stated as an important methodological enhancement.

1.2.3. Economic regulation as a source of spatial heterogeneity

Government economic regulation is a powerful source of airport spatial heterogeneity. Different regulation approaches, utilised in the European countries, lead to adjustment of managerial objective functions for all national airports and to airports' spatial similarities.

Economic regulators are basically used to prevent abusing of dominance by monopolies. Despite the liberalisation of the European air market, many airports still have significant market power and can be considered as spatial natural monopolies or oligopolies. European airport charges have traditionally been regulated, and European Union (EU) authorities continue this practice. Commission Regulation No 1794/2006 [10] defines general principles of air services changes and postulates that "in accordance with the overall objective of improving the cost efficiency of air navigation services, the charging scheme should promote the enhancement of cost and operational efficiencies". Flaming academic debates are related with types of airport activity, which should be regulated. As we described in the paragraph 1.1.1, the airport business is very diverse and include different types of aeronautical and non-aeronautical activities. A single-till regulation approach includes non-aeronautical revenues into the price-cap formula, when a dual-till approach, in contrast, tries to restrict only aeronautical revenues because they are the only ones having a monopolistic nature. A good review and analysis of single-till and dual-till regulation can be found in [106].

As regulation is considered as a replacement for competitive mechanisms, its influence on airport efficiency became a point of many academic and commercial studies during last years. There are several empirical evidences of interrelation between regulation and airport efficiency, but their conclusions are inconsistent. Some researchers tested a direct effect of regulation. Barros and Marques[20] included a dummy variable for regulated airports into a frontier definition of the SF model. They assumed a different cost frontier for regulated airports, and discovered that regulation contributes to a cost control. This effect was also analysed by the same authors for a sample of Japanese airports[37], but regulation was found insignificant for frontier's position in that case. Bel and Fagenda[107] investigated an influence of regulation on airport pricing for a sample of European airports and concluded that neither regulation form (rate of return or price-cap), nor regulated activities (single-till or dual-till) are significant for explaining airport charges. Gitto and Mancuso[88] estimated a two-stage DEA model for Italian airports and investigated an influence of the dual-till approach on airport efficiency scores. They discovered a significant positive effect of the dual-till approach in a financial model and an insignificant influence in a physical model. Adler and Liebert[27] also used a two-stage DEA model for discovering an influence of different regulation forms (unregulated, cost-based single-till and double-till, price-cap single-till and double-till) on airport efficiency. The authors investigated regulation effects for different levels of competition and concluded that in "weakly competitive conditions, dual-till price caps appears to be the most appropriate form of economic regulation".

Despite the recent enhancement of regulation, it can't be a perfect replacement for a competitive market. According to Starkie[108], there is "a trade-off between living with imperfect regulation or with imperfect markets".

1.3. Review of spatial competition between airports

1.3.1. Theoretical background of spatial competition

Spatial dependence is another theoretical aspect of spatial effects. It related with interactions between economic units, located close one to another. Presence of spatial dependence can be substantiated by different factors; spatial competition is one of the most intuitively important for the airport industry.

Competition among airports (for passengers, for airlines, etc.) is different by its nature and has various sources and effects. To the best of our knowledge, one of the most under-researched aspects of airport competition is a spatial one.

Spatial competition is mainly concerned with a locational interdependence among economic agents. The theory of spatial competition is well established and there are a significant number of its applications in different economic areas. Recently models of spatial competition were applied to movie theatres, gas stations, retail places, hospitals, country regions and others, but the airport industry is still weakly covered. Open airport market and increasing number of airports from one side and airports unalterable locations from another create good background for spatial completion in this sector.

A study, frequently cited as a pioneering in the area of spatial competition, was presented by Hotelling in 1929[109]. Hotelling considered a basic case of two firms producing homogeneous goods in different locations on a line and stated a key question about competition among firms and their efforts to differentiate from each other. Later the idea of Hotelling's model was developed in different ways. D'Aspermont et al.[110] introduced quadratic transportation costs for the model, which allowed an equilibrium solution. Salop[111] enhanced the model by replacing the linear locational structure with a two-dimensional circular one. A limitation of homogeneous goods, inadmissibly restrictive for the airport industry, also was addressed. Irmen and Thisse[112] introduced a multi-dimensional model where dimensions can have different weights. They proved that in the equilibrium point a firm differentiate itself from competitors in one dimension, but locate in the centre (close to other firms) for all other dimensions.

Correctness of Irmen and Thisse's model has several corroborations in the airport industry. A set of dimensions can include a price segment of served airlines (from LCC to regular and elite), traffic types (from cargo to connecting or direct passenger flights), flight destinations (from domestic to short- and long-haul international), and airport geographical location. Looking

at the European airport industry, we can discover several examples, where airports are differentiated in one of these dimensions, but located closely in others. There are European cities with major and secondary airports (London, Paris, Berlin), where the secondary airport is typically served by LCC (and differentiated in this dimension). Another example is airports in Baltic States' capital cities (Riga, Tallinn, Vilnius), which are differentiated geographically and don't have to distance themselves from each other for other dimensions.

A mode of airport competition is also a subject of academic researches[113], [114]. Biscia and Mota[115] presented an extensive review of studies on both quantity-based Cournot competition and price-based Bertrand competition in spatial settings.

1.3.2. Empirical studies on airports' spatial competition

Empirical estimation of spatial competition among airports is weakly covered by researches. There are two different ways in which airports can compete spatially:

- as departure points for local population; and
- as destination points for tourists and businesses.

Estimation of the first aspect of spatial competition among airports is usually based on the conception of catchment areas. Airport industry researches define airport's catchment area as a geographical zone containing potential passengers of the airport. Usually the geographical definition of airport's catchment area is supplemented with demographic indicators such as population, employment, income and others[116].

Catchment area's radius can be defined in different ways:

- by geographical distance;
- by travel time;
- by travel cost.

These metrics are used linearly or with time (distance) decay functions.

Several empirical researches used overlapping catchment areas as an indicator of spatial competition among neighbour airports. Starkie[108] studied competition between airports for hinterlands as a degree of the airports' catchment areas overlapping (Fig. 1.2) and later applied this approach in his further researches[117], [118]. Analysis of overlapping catchment areas was supplemented by additional characteristics of airport services like flights frequency, destinations, etc.

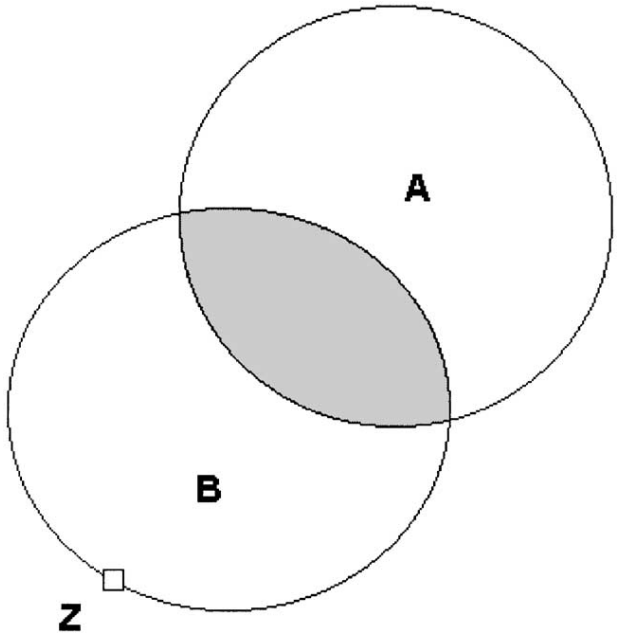


Fig. 1.2. Competition and catchment areas
Source: Starkie[108]

Strobach[119] constructed an index of spatial airport competition for a particular destination point using a set of factors, weighted by their (author-defined) importance. The factors include transport accessibility (distance and time values for private transport and cost and time values for public transport), traffic characteristics (frequency of flights to a selected direction, minimum connecting time, numbers of gates and check-ins), and characteristics of convenience (parking spaces, a terminal area, an area of shopping and services). Malina[120] suggested a substitution coefficient, which "defined as the share of inhabitants within the relevant regional market of an airport that consider another airport (...) to be a good substitute from their perspective as well". Hancioglu[121] investigated competition between Dusseldorf and Cologne/Bonn airports using Malina's airports substitution coefficient, mainly based on overlapping catchment areas, and a custom survey of passengers' origin regions. The author of this thesis [101] suggested constructing multiple catchment areas of an airport for different flight destinations. Bel and Fagenda[122], and Adler and Liebert[27] used number of nearby airports as a simple indicator of competition pressure.

Another popular approach to estimation of completion pressure is interviews with experts and airport management[43], [123], [124]. This approach is very useful for initial analysis of the competition pressure, but has obvious shortcomings of subjectivity and quantitative measurement.

The second way of spatial competition among airports is based on their function to be an intermediate destination point. Leisure and business travellers manage their trips and define intermediate connection points (including airports). This subject of their choice is wider than

selection between two (or more) airports in a destination city and relates to trip's route as whole. For example, for a saving trip from London to Moscow travellers can choose between Riga and Tallinn airports as an airline-railway transfer point. Note that the essence of this way of competition is not necessary spatial, but spatial effect can take place in some cases. To the best of our knowledge, there are no studies containing empirical estimation of this aspect of spatial competition between airports.

1.3.3. Spatial competition and airports efficiency

There are few empirical studies of a relationship between spatial competition and efficiency of airports.

Borins and Advani[43] used interviews with airport managers to estimate levels of competition of two types – transferring traffic and catchment areas. Estimated competition levels were included into two classical regression models with passenger and airline orientations. Both competition types are found significantly positive in both models, so the authors concluded a positive influence of competition on airports activity.

Jing[36] analysed efficiency of Asian cargo airports using the SF approach and including competition into consideration. A suggested competitiveness index was constructed on the base of airports ranking by locational, facility, service quality, charges, staff quality, connectivity, and market environment factors. Although airport's geographical location was included into the index, spatial effects are not examined in the paper.

The author of this thesis[101] suggested index of competition, based on overlapping catchment areas, included it into the SF model, and discovered a positive effect of a competition pressure on efficiency for a sample of European airports. Non-linear spatial interdependence was investigated in the author's further research[102] and a multi-tier model of competition and cooperation effects was suggested. The model estimates provide both positive and negative effects depending on a distance.

Scotti et al.[8], [41], [44] suggested an index of competition between two airports on the base of a share of population living in an overlapped region of the airports' catchment areas. A competition index was calculated separately for every destination point (exact or reasonably close) and combined into the general competition index using available seats shares as weights. The suggested index was included in a set of inefficiency determinants of a multi-output SF model. Estimating parameters of this model for a sample of Italian airports, the authors concluded a significant negative relationship between competition pressure and airport efficiency. Authors explained this fact by overcapacity of airports. Airports, acting in a more competitive environment, captured limited benefits of air transport post-liberalisation traffic

growth, when monopolistic airports easier filled their capacity and improved their technical efficiency.

Adler and Liebert[45] investigated an influence of competition on airport efficiency using a two-stage DEA model. A level of competition was included into the second stage regression as number of significant airports within a catchment area and showed up as a significant factor for results of different regulation forms. The spatial specification of the second stage regression was tested by author, but solely for justifying of the model's robustness.

1.4. Conclusions

During last two decades airport benchmarking attracted a significant attention of the scientific community. Many theoretical and practical studies, addressed to this problem, are recently published, but a formal problem specification and a preferred methodological base are still a matter of discussions. The problem complexity is mainly related with a high level of airport business heterogeneity, based on different specifications of airport resources and outputs. Passengers and cargo transferred by an airport, airline movements served, environmental emission and noise, non-aviation services, and other airport activity aspects are included into studies either as resources or as outputs of the business.

A range of quantitative methods, used for airport benchmarking, is reasonably wide. Productivity indicators (PFP and TFP), deterministic (DEA, FDH) and stochastic (SFA) frontier approaches are widely used. PFP indexes are frequently used for initial analysis of airport efficiency, as they reflect only a particular activity aspect. Modern frontier-based approaches (DEA and SFA) become popular for estimation of overall airport efficiency. The majority of airport studies utilise the DEA approach to benchmarking, but during last five years number of SFA applications is increased significantly. This growing interest to SFA is based on recent theoretical SFA developments, which allow modelling a heterogeneous nature of airport production, and a growing level of data availability.

In this chapter we paid special attention to analysis of spatial effects in the airport industry of their relationships with airport efficiency. Spatial heterogeneity and spatial dependence are two types of spatial effects, which are widely acknowledged in the airport industry. Consideration of spatial effects is, in our opinion, a required enhancement of airport benchmarking procedures.

Spatial heterogeneity is based on uneven distribution of efficiency-related factors within a geographic area. These factors, like climate features, economic and legislative environments, and population habits, can significantly affect airport productivity and must be considered in airport benchmarking.

Spatial dependence is the second type of spatial effects, related with interactions between neighbour economic units. Presence of spatial dependence can be substantiated by different factors; spatial competition is one of the most intuitively important for the airport industry. Despite a limited nature of airport competition, there are several studies with empirical evidences of its presence. The theory of spatial competition is well-developed, but number of its empirical applications in the airport industry is very limited, which creates a direction for further researches.

Finally, a relationship between spatial effects and efficiency of airports is also weakly researched. A small number of empirical studies don’t allow make a comprehensive conclusion about the subject. The methodological base in this area is also scanty, so influence of spatial effects on airports efficiency is an extensive and complicated research topic. We conclude that application of spatial econometrics will enhance the methodological base and lead to practically important results.

2. STOCHASTIC FRONTIER ANALYSIS (SFA) AND A PROBLEM OF SPATIAL EFFECTS INCORPORATION

2.1. Theoretical background of SFA

A process of production in classical economics is defined as the usage of material and immaterial resources for making goods and services[125]. Further in this chapter we will refer a company as a production unit, which uses a set of resources (inputs) to produce a set of goods and services (outputs).

We consider a company, which uses K inputs, indexed k = 1, 2, …, K, to produce M outputs, indexed m = 1, 2, …, M. Input and output bundles can be presented in a vector form as:

$$x = (x_1, x_2, ..., x_K),$$
$$y = (y_1, y_2, ..., y_M).$$

The production process can be defined as transforming of an input vector x into an output vector y. Technological limits of production are usually described as a set of pairs of input and output vectors, which are possible in the sense that a company can produce an output vector using a given input vector[126]. This set of input and output pairs is well known as a production possibility set and we will denote it by PPS:

$$PPS = \{x, y : x \text{ can produce } y\}$$

The set of feasible outputs for an input vector can be defined as:

$$P(x) = \{y : (x, y) \in PPS\}$$

This set includes all output vectors y, which are feasible for a given input vector x.

Definition of efficiency of company's activity strictly depends on goal of this activity. Most widely used goals of a company are maximisation of the output vector given by a fixed input vectors (output-oriented) and minimisation of the input vector given by a fixed output vector (input-oriented). Efficiency, measured on the base of these production-oriented approaches, is called technical. There are a number of alternative goal specifications: revenue maximisation, cost minimisation, profit maximisation and some others. Duality of different approaches is widely acknowledged in the production theory[126] under some not very restrictive assumptions about the PPS (for example, a free disposal assumption). These dualities are very practical; they allow researchers to consider a task, related to a specific approach, and transfer the results on other approaches. Further in this chapter we will consider the output-oriented production approach whereas other approaches are very similar in terms of logic.

An output vector is called technically efficient if, and only if (Koopmans's definition, [127]):

$$y_{eff} : y_{eff} \in P(x) \Rightarrow \forall_{y'>y}\, y' \notin P(x)$$

The term $y' > y$ denotes that y precedes y': a value of at least one component in y' is more than its value in y and values of other components in y' is not less than in y. So technical efficiency means that given an input vector there are no feasible output vectors exceeding y_{eff} in any component.

Expanding this concept to all feasible set of input vectors, a production possibility frontier is defined as a function:

$$f(x) = \{y : y \in P(x), \forall_{y'>y}\, y' \notin P(x)\} \tag{2.1}$$

In case of a single output production process, the production possibility frontier can presented as:

$$f(x) = \max_{y} P(x)$$

Koopmans's definition of technically efficient output vectors is very general and can be applied to outputs of different nature. A more practically convenient definition of technical efficiency of output vector y was presented by Debreu[128] and Farrell[129]:

$$TE(x, y) = \left[\sup_{\theta} \{\theta : \theta y \le f(x)\} \right]^{-1} \tag{2.2}$$

This definition is closely related with a distance function, introduced in Shephard's works on multi-output production[130].The main difference with Koopmans's definition is in direction of output vector increasing. Koopmans's definition allows increasing of any component of y, while the Debreu-Farrell definition considers only equiproportional (radial) increase of y.

Later the Debreu-Farrell definition was extended by Luenberger [131] and Campbers, Chung, and Fare[132], who introduced a directional technology distance function.

See Fig. 2.1 for illustration of different definitions of technical efficiency.

Further in this paper we will follow the Debreu-Farrell definition for a reason of simplicity. All discussed features can be extended to more general definitions of technical efficiency. According to the Debreu-Farrell definition, values of the technical efficiency should satisfy the following properties:

1. $0 \le TE(x,y) \le 1$
2. $TE(x,y_{eff}) = 1$
3. $TE(x,y)$ is non-decreasing in y.
4. $TE(x,\lambda y) = \lambda TE(x,y)$

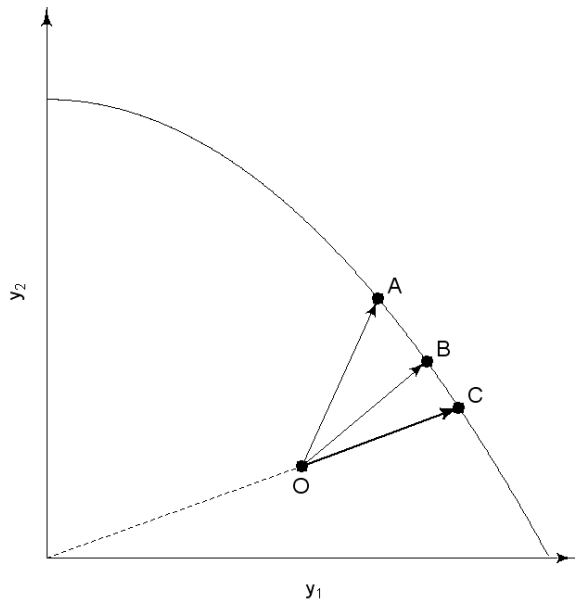


Fig. 2.1. Alternative definitions of the technical efficiency:
OA – an arbitrary directional distance, OB – Koopmans's (closest) distance, OC – Debreu-Farrell's (radial) distance

So a value of technical efficiency equals to 1 for a company, located on the production possibility frontier (produced a maximum possible vector of outputs given by its input vector). Companies, which produce less than maximum possible outputs, feasible with their inputs, are qualified as inefficient.

The Debreu-Farrell definition of the technical efficiency can be presented in a form of equation:

$$y = f(x) \cdot TE(x, y) \tag{2.3}$$

So, given *x* and *y*, tasks of construction of production frontier *f(x)* and technical efficiency *TE(x,y)* are dual to each other. This fact is widely covered in theoretical literature; see [133] for an extensive review.

For estimation purposes the technical efficiency term is usually transformed as:

$$TE(x, y) = \exp(-u), u \geq 0. \tag{2.4}$$

After this transformation properties (1-3) for technical efficiency values are satisfied automatically. The term *u* is an inverse to the technical efficiency value, so it is frequently noticed as an inefficiency term.

Thus the equation (2.3) can be presented as:

$$y = f(x) \cdot \exp(-u) \tag{2.5}$$

This model assumes that the production frontier *f(x)* is deterministic. This assumption ignores the fact that production of a company can be affected by random disturbances. Presence of these random disturbances in practice is widely acknowledged and considered as a background

for econometric analysis[134]. Random disturbances are usually explained by influence of a large set of factors, generated both from company's internal and external environment. Introducing the random disturbances v into the formula (2.5), we consider a classical stochastic frontier (SF) model:

$$y = f(x) \cdot \exp(v) \cdot \exp(-u) \tag{2.6}$$

For econometric estimation of this model we assume that we have a sample of n companies, indexed $i = 1, 2, \ldots, n$. Values of output (y_i) and input (x_i) vectors are available for each company, while values of random disturbances (v_i) and inefficiencies (u_i) are not observable. Supposing that the production possibility frontier $f(x)$ is common for all companies in the sample and depends on a vector of parameters β, we receive a cross-sectional specification of the stochastic frontier model:

$$y_i = f(x_i, \beta) \cdot \exp(v_i) \cdot \exp(-u_i) \tag{2.7}$$

When a production process is described only by one output ($M = 1$), the specification (2.7) represents a standard econometric model, which parameters can be estimated. This approach is frequently used in cases when the single-output assumption is appropriate for a real production process or when production outputs can be aggregated. The model is frequently presented in the logarithmic form, which is more convenient in practice:

$$\ln y_i = \ln f(x_i, \beta) + v_i - u_i \tag{2.8}$$

Models with multiple outputs ($M > 1$) production require a transformation to become econometrically estimatable. A popular transformation[135], [136] utilises the property 4 (homogeneity of degree 1 in outputs) of technical efficiency. Selecting an arbitrary output (following Coelli and Perelman, we use the last output y_M) and putting λ to $1/y_M$ we have:

$$TE(x, y/y_M) = \frac{1}{y_M} TE(x, y) \tag{2.9}$$

Using (2.4) representation of the technical efficiency:

$$TE(x, y/y_M) = \frac{1}{y_M} \exp(-u)$$

And finally

$$[y_M]^{-1} = TE(x, y/y_M) \exp(u) \tag{2.10}$$

Embedding random disturbances into the model and introducing parameters of technical efficiency β (dual to the parameters of the production possibility frontier), we receive a specification of a multi-output cross-sectional stochastic frontier model:

$$[y_{Mi}]^{-1} = TE(x_i, y_i / y_{Mi}) \cdot \exp(v_i) \cdot \exp(u_i) \tag{2.11}$$

In this form the model can be estimated using standard econometric techniques.

Another approach to specification of econometric model for a multi-output case is presented by Lothgren[137] and called stochastic ray production frontier. In this research we use the presented Coelli and Perelman's approach.

The model (2.11) in the logarithmic form is:

$$-\ln y_{Mi} = \ln TE(x_i, y_i / y_{Mi}) + v_i + u_i$$

Estimation of the models requires a functional form assumption – for the production possibility frontier $f(x_i, \beta)$ in the single-output model (2.8) and for the technical efficiency $TE(x_i, y_i/y_{Mi}, \beta)$ in the multi-output model (2.11). There is a set of widely known theoretical production functions: a Cobb-Douglas function, a translog function, a Diewert function, a CES (constant elasticity of substitution) function.

The Cobb-Douglass function is one of the simplest forms:

$$\ln f(x_i, \beta) = \beta_0 + \sum_{j=1}^{K} \beta_j \ln x_{ji}$$

All elasticities of substitution between inputs in the Cobb-Douglas function are equal to 1.

The translog production function is more flexible in terms of substitution elasticity:

$$\ln f(x_i, \beta) = \beta_0 + \sum_{j=1}^{K} \beta_j \ln x_{ji} + \sum_{j=1}^{K} \sum_{k=1}^{K} \beta_{jk} \ln x_{ji} \ln x_{ki}$$

Elasticity of substitution in the translog production function is not fixed to 1, but can be estimated.

Other popular functions also differ in terms of elasticity of substitution: Diewert function fixes elasticity to 0, CES function fixes elasticity to an estimatable constant. This research is limited with consideration of the Cobb-Douglas functions.

Thus specifications of the stochastic frontier model, used in this research, are:

1. Single-output Cobb-Douglas stochastic frontier:

$$\ln y_i = \beta_0 + \sum_{k=1}^{K} \beta_k \ln x_{ki} + v_i - u_i \tag{2.12}$$

2. Multi-output Cobb-Douglas stochastic frontier:

$$-\ln y_{Mi} = \beta_0 + \sum_{k=1}^{K} \beta_k \ln x_{ki} + \sum_{m=1}^{M-1} \beta_j^o \ln(y_{mi}/y_{Mi}) + v_i + u_i \tag{2.13}$$

For simplicity of further model specification we will use the presented specifications in the matrix form. Folding the model (2.8) by *i*, we receive:

$$Y = X\beta + v - u \tag{2.14}$$

This form is general for both specifications presented above; the matter is in definition of matrices.

The single-output Cobb-Douglas stochastic frontier:

$$Y = \{Y_i\} = (-\ln y_1, -\ln y_2, ..., -\ln y_n)^T$$

$$X = \{X_i\}^T = \begin{bmatrix} 1 & ... & 1 \\ \ln x_{11} & ... & \ln x_{1n} \\ ... & ... & ... \\ \ln x_{K1} & ... & \ln x_{Kn} \end{bmatrix}^T_{(1+K)\times n}$$

$$\beta = (\beta_0, \beta_1, ..., \beta_K)^T$$

$$v = (v_1, v_2, ..., v_n)^T$$

$$u = (u_1, u_2, ..., u_n)^T$$

The multi-output Cobb-Douglas stochastic frontier:

$$Y = \{Y_i\} = (-\ln y_{M1}, -\ln y_{M2}, ..., -\ln y_{Mn})^T$$

$$X = \{X_i\}^T = \begin{bmatrix} 1 & ... & 1 \\ lnx_{11} & ... & lnx_{1n} \\ ... & ... & ... \\ lnx_{K1} & ... & lnx_{Kn} \\ ln(y_{11}/y_{M1}) & ... & ln(y_{1n}/y_{Mn}) \\ ... & ... & ... \\ ln(y_{K1}/y_{M1}) & ... & ln(y_{Kn}/y_{Mn}) \end{bmatrix}^T_{(1+K+(M-1))\times n}$$

$$\beta = (\beta_0, \beta_1, ..., \beta_K, \beta_1^o, ..., \beta_{M-1}^o)^T$$

$$v = (v_1, v_2, ..., v_n)^T$$

$$u = -(u_1, u_2, ..., u_n)^T$$

It should be noted that the output ratios are included into the matrix of explanatory variables *X* for the multi-output frontier. Usually explanatory variables are supposed to be exogenous, but in this case the endogeneity problem could arise. The problem arises if the output ratios are correlated with random disturbances *v* and inefficiency *u* (for example, if the inefficiency has different effects on different outputs). A comprehensive treatment for this

problem is discussed in [138]; also Kumbhakar[139] summarises different approaches to estimation of the multi-output stochastic frontier.

2.2. Review of the maximum likelihood estimator of the SF model parameters

A wide range of statistical methods can be applied to estimate parameters of the production frontier and inefficiency terms of the model (2.14):

- Method of moments (MOM) estimator
- Maximum likelihood estimator (MLE)
- Generalised maximum entropy (GME) estimator
- Bayesian estimator[140], [141]

MOM estimator of the SF model includes two steps: calculation of consistent estimates of frontier parameters using the ordinary least squares method and further estimation of the inefficiency terms' and random disturbances' parameters and intercept using sample moments. This procedure is well developed for different specifications of the model's inefficiency term [142], [143].

GME is a modern statistical estimation technique, which also can be applied to the SF model[57], [144]. This technique utilises information from every sample observation (instead of sample moments only in MOM) and allows receiving more robust estimates for ill-posed models and small samples.

Both MOM and GME don't require any additional assumptions about the structure of the random disturbances v and the inefficiency term u. If distribution laws of v and u can be defined, the most natural choice for estimation of the SF model parameters is MLE. This popular statistical approach utilises an assumption about v and u distributions and provides consistent and asymptotically efficient estimates. This research is mainly based on the ML approach.

The distribution of the random disturbances v is usually set to independent identically distributed (IID) normal with zero mean and constant deviation σ_v:

$$v_i \sim N\left(0, \sigma_v^2\right)$$

The matrix form:

$$v \sim MVN\left(0_n, \sigma_v^2 I_n\right),$$

where 0_n is a vector of n zeros, I_n is an $n \times n$ identity matrix.

Distribution of the inefficiency term u can be selected from a set of appropriate distribution laws of non-negative random variables. There are several specifications of the SF model, based on different distributions of the inefficiency term u:

- half-normal distribution [32]:

  $u_i \sim N^+\left(0, \sigma_u^2\right),$

- truncated normal distribution [145]:

  $u_i \sim TN_{0,+\infty}\left(\mu, \sigma_u^2\right),$

- exponential distribution[33]:

  $u_i \sim Exp(\lambda),$

- gamma distribution [146]:

  $u_i \sim Gamma\,(k, \theta).$

Note that the truncated normal distribution is a generalisation of the half-normal, and the gamma distribution is a generalisation of the exponential.

Taking advantages and shortcoming of different distribution specifications, in this research we concentrate on the specification with the truncated normal distribution. The probability density function for truncated normal distribution is (truncation limits are set to 0 and +∞ to match the non-negativity requirement of *u*):

$$f(u_i) = \begin{cases} \left[\Phi\left(\dfrac{\mu}{\sigma_u}\right)\right]^{-1} \dfrac{1}{\sqrt{2\pi}\sigma} \exp\left(-\dfrac{(u_i-\mu)^2}{2\sigma_u^2}\right), u_i \geq 0 \\ 0, u_i < 0 \end{cases} \tag{2.15}$$

Example plots of this function are presented on the Fig. 2.2.

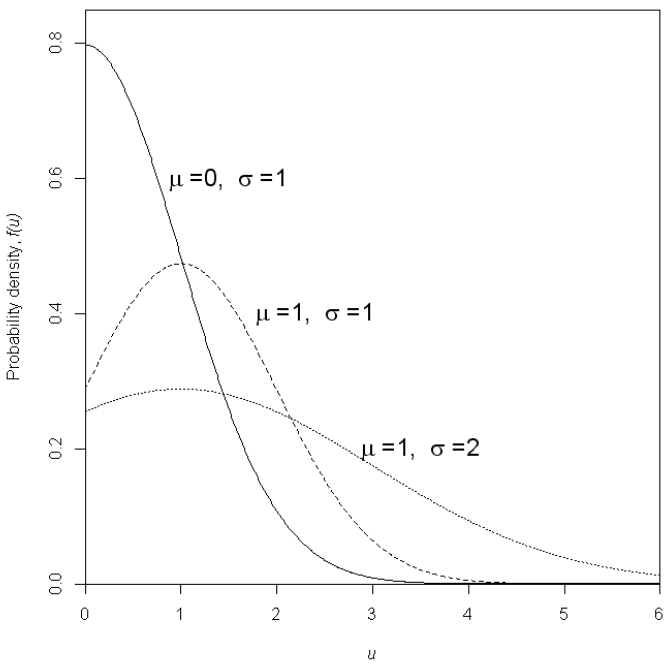


Fig. 2.2. Plots of truncated normal probability density functions

If μ=0, then the density function is folding to the half-normal density.

The composed error term of the SF model (2.14) is constructed as a difference of random variables with normal and truncated normal distributions:

$$\varepsilon_i = v_i - u_i \tag{2.16}$$

The density distribution function of a sum of normal and truncated normal distribution is well known[31]:

$$f(\varepsilon_i) = \frac{1}{\sigma}\left[\Phi\left(\frac{\mu}{\sigma_u}\right)\right]^{-1} \varphi\left(\frac{\varepsilon_i + \mu}{\sigma}\right)\Phi\left(-\lambda\frac{\varepsilon_i + \mu}{\sigma} + \frac{\mu}{\sigma\lambda}\right), \tag{2.17}$$

where

φ and Φ are the standard normal density function and cumulative distribution function respectively,

$$\sigma = \sqrt{\sigma_v^2 + \sigma_u^2},$$

$$\lambda = \frac{\sigma_u}{\sigma_v},$$

$$\sigma_u = \frac{\lambda\sigma}{\sqrt{1+\lambda^2}}.$$

An alternative parameterisation[147] with very similar computational properties uses γ instead of λ:

$$\gamma = \frac{\sigma_u^2}{\sigma^2}. \tag{2.18}$$

This density function is known as an extended skew normal distribution function, introduced by Azzalini[148], [149] (up to re-parameterisation, discussed in [150]). Example plots of this function are presented on the Fig. 2.3.

The log-likelihood function for the SF model with the truncated normal distribution of u and a sample of n observation is[145]:

$$\ln L = -\frac{n}{2}\ln 2\pi - n\ln\sigma - n\ln\Phi\left(\frac{\mu}{\sigma_u}\right) + \sum_{i=1}^{n}\ln\Phi\left(-\lambda\frac{e_i + \mu}{\sigma} + \frac{\mu}{\sigma\lambda}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(e_i + \mu)^2, \tag{2.19}$$

$$e_i = Y_i - \beta X_i$$

Note that both random disturbances and inefficiency terms are supposed to be independent one from each other and for different sample observations.

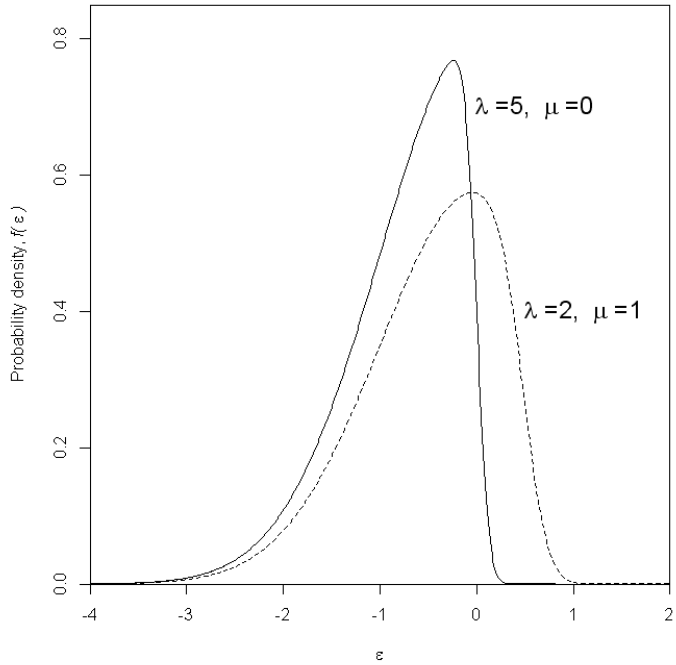


Fig. 2.3. Plots of extended skew normal probability density functions

Given estimates of the SF model parameters β, σ_v, σ_u, inefficiency terms u_i can be estimated[151] as a conditional expected value $E(u_i|\varepsilon_i)$ and further the technical efficiency can be estimated as $TE_i = exp(-u_i)$ by u definition.

A conditional distribution of u_i by ε_i is a truncated normal distribution:

$$u_i|\varepsilon_i \sim TN_{0,+\infty}\left(\tilde{\mu}_i, \tilde{\sigma}^2\right), \tag{2.20}$$

where

$$\tilde{\mu}_i = \frac{\mu\sigma_v^2 - \varepsilon_i\sigma_v^2}{\sigma^2},$$

$$\tilde{\sigma} = \frac{\sigma_v\sigma_u}{\sigma}.$$

Moments of the truncated normal distribution are well known[134], so point and interval estimates for technical efficiency can be easily calculated[152].

2.3. Review of existing approaches to modelling of spatial effects in SFA

The classical SF model is based on a core statistical assumption of independence of observations in the sample. Under this assumption inputs, outputs and efficiencies of all sample companies are considered as not dependent. In practice, this assumption is frequently violated due to different links connecting companies in any economies. These links can be based on common markets and customers, common suppliers, common economical and political

environments, competition and cooperation, and other economic relationships. One of the possible ways for identification of these links is based on companies' geographical location; in this case the links are called spatial effects. Closely located companies can influence one to another or experience common area-specific difficulties. Presence of spatial effects violates the independency assumption in different manners:

- Activity of a company can be affected by output and input values of neighbour companies (this spatial effect is called spatial dependence). Many scientific theories rely on presence of spatial dependence. For example, regional science is almost completely based on spatial processes in transportation, agriculture, industry and other fields; spatial competition is a traditional component of the economic theory; spatial relationships are accepted in biology, ecology and other natural sciences.

  From the econometric point of view, these effects can be separated to endogenous and exogenous[153].

  Endogenous effects represent a relationship between outputs of neighbour companies that is an output of a given company is determined by outputs of its competitors. Note that spatial effects can be asymmetric (effect of a company i on a company j is not equal to effect of the company j on the company i), so the number of endogenous spatial effects for M outputs equals to $M(M-1)$.

  Exogenous effects represent a relationship between outputs of a given company and input of neighbour companies. Exogenous effects are usually explained by a common market of companies' inputs, where shortage of an input in one company leads to higher level of this input utilisation in neighbour companies. A number of exogenous spatial effects for M outputs and K inputs equals to MK.

- Activity of a company can be affected by area-specific factors. Many influencing factors are unevenly distributed over the space (this effect is called spatial heterogeneity). Usually factors are distributed continuously, so it can be assumed that they have similar effects on companies, located closely one to another. There are a lot of factors of this nature – weather conditions, economic environment, ecosystems, and others. Some of them can be observed easily, but a considerable part of this influence is directly unobserved or hard to measure. A number of spatial heterogeneity effects is 1.

Potential problems of spatial effects in SFA were noted in early frontier researches. Farrell[129] constructed a production frontier agricultural firms in US and noted apparent differences in efficiency, shown up due to factors like climate, location and fertility. Though the problem is stated, it is rarely attended by researches. Only a limited number of researches pay attention for spatial distribution of estimated efficiency scores. Among a few others, Fahr and

Sunde[50] discovered significant spatial autocorrelation of regional efficiency of job creation in Western Germany; Bragg[154] analysed spatial relationships of efficiency scores in the Maine dairy industry; Igliori [155] noted apparent spatial heterogeneity in agricultural production on regional level in Brazil; Hadley[156] revealed the same patterns on a firm level in England and Wales.

Distinguishing inefficiency from heterogeneity (of different natures) in SF models has become a popular point of scientific interest during the last decade. Note that this is econometrically impossible to separate company-specific inefficiency and unobserved heterogeneity having only cross-sectional data and making no assumptions about nature of heterogeneity.

2.3.1. Approaches to estimation of spatial effects

The mainstream solution of this problem is directed to analysis of panel data. Using panel data (repeated observations of the same set of producers over time) separation of inefficiency and heterogeneity becomes technically possible, but require additional assumptions. Schmidt and Sickles[157] developed an estimator, assumed time-invariant inefficiency (which can be substantiated for short panels). Battese and Coelli[158] discussed an SF model specification, where inefficiency is not time-invariant, but changes over time by a functional form (common for all companies).

An opposite assumption can be appropriate for long panels: all time-invariant producer-specific effects are considered as heterogeneity and all production variations over time are considered as inefficiency. Economic validity of this assumption is a matter of every particular application. Under this assumption, Kumbhakar[159] proposed a random-effects model to separate inefficiency and factors that are outside producer's control, Greene[160], [161] suggested "true" fixed- and random-effects model specifications for separate estimation of unobserved heterogeneity. Recently the proposed model has been generalised by Wang and Ho[162], utilising though the same principles. This approach was applied (among a few others) by Abrate et al.[163] to analysis of water industry in Italy. Kopsakangas-Savolainen and Svento[164] executed a comparative analysis of different model specifications and estimators within this approach.

Ahn and Sickles[165] and Tsionas[166] introduced dynamic SF models, where inefficiency is considered as a stationary process with a long-run equilibrium. Development of the latter approach and related heterogeneity issues are extensively discussed in [167].

All approaches based on panel data require a significant number of time points in a data set (a long panel). Unfortunately, the majority of panels in practice are relatively short, which leads to obvious estimation difficulties.

Another direction in distinguishing heterogeneity from inefficiency is based on assumption about a known structure of heterogeneity. One of the most natural and theoretically well-grounded forms of this assumption is based on a spatial structure of heterogeneity. It can be assumed that heterogeneity is explained by spatial settings and common for neighbour companies, when the efficiency itself is company-specific. Under this assumption distinguishing becomes econometrically possible. Spatial heterogeneity is acceptable in many real-world data sets, and usage of this information allows separating at least a part of heterogeneity from inefficiency values.

The methodological part of this research is devoted to integrating spatial effects into SFA.

2.3.2. Observed spatial effects in SF models

Spatial effects can be conventionally separated into two types – observed and unobserved. Observed spatial effects can be presented as a set of measurable factors, but the unobserved are based on factors, which can't be monitored or even identified. Observed spatial heterogeneity is considered as a necessary component of the SF model, starting from earlier applications[129].

Observed spatial effects are often controlled by introducing:

- dummy variables for regional divisions (countries, states, city districts, economic areas, etc.);
- distance-related locational factors (distance to a city centre, to a nearest service provider, to transport nodes, etc.);
- area-specific exogenous factors (weather conditions, population income and other characteristics, soil types, etc).

There are many empirical SFA applications, where observed spatial effects are included into consideration. Handley[156] introduced dummy variables for regional heterogeneity in UK farming efficiency; Schettini[168] applied the same technique for analysis industrial performance of Brazilian regions; Perelman and Serebrisky[80] used continental dummies for world airport benchmarking. Feng et al.[169] included a distance and cost of travel to the nearest high-speed railway station as a factor of regional development. Misra[170] analysed competition between public schools in United States and schools' efficiency using distances between nearest competitors.

A wide range of application areas, where observed spatial effects are utilised, supports our conclusion about empirical necessity of spatial components in SF models. Generally a spatial

structure can't be completely described using a set of observed factors, so we expect presence of unobserved spatial effects in all these cases.

Company's geographical location is the only information that can be used for dealing with unobserved spatial dependence and heterogeneity. The famous geographical Tobler's Law[171] says that "everything is related to everything else, but near things are more related than distant things", so it is usually expected that a power of spatial effects can be introduced by a distance between companies. It is worth to note that the meaning of "distance" can be different – geographical distance, economic links, infrastructure connections (roads, etc). In any case this distance is considered as exogenous to the model within the paradigm of spatial econometrics.

2.3.3. Principles of spatial econometrics

Spatial econometrics[11], [172] provides a extensive set of treatments for accounting for unobserved spatial effects in regression models.

So let we have a distance between every two companies *i* and *j* in the sample, captured from the spatial structure. Exogenous and non-stochastic distances are required for consistent estimation of model parameters in cross-sectional settings[173]. A higher distance between companies generally means weaker spatial relationship, so an inverse distance is frequently used and called a spatial weight w_{ij}. So for *n* producers in a sample, a matrix of spatial weights (called a contiguity matrix) can be constructed:

$$W = \{w_{ij}\}_{n \times n}$$

All main diagonal elements of *W* are conventionally put to zeros to exclude self-dependency.

Specification of the matrix *W* is usually under researcher's responsibility and determines a power of interrelation between objects on the base of their geographical locations. There are some different approaches to specification of the matrix *W*, based on different types of geographical objects:

- Objects have geographical areas with borders. In this case it can be noted if two objects are adjacent. For example, countries, regions within a country, districts of a city, agricultural firms can be considered as objects with an area. The matrix *W*, based on this principle, is denoted as contiguity-based.
- Objects are geographical points. These objects cannot be specified as neighbours; a distance between objects is used instead. The matrix *W*, based on this principle, is denoted as distance-based.

There are different approaches to define the spatial weights matrix for objects of first and second types.

The easiest form of a contiguity-based matrix is a binary matrix of neighbourhood. In this approach a matrix item equals to 1 if two objects are adjacent (have a common border) and equals to 0 otherwise.

Construction of a distance-based matrix strictly depends on the definition of the term "distance". Usually a distance is considered as a geographical distance, but also can be estimated in different ways[174]:

- An exact physical distance in kilometres between two objects. For relatively close objects the distance can be calculated as Euclidean distance for objects' coordinates, but if objects are relatively far one from another (so the spherical form of the Earth becomes significant), a great circle distance should be used.
- Time required for a trip from one object to another. This metric is better than the previous one in case when accessibility should be included into consideration.
- Travel cost is also often used as a distance metric.

Travel time and cost should be used when we suppose a spatial structure related with human activities. So if a distance between airports is considered in context of their competition for passengers, travel time becomes a good metric for a distance. When spatial heterogeneity (for example, weather conditions) should be included into the model, the geographical distance becomes more convenient.

Power of spatial interdependence can be non-linearly reduced with a distance between objects. In this case a kind of distance decay function should be considered.

Finally, spatial influence can be limited with a predefined distance value h, so objects located far than h kilometres one from another have 0 values for their positions in the contiguity matrix.

As it was mentioned earlier in this chapter, spatial effects are not symmetric, so $w_{ij} \neq w_{ji}$, when a geographical distance between spatial objects is obviously symmetric. This fact is partially levelled by row-standardisation of the spatial weights matrix. The row-standardisation procedure specifies a standardised spatial weight as a ratio of w_{ij} to the sum of all spatial weight by row (by j). Generally this procedure makes a company, which has a small number of neighbours, "closer" to its neighbours. Validity of this assumption is a matter of application and may not be appropriate in some situations. Row-standardisation makes the spatial weights matrix stochastic (taking that there are no companies without neighbours in the sample). Dealing with stochastic matrixes is easier from mathematical point of view, therefore row-standardisation of the spatial weights matrix is widely acknowledged as a necessary procedure.

An aggregate spatial influence of neighbour producers can be presented as a weighted sum of neighbour parameter values. This influence is called a spatial lag and expressed as[173]:

$$[Wy]_i = \sum_{j=1}^{n} w_{ij} Y_j \tag{2.21}$$

A general spatial regression model is expressed (in linear form) as[153], [175]:

$$\begin{aligned} Y &= \rho_Y W_Y Y + X\beta + W_X X\beta^{(s)} + v, \\ v &= \rho_v W_v v + \tilde{v}, \end{aligned} \tag{2.22}$$

where

- W_Y, W_X, W_v are spatial weights matrixes for output-output (endogenous spatial effects), output-input (exogenous spatial effects) and error-error (spatial heterogeneity effects) relationships accordingly;
- $\rho_{\boldsymbol{Y}}$, $\beta^{(s)}$, $\rho_{\boldsymbol{v}}$ are unknown parameters of spatial dependence in outputs, between inputs and outputs, and in error term accordingly;
- $\tilde{\boldsymbol{v}}$ is a vector of IID symmetric disturbances.

According to the general spatial model specification (2.22), output of producers $\boldsymbol{y}$ is influenced by:

- its own inputs X with parameters β;
- spatially weighted outputs of neighbour producers $W_Y Y$ with a parameter $\rho_{\boldsymbol{Y}}$ (endogenous spatial dependence);
- spatially weighted inputs of neighbour producers $W_X X$ with parameters $\beta^{(s)}$ (exogenous spatial dependence).

Also the model includes spatial effects in random disturbances ($W_v v$ with a parameter ρ_v), which expresses spatial heterogeneity. Spatial weights matrixes W_Y, W_X, and W_v are generally different, but in empirical researches are frequently put to be the same $W_Y = W_X = W_v$ subject to potential identification problems.

2.3.4. Estimation of the general spatial regression model

A problem of estimation of the general spatial model is very well researched[175], so a lot of empirical applications appears over last thirty years[46]. Nevertheless, there are many researches, where spatial effects are ignored, which can lead to estimation problems. Note that spatial dependence and spatial heterogeneity are included into the general model in different ways, so consequences of their ignorance are very different:

- Ignored spatial dependence can be considered as a classical omitted-variable problem[134]: when an important explanatory variable is missed in an econometric model, estimates of the model's parameters become biased. Thus effects of all factors, included into the model, will be over- or underestimated in presence of spatial

dependence. Estimation of the SF model suffers from this problem like any other regression model, but the bias appears not only in frontier parameters, but also in efficiency estimates, dual to them.

- Ignored spatial heterogeneity doesn't lead to a biased estimates of classical regression model parameters, but just increases their variance (make them inefficient). It directly affects the error term of models and leads to their correlation (the well-known autocorrelation problem[134]). SF models affected by spatial heterogeneity exactly in the same way, but the error term of these models includes the inefficiency component *u*, which frequently a first-priority matter of research interests. So in this case spatial heterogeneity is incorrectly included into companies' efficiency values.

A wide range of statistical methods is used for estimation of spatial model parameters. The most popular are MLE[11], [176], two-step least squares[177], and generalised method of moments[49], [178].

This research is mainly related with ML estimation, so we pay attention to this approach. Application of MLE requires definition of the distribution law for the random disturbances. The usual assumption is the normal probability distribution of $\tilde{v}$:

$$\tilde{v}_i \sim N\left(0, \sigma_{\tilde{v}}^2\right)$$

The matrix form (taking IID property of $\tilde{v}$):

$$\tilde{v} \sim MVN\left(0_n, \sigma_{\tilde{v}}^2 I_n\right)$$

The general spatial regression model (2.22) can be transformed as:

$$Y = \left(I_n - \rho_Y W_Y\right)^{-1}\left(X\beta + W_X X\beta^{(s)}\right) + \left(I_n - \rho_Y W_Y\right)^{-1}\left(I_n - \rho_v W_v\right)^{-1}\tilde{v}$$

Let assume that the matrixes ($I_n - \rho_Y W_Y$) and ($I_n - \rho_v W_v$) should be non-singular. This assumption is usual for practical applications and mathematically proved for special types of spatial weights matrixes W_Y and W_v, and parameters ρ_Y and ρ_Y; see, for example, [179] for mathematical conditions of matrixes' non-singularity.

Applying properties of the multivariate normal distribution, we have:

$$\left(I_n - \rho_Y W_Y\right)^{-1}\left(I_n - \rho_v W_v\right)^{-1}\tilde{v} \sim MVN\left(0_n, \Sigma_v\right), \tag{2.23}$$

where

$$\Sigma_v = \sigma_{\tilde{v}}^2\left(\left(I_n - \rho_Y W_Y\right)^{-1}\left(I_n - \rho_v W_v\right)^{-1}\right)^T\left(I_n - \rho_Y W_Y\right)^{-1}\left(I_n - \rho_v W_v\right)^{-1}$$

Thus the random disturbances of the model (2.22) have multivariate normal distribution, and the task model estimation adds up to estimation of parameters of the multivariate normal random variable. The log-likelihood function can be easily presented in this case:

$$\ln L = -\frac{n}{2}\ln 2\pi - n\ln\sigma_{\tilde{v}} - \frac{1}{2\sigma_{\tilde{v}}^2} e^T \Sigma_{\tilde{v}}^{-1} e + \ln\det\left(I_n - \rho_Y W_Y\right) + \ln\det\left(I_n - \rho_v W_v\right) \tag{2.24}$$

$$e = Y - \left(I_n - \rho_Y W_Y\right)^{-1}\left(X\beta + W_X X\beta^{(s)}\right)$$

The most computationally difficult component of the log-likelihood function (2.24) is determinant of ($I_n - \rho_Y W_Y$) and ($I_n - \rho_v W_v$) (called a spatial determinant). There are a number of approaches used to speed up its computation:

- based on eigenvalues (for a symmetric spatial weights matrix)[180],

  $$\ln\det\left(I_n - \rho_Y W_Y\right) = \sum_{i=1}^{n} \ln\left(1 - \rho_Y \omega_i\right)$$

  ω_i are eigenvalues of W_Y
- Cholesky or LU decomposition for sparse matrices[181],
- Chebyshev approximation[182],
- Characteristic polynomial approach[183].

There are no technical obstacles for simultaneous estimation of all spatial effects included into the general spatial regression model. But often estimated parameters cannot be analysed, because different types of effects cannot be distinguished one from another. This problem is well-known as Manski's reflection problem[184]. Identification of the parameters depends on the definition of the spatial weights matrixes. Lee[185] presented an example of spatial weights specification for which all spatial effects can be identified, but identification of parameters in the general case is not natural. Probably the best approach to solve the problem is to reduce the general model, removing some less probable spatial effects.

2.4. Review of empirical applications of SFA with spatial effects

Incorporating of spatial econometric principles into SFA is covered by a very limited number of researches. Fahr and Sunde[50] constructed the SF model to estimate efficiency of job creation in UK regions. The authors included a spatial lag of unemployment (weighted number of persons, unemployed in neighbour regions) into explanatory variables, and discovered its significant influence on hiring in a given region. Inclusion of spatial lags of input variables ($\boldsymbol{W_X X}$ in the general model (2.22)) doesn't lead to estimation difficulties; one of developed estimators can be applied. Disregarding other spatial effects in the model, conclusions about influence of input spatial lags are arguable.

Barrios [51] was the first (to the best of our knowledge) to embed the output spatial lag ($W_y y$) into the SF model:

$$Y = \rho_Y W_Y Y + X\beta + W_X X\beta^{(s)} + v - u. \tag{2.25}$$

Parameter estimation of this model is affected by a well-known endogeneity problem – weighted outputs are included into explanatory variables, but obviously correlated with the random disturbances. In presence of endogeneity the ordinary least squares estimator provides biased parameter estimates and shouldn't be applied. Barrios suggested a backfitting estimation algorithm, similar to Cochrane-Orcutt's iterative procedure[186]. The developed model included both input and output spatial lags and was applied to data on rural household production. Later the model and the estimation algorithm were extended by the author for a case of panel data[187], [188]. Affuso[53] applied the SF model with output spatial lags to evaluate an agricultural extension project in Tanzania. Unlike Barrios, Affuso used a derived MLE for the model. Following Affuso, the author of this thesis applied[102] similar specification of the model and MLE to airport benchmarking.

Spatial heterogeneity in SF models is also covered by a number of recent researches. Druska and Horrace[49] embedded spatial dependence into the symmetrical error term (***v***) of the SF model:

$$v = \rho_v W_v v + \tilde{v}. \tag{2.26}$$

The authors didn't make any assumptions about distributions of the model's random terms and derived a generalised MOM estimator of model parameters. The suggested estimator requires panel data and generally based on the assumption of time-invariant inefficiency. Authors applied the developed method to a panel of Indonesian rice farms, and discovered significant spatial heterogeneity in production.

Lin et al.[54], [55] also used the same specification of the SF model and suggested a MLE both for cross-sectional and panel settings. The author of this thesis[189] derived a similar MLE and applied it to a data set of European airports. In that research distinguishing inefficiency from spatial heterogeneity was also modelled with Monte-Carlo simulation and discussed. Recently a quasi-maximum likelihood estimator for the same model specification was presented by Simwaka [190], but disregarding frontier-specific error term structure.

Glass et al.[59] used this specification of the model (together with spatial dependency in production outputs) for panel data. Derived MLE was applied to analysis of a country-level translog production function.

In addition to the random disturbances *v* the SF model includes an inefficiency term *u*, which also can be a subject of spatial dependence.

Observed spatial heterogeneity can be included into inefficiency specification as any other explanatory variables associated with inefficiency[191]:

$$u_i = z_i\delta + \tilde{u}_i\,, \tag{2.27}$$

where

z_i is a vector of explanatory variables associated with inefficiency of the producer i;

δ is a vector of unknown coefficients;

$\tilde{u}_i$ is a random variable, truncated at the point $-z_i\delta$, so $\tilde{u}_i \geq -z_i\delta$.

Definition of explanatory variables $\boldsymbol{z}$ can include regional dummies, distances, area-specific and other factors, related with observed spatial structure. Also spatial lags of inputs can be included into this vector. For example, Barrios and Lavado[51], [187] utilised this approach to investigate influence of neighbour farm incomes on efficiency of a given farm in Philippines. Igliori[155] used spatially weighted road infrastructure and educational characteristics to explain inefficiency of agricultural production in the Brazilian Amazon (but didn't discover significant spatial effects of this type). Later Schettini et al.[52] included a spatial lag of employment into efficiency determinants of regional production in Brazil and found its significant influence in some industrial sectors.

Frequently for estimation purposes a known distribution is assumed for the inefficiency term. In this case the spatial dependence can be associated with these distribution parameters. For example, for the truncated normal distribution $u_i \sim TN_{0,+\infty}\left(\mu, \sigma_u^2\right)$, unobserved spatial heterogeneity can be included directly into parameters μ and σ_u. Schmidt et al.[192] suggested conditional autoregressive dependence for the mean parameter μ and developed a Bayes estimator for this case. The proposed model was applied to a data set of Brazilian farms.

A classical spatial lags structure also can be applied to the inefficiency term:

$$u = \rho_u W_u u + \tilde{u}, \tag{2.28}$$

where

W_u is a spatial weights matrix,

ρ_u is an unknown parameter of spatial dependence.

Areal et al.[56] utilised this specification, derived a Bayes estimator for the model and applied it to a sample of dairy farms in England and Wales. The parameter of inefficiency spatial dependence ρ_u is found statistically significant in all considered models. Tonini and Pede[57] derived a GME estimator for a similar model specification and also discovered significant spatial dependency in agricultural productivity of European countries.

Fusco and Vidoli[60] also used this specification of the inefficiency term for analysis of spatial heterogeneity in agricultural sector in Italy. The model was estimated using the derived MLE.

Another specification of spatial inefficiency was recently proposed by Mastromarco et al. [58]. Company's inefficiency is supposed to be explained by its spatial lag in the previous point of time, and a "distance" between producers is defined as a difference of their previous inefficiency values. The approach was applied to macroeconomic productivity of OECD countries and the authors discovered significant spatial spillovers.

It should be noted that estimation is an important technical problem for all studies, where spatial effects are included into the SF model. Generally, a researcher, who suggests an approach to integrating spatial effects into the SF model, has to develop a software tool for its empirical application. Obviously that absence of a unified tool is a great obstacle for empirical researches in this area.

2.5. Conclusions

In this chapter we presented an overview of production theory basic concepts and the stochastic frontier analysis as a comprehensive tool for production modelling. A special attention was paid to integrating of spatial relationships into the stochastic frontier model.

Mathematical formalisation for the problem of estimation of production possibility frontier parameters and technical efficiency is stated. Single- and multi-output production processes are discussed and their representation in a form of an econometric model is presented. We also attended a problem of econometric estimation of the stochastic frontier model parameters.

We discussed a problem of integrating of spatial dependencies into econometric models, presented approaches based on observed and unobserved spatial components. Special attention was devoted to spatial econometrics, an extensive treatment for analysis of spatial relationships. Theoretical and empirical researches on integrating spatial effects into the stochastic frontier model were analysed. Based on the analysis, the following conclusions were made:

1. Despite the fact that the importance of spatial relationships for the stochastic frontier analysis is widely acknowledged in literature, number of researches, where spatial effects are included into consideration, is very limited. Mainly researchers ignore the presence of spatial effects or include them in an observed form only (via regional dummy variables, distances or observed location-specific conditions).
2. Theories of stochastic frontier analysis and spatial econometrics are very well developed, but there are almost no systematic researches on merging their principles.
3. There is no general formulation of the stochastic frontier model with different types of spatial effects (spatial dependence, spatial heterogeneity). This leads a significant number of private-case models, formulated and estimated by different researchers.

4. As a consequence of the previous conclusion, there are no unified software tools for analysis of stochastic frontier models with spatial components. Researchers in this area have to implement their own algorithms in a form of software packages, rarely available to the public for further usage.

Following the presented conclusions, a task of formulation of a general stochastic frontier model with spatial effects and development of methods for its parameters estimation can be considered as an important research target. Also development of a public software package for estimation of a stochastic frontier model with spatial effects seems to be empirically meaningful.

3. SPATIAL STOCHASTIC FRONTIER (SSF) MODEL AND ITS PARAMETERS ESTIMATION

3.1. Formal specification of the proposed SSF model

Existence of spatial interactions is widely acknowledged in different sciences: economics, social science, regional sciences, biology, chemistry and others. For example, production of a company may be affected by production of its competitors, acting on the same market; purchases of a customer may depend on his neighbourhood and social interactions; air pollution in a specific region may be affected by activities in neighbour regions.

In the context of the SF model, described in the Chapter 2, we specify a hypothesis about existence of the following four types of spatial effects:

Type 1. Endogenous spatial effects

Type 2. Exogenous spatial effects

Type 3. Spatially correlated random disturbances

Type 4. Spatially related efficiency

Note that first three types of spatial effects are well known[153] in spatial econometrics, but spatial effects in efficiency is a relative novelty.

Endogenous spatial effects represent a relationship between outputs (or, more generally, decisions) of a company and outputs of its neighbours. Existence of these effects is well-grounded and supported by theories in different science areas:

- In economic non-cooperative games[193], a strategy of an individual depends on behaviour of other game participants. Generally, a solution of non-cooperative games is described in terms of equilibriums, where an output of an agent is determined by a joint production of its neighbours. For example, in oligopoly models this relationship is introduced in a form of a reaction function.
- In customer demand models[194], consumption of a particular customer depends on demand of other customers in a reference group.
- Spatial effects between individuals play a central role in sociology and social psychology; interactions between an individual and his neighbourhood are supposed to be the main factor, affecting his decisions and behaviour.
- In ecology, spatial spillovers of activity in a region are generally admitted. For example, air pollution in a particular area is affected by human and natural activity in its neighbourhood.

Exogenous spatial effects represent a relationship between an output of a company and inputs (resources) of its neighbours. These effects can be explained by indirect flow of resources into neighbourhood, where the production process and output registration can be separated in space. Let consider a simple example from regional economics, where a general volume of customer spending in a particular region (output) is defined by an average income this region (input). In case of economic integration of regions, this is very likely that a significant share of customer earns their income in one region and spends them in a neighbour region. For example, this behaviour is intrinsic for labour market of capital city and its outskirts or, on a higher level, for central and provincial districts and countries. Theories of migration and regional convergence are also based on exogenous spatial effects, supposing that resources flow from less attractive regions with more attractive. Similarly, exogenous spatial effects may be observed in biology (population in a particular region may depend on explanatory factors in its neighbourhood due to migration, etc.).

The third type of spatial effects, spatially correlated random disturbances, isn't based on a theoretical model, but usually is consistent with the modelling theory. Suppose a model where observations are affected by an unobserved factor, which has a spatial nature. For example, a theory of hedonic prices on real estate market states the influence of many area specific factors of house prices – air and water pollution, noise, aesthetic sights and closeness to natural attractors, a subjective sense of security and others. Also environment plays an important role in a theory of production. For example, production of agricultural farms depends on weather conditions, plat pests and other factors. Some of these factors can be unobservable from its nature; some of them are just not available in a research sample. Obvious 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 [49], [56], [57], [60], [187], among a few others. From the practical point of view, distinguishing between inefficiency and heterogeneity of the production frontier is quite challenging[161], [195]. Usually a negative impact of factors under company control is considered as inefficiency, while a negative or positive impact of factors outside of company control is interpreted as heterogeneity. Thus the reasoning of spatially related efficiency is very similar to grounds of endogenous spatial effects, matter of control and impact direction. Possible reasons of spatially related efficiency contain the following points:

- An agent may emulate behaviour of neighbour agents, including inefficient components. For example, a company may reproduce a production process of other companies in this

area due to shared professionals; an individual may copy behaviour patterns of his colleagues; regional and state governments may adopt similar laws and practices.

- Local policies and other market power restrictions lead to weaker competition pressure, affecting companies, and indirectly decrease efficiency of all companies in an area.
- Neighbour companies use the same infrastructure and labour resources and may suffer from similar problems. These effects can be specifies as inefficiencies if they can be considered as controlled by a company. For example, a level of staff education can be low in a particular area, but generally can be improved by company management.

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

$$Y_i = \rho_Y \sum_{j=1}^{n} w_{Y,ij} Y_j + \sum_{k=1}^{K} X_{ki}\beta_k + \sum_{k=1}^{K}\left(\beta_k^{(s)} \sum_{j=1}^{n} w_{X,ij} X_{kj}\right) + v_i - u_i, \tag{3.1}$$

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

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

where

i is a company index, $i = 1,\ldots, n$,

Y_i is an output of a company i,

X_{ki} are inputs of a company i, $k = 1,\ldots,K$,

$w_{Y,ij}$ are spatial weights for spatial endogenous effects between companies i and j, $j = 1,\ldots, n$,

$w_{X,ij}$ are spatial weights for spatial exogenous effects between companies i and j, $j = 1,\ldots, n$,

$w_{v,ij}$ are spatial weights for spatially correlated random disturbances of companies i and j,

$w_{u,ij}$ are spatial weights for spatially related efficiency of companies i and j, $j = 1,\ldots, n$,

v_i is a random disturbance term,

u_i is an inefficiency of a company i,

β_k are coefficients, representing direct effects of inputs, $k = 1,\ldots,K$,

$\beta^{(s)}_k$ are coefficients, representing spatial exogenous effects of inputs, $k = 1,\ldots,K$,

ρ_Y is a coefficient, representing spatial endogenous effects,

ρ_v is a coefficient, representing spatially correlated random disturbances,

ρ_u is a coefficient, representing spatially related efficiency,

$\tilde{v}_i$ are independent identically distributed (IID) random disturbances,

$\tilde{u}_i$ are IID inefficiency levels.

Folding the model by *i*, *j*, and *k*, we formulate the model in the matrix form:

$$\begin{aligned} &Y = \rho_Y W_Y Y + X\beta + W_X X\beta^{(s)} + v - u, \\ &v = \rho_v W_v v + \tilde{v}, \\ &u = \rho_u W_u u + \tilde{u}. \end{aligned} \tag{3.2}$$

Utilising a usual property of non-singularity of $(I_n - \rho_Y W_Y)$, $(I_n - \rho_v W_v)$, and $(I_n - \rho_u W_u)$ matrixes, spatial operators can be introduced:

$$\begin{aligned} &S_Y = S(\rho_Y, W_Y) = (I_n - \rho_Y W_Y)^{-1}, \\ &S_v = S(\rho_v, W_v) = (I_n - \rho_v W_v)^{-1}, \\ &S_u = S(\rho_u, W_u) = (I_n - \rho_u W_u)^{-1}. \end{aligned} \tag{3.3}$$

Using spatial operators, the model's error components can be presented as:

$$\begin{aligned} &v = (I_n - \rho_v W_v)^{-1}\tilde{v} = S_v \tilde{v}, \\ &u = (I_n - \rho_u W_u)^{-1}\tilde{v} = S_u \tilde{u}, \end{aligned} \tag{3.4}$$

and finally the model can be formulated as

$$Y = S_Y\left(X\beta + W_X X\beta^{(s)} + S_v\tilde{v} - S_u\tilde{u}\right). \tag{3.5}$$

Note that the presented model specification can be generalised in different ways:

- The production frontier function can be included into the model in a non-linear form. In this research we consider only Cobb-Douglas and translog forms of the production frontier, which can be easily linearised (2.12).
- The specification includes spatial dependence of the first order, so only direct spatial effects between two companies are considered. Indirect (higher-order) spatial effects, which represent relationships between two companies via intermediate neighbours, are not included into the model.
- All spatial components are included in a form of spatial lags (3.3) (an autoregressive form). More general specifications with autoregressive and moving average (ARMA) terms are also possible, but not used in this research. Spatial ARMA process[196] represents a highly complicated spatial pattern and rarely used in practice. Note that the first-order autoregressive process AR(1) can be represented in the following form (using an expansion of the spatial operator into infinite series):

  $$v = (I_n - \rho_v W_v)^{-1}\tilde{v} = \left(I_n + \rho_v W_v + \rho_v^2 W_v^2 + \rho_v^3 W_v^3 + \ldots\right)\tilde{v} = \tilde{v} + \rho_v W_v \tilde{v} + \rho_v^2 W_v^2 \tilde{v} + \rho_v^3 W_v^3 \tilde{v} + \ldots,$$

  which represents the moving average process MA(∞). This invertibility property of AR and MA processes is well-known in time series analysis[134].

Considering possible ways of model generalisation, the model (3.2) can be referenced as a linear stochastic frontier model with first-order spatially autoregressive dependent variable, explanatory variables, random disturbances, and inefficiency terms. The model will be referred as the **SSF(1,1,1,1) model**, where SSF is used for spatial stochastic frontier, and parameters in brackets represent orders of spatial autoregressive terms in a dependent variable, explanatory variables, random disturbances, and inefficiency terms respectively.

A mainstream approach to specification of spatial econometric models is stepwise-forward, which starts with no spatial effects in the model and further extends the model with appropriate spatial effects. Thus, there are a number of restricted specifications of the SSF(1,1,1,1) model, which can be used in practice. A list of useful restricted specifications is presented below:

The SSF(0,0,0,0) gives a classical stochastic frontier model without spatial effects:

$$Y = X\beta + \tilde{v} - \tilde{u}. \tag{3.6}$$

The SSF(1,0,0,0) model:

$$Y = \rho_Y W_Y Y + X\beta + \tilde{v} - \tilde{u}. \tag{3.7}$$

The SSF(1,1,0,0) (spatial Durbin) model:

$$Y = \rho_Y W_Y Y + X\beta + W_X X\beta^{(s)} + \tilde{v} - \tilde{u}. \tag{3.8}$$

The SSF(0,0,0,1) model:

$$\begin{aligned} Y &= X\beta + \tilde{v} - u, \\ u &= \rho_u W_u u + \tilde{u}. \end{aligned} \tag{3.9}$$

The SSF(0,0,1,0) model:

$$\begin{aligned} Y &= X\beta + v - \tilde{u}, \\ v &= \rho_v W_v v + \tilde{v}. \end{aligned} \tag{3.10}$$

Selection of an appropriate model specification is usually implemented on the base of statistical tests, but also can be enhanced by knowledge of the domain area. Some types of spatial effects are hardly probable in specific spatial settings. For example, a model of production that includes technical provisions of a company (a number of machines, working area, etc.) as explanatory variables, unlikely contains exogenous spatial effects.

3.2. Derivation of estimator of the SSF model parameters

A set of methods, used to estimation of a classical stochastic frontier model, includes MLE, MOM, GME estimators, and Bayesian estimator. Objectives of the estimators include estimation of the production frontier parameters β and values of the technical inefficiency u_i for each

company in the sample. The SSF(1,1,1,1) model also requires estimation of coefficients ρ_Y, $\beta^{(s)}$, ρ_v, and ρ_u for spatial effects of four types.

3.2.1. MLE for the SSF model parameters

Classical estimators, discussed in the paragraphs 2.2 and 2.3.4, provide consistent parameters' estimates in case of absence of spatial effects in real data generating process (DGP). Presence of spatial effects in a DGP leads to unfitness of these estimators. Endogenous and exogenous spatial effects and spatially related efficiency (types 1, 2, and 4) in a DGP result in biased and inconsistent estimates both for the production frontier parameters and inefficiency values. Spatially correlated random disturbances (type 3) lead to inefficient estimates of the production frontier parameters and inconsistent estimates for individual inefficiency values. Taking that technical efficiency is usually the main objective of SFA, we can conclude that presence of spatial effects of any types doesn't allow using classical estimators. Thus a specialised estimator should be developed and applied in this case. In this research we develop a maximum likelihood estimator for the SSF model.

MLE requires additional assumptions about distributions of the random terms. We consider the following assumptions:

1. $\tilde{v}_i$ are IID normal with zero mean, $\tilde{v}_i \sim N\left(0, \sigma_{\tilde{v}}^2\right)$,
2. $\tilde{u}_i$ are IID non-negative truncated normal, $\tilde{u}_i \sim TN_{0,+\infty}\left(\mu, \sigma_{\tilde{u}}^2\right)$,
3. $\tilde{v}_i$ and $\tilde{u}_i$ are distributed independently of each other, and of the explanatory variables.

The Assumption 1 is conventional for econometric models and can be presented in the matrix form:

$$\tilde{v} \sim MVN\left(0_n, \sigma_{\tilde{v}}^2 I_n\right).$$

The Assumption 2 is usual for stochastic frontier models[145] and frequently used in practice. Another popular assumption, a half-normal distribution of inefficiency[32], is a private case of the truncated normal with $\mu = 0$.The matrix form of this assumption is:

$$\tilde{u} \sim MVTN_{0,+\infty}\left(\mu, \sigma_{\tilde{u}}^2 I_n\right).$$

Note that parameters μ and $\sigma_{\tilde{u}}^2$ are related to the mean and variance of a normal distribution before truncation.

The Assumption 3, especially the statement of independence from explanatory variables (inputs), looks problematic. Generally, if a company has information about its inefficiency, it can update its production process and change inputs values. In this research we accept this assumption as a matter of simplification.

This is well-known that a linear transformation of a multivariate normal random vector has a multivariate normal distribution with transformed parameters. Using that

$$v = S_v \tilde{v},$$
$$u = S_u \tilde{u},$$

we obtain the distribution of v and u as:

$$v \sim MVN(0_n, \Sigma_v), \tag{3.11}$$
$$\Sigma_v = \sigma^2_{\tilde{v}} S_v S_v^T$$
$$u \sim MVTN_{0,+\infty}(\mu, \Sigma_u),$$
$$\Sigma_u = \sigma^2_{\tilde{u}} S_u S_u^T$$

The SSF model can be presented in the form

$$Y = \rho_Y W_Y Y + X\beta + W_X X\beta^{(s)} + \varepsilon, \tag{3.12}$$

where $\varepsilon = v - u$ is a composed error term. The distribution of ε is derived in the following theorem.

Theorem 1.

Let we have two independent multivariate random variables:

- $v = (v_1, v_2, \ldots, v_n)$ with the multivariate (n-variate) normal distribution with a zero mean and a covariance matrix Σ_v,
  $v \sim MVN(0_n, \Sigma_v)$
- $u = (u_1, u_2, \ldots, u_n)$ with the multivariate (n-variate) truncated normal distribution with a mean μ and a covariance matrix Σ_u and $(0, +\infty)$ truncation interval
  $u \sim MVTN_{0,+\infty}(\mu, \Sigma_u)$

Then an n-variate random variable $\varepsilon = v - u$ has the closed skew normal (CSN) distribution

$$\varepsilon \sim CSN_{n,n}(\mu', \Sigma', \Gamma', \nu', \Delta'), \tag{3.13}$$

where

$$\mu' = -\mu,$$
$$\Sigma' = \Sigma_v + \Sigma_u,$$
$$\Gamma' = -\Sigma_u (\Sigma_v + \Sigma_u)^{-1},$$
$$\nu' = -\mu,$$
$$\Delta' = \left(\Sigma_v^{-1} + \Sigma_u^{-1}\right)^{-1}.$$

and the probability density function of ε is:

$$f_\varepsilon(\varepsilon)=[\Phi_n(0,-\mu,\Sigma_u)]^{-1}\Phi_n\left(-\Sigma_u(\Sigma_v+\Sigma_u)^{-1}(\varepsilon+\mu),-\mu,\left(\Sigma_v^{-1}+\Sigma_u^{-1}\right)^{-1}\right)\varphi_n(\varepsilon,-\mu,\Sigma_v+\Sigma_u), \quad (3.14)$$

where

φ_n is the standard MVN probability density function,

Φ_n is the standard MVN cumulative distribution function.

Proof of the Theorem 1.

Firstly, we proof that a random variable with the closed skew normal distribution and parameters μ',Σ',Γ',υ',Δ', defined in the Theorem's proposition, has the specified probability density function. Next the probability of the composed random variable $\varepsilon = v - u$ will be constructed, following the procedure from [31].

The closed skew normal probability density function is [197]:

$$x \sim CSN_{n,n}(\mu',\Sigma',\Gamma',\nu',\Delta'), \quad (3.15)$$
$$f_x(x)=\left[\Phi_n\left(0,\nu',\Delta'+\Gamma'\Sigma'\Gamma'^T\right)\right]^{-1}\Phi_n(\Gamma'(x-\mu'),\nu',\Delta')\varphi_n(x,\mu',\Sigma')$$

Putting the parameters, defined in the theorem proposition as:

$$\mu'=-\mu,$$
$$\Sigma'=\Sigma_v+\Sigma_u,$$
$$\Gamma'=-\Sigma_u(\Sigma_v+\Sigma_u)^{-1},$$
$$\nu'=-\mu,$$
$$\Delta'=\left(\Sigma_v^{-1}+\Sigma_u^{-1}\right)^{-1}.$$

and taken that

$$\Delta'+\Gamma'\Sigma'\Gamma'^T=\Delta'+\Gamma'\Sigma'\Gamma'=\left(\Sigma_v^{-1}+\Sigma_u^{-1}\right)^{-1}+\Sigma_u(\Sigma_v+\Sigma_u)^{-1}(\Sigma_v+\Sigma_u)\Sigma_u(\Sigma_v+\Sigma_u)^{-1}=$$
$$=\left(\Sigma_v^{-1}+\Sigma_u^{-1}\right)^{-1}+\Sigma_u(\Sigma_v+\Sigma_u)^{-1}\Sigma_u=\Sigma_v(\Sigma_v+\Sigma_u)^{-1}\Sigma_u+\Sigma_u(\Sigma_v+\Sigma_u)^{-1}\Sigma_u=$$
$$=(\Sigma_v+\Sigma_u)(\Sigma_v+\Sigma_u)^{-1}\Sigma_u=\Sigma_u,$$

the probability density function of ε is:

$$f_\varepsilon(\varepsilon)=[\Phi_n(0,-\mu,\Sigma_u)]^{-1}\cdot\Phi_n\left(-\Sigma_u(\Sigma_v+\Sigma_u)^{-1}(\varepsilon+\mu),-\mu,\left(\Sigma_v^{-1}+\Sigma_u^{-1}\right)^{-1}\right)\cdot \quad (3.16)$$
$$\cdot\varphi_n(\varepsilon,-\mu,\Sigma_v+\Sigma_u)$$

This probability density function exactly matches the form, stated in the Theorem.

Multivariate normal and truncated normal probability density functions have the following forms.

The multivariate normal probability density function:

$$v \sim MVN(0_n, \Sigma_v) \tag{3.17}$$
$$f_v(v) = \varphi_n(v, 0, \Sigma_v) = \frac{1}{(2\pi)^{n/2}} \cdot (\det(\Sigma_v))^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2} v^T \Sigma_v^{-1} v\right\}$$

The multivariate truncated normal probability density function[198]:

$$u \sim MVTN_{0,+\infty}(\mu, \Sigma_u) \tag{3.18}$$
$$f_u(u) = [\Phi_n(0, -\mu, \Sigma_u)]^{-1} \varphi_n(u, \mu, \Sigma_u),$$

Given the independence assumption, the joint density function of *v* and *u* is the product of their individual density functions:

$$f_{uv}(u, v) = f_u(u) f_v(v) = [\Phi_n(0, -\mu, \Sigma_u)]^{-1} \varphi_n(u, \mu, \Sigma_u) \varphi_n(v, 0, \Sigma_v) \tag{3.19}$$

By the Theorem's statement $v = \varepsilon + u$, so:

$$f_{u\varepsilon}(u, \varepsilon) = f_u(u) f_v(\varepsilon + u) = [\Phi_n(0, -\mu, \Sigma_u)]^{-1} \varphi_n(u, \mu, \Sigma_u) \varphi_n(\varepsilon + u, 0, \Sigma_v) \tag{3.20}$$

$$\begin{aligned} f_{u\varepsilon}(u, \varepsilon) &= [\Phi_n(0, -\mu, \Sigma_u)]^{-1} \cdot \\ &\cdot \frac{1}{(2\pi)^{n/2}} \cdot (\det(\Sigma_u))^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}(u-\mu)^T \Sigma_u^{-1} (u-\mu)\right\} \cdot \\ &\cdot \frac{1}{(2\pi)^{n/2}} \cdot (\det(\Sigma_v))^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}(\varepsilon+u)^T \Sigma_v^{-1} (\varepsilon+u)\right\} = \\ &= [\Phi_n(0, -\mu, \Sigma_u)]^{-1} (2\pi)^{-n} (\det(\Sigma_u))^{-\frac{1}{2}} (\det(\Sigma_v))^{-\frac{1}{2}} \cdot \\ &\cdot \exp\left\{-\frac{1}{2}\left((u-\mu)^T \Sigma_u^{-1} (u-\mu) - (\varepsilon+u)^T \Sigma_v^{-1} (\varepsilon+u)\right)\right\} \end{aligned} \tag{3.21}$$

The power of the exponent can be regrouped as:

$$\begin{aligned} &(u-\mu)^T \Sigma_u^{-1} (u-\mu) - (\varepsilon+u)^T \Sigma_v^{-1} (\varepsilon+u) = \\ &= u^T \Sigma_u^{-1} u - 2u^T \Sigma_u^{-1} \mu + \mu^T \Sigma_u^{-1} \mu + \varepsilon^T \Sigma_v^{-1} \varepsilon + 2u^T \Sigma_v^{-1} \varepsilon + u^T \Sigma_v^{-1} u = \\ &= u^T \left(\Sigma_v^{-1} + \Sigma_u^{-1}\right) u + 2u^T \left(\Sigma_v^{-1} \varepsilon - \Sigma_u^{-1} \mu\right) + \mu^T \Sigma_u^{-1} \mu + \varepsilon^T \Sigma_v^{-1} \varepsilon = \\ &= u^T \left(\Sigma_v^{-1} + \Sigma_u^{-1}\right) u + 2u^T \left(\Sigma_v^{-1} + \Sigma_u^{-1}\right)\left(\Sigma_v^{-1} \varepsilon - \Sigma_u^{-1} \mu\right)\left(\Sigma_v^{-1} + \Sigma_u^{-1}\right)^{-1} + \mu^T \Sigma_u^{-1} \mu + \varepsilon^T \Sigma_v^{-1} \varepsilon \end{aligned} \tag{3.22}$$

Following [31], we transform this expression to the quadratic form as:

$$(u + A\varepsilon + D\mu)^T B^{-1} (u + A\varepsilon + D\mu) + (\varepsilon + \mu)^T C^{-1} (\varepsilon + \mu), \tag{3.23}$$

where *A*, *B*, *C*, *D* are matrixes.

The presented quadratic form can be regrouped as:

$$
\begin{aligned}
&(u+A\varepsilon+D\mu)^T B^{-1}(u+A\varepsilon+D\mu)+(\varepsilon+\mu)^T C^{-1}(\varepsilon+\mu)= \\
&=u^T B^{-1}u+2u^T B^{-1}A\varepsilon+2u^T B^{-1}D\mu+(D\mu)^T B^{-1}D\mu+ \\
&+\varepsilon^T\left(A^T B^{-1}A+C^{-1}\right)\varepsilon+2\varepsilon^T A^T B^{-1}D\mu+2\varepsilon^T C^{-1}\mu+\mu^T C^{-1}\mu= \\
&=u^T B^{-1}u+2u^T B^{-1}A\varepsilon+2u^T B^{-1}D\mu+\mu^T\left(D^T B^{-1}D+C^{-1}\right)\mu+ \\
&+2\varepsilon^T\left(A^T B^{-1}D+C^{-1}\right)\mu+\varepsilon^T\left(A^T B^{-1}A+C^{-1}\right)\varepsilon
\end{aligned}
\tag{3.24}
$$

Equating corresponding terms of equations (3.34) and (3.36), we construct a system of equations:

$$
\begin{cases}
\Sigma_v^{-1}+\Sigma_u^{-1}=B^{-1}, \\
\Sigma_v^{-1}=B^{-1}A, \\
B^{-1}D=-\Sigma_u^{-1}, \\
D^T B^{-1}D+C^{-1}=\Sigma_u^{-1}, \\
A^T B^{-1}D+C^{-1}=0, \\
A^T B^{-1}A+C^{-1}=\Sigma_v^{-1}.
\end{cases}
$$

Using straightforward methods of matrix algebra, the system is solved for *A*, *B*, *C*, *D*:

$$
\begin{cases}
A=B\Sigma_v^{-1}, \\
B=\left(\Sigma_v^{-1}+\Sigma_u^{-1}\right)^{-1}, \\
C=\Sigma_v+\Sigma_u, \\
D=-B\Sigma_u^{-1}.
\end{cases}
\tag{3.25}
$$

So the joint probability density function (3.35) is:

$$
\begin{aligned}
&f_{u\varepsilon}(u,\varepsilon)=\left[\Phi_n(0,-\mu,\Sigma_u)\right]^{-1}(2\pi)^{-n}\left(\det(\Sigma_u)\right)^{-\frac{1}{2}}\left(\det(\Sigma_v)\right)^{-\frac{1}{2}}\cdot \\
&\cdot\exp\left\{-\frac{1}{2}\left((u+A\varepsilon+D\mu)^T B^{-1}(u+A\varepsilon+D\mu)+(\varepsilon+\mu)^T C^{-1}(\varepsilon+\mu)\right)\right\}= \\
&=\left[\Phi_n(0,-\mu,\Sigma_u)\right]^{-1}(2\pi)^{-n}\left(\det(\Sigma_u)\right)^{-\frac{1}{2}}\left(\det(\Sigma_v)\right)^{-\frac{1}{2}}\cdot \\
&\cdot\exp\left\{-\frac{1}{2}(u+A\varepsilon+D\mu)^T B^{-1}(u+A\varepsilon+D\mu)\right\}\exp\left\{-\frac{1}{2}(\varepsilon+\mu)^T C^{-1}(\varepsilon+\mu)\right\}= \\
&=\left[\Phi_n(0,-\mu,\Sigma_u)\right]^{-1}\varphi_n(u+A\varepsilon+D\mu,0,B)\varphi_n(\varepsilon,-\mu,C)
\end{aligned}
\tag{3.26}
$$

Replacing *A*, *B*, *C*, *D* with their expressions (3.25) and taken that

$$
\begin{aligned}
&A\varepsilon+D\mu=\left(\Sigma_v^{-1}+\Sigma_u^{-1}\right)^{-1}\Sigma_v^{-1}\varepsilon-\left(\Sigma_v^{-1}+\Sigma_u^{-1}\right)^{-1}\Sigma_u^{-1}\mu=\Sigma_u(\Sigma_v+\Sigma_u)^{-1}\varepsilon-\Sigma_v(\Sigma_v+\Sigma_u)^{-1}\mu= \\
&=\Sigma_u(\Sigma_v+\Sigma_u)^{-1}\varepsilon+\Sigma_u(\Sigma_v+\Sigma_u)^{-1}\mu-\Sigma_u(\Sigma_v+\Sigma_u)^{-1}\mu-\Sigma_v(\Sigma_v+\Sigma_u)^{-1}\mu= \\
&=\Sigma_u(\Sigma_v+\Sigma_u)^{-1}(\varepsilon+\mu)-(\Sigma_v+\Sigma_u)^{-1}(\Sigma_v+\Sigma_u)\mu= \\
&=\Sigma_u(\Sigma_v+\Sigma_u)^{-1}(\varepsilon+\mu)-\mu,
\end{aligned}
$$

the joint probability density function is

$$f_{u\varepsilon}(u,\varepsilon)=[\Phi_n(0,-\mu,\Sigma_u)]^{-1}\varphi_n\left(\Sigma_u(\Sigma_v+\Sigma_u)^{-1}(\varepsilon+\mu),\mu,\left(\Sigma_v^{-1}+\Sigma_u^{-1}\right)^{-1}\right)\varphi_n(\varepsilon,-\mu,\Sigma_v+\Sigma_u) \quad (3.27)$$

The marginal density function of ε is obtained by integrating u out of the joint probability density function, which yields

$$f_\varepsilon(\varepsilon)=\int_{-\infty}^{+\infty} f_{u\varepsilon}(u,\varepsilon)du=$$

$$=\int_0^{+\infty}[\Phi_n(0,-\mu,\Sigma_u)]^{-1}\varphi_n\left(\Sigma_u(\Sigma_v+\Sigma_u)^{-1}(\varepsilon+\mu),\mu,\left(\Sigma_v^{-1}+\Sigma_u^{-1}\right)^{-1}\right)\varphi_n(\varepsilon,-\mu,\Sigma_v+\Sigma_u)du=$$

$$=[\Phi_n(0,-\mu,\Sigma_u)]^{-1}\varphi_n(\varepsilon,-\mu,\Sigma_v+\Sigma_u)\int_0^{+\infty}\varphi_n\left(\Sigma_u(\Sigma_v+\Sigma_u)^{-1}(\varepsilon+\mu),\mu,\left(\Sigma_v^{-1}+\Sigma_u^{-1}\right)^{-1}\right)du=$$

$$=[\Phi_n(0,-\mu,\Sigma_u)]^{-1}\Phi_n\left(-\Sigma_u(\Sigma_v+\Sigma_u)^{-1}(\varepsilon+\mu),-\mu,\left(\Sigma_v^{-1}+\Sigma_u^{-1}\right)^{-1}\right)\varphi_n(\varepsilon,-\mu,\Sigma_v+\Sigma_u)$$

So the probability density function of ε:

$$f_\varepsilon(\varepsilon)=[\Phi_n(0,\mu,\Sigma_u)]^{-1}\Phi_n\left(-\Sigma_u(\Sigma_v+\Sigma_u)^{-1}\varepsilon,-\mu,\left(\Sigma_v^{-1}+\Sigma_u^{-1}\right)^{-1}\right)\varphi_n(\varepsilon,-\mu,\Sigma_v+\Sigma_u) \quad (3.28)$$

This function exactly match (3.14), so the Theorem is proved completely.

End of the Proof.

The Theorem 1 states that the composed error term of the SSF model has the closed skew normal distribution[197] $CSN_{n,n}$ with the specified parameters. Note that the derived probability distribution function is reduced to (2.17), when random disturbances and inefficiencies are independent and identically distributed.

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:

$$\begin{aligned}
&\ln L\left(\beta,\beta^{(s)},\sigma_{\tilde{v}}^2,\sigma_{\tilde{u}}^2,\mu,\rho_Y,\rho_v,\rho_u\right)=-\ln\Phi_n(0,-\mu,\Sigma_u)+ \quad (3.29)\\
&+\ln\Phi_n\left(-\Sigma_u(\Sigma_v+\Sigma_u)^{-1}(e+\mu),-\mu,\left(\Sigma_v^{-1}+\Sigma_u^{-1}\right)^{-1}\right)+\ln\varphi_n(e,-\mu,\Sigma_v+\Sigma_u),\\
&e=Y-\rho_Y W_Y Y-X\beta-W_X X\beta^{(s)},\\
&\Sigma_v=\sigma_{\tilde{v}}^2\left((I_n-\rho_v W_v)^{-1}\right)^T(I_n-\rho_v W_v)^{-1},\\
&\Sigma_u=\sigma_{\tilde{u}}^2,\left((I_n-\rho_u W_u)^{-1}\right)^T(I_n-\rho_u W_u)^{-1}.
\end{aligned}$$

The log-likelihood function is maximised to obtain consistent maximum likelihood estimates for all parameters.

The main problem with maximisation of the log-likelihood function is related with calculation of the multivariate normal cumulative distribution function Φ_n. This function has no

analytical representation as well as no analytical gradients. Although numeric methods helps in this case, but maximisation takes a long time and hardly can be used in practice. Another important problem with the MLE relates with the fact[199] that with non-zero probability, the maximum likelihood estimates are not converged even in the univariate case of simple skew normal distribution. The problem is softened by a special kind of reparametrisation and penalisation of the likelihood function, but it can grow in the multivariate case of the more complex closed skew normal distribution. An alternative estimation technique can be based on utilisation of the expectation-maximization (EM) algorithm, but to the best of our knowledge currently the EM algorithm is only applied to the a not closed form of the multivariate skew normal distribution[200].

The only (known to us) alternative approach to estimation of the closed skew normal parameters was presented by Flecher et al.[201]. The approach is based on the weighted method of moments and allows enhancing of parameter estimates for small samples. According to the authors, this approach is outperform the MLE at least in univariate and bivariate cases and can be used to initialise the MLE algorithm.

3.2.2. Estimation of individual efficiency values

The second step of estimation is obtaining estimates of the company-specific inefficiency values u_i. From the MLE procedure we have estimates of the composed error term ε_i, which obviously contains information about u_i. To extract the information about u_i, the conditional distribution of u_i given ε_i can be applied. We apply this procedure, following Jondrow et al.[151]. In the proof of the Theorem 1 the joint distribution function of u and ε was derived:

$$f_{u\varepsilon}(u,\varepsilon) = [\Phi_n(0,-\mu,\Sigma_u)]^{-1} \cdot \\ \cdot \varphi_n\left(-\Sigma_u(\Sigma_v+\Sigma_u)^{-1}(\varepsilon+\mu),-\mu,\left(\Sigma_v^{-1}+\Sigma_u^{-1}\right)^{-1}\right)\varphi_n(\varepsilon,-\mu,\Sigma_v+\Sigma_u) \tag{3.30}$$

The conditional distribution of u_i given ε_i is

$$f_{u|\varepsilon}(u|\varepsilon) = \frac{f_{u\varepsilon}(u,\varepsilon)}{f_\varepsilon(\varepsilon)} = \\ = \frac{[\Phi_n(0,-\mu,\Sigma_u)]^{-1}\varphi_n\left(\Sigma_u(\Sigma_v+\Sigma_u)^{-1}(\varepsilon+\mu),\mu,\left(\Sigma_v^{-1}+\Sigma_u^{-1}\right)^{-1}\right)\varphi_n(\varepsilon,-\mu,\Sigma_v+\Sigma_u)}{[\Phi_n(0,-\mu,\Sigma_u)]^{-1}\Phi_n\left(\Sigma_u(\Sigma_v+\Sigma_u)^{-1}(\varepsilon+\mu),\mu,\left(\Sigma_v^{-1}+\Sigma_u^{-1}\right)^{-1}\right)\varphi_n(\varepsilon,-\mu,\Sigma_v+\Sigma_u)} = \\ = \frac{\varphi_n\left(\Sigma_u(\Sigma_v+\Sigma_u)^{-1}(\varepsilon+\mu),\mu,\left(\Sigma_v^{-1}+\Sigma_u^{-1}\right)^{-1}\right)}{\Phi_n\left(-\Sigma_u(\Sigma_v+\Sigma_u)^{-1}(\varepsilon+\mu),-\mu,\left(\Sigma_v^{-1}+\Sigma_u^{-1}\right)^{-1}\right)} \tag{3.31}$$

for $u \geq 0$. The derived function exactly matches the multivariate truncated normal probability density function, so

$$u|\varepsilon \sim MVTN_{0,+\infty}\left(\mu_{u|\varepsilon}, \Sigma_{u|\varepsilon}\right), \tag{3.32}$$

where

$$\mu_{u|\varepsilon} = \mu - \Sigma_u\left(\Sigma_v + \Sigma_u\right)^{-1}\left(\varepsilon + \mu\right)$$

$$\Sigma_{u|\varepsilon} = \left(\Sigma_v^{-1} + \Sigma_u^{-1}\right)^{-1}$$

Note that the presented formulas are reduced to (2.20) when random disturbances and inefficiencies are independent and identically distributed, $\Sigma_v = \sigma_v^2 I_n, \Sigma_u = \sigma_u^2 I_n$:

$$\mu_{u|\varepsilon} = \mu - \sigma_u^2\left(\sigma_v^2 I_n + \sigma_u^2 I_n\right)^{-1}\left(\varepsilon + \mu\right) = \mu - \frac{\sigma_u^2}{\sigma_v^2 + \sigma_u^2}\left(\varepsilon + \mu\right) I_n = \frac{\mu\sigma_v^2 - \sigma_u^2\varepsilon}{\sigma_v^2 + \sigma_u^2},$$

$$\Sigma_{u|\varepsilon} = \left(\sigma_v^{-2} I_n + \sigma_u^{-2} I_n\right)^{-1} = \left(\frac{1}{\sigma_v^2} + \frac{1}{\sigma_u^2}\right)^{-1} I_n = \frac{\sigma_v^2\sigma_u^2}{\sigma_v^2 + \sigma_u^2} I_n.$$

Given the conditional distribution of u, a vector of point estimates $\hat{u}$ can be found as a conditional expected value:

$$\hat{u} = E\left(u|\varepsilon\right). \tag{3.33}$$

Confidence intervals also can be constructed using the conditional variance. Corresponding theoretical moments of the multivariate truncated normal distribution are well-known[202].

3.2.3. Identification of the SSF model parameters

One of the most important issues of a spatial econometric model concerns identification of their parameters. The notable reflection problem[184] specifies that different types of spatial effects, included into the model, cannot be distinguished one from another under some conditions. SF models are also suffer from the identification problem; for example, Greene [203] notes that parameters μ and σ_u of the truncated normal inefficiency are weakly identified and the model is extremely volatile. The proposed SSF model is affected by weak identification to a greater degree.

Let consider the SSF(1, 1, 1, 1) model in the following form:

$$Y_i = \rho_Y \sum_{j=1}^{n} w_{Y,ij} Y_j + \sum_{k=1}^{K} X_{ki}\beta_k + \sum_{k=1}^{K}\left(\gamma_k \sum_{j=1}^{n} w_{X,ij} X_{kj}\right) + v_i - u_i, \tag{3.34}$$

$$v_i = \rho_v \sum_{j=1}^{n} w_{v,ij} v_j + \tilde{v}_i,\ \tilde{v}_i \sim N\left(0, \sigma_{\tilde{v}}^2\right),$$

$$u_i = \rho_u \sum_{j=1}^{n} w_{u,ij} u_j + \tilde{u}_i,\ \tilde{u}_i \sim TN_{0,+\infty}\left(\mu, \sigma_{\tilde{u}}^2\right).$$

The expected value of the output Y_i given a vector of inputs $X_i = (X_{1i}, X_{1i}, \ldots, X_{ki})$ is:

$$E(Y_i|X_i)=\rho_Y\sum_{j=1}^{n}w_{Y,ij}E(Y_j|X_{ki})+\sum_{k=1}^{K}X_{ki}\beta_k+\sum_{k=1}^{K}\left(\gamma_k\sum_{j=1}^{n}w_{X,ij}X_{kj}\right)+E(v_i|X_{ki})-E(u_i|X_{ki}). \tag{3.35}$$

Assuming that

- the matrix W is row-standardised
- random disturbances and inefficiencies are independent from the inputs,
- expected value of random disturbances is conventionally zero,

the expression folds to:

$$E(Y_i|X_{ki})=\rho_Y E(Y_j|X_{ki})+\sum_{k=1}^{K}X_{ki}\beta_k+\sum_{k=1}^{K}\left(\gamma_k\sum_{j=1}^{n}w_{X,ij}X_{kj}\right)-E(u_i). \tag{3.36}$$

An expected value of the inefficiency u is presented as:

$$E(u_i)=\rho_u\sum_{j=1}^{n}w_{u,ij}E(u_j)+E(\tilde{u}_i), \tag{3.37}$$

or, for row-standardised spatial weights,

$$E(u_i)=\rho_u E(u_i)+E(\tilde{u}), \tag{3.38}$$
$$E(u_i)=\frac{1}{1-\rho_u}E(\tilde{u}).$$

An expected value of the truncated normal $\tilde{u}$ term is well-known:

$$E(\tilde{u})=\mu-\frac{\varphi\left(\frac{b-\mu}{\sigma_u}\right)-\varphi\left(\frac{a-\mu}{\sigma_u}\right)}{\Phi\left(\frac{b-\mu}{\sigma_u}\right)-\Phi\left(\frac{a-\mu}{\sigma_u}\right)}\sigma_u=\mu-\frac{\varphi(+\infty)-\varphi\left(\frac{-\mu}{\sigma_u}\right)}{\Phi(+\infty)-\Phi\left(\frac{-\mu}{\sigma_u}\right)}\sigma_u= \tag{3.39}$$
$$=\mu-\sigma_u\varphi\left(\frac{\mu}{\sigma_u}\right)\Big/\Phi\left(\frac{\mu}{\sigma_u}\right)$$

So the final expression for the expected value of the output Y is:

$$E(Y_i|X_{ki})=\frac{1}{1-\rho_Y}\sum_{k=1}^{K}X_{ki}\beta_k+\frac{1}{1-\rho_Y}\sum_{k=1}^{K}\left(\gamma_k\sum_{j=1}^{n}w_{X,ij}X_{kj}\right)- \tag{3.40}$$
$$-\frac{1}{1-\rho_Y}\frac{1}{1-\rho_u}\left(\mu-\sigma_u\varphi\left(\frac{\mu}{\sigma_u}\right)\Big/\Phi\left(\frac{\mu}{\sigma_u}\right)\right)$$

Separating a usual constant β_0 from the frontier, the intercept in the expected value is expressed as:

$$\frac{\beta_0}{1-\rho_Y}-\frac{1}{1-\rho_Y}\frac{1}{1-\rho_u}\left(\mu-\sigma_u\varphi\left(\frac{\mu}{\sigma_u}\right)\Big/\Phi\left(\frac{\mu}{\sigma_u}\right)\right).$$

Obviously that parameters β_0, ρ_Y, ρ_u, μ and σ_u can co-vary to produce identical results in the expectation of Y, which make it difficult to identify their specific contribution.

Let consider three data generating process specifications to illustrate the identification problems (the SSF(0,0,0,1) model is analysed for simplicity reasons). A function form of a frontier is identical for all 3 processes:

$$Y=\alpha+5+10\log(x)+\log(x)^2, \tag{3.41}$$

where α is a process-related shift of the intercept.

The frontier functional form doesn't make a difference here and included for research reproducibility only; selection of the DGP frontier specification is explained in the paragraph 3.4.2. Three considered DGP specifications are:

1. DGP A: positively spatially related inefficiencies, a small variance of random disturbances and no frontier shift:
   $\alpha=0,$
   $\sigma_u=2.5,$
   $\sigma_v=0.5,$
   $\rho_u=0.5.$
2. DGP B: negatively spatially related inefficiencies, a small variance of random disturbances and a shifted down frontier:
   $\alpha=-3,$
   $\sigma_u=2.5,$
   $\sigma_v=0.5,$
   $\rho_u=-0.5.$
3. DGP C: independent inefficiencies, a high variance of random disturbances and a shifted down frontier:
   $\alpha=-3,$
   $\sigma_u=0.5,$
   $\sigma_v=1.5,$
   $\rho_u=0.$

Note that processed B and C have identical frontiers, which is located below the DGP A frontier.

Simulated data and true frontiers for the processes are presented on the Fig. 3.1 (source codes for the simulations are provided in the Appendix 2).

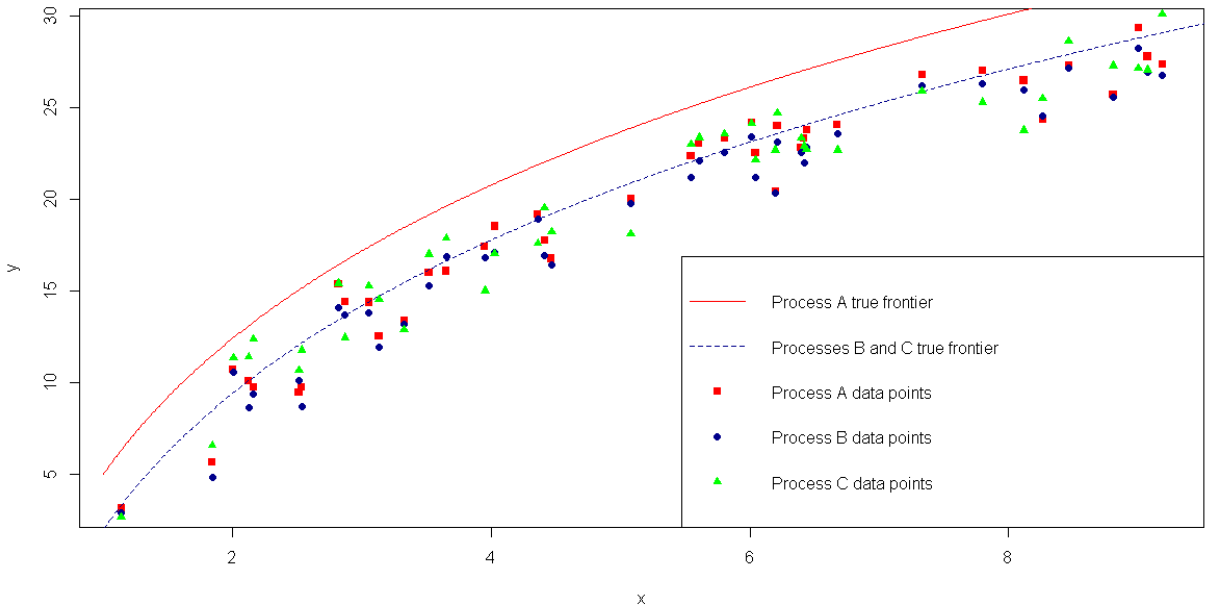


Fig. 3.1. Simulated data and true frontiers for sample DGP specifications

Expected values of the dependent variable for all three DGP are almost identical, although explained by different factors. The DGP A describes a classical stochastic frontier process, where almost all data are located under the frontier due to inefficiency. A positive spatial effect in the DGP B increases the output of all units, which is compensated by a lower frontier position. A similar effect is produced in the DGP C with smaller inefficiency in data, but higher values of random disturbances. Data points for different DGP specifications, presented on the Fig. 3.1, have a very similar pattern and it is almost impossible to distinguish them without a spatial structure. Nevertheless, when a spatial structure is provided, spatial patterns can be easily discovered. The Table 3.1 contains results of the Moran's I tests for residuals (an extended simulated sample of 300 units is used to reach the statistical significance) and discovers simulated spatial dependencies.

Table 3.1. Results of the Moran's I test for spatial correlation in simulated data

	Moran's I	Moran's I two-sided significance	Conclusion
DGP A	0.198	0.000	Positive spatial correlation
DGP B	-0.083	0.008	Negative spatial correlation
DGP C	-0.039	0.239	No spatial correlation

Generally, identification of the model parameters depends on specification of spatial weight matrixes. Whether parameters of the SSF(1,1,1,1) model are identified for spatial weights, specified in an application, needs to be investigated. An extensive simulation study on different

spatial weights matrix specification in classical spatial regression models was presented by Stakhovych and Bijmolt[204], but likely the SSF model has some specifics. We suppose that usage of different spatial weight matrixes for the dependent variable, explanatory variables, random disturbances, and inefficiency terms should improve model parameter identification, but this statement require additional research.

3.3. Implementation of the MLE of the SSF model parameters

3.3.1. Review of R and the spfrontier package

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

1. Multivariate normal probability density and distribution functions calculation is required for the likelihood function (3.29). Note that number of dimensions matches the sample size *n* and can be very significant. Computation of multivariate normal functions is well researched[205] and implemented in many software packages.
2. Multivariate truncated normal probability density and distribution functions calculation is straightforward on the base of multivariate normal functions.
3. Moments for multivariate truncated normal random variables are required for estimation of technical efficiency (3.32).
4. The proposed MLE also requires extensive matrix algebra (3.13). In practice, the matrixes contain a large percent of zero values (sparse), so implementation of sparse matrix algebra algorithms is helpful.
5. Maximisation of the likelihood function requires implementation of modern optimisation algorithms (quasi-Newton BFGS, Nelder-Mead, SANN, or others).

R[206] is one of popular software tools, where all of the required core algorithms are implemented. R is a freely available environment (under the GNU license) for statistical computing, which provides a wide set of statistical and graphical techniques. The Comprehensive R Archive Network (CRAN) contains a large number of packages, implementing particular statistical tools and algorithms. A list of R packages, which implement the required functions, is presented in the Table 3.2.

Relying on the required functions, we chose the R environment as a base for implementation of the derived MLE functions. The developed software package is named *spfrontier* and available in the official CRAN archive[61]. The main estimator of the SSF model is implemented as a function of the same name *spfrontier*. The function encapsulates all algorithms, required for the MLE estimator; a list of arguments is presented in the Table 3.3.

Table 3.2. R packages related to the SSF model estimation

Package	Purpose
mvtnorm	Multivariate Normal Density function Multivariate Normal Distribution function Multivariate Normal Random number generator
tmvtnorm	Truncated Multivariate Normal Density function Truncated Multivariate Normal Distribution function Moments For Truncated Multivariate Normal Distribution Truncated Multivariate Normal Random number generator
ezsim	Framework to conduct simulation
moments	Moments, cumulants, skewness, kurtosis and related tests
Matrix	Sparse and Dense Matrix Classes and Methods
spdep	Spatial dependence: statistics and models
frontier	Stochastic Frontier Analysis
optim (stats)	General-purpose optimization based on Nelder–Mead, quasi-Newton and conjugate-gradient algorithms.

Table 3.3. Arguments of the *spfrontier* function

Argument	Description
formula	an object of class ‘formula’: a symbolic description of the model to be fitted.
data	data frame, containing the variables in the model.
W_y	a spatial weight matrix for spatial lag of the dependent variable, W_Y.
W_v	a spatial weight matrix for spatial lag of the symmetric error term, W_v.
W_u	a spatial weight matrix for spatial lag of the inefficiency error term, W_u.
initialValues	an optional vector of initial values, used by maximum likelihood estimator. If not defined, the proposed method of initial values estimation is used.
inefficiency	a distribution for inefficiency error component. Possible values are ‘half-normal’ (for half-normal distribution) and 'truncated' (for truncated normal distribution). By default set to ‘half-normal’.
logging	an optional level of logging. Possible values are ‘quiet’, ’warn’, ’info’, and ’debug’. By default set to ‘quiet’.
onlyCoef	Logical, allows calculating only estimates for coefficients (with inefficiencies and other additional statistics). Developed generally for testing, to speed up the process.
control	an optional list of control parameters, passed to optim estimator from the stats package.

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;
- a vector of residuals.

Together with implementation of the SSF model estimator, the *spfrontier* package includes all data sets, used in this research, which ensures research reproducibility.

Official documentation of the *spfrontier* package is available in the Appendix 3 and online. The package is also enhanced with demo files and simulation tests.

The following paragraphs of this chapter describe some critical aspects of the MLE implementation.

3.3.2. *Calculating initial values for the MLE*

Selection of the initial parameter values is extremely important for numeric maximisation of the likelihood function, especially if this function is not convex. The following procedure of initial values searching was suggested and implemented:

1. If the model specification contains only exogenous spatial components that is the SSF(0,1,0,0) model:

   $$Y = X\beta + W_X X\beta^{(s)} + v - u,$$

   a corresponding model with a symmetric error term is considered and ordinary least square estimates for its parameters β and $\beta^{(s)}$ are obtained:

   $$\hat{\beta}_{ols}, \hat{\beta}^{(s)}_{ols}.$$

   Method of moments can be used to obtain initial values for variance of random disturbances $\sigma_{\tilde{v}}$ and inefficiency $\sigma_{\tilde{u}}$. Assuming that the inefficiency term is half-normal (μ = 0), the second and third theoretical moments of ε are:

   $$\nu_2 = \sigma^2_{\tilde{v}} + \frac{\pi - 2}{\pi}\sigma^2_{\tilde{u}} \tag{3.42}$$
   $$\nu_3 = \sigma^3_{\tilde{u}}\sqrt{\frac{2}{\pi}}\frac{\pi - 4}{\pi}$$

   Corresponding sample moments of the OLS residuals e_{ols} are:

   $$m_2 = \frac{1}{n} e_{ols}{}^T e_{ols}, \tag{3.43}$$
   $$m_3 = \frac{1}{n} e_{ols}{}^T e_{ols} e_{ols},$$

   where

   $$e_{ols} = Y - X\hat{\beta}_{ols} - W_X X\hat{\beta}^{(s)}_{ols}.$$

   Thus initial estimates for standard deviations are:

   $$\sigma_{\tilde{u}} = \sqrt[3]{m_3\sqrt{\frac{\pi}{2}}\frac{\pi}{\pi - 4}}, \tag{3.44}$$
   $$\sigma_{\tilde{v}} = \sqrt{m_2 - \frac{\pi - 2}{\pi}\sigma^2_{\tilde{u}}}.$$

   The algorithm provides initial estimates for $\beta, \beta^{(s)}, \mu, \sigma^2_{\tilde{v}}$, and $\sigma^2_{\tilde{u}}$.

2. If the model specification contains endogenous and exogenous spatial components that is the SSF(1,1,0,0) model, spatial lags of the dependent variable $W_Y Y$ is included into the model as exogenous variable and the composed SSF(0,1,0,0) model is estimated with

the proposed MLE (3.29), using the step 1 for initial values. The parameter μ is estimated as a sample mean of the residuals. The algorithm provides estimates to ρ_Y and $\beta, \beta^{(s)}, \mu, \sigma_{\tilde{v}}^2,$ and $\sigma_{\tilde{u}}^2$. Note that these estimates are inconsistent due to endogeneity of an explanatory variable.

3. If the model specification contains all types of spatial components, that the SSF(1,1,1,1) model, then spatially correlated random disturbances and spatially related inefficiency are temporarily omitted and the SSF(1,1,0,0) model is estimated using the initial values from the step 3. Next an ancillary regression is estimated with OLS:

   $e = \rho We + v,$

   where e is a vector of residuals of the SSF(1,1,0,0) model. Estimated coefficient ρ is used as an initial value for the ρ_v parameter. An initial estimate for ρ_u is considered as 0 (no spatially related inefficiency).
4. Finally, when initial values are obtained, they are improved by a grid search. The intervals for the grid search
   - $(\beta - 3\sigma_{\tilde{v}}, \beta + 3\sigma_{\tilde{v}})$ for the parameters β,
   - $(\beta^{(s)} - 3\sigma_{\tilde{v}}, \beta^{(s)} + 3\sigma_{\tilde{v}})$ for the parameters $\beta^{(s)}$,
   - $(0.5\sigma_{\tilde{v}}, 1.5\sigma_{\tilde{v}})$ for the parameter $\sigma_{\tilde{v}}$,
   - $(0.5\sigma_{\tilde{u}}, 1.5\sigma_{\tilde{u}})$ for the parameter $\sigma_{\tilde{u}}$,
   - $(-0.99, 0.99)$ for the parameter ρ_u.

Note that the suggested procedure is empirical to a considerable degree, so a well theoretically grounded alternative is called.

3.3.3. Estimation of parameters and their variance

In addition to theoretical issues of MLE of the skew normal distribution parameters, there are some computational problems. The presented log-likelihood function (3.29) obviously is not convex and not smooth. Frequently used Olsen's transformation[207] of the likelihood function parameters makes it smoother and computationally easier:

$$\eta = \left(\sigma_{\tilde{v}}^2 + \sigma_{\tilde{u}}^2\right)^{-1/2}, \tag{3.45}$$

$$\lambda = \frac{\sigma_{\tilde{u}}}{\sigma_{\tilde{v}}},$$

$$\gamma = \eta\beta,$$

$$\gamma^{(s)} = \eta\beta^{(s)},$$

$$\omega = \eta\left(Y - \rho_Y W_Y Y\right) - X\beta - W_X X\beta^{(s)} = \eta\varepsilon.$$

Analytical gradients of the log-likelihood function are highly convenient for computational optimisation. Unfortunately, the log-likelihood function includes the multivariate normal cumulative distribution function, which has no analytical gradients. Absence of analytical gradients makes optimisation computationally harder, but still available for relatively small samples (see the paragraph 3.2). Numeric optimisation methods allow calculating of numeric estimates of the gradient and a Hessian matrix, which is necessary for hypothesis testing:

$$H(\theta)=\left\{\frac{\partial^2 \ln L(\theta)}{\partial\theta_i \partial\theta_j}\right\}, \tag{3.46}$$

where $\theta=\left(\gamma,\gamma^{(s)},\eta,\lambda,\mu,\rho_Y,\rho_v,\rho_u\right)^T$ is a vector of parameters of the log-likelihood function.

Given the Hessian matrix, a variance-covariance matrix $Var(\Theta)$ of the parameters can be estimated as:

$$Var\left(\hat{\theta}\right)=\left(-E\left(H\left(\hat{\theta}\right)\right)\right)^{-1}. \tag{3.47}$$

Numeric Hessian allows estimating a variance-covariance matrix of transformed parameters, so a final inverse transformation is necessary. The appropriate estimator of the variance-covariance matrix is the sandwich estimator[134]:

$$Var\left(\hat{\theta}_{ini}\right)=G^{-1}\left(\hat{\theta}\right)Var\left(\hat{\theta}\right)G^{-1}\left(\hat{\theta}\right), \tag{3.48}$$

where $\theta_{ini}=\left(\beta,\beta^{(s)},\sigma_{\tilde{v}}^2,\sigma_{\tilde{u}}^2,\mu,\rho_Y,\rho_v,\rho_u\right)^T$ is a vector of initial parameters and

$$G\left(\hat{\theta}\right)=\frac{\partial\hat{\theta}_{ini}}{\partial\hat{\theta}}. \tag{3.49}$$

An inverse transformation of the parameters is expressed as:

$$\begin{aligned}
&\beta=\gamma/\eta, \\
&\beta^{(s)}=\gamma^{(s)}/\eta, \\
&\sigma_v=\frac{1}{\eta\sqrt{1+\lambda^2}}, \\
&\sigma_u=\frac{\lambda}{\eta\sqrt{1+\lambda^2}},
\end{aligned} \tag{3.50}$$

so we obtained an expression for G:

$$G(\theta)=\frac{\partial\theta_{ini}}{\partial\theta}=\frac{\partial\left(\frac{\gamma}{\eta},\frac{\gamma^{(s)}}{\eta},\frac{1}{\eta\sqrt{1+\lambda^2}},\frac{\lambda}{\eta\sqrt{1+\lambda^2}},\mu,\rho_Y,\rho_v,\rho_u\right)^T}{\partial\left(\gamma,\gamma^{(s)},\eta,\lambda,\mu,\rho_Y,\rho_v,\rho_u\right)^T}= \tag{3.51}$$

$$=\begin{bmatrix} \frac{1}{\eta} & \frac{1}{\eta} & -\frac{\gamma}{\eta^2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{\gamma^{(s)}}{\eta^2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{\eta^2\sqrt{1+\lambda^2}} & -\frac{\lambda}{\eta\left(\sqrt{1+\lambda^2}\right)^3} & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{\lambda}{\eta^2\sqrt{1+\lambda^2}} & \frac{1}{\eta\left(\sqrt{1+\lambda^2}\right)^3} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Computation of the variance-covariance matrix of the parameters requires a non-singular Hessian, which is not always satisfied in practice. A general treatment in case of a non-singular Hessian is reformulation of the model.

3.4. Validation of the proposed MLE for the SSF model

3.4.1. Compliance of obtained estimates with existing software results

The classical stochastic frontier model without spatial effects can be considered as a private case of the SSF model. Thus the estimates for the SSF(0,0,0,0) model parameters, calculated with the proposed estimator, should exactly match the result of the classical model estimation. For comparison of results we used the Frontier 4.1 package[208]. This package is a single purpose software tool, specifically designed for the estimation of the classical stochastic frontier model with different specifications of the inefficiency term. We used an R package *frontier*[209] as a handy wrapper for the Frontier 4.1 package.

A sample dataset, provided Frontier 4.1 package, is used for calculations. The dataset contains cross-sectional data of 60 firms and includes three variables, typical for production functions: output, labour, and capital. The Cobb-Douglas form is used as a functional specification of the production frontier.

The experiment includes two different specifications of the inefficiency term: half-normal and truncated normal (with a constant mean). Results of estimation are presented in the Table 3.4.

Table 3.4. Comparison of *frontier* and *spfrontier* estimators

	Half-normal inefficiencies				Truncated normal inefficiencies			
	spfrontier		**frontier**		**spfrontier**		**frontier**	
	Estimate	Std.Error	Estimate	Std.Error	Estimate	Std.Error	Estimate	Std.Error
Intercept, β_0	0.5616	0.2026	0.5616	0.2026	0.4655	0.2276	0.4764	0.2141
log(capital) , β_1	0.2811	0.0475	0.2811	0.0476	0.2832	0.0479	0.2826	0.0479
log(labour) , β_2	0.5365	0.0452	0.5365	0.0453	0.5410	0.0453	0.5404	0.0457
σ_v	0.2098	0.0513	0.2098*		0.2274	0.0515	0.2243*	
σ_u	0.4159	0.0926	0.4159**		0.8903	1.9835	0.7053**	
μ					-2.6491	14.9390	-1.4106	2.5990
Log likelihood	-17.0272		-17.0272		-16.7857		-16.7957	

*, ** calculated by the author using reparamestrisation formulas (3.50).

Estimates, calculated by *frontier* and *spfrontier* packages for a model with half-normal inefficiencies, are matched perfectly. Analysis of the model with truncated normal inefficiencies is not so straightforward. Both estimators provide similar estimates of the production function parameters β and a variance of random disturbances σ_v, but estimates for inefficiency parameters μ and σ_u differ significantly. This problem is well-known in literature[203]. The model with truncated normal distribution of inefficiency is extremely unstable and parameters μ and σ_u are weakly identified. Note that from (3.39) the expected value of the truncated normal inefficiency term is presented as:

$$E(u_i) = \mu + \sigma_u \, \varphi\left(\frac{\mu}{\sigma_u}\right) \Big/ \Phi\left(\frac{\mu}{\sigma_u}\right), \tag{3.52}$$

so different combinations of μ and σ_u can deliver the same expected value to u_i (which is the only moment of u_i used in the MLE). The problem is illustrated on the Fig. 3.2, containing contours of the likelihood function for different values of μ and σ_u for the sample dataset. The almost flat area in the middle represents combinations of μ and σ_u, which deliver very close values to the likelihood function. The chart clarifies that these different results of the *frontier* and *spfrontier* estimators are a matter of optimization algorithm's precision settings.

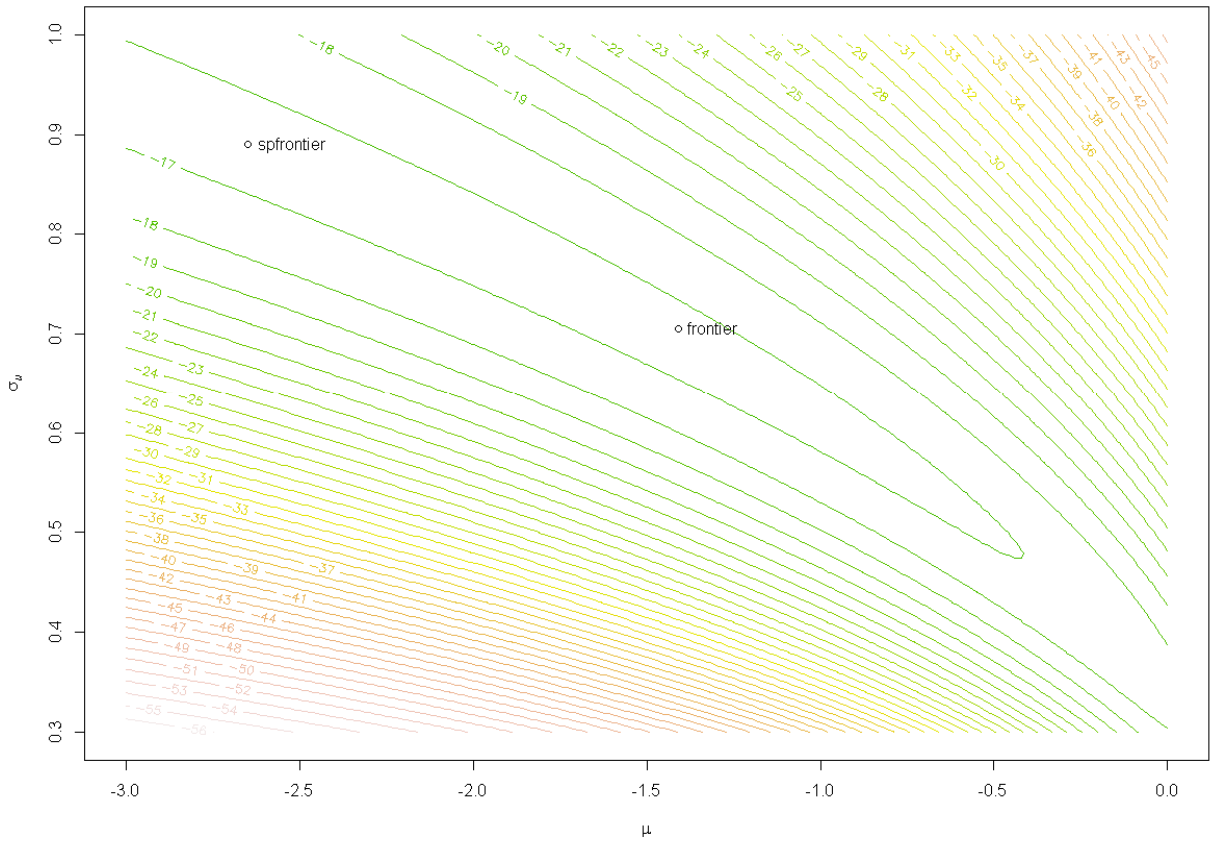


Fig. 3.2. Contours of the SF likelihood function for μ and σ_u

3.4.2. Simulation testing of the proposed MLE

The finite sample performance of the proposed MLE estimator is investigated via a set of Monte Carlo simulation tests.

A data generated process DGP, used in this research, is described with the following parameters:

- A vector of parameters β^*, which define the form of the production frontier. The production frontier function is supposed to be linearised. An exact form of the production frontier differs between researches. Banker and Natarajan[210] discussed different production frontier functional specifications (a third-order polynomial function, Cobb-Douglas and translog single-input production functions), which are stated to be continuous, monotonic increasing, and concave over the relevant range of inputs used in the simulations. For all our simulation tests we used a single-input translog production function:

$$f\left(X^*, \beta_0^*, \beta_1^*, \beta_2^*\right) = \beta_0^* + \beta_1^* \log\left(X^*\right) + \beta_2^* \log\left(X^*\right)^2$$

- True DGP values of parameters β^* were predefined as

$\beta_0^* = 5,$
$\beta_1^* = 10,$
$\beta_2^* = 1$

for all executed tests.

- Distribution of the input X^* components. The input of the production possibility frontier are supposed to be uniformly distributed on the [1,10] range:

  $X^* \sim U(1,10)$

  The simulated production function is monotonic increasing and concave over the specified range of the input.

- Parameters ρ_Y^*, ρ_v^*, and ρ_u^* for spatial effects of four types in the SSF model (3.1). Zero values for specified parameters mean absence of corresponding spatial effects in DGP. Estimation of spatial exogenous effects, represented in the SSF model as $\beta^{(s)}$, doesn't differ from independent inputs specification and are not considered in these simulations.

- Spatial weights matrixes W_Y, W_v, and W_u. Artificial rook- and queen-style spatial weight matrixes are used: rook-style for W_v and queen-style for W_Y and W_u.

- Distribution of the symmetric random disturbances are conventionally put to normal with zero mean and a specified standard deviation $\sigma_{\tilde{v}}^*$:

  $\tilde{v}^* \sim MVN\left(0_n, \sigma_{\tilde{v}}^{*2} I_n\right)$

  We consider a "low noise" scenario, putting a true DGP value of the parameter $\sigma_{\tilde{v}}^*$ to 0.5 for all executed tests:

  $\sigma_{\tilde{v}}^* = 0.5.$

- Distribution of the inefficiency term are conventionally put to truncated normal with a specified mean μ^* and a specified standard deviation $\sigma_{\tilde{u}}^*$:

  $\tilde{u}^* \sim MVTN_{0,+\infty}\left(\mu^*, \sigma_{\tilde{u}}^{*2}\right)$

  A ratio λ of standard deviations of random disturbances and inefficiencies is a critical point of the stochastic frontier model. We consider a scenario with significant inefficiency in data, putting a true DGP value of the parameter $\sigma_{\tilde{u}}^*$ to 2.5 for all executed tests:

  $\sigma_{\tilde{u}}^* = 2.5,$

  so the ratio λ is

  $\lambda = \sigma_{\tilde{u}}^* / \sigma_{\tilde{v}}^* = 5.$

So a sample model of the DGP can be summarised as:

$$DGP\left(\beta_0^*, \beta_1^*, \beta_2^*, \rho_Y^*, \rho_v^*, \rho_u^*, \sigma_X^*, \sigma_{\tilde{v}}^*, \mu^*, \sigma_{\tilde{u}}^*\right): \tag{3.53}$$
$$Y^* = \left(I_n - \rho_Y^* W_Y\right)^{-1}\left(\beta_0^* + \beta_1^* \log\left(X^*\right) + \beta_2^* \log\left(X^*\right)^2 + \left(I_n - \rho_v^* W_v\right)^{-1}\tilde{v} - \left(I_n - \rho_u^* W_u\right)^{-1}\tilde{u}\right),$$
$$X^* \sim U(1,10),$$
$$\tilde{v}^* \sim N\left(0, \sigma_{\tilde{v}}^{*2}\right),$$
$$\tilde{u}^* \sim N\left(\mu^*, \sigma_{\tilde{u}}^{*2}\right).$$

Putting predefined parameters we have the final definition of the DGP:

$$DGP\left(\rho_Y^*, \rho_v^*, \rho_u^*, \mu^*\right): \tag{3.54}$$
$$Y^* = \left(I_n - \rho_Y^* W_Y\right)^{-1}\left(5 + 10\log\left(X^*\right) + \log\left(X^*\right)^2 + \left(I_n - \rho_v^* W_v\right)^{-1}\tilde{v} - \left(I_n - \rho_u^* W_u\right)^{-1}\tilde{u}\right),$$
$$X^* \sim U(1,10),$$
$$\tilde{v}^* \sim N\left(0, 0.5^2\right),$$
$$\tilde{u}^* \sim N\left(\mu^*, 2.5^2\right).$$

Data sets, simulated from the constructed DGP for different types of spatial effects, are illustrated on the Fig. 3.1. Endogenous spatial effects are modelled as $\rho_Y = 0.2$; all other spatial effects are simulated as $\rho = 0.4$, that is positive spatial relationships.

A list of executed simulation experiments is presented in the Table 3.5.

Table 3.5. List of executed simulation experiments

Simulation Experiment	DGP	Estimator	Sample size, n	Simulations, runs
SimE1	DGP: $\mu^* = 0, \rho_Y^* = 0, \rho_v^* = 0, \rho_u^* = 0$	SSF(0,0,0,0), half-normal	50, 100, 200, 300	100
SimE2	DGP: $\mu^* = 1, \rho_Y^* = 0, \rho_v^* = 0, \rho_u^* = 0$	SSF(0,0,0,0), truncated-normal	50, 100, 200, 300	100
SimE3	DGP: $\mu^* = 0, \rho_Y^* = 0.2, \rho_v^* = 0, \rho_u^* = 0$	SSF(1,0,0,0), half-normal	50, 100, 200, 300	100
SimE3b	DGP: $\mu^* = 0, \rho_Y^* = 0.2, \rho_v^* = 0, \rho_u^* = 0$	SSF(0,0,0,0), half-normal	50, 100, 200, 300	100
SimE4	DGP: $\mu^* = 1, \rho_Y^* = 0.2, \rho_v^* = 0, \rho_u^* = 0$	SSF(1,0,0,0), truncated-normal	50, 100, 200, 300	100
SimE5	DGP: $\mu^* = 0, \rho_Y^* = 0, \rho_v^* = 0.4, \rho_u^* = 0$	SSF(0,0,1,0), half-normal	50, 100, 200, 300	100
SimE5b	DGP: $\mu^* = 0, \rho_Y^* = 0, \rho_v^* = 0.4, \rho_u^* = 0$	SSF(1,0,0,0), half-normal	50, 100, 200, 300	100
SimE6	DGP: $\mu^* = 0, \rho_Y^* = 0, \rho_v^* = 0, \rho_u^* = 0.4$	SSF(0,0,0,1), half-normal	50, 100, 200, 300	100
SimE6b	DGP: $\mu^* = 0, \rho_Y^* = 0, \rho_v^* = 0, \rho_u^* = 0.4$	SSF(1,0,0,0), half-normal	50, 100, 200, 300	100

The estimator validity is measured using the following statistics:

- absolute and relative bias of estimates;
- standard deviation and root-mean-square deviation (RMSD) of estimates, defined for a parameter θ as:

$$RMSD(\hat{\theta}) = \sqrt{\frac{\sum_{r=1}^{runs}\left(\hat{\theta}_r - \theta^*\right)^2}{runs}}, \tag{3.55}$$

  where θ^* is a true value of the coefficient, $\hat{\theta}_r$ is an estimate of the coefficient θ in the simulation run *r;*
- estimate's confidence intervals to test estimate convergence to parameter's true value for larger samples (estimate consistency);
- kernel density estimation of estimates' empirical probability density functions.

Source codes for simulation studies are presented in the Appendix 4. All simulation experiments are executed within the Amazon Elastic Cloud environment, using Bioconductor Amazon Machine Image (AMI). A complete description of the environment is presented in the Appendix 5; simulation commands are included into the *spfrontier* package. Note that number of computers used for simulations are critical for research reproducibility. Detailed results of all simulation experiments are provided in the Appendix 6.

Spatially related efficiency is one of the key components of the introduced SSF model, so let pay special attention to the SimE6 simulation experiment, which deals with the SSF(0,0,0,1) model that is a model with spatially related efficiency included both into DGP and the estimator. Complete results for this simulation experiment can be found in the Appendix 6; here we discuss some critical aspects.

A short summary of the SimE6 experiment is presented in the Table 3.6.

Estimates of the frontier parameters β_1 and β_2 are unbiased (with respect of their standard deviations) and consistent for all sample volumes. Although standard deviations of these parameters are significantly decreasing for larger samples, a sample of 100 looks quite appropriate to their correct and statistically significant identification.

Estimates of the frontier intercept β_0 and the standard deviation σ_u of inefficiency are also statistically unbiased and consistent, but slightly suffer from the identification problem, discussed in the paragraph 3.2.3. The estimator identifies slightly lower positions of the frontier (-4.67%, -1.45%, and -2.35% bias of the intercept's estimate for sample volumes of 100, 200, and 300 respectively) in correspondence with slightly smaller standard deviations of inefficiency (-0.09%, -0.07%, and -0.05% respectively).

Table 3.6. Summary results of the simulation study SimE6

n	*Parameter*	*True Values*	*Mean*	*Bias*	*Bias, %*	*SD*	*RMSD*
50	β_0	5	5.5095	0.5095	3.5904	3.6263	0.1019
	β_1	2	1.9907	-0.0093	0.0755	0.0761	-0.0047
	β_2	3	3.0006	0.0006	0.0673	0.0673	0.0002
	σ_v	0.1	0.1317	0.0317	0.0518	0.0608	0.3171
	σ_u	0.5	0.3706	-0.1294	0.1559	0.2026	-0.2587
	ρ_u	0.4	0.3297	-0.0703	0.3409	0.3481	-0.1759
300	β_0	5	5.6143	0.6143	4.6745	4.7147	0.1229
	β_1	2	2.0028	0.0028	0.0514	0.0514	0.0014
	β_2	3	3.0082	0.0082	0.0599	0.0605	0.0027
	σ_v	0.1	0.1392	0.0392	0.0505	0.0639	0.3922
	σ_u	0.5	0.4014	-0.0986	0.1575	0.1858	-0.1972
	ρ_u	0.4	0.2641	-0.1359	0.2882	0.3186	-0.3397
200	β_0	5	4.9783	-0.0217	1.4523	1.4525	-0.0043
	β_1	2	2.0074	0.0074	0.0624	0.0628	0.0037
	β_2	3	2.995	-0.005	0.0533	0.0535	-0.0017
	σ_v	0.1	0.1459	0.0459	0.0369	0.0589	0.4592
	σ_u	0.5	0.4284	-0.0716	0.1319	0.1501	-0.1432
	ρ_u	0.4	0.244	-0.156	0.2597	0.3029	-0.3901
300	β_0	5	5.2245	0.2245	2.3552	2.3659	0.0449
	β_1	2	1.9964	-0.0036	0.0371	0.0372	-0.0018
	β_2	3	3.0018	0.0018	0.0359	0.0359	0.0006
	σ_v	0.1	0.1418	0.0418	0.0326	0.053	0.4176
	σ_u	0.5	0.4418	-0.0582	0.1115	0.1258	-0.1164
	ρ_u	0.4	0.2713	-0.1287	0.2718	0.3007	-0.3217

Estimates of the frontier intercept β_0 and the standard deviation σ_u of inefficiency are also statistically unbiased and consistent, but slightly suffer from the identification problem, discussed in the paragraph 3.2.3. The estimator identifies slightly lower positions of the frontier (-4.67%, -1.45%, and -2.35% bias of the intercept's estimate for sample volumes of 100, 200, and 300 respectively) in correspondence with slightly smaller standard deviations of inefficiency (-0.09%, -0.07%, and -0.05% respectively).

The most important parameter for this research is ρ_u, representing an effect of the spatially related inefficiencies in the sample. Generally, the conclusions about its estimates are positive – the effect (positive relationship between neighbour objects) was correctly identified (statistically unbiased), and estimates' standard deviations decrease for larger samples (consistency). This conclusion is based on the Table 3.6 values and their visual representation on the Fig. 3.3.

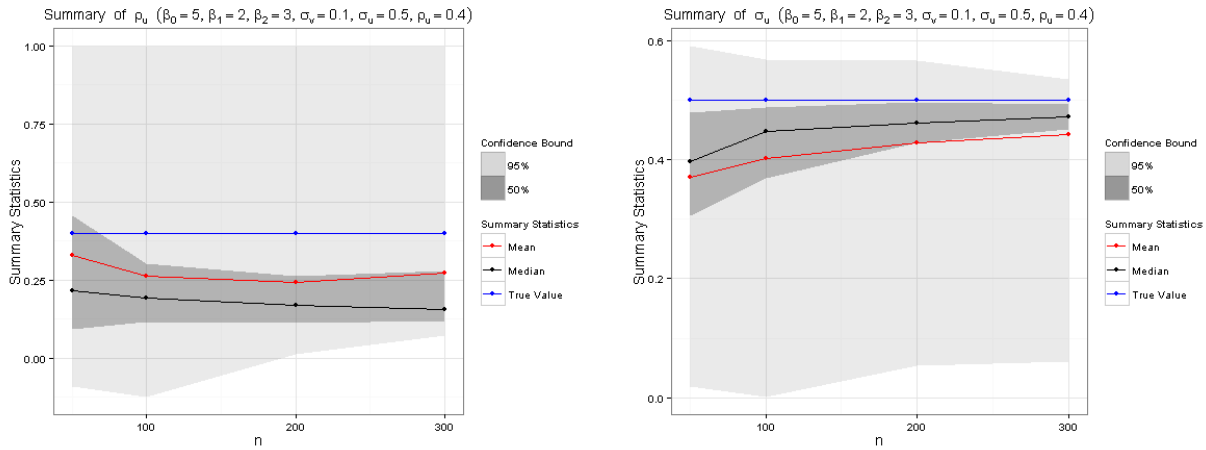


Fig. 3.3. Summary statistics plots for ρ_u and σ_u parameters in SimE6

However, a significant bias percentage for the parameter ρ_u estimates can be noted. The empirical kernel density of estimates is presented on the Fig. 3.4. Empirical kernel density plots for ρ_u and σ_u parameters in SimE6, can be used to clarify this bias.

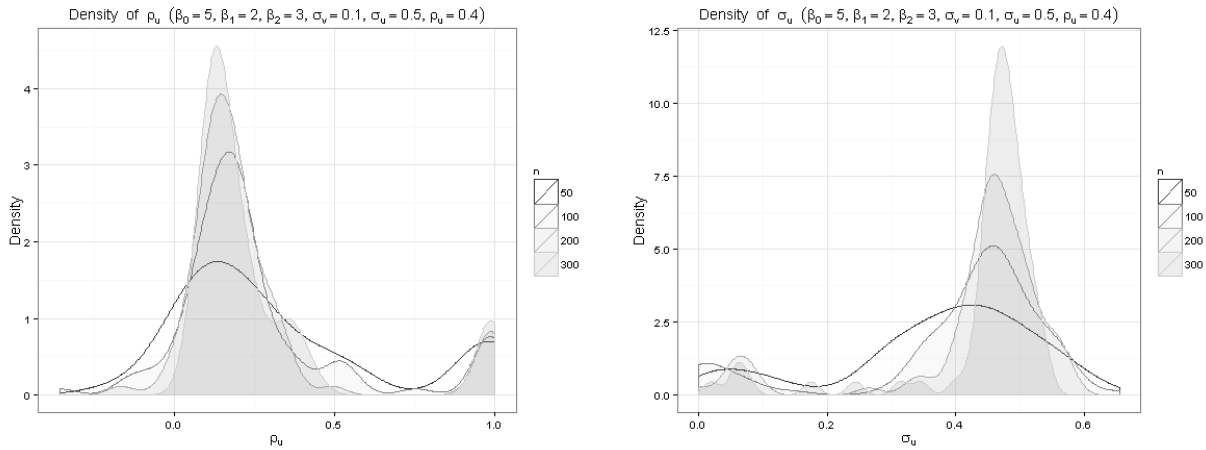


Fig. 3.4. Empirical kernel density plots for ρ_u and σ_u parameters in SimE6

Note a significant peak for estimates of ρ_u, located close to 1, which lead to a detected bias of estimates. This peak is related with a local maximum point of the likelihood function, which is interpreted as a global one by the numeric optimisation algorithm (Nelder-Mead). Local optimums are a basic problem of numeric optimisation and it cannot be avoided completely. A usual recommendation in this case is to provide an optimisation algorithm with initial values, located closer to the global maximum. The developed *spfrontier* module supports user-defined initial values and also allows managing the grid search for more careful initial values identification. Also it can be noted that the density of local optimums (peaks in the negative area) decreases for larger samples (200 and 300 objects), which leads to more convenient results. Probably, the problem will be solved completely for larger samples, but unfortunately this assumption cannot be tested currently tested due to floating point numbers precision limits in the

specified testing environment. Except of this problem, the estimator demonstrates good statistical performance and can be used for relatively moderate samples.

A detailed description of results of all executed simulation experiments is presented in the Appendix 6. Main conclusions are summarised for all experiments in the Table 3.7.

Table 3.7. Summary conclusions for the executed simulation studies

Simulation Experiment	*Main Conclusions*
SimE1	- unbiased estimates for frontier and inefficiency parameters; - consistent estimates both for frontier and inefficiency parameters
SimE2	- unbiased estimates for frontier and inefficiency parameters; - consistent estimates both for frontier and inefficiency parameters; - weak identification of σ_u and μ, especially for small samples.
SimE3	- unbiased estimates for frontier and inefficiency parameters; - consistent estimates both for frontier and inefficiency parameters; - unbiased and consistent estimates for endogenous spatial effects parameter ρ_Y.
SimE3b	- biased and inconsistent estimates for frontier intercept and random disturbances' standard deviations (as expected due to missed endogenous spatial effects in the estimator).
SimE4	- unbiased estimates for frontier and inefficiency parameters; - consistent estimates both for frontier and inefficiency parameters; - unbiased and consistent estimates for endogenous spatial effects parameter ρ_Y; - weak identification of σ_u and μ.
SimE5	- unbiased and consistent estimates for frontier parameters; - consistent estimates for the spatially correlated random disturbances parameter ρ_v; - large sample variance of the spatially correlated random disturbances parameter ρ_v and inefficiency standard deviation σ_u for small samples. So this is not recommended to apply MLE estimator of the SSF model for small samples; - estimation of the model for samples of 1000 or more objects is impossible in the specified environment due to double-precision floating-point limits; - model estimation takes a long time in a relatively powerful environment.
SimE5b	- unbiased and consistent estimates for frontier parameters, except σ_v and σ_u; - unbiased and consistent estimates for absent endogenous spatial effects parameter ρ_Y, so there is no replacement of spatially correlated random disturbances with endogenous spatial effects;
SimE6	- unbiased and consistent estimates for frontier parameters; - consistent estimates for the spatially related efficiency parameter ρ_u; - large sample variance of the spatially related efficiency parameter ρ_u and inefficiency standard deviation σ_u for a small sample of 100 objects. So this is not recommended to apply MLE estimator of the SSF model for small samples; - potential falling of the algorithm into local extremum points requires additional attention to initial values; - estimation of the model for samples of 1000 or more objects is impossible in the specified environment due to double-precision floating-point limits; - model estimation takes a long time in a relatively powerful environment.
SimE6b	- unbiased and consistent estimates for frontier parameters, except σ_v and σ_u; - unbiased and consistent estimates for absent endogenous spatial effects parameter ρ_Y, so there is no replacement of spatially related inefficiencies with endogenous spatial effects.

Summarising the Table 3.7, it can be stated that the simulation experiment results match our initial expectations:

1. The developed estimator provides unbiased and consistent estimates for classical non-spatial specifications of the stochastic frontier model (experiments SimE1 and SimE2). This fact ensures that the estimator can be applied to non-spatial models in case when spatial effects are non-realistic or as a simple comparison base for spatial models.
2. Endogenous spatial effects can be well identified on the base of limited samples (experiments SimE3 and SimE4); estimation of spatially correlated random disturbances and spatially related efficiency requires larger samples (experiments SimE5, SimE6).
3. Some parameters of the spatial stochastic frontier models are weakly identified and can be distinguished from each other (experiments SimE2, SimE4, SimE5b, and SimE6b). A problem of weak identification of mean μ and standard deviation σ_u of the truncated normal inefficiency is discussed in the paragraph 3.4.1; similar problems are discovered for the effect of spatially correlated random disturbances ρ_v and their standard deviation σ_v (experiment SimE5), and the effect of spatially related efficiency ρ_u, their standard deviation σ_u, and the frontier intercept (experiment SimE6).
4. Different types of spatial effects can be confidently distinguished from each other. Simulation experiment SimE5b shows that if spatially correlated random disturbances present in data, but forcibly excluded from the model, they are not recognised by the estimator as endogenous spatial effects. Similarly (experiment SimE6b), spatially related efficiencies aren't recognised as endogenous spatial effects.

3.5. Conclusions

This chapter contains a detailed description of the spatial stochastic frontier model, proposed by the author. Four types of spatial effects, possibly important in SFA, are spatial exogenous effects, spatial endogenous effects, spatially correlated random disturbances, and spatially related efficiency. We presented reasoning for these spatial effects as phenomena in different branches of knowledge and proposed the SSF model, which includes all four types of spatial effects.

The model can be considered as integration modern principles of spatial econometrics into the classical stochastic frontier analysis. In this chapter the SSF model is stated in a reasonably general form, where influence of spatial effects is included as first-order spatial lags. A number of practically effective private cases of the SSF model are also discussed. Specification of the SSF model is an important component of this research novelty.

A special attention is devoted to the problem of model parameter identification. Parameter identification is one of important issues, frequently noted both in spatial econometrics and stochastic frontier modelling literature. The SSF model as a combination of stochastic frontier and spatial regression models also suffer from weak parameter identification. In this chapter we presented a theoretical justification of parameter identification problem and illustrated it with real and simulated data examples.

One of the main practical results of this research is a derived maximum likelihood estimator of the SSF model parameters. A distribution law of the composed error term of the SSF model is derived and stated as a private case of the closed multivariate skew normal distribution. Using the derived distribution of the model's error term, the likelihood function is specified and a related estimator is constructed. Estimation of individual inefficiency values is one of the main benefits of the classical stochastic frontier models, so we also derived formulas for estimates of individual inefficiency values in the SSF model.

The derived MLE of the SSF model parameters is implemented as a package for CRAN R software, called *spfrontier*. The package includes all derived algorithms for the SSF model estimation and accepted and published in the official CRAN archive. In this chapter we also presented several specific issues, used in package implementation, like initial values selection and estimates variance's calculation. The package can be considered as a part of the practical utility of this research.

The derived MLE and the developed package are validated. We compared estimates of a private case of the SSF model with popular software that designed for classical stochastic frontier model and found them almost identical. Also we organised a set of simulation experiments, which allows investigating of the SSF model estimate properties for different specifications and sample sizes. According to the executed simulations, the derived estimator provides statistically unbiased and consistent estimates and allows confidently distinguish between different types of spatial effects; a range of other practically useful conclusions can be found in the chapter.

4. EMPIRICAL STUDY OF THE EUROPEAN AIRPORT INDUSTRY

4.1. Description of the research methodology

4.1.1. Collection of data sets

Taking features of airports data, discussed in the paragraph 1.1.1, into account, we formulated the following critical principles of compiling research data sets:

- Consistency, so data set variables are calculated using the same methodology for all objects in a sample. This requirement, usual for regression analysis, plays an important role for frontier approaches to efficiency estimation. A required data set should include all variables, necessary for at least one of frontier definitions (physical or financial).
- Geographical completeness of a dataset, so all neighbour airports are presented in the dataset. This requirement is inherited from spatial econometrics, where presence of a complete spatial structure is considered as an essential requirement.
- Availability of individual airport data. Frequently an operator company, which manages several airports, provides information in an aggregated form. Disaggregated information about sample airports is an essential requirement for this research.

Due to a lack of a data set of European airports, which satisfy all critical principles, we constructed a database with airports information to be used in this research (the entity-relationship diagram of the database is presented in the Appendix 7). Collected data is received from public data sources only; no private information is used.

A list of utilised data sources includes:

- The Eurostat (the Statistical Office of the European Community) database[211] (referred as Eurostat) is mainly used as a source of information about airport activities (PAX, ATM, cargo) and infrastructure facilities (check-in desks, gates, runways, and parking spaces).
- Individual airports' annual reports (referred as Reports) as a supplementary source of airport activity and infrastructure facilities information.
- The Digital Aeronautical Flight Information File database[212] (referred as DAFIF) as a source of airports' geographical coordinates.
- Google Maps as a supplementary source of geographical information (used mainly for presentation purposes).
- The OpenFlights/Airline Route Mapper Route database[213] (referred as OpenFlights) as a source of routes, served by airports.

- The Gridded Population of the World database from the Centre for International Earth Science Information Network[214] (referred as CIESIN) as a source of population counts in 2005, adjusted to match totals. The raster data contains information about Europe population with 2.5 arc-minutes (~5 kilometres) resolution.

Also we collected some data from country-specific data sources:

- The auditor's report, provided by Spanish airports operator (referred as AENA) as a source of Spanish airports' financial information.
- UK airports' annual financial statements (referred as Financial Statements), ordered from Companies House, a UK registry of company information, as a source of UK airports' financial information.
- Data set, collected by Tsekeris[215], as a separate source of Greece airports' information (referred as Tsekeris).

Finally, 4 data sets are compiled:

- European airports data set, 359 European airports, 2008-2012;
- Spanish airports data set, 38 Spanish airports, 2009-2010;
- UK airports data set, 48 UK airports, 2011-2012;
- Greek airports data set, 42 Greek airports, 2007.

Later in this paragraph we present a complete technical description of the collected data sets. All collected data sets are publicly available as a part of the *spfrontier* package, developed by the author. A complete description of the *spfrontier* package is presented in the chapter 3.3.

4.1.2. Specification of the contiguity matrix

All spatial techniques, used in this research, require formulation of the contiguity matrix *W*, whose components w_{ij} are metrics of spatial relation between objects *i* and *j*. A correct specification of the contiguity matrix is a complicated task itself, so there are a number of different approaches, which can be applied depending on the application area and spatial object types. Two frequently used types of spatial objects are objects with area and borders and point objects (without area). A discussion about alternative approaches to specification of the contiguity matrix for different object types is presented in the paragraph 2.3.3.

In this research we consider airports as point objects and specify a spatial weight between airports *i* and *j* as a distance linear decay. We used a great circle geographical distance as a metric of relation between airports, with linear distance decay function:

$$w_{ij} = \frac{1}{\text{distance}\left(airport_i, airport_j\right)} \tag{4.1}$$

Frequently for calculation purposes the matrix is row-standardised, so matrix values are divided by row sums. This approach is widely acknowledged as a standard in practice of spatial econometrics. Nevertheless the row-standardisation procedure is mainly grounded on computational issues (calculations for row-standardised matrixes are simpler), but not on real process specifications. An interesting discussion on row-standardisation effects can be found at [216]. Technically row-standardisation makes an airport, which has a small number of neighbours, "closer" to its neighbours. In our opinion, this fact is very arguable for the airport industry, so we decided to use non-standardised weights in this research.

4.1.3. PFP indexes and spatial correlation testing

PFP indexes are one of the simplest approaches to analysis of airport efficiency. This approach is not related to overall airport's efficiency, but reflects a particular aspect of its activity. In this research we used a number of PFP indexes, separated into two general groups – technical and economic.

Technical PFP indexes:

- ATM/PAX/WLU per runway,
- ATM/PAX/WLU per route,
- PAX per capita in 100 km area around an airport.

First three indexes (per runway) are usual infrastructure productivity indicators. In our samples, numbers of airport infrastructure elements (runways, gates, check-in facilities) are highly correlated, so one of them (runways) is selected for this thesis arbitrary. These indicators have a problem, based on data availability and compatibility. Statistics on infrastructure is not available for all airports in our datasets, and values can be hardly compatible, where they are available. We use just number of runways for these indicators, but runways themselves can be quite different – by length, surface, or area. To override (at least partly) these problems, we introduced route-based indicators. Number of routes, served by an airport, is available for all sample objects from the OpenFlights database[213], and generally compatible. The last indicator is not really technical, but utilise number of inhabitants around an airport as its "resource".

Economic PFP indexes:

- WLU per employee cost;
- Revenue per WLU/ATM;
- EBITDA per WLU/ATM;
- EBIDTA per revenue.

Economic PFP indexes are used for smaller data sets (Spanish and UK airports), where financial data is publicly available.

We applied the following statistical procedures to discover spatial relationships between values of selected PFP indicators:

- Moran's I test,
- Geary's C test,
- Mantel permutation test.

All used approaches are well-known in spatial data analysis and their formal description can be found in literature (e.g., [47], [217]).

4.1.4. Spatial model specifications

A general specification of the model, investigated in this research, can be formulated as an SSF(1,0,1,1) model with half-normal inefficiency:

$$\begin{aligned} &Y = \rho_Y W_Y Y + X\beta + v - u, \\ &v = \rho_v W v + \tilde{v}, \tilde{v} \sim MVN\left(0, \sigma_{\tilde{v}}^2 I\right), \\ &u = \rho_u W u + \tilde{u}, \tilde{u} \sim TMVN\left(0, \sigma_{\tilde{u}}^2 I\right). \end{aligned} \tag{4.2}$$

All used notations are described in the chapter 3 of this thesis.

A set of analysed private cases of the model include (all models are matter of parameter restrictions):

1. OLS regression model:

   $\rho_Y = 0, \rho_v = 0, \rho_u = 0, \sigma_{\tilde{u}}^2 = 0$

2. Spatial autoregressive (SAR) model:

   $\rho_v = 0, \rho_u = 0, \sigma_{\tilde{u}}^2 = 0$

3. Spatial error model (SEM):

   $\rho_Y = 0, \rho_u = 0, \sigma_{\tilde{u}}^2 = 0$

4. Classical non-spatial stochastic frontier (SF) model:

   $\rho_Y = 0, \rho_v = 0, \rho_u = 0$

5. SSF(1,0,0,0) model

   $\rho_v = 0, \rho_u = 0$

6. SSF(0,0,1,0)

   $\rho_Y = 0, \rho_u = 0$

7. SSF(0,0,0,1)

   $\rho_Y = 0, \rho_v = 0$

8. SSF(1,0,1,0)

   $\rho_u = 0$

9. SSF(1,0,0,1)

$$\rho_v = 0$$

An inheritance diagram of the evaluated models is presented on the Fig. 4.1.

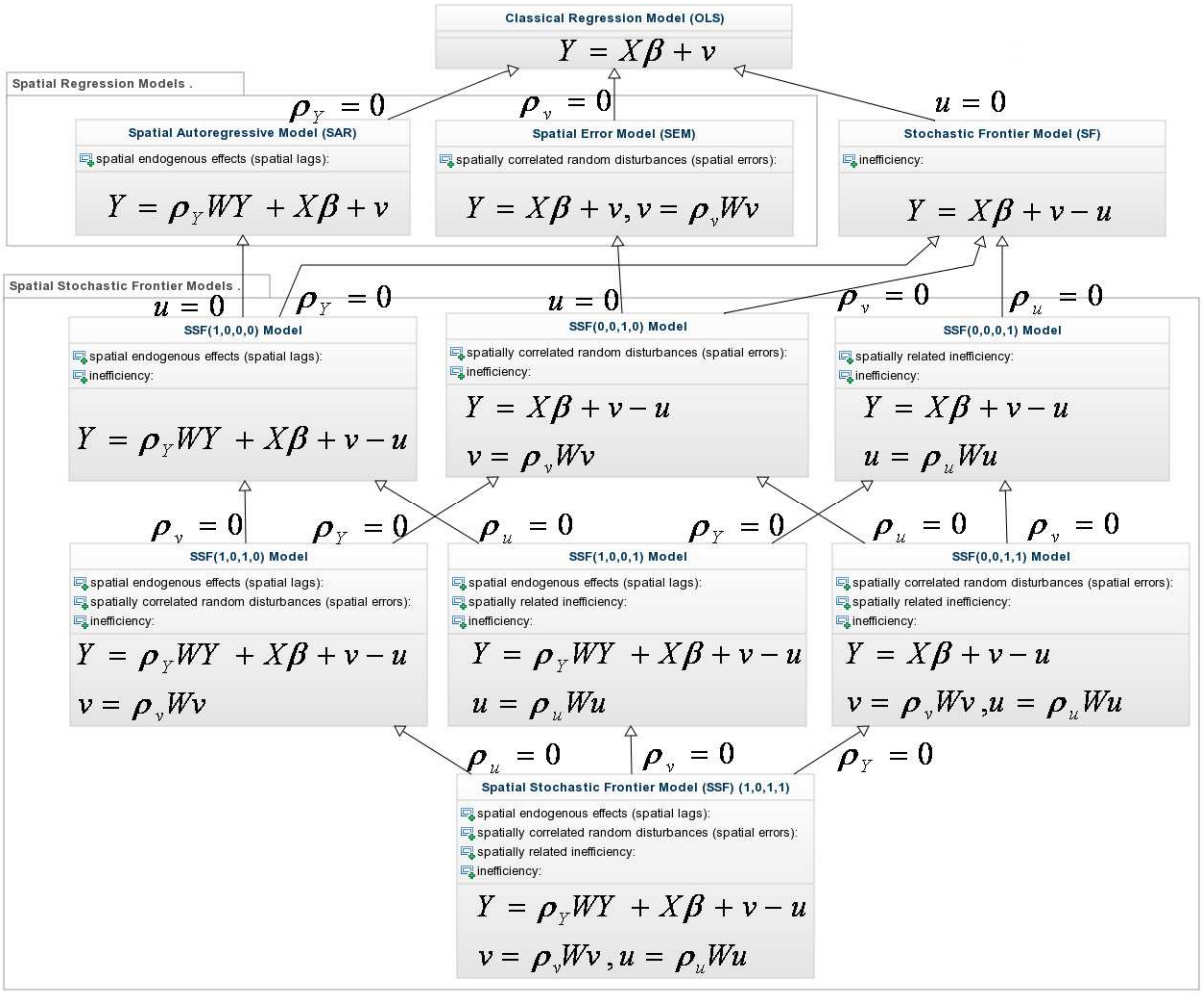


Fig. 4.1. Inheritance diagram of the research models

Also standard econometric techniques are used for multicolleniarity diagnostics (VIF), model comparison (likelihood ratio tests), and others. Spatial correlation of model residuals is tested using classic and robust Lagrange multiplier diagnostics[11].

4.2. Empirical analysis of European airports

4.2.1. Data set description

This data set includes information about airports in Europe in 2008-2012. Mainly the data set is based on information, received from the Eurostat and Open Flights databases, and includes indicators of airports' traffic and infrastructure. The panel is unbalanced with the most complete data for 2011. Consistent financial information is not available for all European airports and not included into this data set. A list of data set variables is presented in the Table 4.1 and supplemented with each variable's data source.

Table 4.1. Description of the European airports data set

Country	30 European countries		
Number of airports	359		
Years	2008-2012		
Panel	Unbalanced		
Variables			
	Variable	*Description*	*Source*
	ICAO	ICAO code	DAFIF
	AirportName	Airport official name	DAFIF
	longitude	Airport longitude	DAFIF
	latitude	Airport latitude	DAFIF
	Year	Observation year	
	PAX	A number of carried passengers	Eurostat, Reports
	ATM	A number of air transport movements served by an airport	Eurostat, Reports
	Cargo	A total volume of cargo served by an airport	Eurostat, Reports
	Population100km	A number of inhabitants, living in 100 km around an airport	CIESIN
	Population200km	A number of inhabitants, living in 200 km around an airport	CIESIN
	Island	1 if an airport is located on an island; 0 otherwise	Google Maps
	GDPpc	Gross domestic product per capita in airport's NUTS3 region	Eurostat
	RunwayCount	A number of airport runways	Eurostat, Reports
	CheckinCount	A number of airport check-in facilities	Eurostat, Reports
	GateCount	A number of airport gates	Eurostat, Reports
	ParkingSpaces	A number of airport parking spaces	Eurostat, Reports
	RoutesDeparture	A number of departure routes, served by an airport	OpenFlights
	RoutesArrival	A number of arrival routes, served by an airport	OpenFlights

Summary statistics of the data set variables are presented in the Appendix 8. The data set includes information about almost all significant airports in Europe. Spatial distribution of ATM values in the data set is presented on the Fig. 4.2.

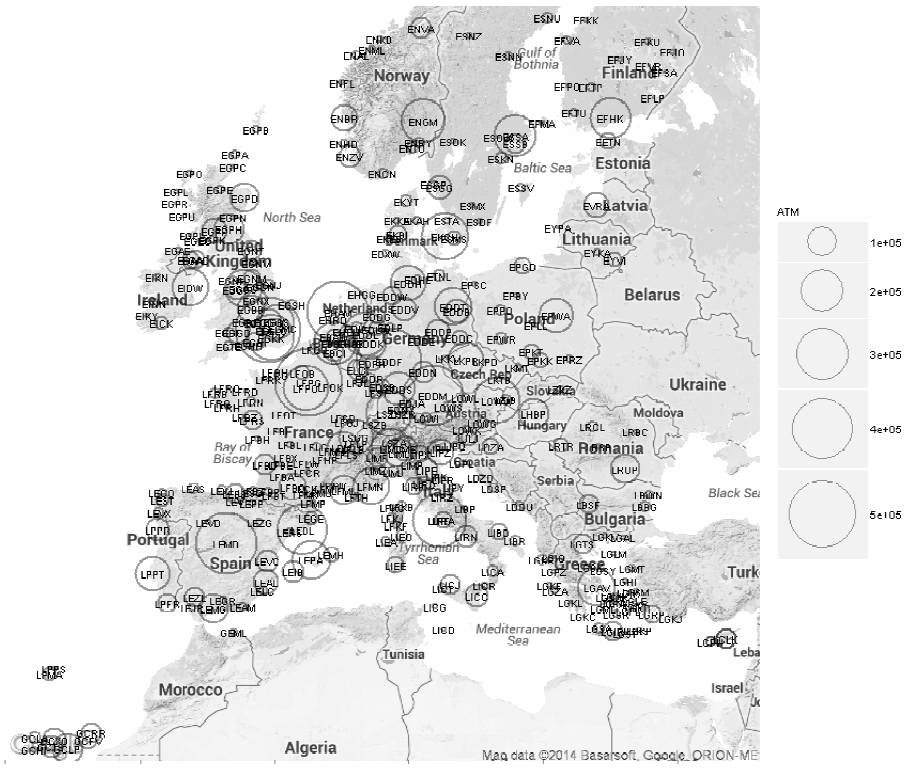


Fig. 4.2. ATM values in the European airports data set, 2011

4.2.2. Spatial analysis of airports' PFP indexes

As financial information is not available in the data set, this research is limited with physical (intermediary) approach to airport activity. According to this approach, outputs of an airport include numbers of ATM (as airports' result for air carriers) and number of carried passengers and volume of served cargo (as a result for population). Served passengers and cargo are frequently joined to the WLU indicator, which is more convenient for single-output modelling approaches. We used WLU for partial factor productivity measures.

The data set includes information about many characteristics, which can be classified as inputs within intermediary approach: numbers of runways, check-ins, gates, parking spaces, terminals. These resources can be considered separately to investigate a role of each infrastructure unit in airport productivity. However the indicators' values are logically correlated, because all infrastructure units are used for serving two general processes – handling passengers and cargo. Formally, this obvious statement is supported by the sample correlation matrix, presented in the Appendix 9. So in terms of benchmarking all infrastructure units can be represented as a joined indicator. Considering alternatives of selecting a natural indicator and an artificially composed one (which can be calculated using factor analysis techniques), we decided to use a total number of routes (both arrival and departure) as a proxy for all infrastructure units. This decision is based on three reasons: high level of correlation between number of served

routes and other infrastructure indicators, availability of data (in the OpenFlights database), and homogeneity of the indicator values.

The first step of research is spatial analysis of PFP indicators. A final list of PFP indicators, used for this data set, includes:

- ATM/PAX/WLU per Runway,
- ATM/PAX/WLU per Route,
- PAX per capita in 100 km.

Descriptive statistics of the PFP indicators are presented in the Appendix 10. A distribution pattern of all indicators' values is very similar, and a typical kernel density is presented on the Fig. 4.3.

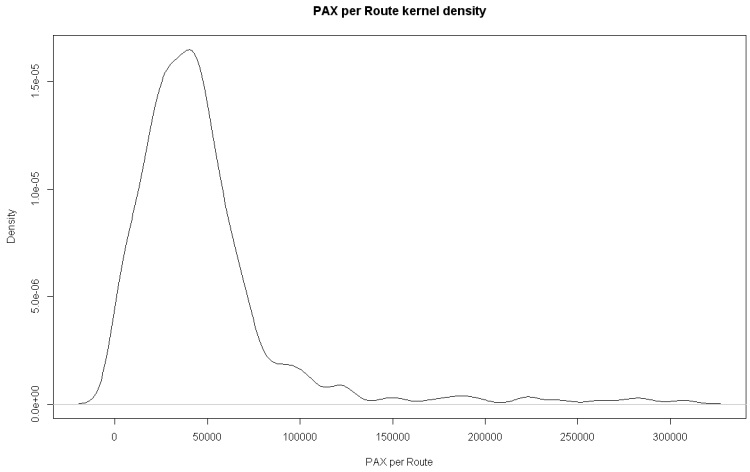


Fig. 4.3. Chart of an empirical kernel density function of the PAX per route ratio

Empirical distributions of all PFP indicators are positively skewed, due to a small number of airports with extremely high vales. Technically these airports can be classified as outliers, but in practice these airports can utilise the same business model. In this case their performance defines an important level, which can be useful for comparison, and this is preferred to keep them in sample. We executed all further tests both for a complete data set and for a dataset with excluded outliers and didn't find a significant difference in conclusions, so this thesis includes results for a complete sample only.

The primary goal of this research is to discover possible spatial patterns in airport benchmarking. The Table 4.2 contains results of Moran's I, Geary's C, and Mantel permutation tests for spatial autocorrelation between all considered PFP indicators.

Note that Moran's I and Mantel tests are designed to identification of global spatial autocorrelation, when Geary's C is sensitive to local autocorrelation.

Table 4.2. Results of spatial autocorrelation testing for PFP indicators of European airports

	Moran's I	*Geary's C*	*Mantel*
ATM per Runway	0.001 (0.578)	1.088** (0.040)	-0.08 (0.982)
WLU per Runway	0.003 (0.491)	1.096* (0.061)	-0.082 (0.980)
PAX per Runway	0.003 (0.491)	1.096* (0.061)	-0.082 (0.982)
ATM per Route	0.006 (0.128)	0.976 (0.511)	0.05* (0.056)
WLU per Route	0.024*** (0.000)	0.952 (0.142)	0.026 (0.191)
PAX per Route	0.024*** (0.000)	0.952 (0.142)	0.026 (0.203)
PAX per capita in 100 km	0.041*** (0.000)	0.707*** (0.000)	0.267*** (0.001)

Coefficients' p-values are presented in brackets.

Significant spatial autocorrelation is discovered for all considered indicators. Significant positive local autocorrelation is discovered for ATM/PAX/WLU per runway indicators, so it can be concluded that airports with higher and lower values of infrastructure performance are spatially clustered. Per-route indicators (WLU/PAX per route) also demonstrate similar spatial patterns, but for global autocorrelation. The only PFP indicator, which demonstrates both global and local positive spatial autocorrelation, is PAX per capita in 100 km. This result is the most expected as population is unevenly distributed over Europe.

Generally, all the conclusions match our expectations: there are a wide set of factors, which affect infrastructure performance of airport and unevenly distributed over space. These factors include country-specific legal features (antitrust laws, government regulation of airport industry, etc.), climate differences (e.g. snow-belt airports) and other issues, discussed in the chapter 1.

Note that executed spatial analysis of PFP indicators allows identification of an aggregate spatial effect in the sample, but doesn't provide information on different types of spatial relationships. Spatial heterogeneity and different types of spatial interactions have different nature, are likely to be oppositely directed and generally should not be aggregated. Further analysis, based on the SSF model, allows getting over this problem, separately identifying different types of spatial effects and enhancing the results.

4.2.1. The SSF analysis of European airports' efficiency and spatial effects

Two different specifications of the frontier are investigated in this research:

1. Single-output frontier, where the only airports' output is PAX. Models, based on this specification of the frontier, will be further referred as Model Europe1.
2. Multi-output frontier with two outputs: PAX and Cargo. These models will be referred as Model Europe2.

Model Europe1: single-output intermediary model

A final frontier specification of the Model Europe1 is formalised using the Cobb-Douglass function and has the following appearance:

$$\log(PAX) = \beta_0 + \rho_Y W \log(PAX) + \beta_1 \log(Routes) + \\ + \beta_2 \log(Population100km) + \beta_3 \log(GDPpc) \tag{4.3}$$

An initial list of explanatory variables included all airport infrastructure characteristics, available in the European airports data set – numbers of runways, check-in facilities, gates, and parking spaces. A high level of correlation between these characteristics leads to the multicolleniarity problem on regression models, so we decided to exclude them from the final specification. The Routes variables, included into the model, should be considered as a proxy for overall airport infrastructure.

Ten different specifications of the model, described in the paragraph 4.1.4, are estimated and analysed. Inheritance of the model specifications is presented on the Fig. 4.1. Calculated estimates of the models' parameters and necessary statistics are summarised in the Table 4.3.

In this research we applied the classical approach to model specification selection. According to this approach, we started with the simplest specification (OLS) and moved up to more complex specifications with inefficiency terms and spatial dependence on the base of statistical tests. A discussion about potential problems of the alternative "specific to general" (Hendry's) approach in spatial models can be found in[218].

An empirical kernel density of OLS residuals is presented on the Fig. 4.4.

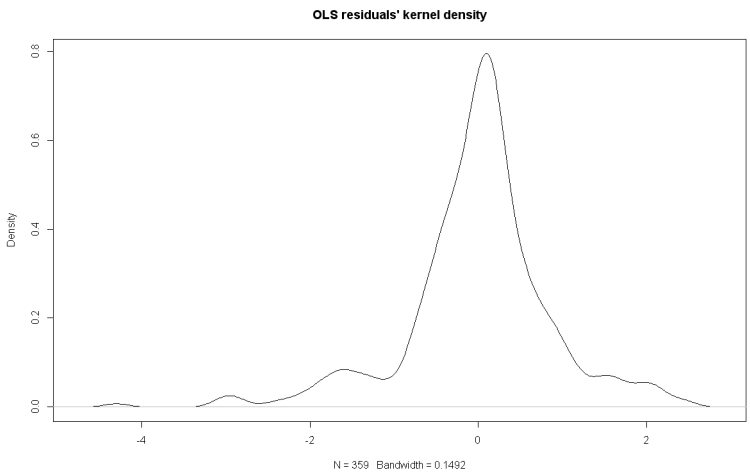


Fig. 4.4. Empirical kernel density of the Model Europe1 OLS residuals

A corresponding value of the OLS residuals skewness equals to –0.659. The asymmetric form of the density plot and the negative skewness can be explained by presence of inefficiency in data.

Table 4.3. Estimation results of the Model Europe1 alternative specifications

Model		*Intercept*	*log(Population100 km)*	*log(Routes)*	*log(GDP pc)*	σ_v	σ_u	ρ_Y	ρ_v	ρ_u
OLS	Estimate	10.776	0.031	1.143	-0.124					
	Std. Error	1.397	0.032	0.033	0.133					
	Sig.	$< 10^{-16}$	0.327	$< 10^{-16}$	0.352					
	Likelihood									-460.319
SAR	Estimate	9.883	0.124	1.118	-0.087			-0.002		
	Std. Error	1.406	0.044	0.033	0.131			0.001		
	Sig.	$< 10^{-16}$	0.005	$< 10^{-16}$	0.504			0.003		
	Likelihood									-456.054
SEM	Estimate	9.847	0.114	1.121	-0.094				0.025	
	Std. Error	1.508	0.045	0.033	0.138				0.002	
	Sig.	$< 10^{-16}$	0.011	$< 10^{-16}$	0.497				$< 10^{-16}$	
	Likelihood									-454.761
SF	Estimate	12.643	0.017	1.069	-0.183	0.569	1.079			
	Std. Error	1.352	0.032	0.035	0.124	0.053	0.100			
	Sig.	$< 10^{-16}$	0.602	$< 10^{-16}$	0.142	$< 10^{-16}$	$< 10^{-16}$			
	Likelihood									-450.748
SSF (1,0,0,0)	Estimate	11.697	0.094	1.058	-0.140	0.584	1.035	-0.001		
	Std. Error	1.390	0.045	0.034	0.125	0.053	0.102	0.001		
	Sig.	$< 10^{-16}$	0.035	$< 10^{-16}$	0.262	$< 10^{-16}$	$< 10^{-16}$	0.016		
	Likelihood									-447.806
SSF (0,0,1,0)	Estimate	11.728	0.025	1.063	-0.090	0.423	1.259		0.053	
	Std. Error	0.000	na*	0.000	na	0.000	0.000		0.000	
	Sig.	$< 10^{-16}$		$< 10^{-16}$		$< 10^{-16}$	$< 10^{-16}$		$< 10^{-16}$	
	Likelihood									-448.782
SSF (0,0,0,1)	Estimate	11.258	0.170	1.020	-0.188	0.409	1.219			0.041
	Std. Error	0.000	na	0.000	na	0.000	0.000			0.000
	Sig.	$< 10^{-16}$		0.000		$< 10^{-16}$	$< 10^{-16}$			$< 10^{-16}$
	Likelihood									-474.037
SSF (1,0,1,0)	***Estimate***	***12.199***	***0.068***	***1.091***	***-0.182***	***0.557***	***1.087***	***-0.001***	***0.043***	
	Std. Error*	***1.390***	***0.045***	***0.034***	***0.125***	***0.053***	***0.102***	***0.001***	***0.000***	
	Sig.	***$< 10^{-16}$***	***0.035***	***$< 10^{-16}$***	***0.262***	***$< 10^{-16}$***	***$< 10^{-16}$***	***0.016***	***$< 10^{-16}$***	
	Likelihood									***-444.253***
SSF (1,0,0,1)	Estimate	12.245	0.128	1.029	-0.216	0.556	1.054	-0.002		0.002
	Std. Error	0.001	0.000	na	na	0.000	0.000	na		na
	Sig.	0.000	0.000			0.000	0.000			
	Likelihood									-449.196
SSF (1,0,1,1)	Estimate	11.950	0.088	1.063	-0.159	0.572	1.028	-0.001	0.022	-0.004
	Std. Error	0.000	0.000	na	na	0.000	0.000	na	0.000	na
	Sig.	$< 10^{-16}$	$< 10^{-16}$			$< 10^{-16}$	$< 10^{-16}$		$< 10^{-16}$	
	Likelihood									-446.259

* "na" values mean that numerical estimates of corresponding standard errors are close to zero or negatives,

** standard errors for this model cannot be calculated numerically due to optimisation method limitations. Standard errors of parent models are presented for reference.

The classical SF model supports this conclusion, demonstrating significant estimate of the inefficiency standard deviation: σ_u = 1.079. A popular ratio metric for comparing standard deviations of the symmetric and inefficiency components equals to:

$$\gamma = \frac{\sigma_u^2}{\sigma_v^2 + \sigma_u^2} = 0.782,$$

so we conclude a significant share of inefficiency in variation of the model outcome.

Results of popular tests for spatial independence of OLS residuals are presented in the Table 4.4.

Table 4.4. Results of spatial independence testing of the Model Europe1 OLS residuals

Test statistic	*Value*	*p-value*	*Conclusion*
Moran's I	0.022	0.000	Positive spatial autocorrelation of OLS residuals
Lagrange multiplier test for spatial lags	8.224	0.004	Positive spatial lag in the OLS model
Lagrange multiplier test for spatial errors	10.404	0.001	Positive spatial errors in the OLS model

All tests support our hypothesis about presence of significant spatial effects in data.

Both classical SAR and SEM models provide statistically significant estimates of their specific types of spatial effects (Table 4.3). Note that the estimate of spatial endogenous effects parameter ρ_Y is significant and negative in the SAR model, which can be described as a negative influence of PAX traffic in neighbour airports on PAX traffic in a given one. Spatial errors are also found significant in the SEM model, but have a positive effect ($\rho_v > 0$). Spatial clustering of model random disturbances supports our hypothesis about spatial heterogeneity in airport industry. Note that both SAR and SEM model don't include inefficiency component in their specifications.

Finally having statistical evidences about presence of inefficiency and spatial effects in the data set, we estimated a number of different specifications of the proposed SSF model.

All estimated SSF model specifications with spatial endogenous effects (SSF(1,0,0,0), SSF(1,0,1,0), SSF(1,0,0,1), and SSF(1,0,1,1)) demonstrate significant negative effects of these types ($\rho_Y < 0$). It means that number of passengers, served by an airport, in average is negatively affected by its neighbour airports. Spatial competition for passengers is one of possible explanations of this phenomenon (see the chapter 1.3 for a corresponding discussion). It would be practically interesting to test a significance of these effects in data on a pre-liberalised airport industry (early nineties in Europe) and analyse its dynamics. These will require a panel data specification of the SSF model and its estimators and can be stated as a direction of further research.

Significant spatial correlations of random disturbances are also discovered in all corresponding model specifications (SSF(0, 0,1,0), SSF(1,0,1,0), and SSF(1,0,1,1)). A direction of these effects is positive as expected, so random disturbances have common parts for all airports, located within a particular area. The result can be explained by all spatial heterogeneity factors, discussed earlier – climate, legislative environment, population habits, etc.

Spatial effects in inefficiency components are not found as significant in all model specifications.

Selection of a model specification, which optimally fits the data, is based on the calculated values of a log-likelihood function (a formal likelihood ratio test can be applied). We selected the SSF(1,0,1,0) with the log-likelihood value –444.253 as the best model specification, and used this for further analysis.

Frontier parameter estimates (which are elasticities of resources in the Cobb-Douglass specification of the frontier function) match our initial expectations. A coefficient β_1 for number of routes equals to 1.091 and states that elasticity of airport infrastructure (represented with the *Routes* variable in the model) have a slightly over the unit elasticity. Significant positive effects of population, living in 100 km area from an airport (*Population100km*), are also easily explained by common sense. GDP per capita (*GDPpc*) in airport's NUTS3 region doesn't affect passenger traffic significantly.

One of the advantages of stochastic frontier approach is estimation of unit-specific efficiency values. We applied formulas, developed for the SSF model in the paragraph 3.2 and implemented in the *spfrontier* package, to estimate efficiency levels of airports in the sample. A complete list of efficiency values is presented in the Appendix 11; their empirical distribution is presented on the Fig. 4.5.

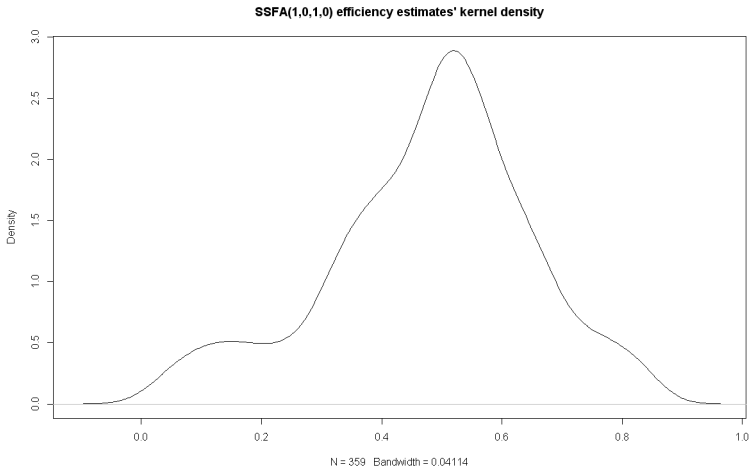


Fig. 4.5. Empirical kernel density of the Model Europe1 SSF(1,0,1,0) efficiency estimates

We conclude a significant level of inefficiency in data: a sample mean of efficiency is 0.479, sample median is 0.502. These values looks underestimated due to several airports with very small efficiency values. Partly this can be explained by incomplete data set with non-random selection of airports. We included all airports, where data is available, in the sample, and availability of data is not the same for European countries. In particular, we analysed a complete list of Greek airports[215], including small regional ones. As a result estimated efficiency values of these small airports are close to zero, due to their distance to the frontier, mostly defined by average-sized airports. Although spatial specification of the model allows correctly handling of

spatial heterogeneity in data, these size-based heterogeneity is not always spatial and so can't be modelled completely within our specification of the model. Separate analysis of regional airports seems to be a further practically important application of the proposed model.

Another interesting point of this research in context of the SSF model development is comparison of efficiency estimates, provide by classical SF and SSF models. The Appendix 11 contains both values for all airports in the sample. In the Table 4.5 we compiled top ten airports with overestimated (SF efficiency values are higher that SSF efficiency values) and underestimated (SF efficiency values are lower than SSF ones).

Table 4.5. Comparison of efficiency estimates of the SF and SSF(1,0,1,0) models

	Country	ICAO	AirportName	PAX	SF efficiency values	SSF(1,0,1,0) efficiency values
Top 10 (underestimated)						
1	France	LFLP	Meythet	42875	0.377	0.515
2	France	LFSD	Longvic	44538	0.391	0.524
3	France	LFRG	St Gatien	119804	0.636	0.738
4	United Kingdom	EGBB	Birmingham	8606497	0.499	0.594
5	Switzerland	LSGG	Geneve Cointrin	13003611	0.399	0.490
6	France	LFMH	Boutheon	108648	0.426	0.514
7	France	LFOK	Vatry	50817	0.425	0.510
8	France	LFLL	Saint Exupery	8318143	0.405	0.490
9	United Kingdom	EGHH	Bournemouth	612499	0.588	0.671
10	France	LFBE	Roumaniere	290020	0.492	0.575
Last 10 (overestimated)						
350	Spain	LEBL	Barcelona	34314376	0.555	0.482
351	Bulgaria	LBSF	Sofia	3465823	0.355	0.281
352	Italy	LICC	Catania Fontanarossa	6771238	0.554	0.480
353	Italy	LIBD	Bari	3700248	0.523	0.448
354	Romania	LROP	Henri Coanda	5028201	0.358	0.276
355	Greece	LGKF	Kefallinia	346397	0.624	0.539
356	Spain	LEAL	Alicante	9892302	0.516	0.430
357	Greece	LGIO	Ioannina	88597	0.570	0.482
358	Greece	LGTS	Makedonia	3958475	0.511	0.419
359	Greece	LGAV	Eleftherios Venizelos Intl	14325505	0.535	0.428

There are two opposite directions of efficiency changes discovered by the SSF model. Firstly, the SSF model provides higher values of airport efficiency, located in a more competitive environment (due to significant negative spatial endogenous effects). At the same time, the SSF model takes spatial heterogeneity into account (which is discovered as positive in this data set), which leads to lower efficiency values. As an aggregate result, the SSF model provided lower efficiency values for relatively isolated airports (Greek, Italian), and higher values for French and UK airports.

Note that the presented results should be considered only as preliminary ones, which discover spatial effects in data, but require more detailed analysis for practical usage.

Model Europe2: multi-output intermediary model

Model Europe2 also utilises the intermediary approach to airport activity and is based on a multi-output frontier with two outputs, PAX and Cargo. The final frontier specification for the ModelEurope2 is formulated as:

$$-\log(PAX) = \beta_0 + \rho_Y W \log(PAX) + \beta_1 \log(Cargo/PAX) + \\ + \beta_2 \log(Routes) + \beta_3 \log(Population100km) + \beta_4 \log(GDPpc) \quad (4.4)$$

See the paragraph 2.1 for a detailed description of the multi-output frontier specification. Note that the dependent variable in the model is negative, so estimated values of the β coefficients have an opposite direction of influence on airports' PAX. Also the composed random term in this case is considered as a sum: $\varepsilon = v + u$, so the model is estimated with a cost-oriented frontier instead of its natural production-oriented frontier. Different specifications of the Model Europe2, described in the paragraph 4.1.4, are estimated and analysed(Table 4.6).

Table 4.6. Estimation results of the Model Europe2 alternative specifications

Model		*Intercept*	*log(Cargo/PAX)*	*log(Population100km)*	*log(Routes)*	*log(GDPpc)*	σ_v	σ_u	ρ_Y	ρ_v	ρ_u
OLS	Estimate	-10.527	0.035	-0.030	-1.159	0.127	0.876				
	Std. Error	1.407	0.026	0.032	0.035	0.132					
	Sig.	0.000	0.171	0.340	$< 10^{-16}$	0.337					
	Likelihood										-459.36
SAR	Estimate	-9.653	0.034	-0.123	-1.133	0.091			-0.002		
	Std. Error	1.413	0.025	0.044	0.035	0.131			0.001		
	Sig.	0.000	0.180	0.006	$< 10^{-16}$	0.483			0.003		
	Likelihood										-455.15
SEM	Estimate	-9.684	0.029	-0.111	-1.134	0.100				0.025	
	Std. Error	1.512	0.025	0.044	0.035	0.138				0.002	
	Sig.	0.000	0.255	0.012	$< 10^{-16}$	0.471				$< 10^{-16}$	
	Likelihood										-454.11
SF	Estimate	-12.384	0.026	-0.017	-1.083	0.182	0.573	1.070			
	Std. Error	1.376	0.026	0.032	0.037	0.125	0.054	0.101			
	Sig.	$< 10^{-16}$	0.304	0.582	$< 10^{-16}$	0.145	$< 10^{-16}$	$< 10^{-16}$			
	Likelihood										-450.22
SSF (1,0,0,0)	Estimate	-11.444	0.026	-0.095	-1.072	0.140	0.587	1.026	-0.001		
	Std. Error	1.412	0.025	0.045	0.036	0.125	0.054	0.103	0.001		
	Sig.	0.000	0.296	0.033	$< 10^{-16}$	0.264	$< 10^{-16}$	$< 10^{-16}$	0.016		
	Likelihood										-447.26
SSF (0,0,1,0)	Estimate	-11.350	0.018	-0.061	-1.145	0.185	0.816	0.310		0.031	
	Std. Error	0.000	0.000	na	0.000	0.000	na	0.000		0.000	
	Sig.	$< 10^{-16}$	$< 10^{-16}$		$< 10^{-16}$	$< 10^{-16}$		$< 10^{-16}$		$< 10^{-16}$	
	Likelihood										-454.23
SSF (1,0,1,0)	***Estimate***	***-11.376***	***0.024***	***-0.095***	***-1.072***	***0.140***	***0.573***	***1.034***	***-0.001***	***0.045***	
	Std. Error	***0.000***	***0.000***	***na***	***0.000***	***na***	***0.000***	***0.000***	***na***	***0.000***	
	Sig.	***$< 10^{-16}$***	***$< 10^{-16}$***		***$< 10^{-16}$***		***$< 10^{-16}$***	***$< 10^{-16}$***		***$< 10^{-16}$***	
	Likelihood										***-443.20***
SSF (1,0,0,1)	Estimate	-11.090	0.051	-0.117	-1.055	0.130	0.568	1.014	-0.002		-0.002
	Std. Error	0.000	0.000	na	na	0.000	0.000	0.000	0.000		0.000
	Sig.	$< 10^{-16}$	$< 10^{-16}$			$< 10^{-16}$	$< 10^{-16}$	$< 10^{-16}$	$< 10^{-16}$		$< 10^{-16}$
	Likelihood										-448.62

* "na" values mean that numerical estimates of corresponding standard errors are close to zero or negatives.

Note that in many model specifications the coefficient for *log(Cargo/PAX)* is found insignificant. Generally it means that taking the *Cargo* variable into the model doesn't improve its quality and this component can be excluded. Excluding of the *log(Cargo/PAX)* from the Model Europe2 reduces it to the Model Europe1. Thus the Model Europe2 is very similar to the Model Europe1 for our data set, and the most of conclusions, described in the previous paragraph, hold true. So we just state the modelling process conclusions without their detailed description:

- Tests for spatial autocorrelation of OLS residuals provide strong evidences for spatial effects. Moran's I value equals to 0.02, Lagrange multiplier diagnostics equal to 8.11 and 9.58 for spatial lags and spatial errors respectively; all values are statistically significant.
- Classical spatial regression models (SAR and SEM) indicate significant spatial endogenous effects and spatially correlated random disturbances in data. Spatial lags are found negative, and spatial errors are found positive, which keeps all the conclusions about spatial competition and spatial heterogeneity made for the Model Europe1.
- The classical stochastic frontier modelling indicates significant inefficiency in data. Skewness of OLS residuals equals to 0.648, and its positive value indicates presence of inefficiency in data for cost-oriented frontiers. A share of inefficiency in a variance of the composed error term is γ=0.77, which also supports the hypothesis about inefficiency in data.
- SSF specifications of the model support presence of both spatial effects and inefficiency in data. Similar to the Model Europe1, the SSF(1, 0, 1, 0) model specification shows the best performance according to the likelihood ratio tests. So we conclude significant negative spatial endogenous effects and also spatial heterogeneity in airport industry.

Summarising executed spatial analysis of European airports, we state that:

1. Significant spatial autocorrelation is discovered for all considered partial factor productivity indicators – ATM/PAX/WLU per runway/per route and PAX per capita in a catchment area. These spatial effects appear due to uneven distribution over space of different performance-related factors like climate and legal and economic environment.
2. Stochastic frontier analysis shows presence of inefficiency in data both for single-output and multi-output frontier specification.
3. Spatial stochastic frontier model SSF(1,0,1,0) is selected as the best model specification for the research data set. This fact supports one of the main assumption of this thesis about advantages of simultaneous consideration of spatial and inefficiency effects.

4. Different types of spatial effects are identified as significant using the SSF model. In particular, we discovered statistically significant negative endogenous spatial effects, which can be explained by spatial competition for passengers and cargo flows between neighbour airports. Spatial correlation between model random disturbances is also estimated as significant and positive, which can be a consequence of unobserved area-specific factors' influence. Finally, spatial effects between inefficiency values are not discovered for the research data set.

4.3. Empirical analysis of Spanish airports

4.3.1. Data set description

This data set includes traffic, infrastructure and financial information about Spanish airports in 2009-2010. The Spanish airport industry is fairly monopolistic; all 47 commercial airports in Spain are managed by a public company AENA, dependent on the Ministry of Transports. Usually the airport operator provides annual reports with aggregated financial information, so it is frequently impossible to receive airport-specific values. Recently disaggregated data on Spanish airports was released to the public by the Ministry of Public Works as a support for debates over management of the public airport system. This data set includes figures from an auditing report, compiled by the Spanish National Accounting Office[219]. This publicly available report provides financial data on 42 out of the 48 public airports in Spain for 2009 and 2010.

The data set includes 38 airports (4 airports were excluded as almost not acting) and is supplemented with traffic and infrastructure data, collected from the Eurostat and Open Flights databases. Besides the main airport, Madrid Barajas, where traffic flows are considerably explained both by economic activity and tourism, there are a wide range of airports, mainly served tourist flows and located near the seaside and on islands. A extensive description of the Spanish airport industry can be found in [220]. The Table 4.7 presents a technical description of the data set.

Summary statistics of the data set variables and a list of sample airports are presented in the Appendix 12. Financial information in the data set includes total revenue, EBITDA, and net profit values, and deprecation and amortization costs.

Table 4.7. Description of the Spanish airports data set

Country	Spain		
Number of airports	38		
Years	2009-2010		
Panel	Balanced		
Variables			
	Variable	*Description*	*Source*
	ICAO	ICAO code	DAFIF
	AirportName	Airport official name	DAFIF
	longitude	Airport longitude	DAFIF
	latitude	Airport latitude	DAFIF
	Year	Observation year	
	PAX	A number of carried passengers	Eurostat, Reports
	ATM	A number of air transport movements served by an airport	Eurostat, Reports
	Cargo	A total volume of cargo served by an airport	Eurostat, Reports
	Population100km	A number of inhabitants, living in 100 km around an airport	CIESIN
	Population200km	A number of inhabitants, living in 200 km around an airport	CIESIN
	Island	1 if an airport is located on an island; 0 otherwise	Google Maps
	RevenueTotal	Airport total revenue	AENA
	EBITDA	Airport earnings before interest, taxes, depreciation, and amortization	AENA
	NetProfit	Airport net profit	AENA
	DA	Airport depreciation and amortization	AENA
	StaffCost	Airport staff cost	AENA
	RunwayCount	A number of airport runways	Eurostat, Reports
	TerminalCount	A number of airport terminals	Eurostat, Reports
	RoutesDeparture	A number of departure routes, served by an airport	OpenFlights
	RoutesArrival	A number of arrival routes, served by an airport	OpenFlights

4.3.2. Spatial analysis of airports' PFP indexes

There are two groups of PFP indicators of Spanish airports' activity discussed in this research – technical and financial indicators. Technical PFP indicators are constructed on the base of physical airport infrastructure and traffic characteristics. We described issues, related with construction of technical PFP indicators, in the paragraphs 4.1.3 for the data set of European airports; for Spanish airports they are fairly similar. The list of technical PFP indicators includes:

- ATM per Route
- WLU per Route
- WLU per capita in 100 km

Availability of complete and comparable financial information is a feature of this data set, so the main point of our interest is financial PFP indicators. We used two main output indicators, Revenue and EBITDA, representing financial results of airport activity. Note that 25 airports in the sample have negative EBITDA values. We also excluded net profit values from analysis, because interests and taxes depend on previous investments and significantly vary for airports in the sample, so net profit doesn't represent efficiency at least for a short term.

A list of considered inputs is limited with a number of routes (Route) and WLU served by an airport and population within 100 km from the airport. The Route variable is highly correlated with infrastructure units of an airport (numbers of gates, check-ins, etc.) and is considered as a replacement variable for all of them (see the paragraph 4.2.1 for a more detailed discussion on this). Number of WLU represents total traffic, served by an airport. Population is a weak representative of airport's general economic and social environment.

Finally we selected the following list of financial PFP indicators:

- WLU per staff cost
- Revenue per Route/WLU
- Revenue per capita in 100 km
- EBITDA per Route/WLU/Revenue
- EBITDA per capita in 100 km

Each indicator represents a particular aspect of airport activity and their meanings are generally self-explaining. Descriptive statistics of the PFP indicators are presented in the Appendix 13.

One of research goals is to discovering possible spatial patterns in airport benchmarking. The Table 4.8 contains results of Moran's I, Geary's C, and Mantel permutation tests for spatial autocorrelation between all considered PFP indicators.

Table 4.8. Results of spatial autocorrelation testing for PFP indicators of Spanish airports

	Moran's I	Geary's C	Mantel
ATM per Route	0.078** (0.027)	1.017 (0.826)	-0.008 (0.446)
WLU per Route	0.092*** (0.009)	0.764** (0.015)	0.105 (0.139)
WLU per capita in 100 km	0.228*** (0.000)	0.687*** (0.000)	0.377*** (0.002)
WLU per staff cost	0.090** (0.011)	0.913 (0.170)	-0.002 (0.448)
Revenue per Route	0.069** (0.034)	0.934 (0.398)	0.119* (0.085)
Revenue per WLU	-0.072 (0.342)	1.059 (0.368)	0.138* (0.072)
Revenue per capita in 100 km	0.190*** (0.000)	0.653*** (0.001)	0.294** (0.015)
EBITDA per Route	0.091*** (0.003)	1.204* (0.096)	-0.052 (0.627)
EBITDA per WLU	0.060** (0.021)	1.307** (0.024)	-0.040 (0.545)
EBITDA per Revenue	0.034 (0.128)	1.164 (0.174)	0.024 (0.310)
EBITDA per capita in 100 km	0.117*** (0.000)	0.641*** (0.006)	0.464*** (0.001)

Coefficients' p-values are presented in brackets.

Significant spatial autocorrelation is discovered almost for all PFP indicators. Spatial effects are found positive for all cases, so values of PFP indicators are clustered. This conclusion is one of the most expected, because of a generally touristic nature of Spanish air traffic flows. Airports, situated on the sea-side and on islands, generally generate more revenue due to their locations (see the Fig. 4.6 for geographical distribution of EBIDTA).

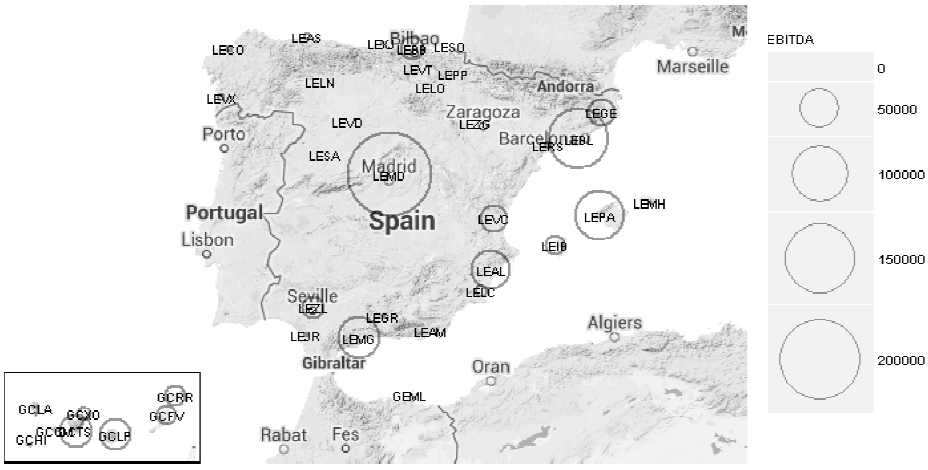


Fig. 4.6. EBITDA in the Spanish airports data set, 2010

Note that seaside and island airports demonstrate significantly higher values of EBIDTA, with exception of capital's Madrid-Barajas airport. This asymmetry in EBIDTA and revenue distribution is a likely background of discovered spatial effects.

Spatial analysis of PFP indicators allows identification of an aggregate spatial effect in the sample, but doesn't provide information on different types of spatial relationships. Further analysis, based on the SSF model, allows separately identifying different types of spatial effects and enhancing the results.

4.3.1. The SSF analysis of Spanish airports' efficiency and spatial effects

We investigated different frontier specifications for this data set, and generally obtained similar results. A final frontier specification, which was selected for presentation in this thesis, is formalised using the Cobb-Douglass function and has the following appearance:

$$\log(Revenue) = \beta_0 + \rho_Y W \log(Revenue) + \beta_1 \log(PAX) + \\ + \beta_2 \log(TerminalCount) + \beta_3 \log(Population100km) \tag{4.5}$$

A general model with this frontier specification is referred as Model Spain.

Our general approach to spatial stochastic frontier analysis of airport contains ten different model specifications, described in the paragraph 4.1.4 and on the Fig. 4.1. Calculated estimates of the models' parameters and necessary statistics are summarised in the Table 4.9.

Table 4.9. Estimation results of the Model Spain alternative specifications

Model		Intercept	log(PAX)	log(TerminalCount)	log(Population100km)	σ_v	σ_u	ρ_Y	ρ_v
OLS	Estimate	-3.753	0.872	0.394	0.066	0.360			
	Std. Error	0.732	0.037	0.219	0.040				
	Sig.	0.000	$< 10^{-16}$	0.082	0.108				
	Likelihood								-12.656
SAR	Estimate	-4.544	0.916	0.285	0.020			0.011	
	Std. Error	0.728	0.038	0.199	0.040			0.005	
	Sig.	0.000	$< 10^{-16}$	0.152	0.619			0.019	
	Likelihood								-10.139
SEM	***Estimate***	***-3.465***	***0.863***	***0.452***	***0.054***				***-0.174***
	Std. Error	***0.483***	***0.024***	***0.188***	***0.022***				***0.072***
	Sig.	$< 10^{-16}$	$< 10^{-16}$	***0.016***	***0.014***				***0.016***
	Likelihood								***-9.755***
SF	Estimate	-3.745	0.872	0.394	0.066	0.341	0.010		
	Std. Error	0.780	0.035	0.207	0.038	0.040	0.451		
	Sig.	$< 10^{-16}$	$< 10^{-16}$	0.057	0.080	$< 10^{-16}$	0.983		
	Likelihood								-12.657
SSF (1,0,0,0)	Estimate	-4.540	0.916	0.284	0.020	0.318	0.010	0.011	
	Std. Error	0.819	0.038	0.199	0.040	0.037	0.465	0.005	
	Sig.	$< 10^{-16}$	$< 10^{-16}$	0.153	0.621	$< 10^{-16}$	0.984	0.020	
	Likelihood								-10.139

Despite the initial assumption about presence of inefficiency in data, supported by previous researches[220], the simple OLS specification is almost perfect for the Model Spain:

$$R^2_{adj} = 0.9602$$

This high goodness of fit value for OLS model can be considered as a first evidence of absence of inefficiency in data. The distribution of OLS residuals, presented on the Fig. 4.7, supports this conclusion.

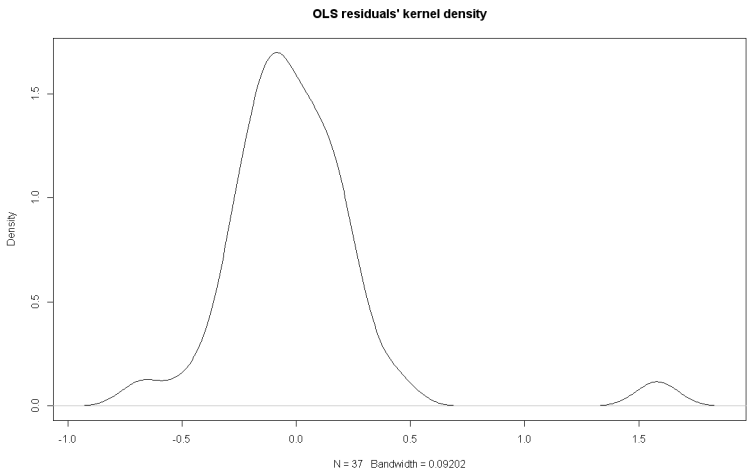


Fig. 4.7. Empirical kernel density of the Model Spain OLS residuals

Residuals are almost symmetric (except of an outlier on the right side) or even right-skewed (a value of sample skewness of OLS residuals is 2.488). Right-skewed OLS residuals for production stochastic frontier also support the conclusion about absence of inefficiency in data.

This fact was finally supported by insignificant estimates of σ_u in classical SF and the SSF(1,0,0,0) models. This result can be easily explained by a natural feature of the stochastic frontier analysis – it estimates inefficiency as a distance to the frontier, constructed on the base of other objects in the same sample. So if all sample objects have the same frontier and similar inefficiency values (even large), the SF model provides the absence of inefficiency. Note that all Spanish airports are managed by the same operator (AENA) and likely use similar principles in traffic handling and revenue forming, the absence of inefficiency in data becomes well-grounded.

So our further analysis was oriented on models with spatial effects and symmetric error terms – SAR and SEM models. The SSF(1,0,0,0) model, also presented in the Table 4.9, has insignificant inefficiency component and reduced to the simpler SAR model.

Results of formal Moran's I and Lagrange multiplier tests for spatial effects in the OLS model residuals are presented in the Table 4.10.

Table 4.10. Results of spatial independence testing of the Model Spain OLS residuals

Test statistic	*Value*	*p-value*	*Conclusion*
Moran's I	-0.111	0.072	Weakly significant negative spatial autocorrelation of OLS residuals
Lagrange multiplier test for spatial lags	4.806	0.020	Positive spatial lag in the OLS model
Lagrange multiplier test for spatial errors	3.263	0.070	Weakly significant positive spatial errors in the OLS model

Analysing estimated spatial models, we note significant spatial effects both in SAR and SEM models. Spatial heterogeneity looks more probable than spatial lags for monopolistic Spanish airport industry. Also the SEM model demonstrated a slightly better statistical performance (-9.755 log-likelihood for the SEM model versus -10.139 for the SAR, which is not a statistically significant difference) and better match our expectations, so the SEM model is selected as the best specification.

Results of the SEM model are generally expected. All three inputs (PAX, TerminalCount, and Population100km) have positive significant elasticities (0.863, 0.452, and 0.054 respectively), which support our choice of traffic, infrastructure, and environment as important resources of an airport. A coefficient p_v for spatial heterogeneity is significantly negative. This fact indicates a chess board-type pattern of spatial distribution of revenue. Generally, this pattern is quite rare for spatial heterogeneity (see a related discussion in [221]) and require additional research.

Summarising executed spatial analysis of Spanish airports, we state that:

- Positive spatial relationships are peculiar to partial factor productivity of Spanish airports, so airports are geographically clustered in respect to the following PFP

indicators: ATM per Route; WLU per Route, per capita in 100 km, per staff cost; Revenue per Route, per capita in 100 km; EBITDA per Route, per WLU, per capita in 100 km.

- Inefficiency is absent in the data set (in respect to the selected specification of the frontier). This result is explained by the comparative SFA approach to inefficiency estimation and monopolistic structure of the Spanish airport industry.
- Significant spatial effects are discovered in data, both in forms of spatial lags and spatial heterogeneity. Spatial heterogeneity is discovered as more statistically significant and the SEM model is preferred for further analysis.

4.4. Empirical analysis of UK airports

4.4.1. Data set description

This data set includes traffic, infrastructure and financial information about UK airports in 2011-2012. UK airports are generally concentrated in the North West of the country, in area with higher population density and economic activity. After a set of airport sales and acquisitions, initiated by UK Competition Commission, airports are generally managed by of different operators (M.A.G., Heathrow Airport ltd., Stansted Airport ltd., Gatwick Airport ltd., London Luton Airport Operations ltd.). Different operators are supposed to act as competitors, enforcing economic efficiency of each other. Government regulation of UK airports is implemented on the base RPI-X approach[222]. We collected financial data on UK airports directly from their publicly available annual reports for 2011 and 2012 years. The UK airports subsample includes 48 airports, and full financial data are available only for 21 of them. The Table 4.11 presents a technical description of the data set.

Summary statistics of the data set variables and a list of sample airports are presented in the Appendix 14.

Financial information in the data set includes total, aviation and non-aviation revenue, EBITDA, deprecation and amortization costs, and staff costs. Spatial distribution of airports' EBITDA is presented on the Fig. 4.8.

A main feature of this data set (in respect to our research) is a relatively separated geographical position of sample objects (the British Isles), which allows considering them as independent from neighbour objects, not included into the sample. Also significant anti-monopolistic efforts of the UK government in the airport industry leaded to a more competitive environment, which is natural for efficiency estimation.

Table 4.11. Description of the UK airports data set

Country	United Kingdom		
Number of airports	48		
Years	2011-2012		
Panel	Balanced		
Variables			
	Variable	*Description*	*Source*
	ICAO	ICAO code	DAFIF
	AirportName	Airport official name	DAFIF
	longitude	Airport longitude	DAFIF
	latitude	Airport latitude	DAFIF
	Year	Observation year	
	PAX	A number of carried passengers	Eurostat, Reports
	ATM	A number of air transport movements served by an airport	Eurostat, Reports
	Cargo	A total volume of cargo served by an airport	Eurostat, Reports
	Population100km	A number of inhabitants, living in 100 km around an airport	CIESIN
	Population200km	A number of inhabitants, living in 200 km around an airport	CIESIN
	Island	1 if an airport is located on an island; 0 otherwise	Google Maps
	RevenueTotal	Airport total revenue	Financial Statements
	RevenueAviation	Airport aviation revenue	Financial Statements
	RevenueNonAviation	Airport non-aviation revenue	Financial Statements
	EBITDA	Airport earnings before interest, taxes, depreciation, and amortization	Financial Statements
	DA	Airport depreciation and amortization	Financial Statements
	StaffCost	Airport staff cost	Financial Statements
	StaffCount	A number of staff employed by an airport	Financial Statements
	RunwayCount	A number of airport runways	Eurostat, Reports
	TerminalCount	A number of airport terminals	Eurostat, Reports
	RoutesDeparture	A number of departure routes, served by an airport	OpenFlights
	RoutesArrival	A number of arrival routes, served by an airport	OpenFlights

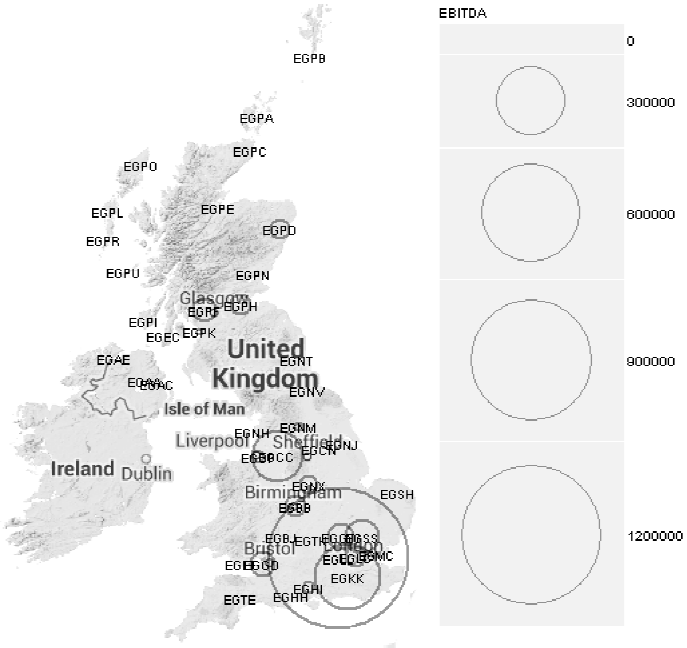


Fig. 4.8. EBITDA in the UK airports data set, 2012

4.4.2. Spatial analysis of airports' PFP indexes

Financial information is available for 21 UK airports. Comparison of financial PFP indicators' values of UK and Spain airports are presented in the Fig. 4.9 in a form of box plots.

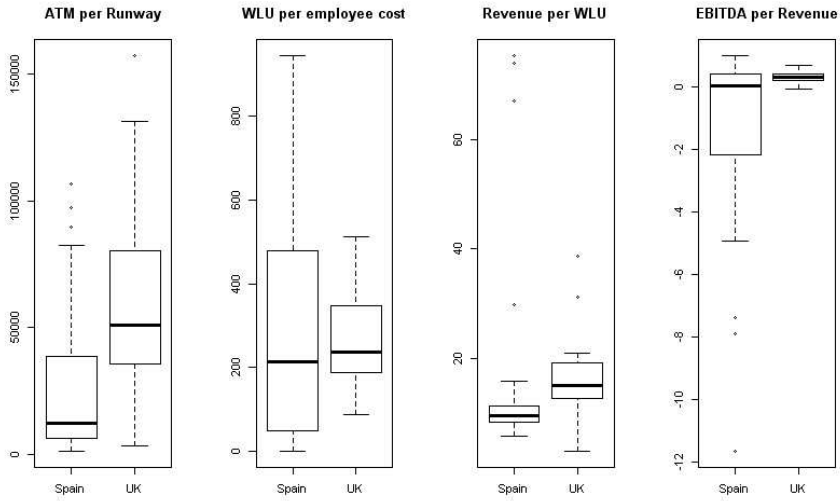


Fig. 4.9. Box plots of UK and Spanish airport PFP indicators

The box plots support our assumption about significant differences in performance of UK and Spanish airports. A level of airport infrastructure loading (ATM per runway) is significantly higher in UK airports; technical performance of employment (WLU per employee cost) is similar in both countries, but has larger variance within the Spanish airports sample. Revenue PFP (revenue per WLU) indicates higher financial performance of UK airports, but with higher variance between them. The most significant difference between two data sets is indicated for EBIDTA per revenue financial ratio. A significant share of Spanish airports provided negative values for EBITDA. These financial losses are explained by reduction of airline tourists flows after the world crisis, and a high level of dependence between these flows and Spanish airports activity (and economy of Spain in general).

A share of the UK airports sample with available financial information is too small for the extensive spatial stochastic frontier analysis (21 airports), so we concentrate on intermediary approach to the airport business and consider physical characteristics of airports and traffic flows.

A list of techincal PFP indicators includes:

- ATM per Runway
- WLU per Runway
- ATM per Route
- WLU per Route
- WLU per capita in 100 km

Descriptive statistics of the PFP indicators are presented in the Appendix 15. The Table 4.12 contains results of statistical tests for spatial autocorrelation between considered PFP indicators.

Table 4.12. Results of spatial autocorrelation testing for PFP indicators of UK airports

	Moran's I	*Geary's C*	*Mantel*
ATM per Runway	0.067* (0.063)	1.023 (0.845)	-0.003 (0.425)
WLU per Runway	0.061* (0.068)	1.136 (0.336)	-0.068 (0.652)
ATM per Route	0.013 (0.227)	0.904 (0.221)	-0.057 (0.812)
WLU per Route	0.079*** (0.001)	0.922 (0.103)	0.148*** (0.005)
WLU per capita in 100 km	-0.005 (0.525)	0.903 (0.137)	0.113 (0.053)

Although spatial autocorrelation is significant for some indicators (ATM/WLU per runway, WLU per route), generally spatial effects are weak. The only indicator with highly significant spatial autocorrelation is WLU per route, representing efficiency of infrastructure usage.

4.4.1. The SSF analysis of UK airports efficiency and spatial effects

The selected specification of the frontier has a standard Cobb-Douglass functional form:

$$\log(PAX) = \beta_0 + \rho_Y W \log(PAX) + \beta_1 \log(Routes) + \\ + \beta_2 \log(Population100km) + \beta_3 Island \tag{4.6}$$

A general model with this frontier specification is referred as Model UK.

A list of selected inputs includes a number of routes (as a proxy for all airport infrastructure units), population in 100 km around an airport (a proxy for economic and social environment), and a dummy for small island airports.

Different specifications of the model, described in the paragraph 4.1.4, are estimated and analysed. Calculated estimates of the models' parameters and necessary statistics are summarised in the Table 4.13.

According to the classical "general-to-specific" approach, we started with the simplest OLS model and enhanced it with inefficiency components and spatial effects if necessary. An empirical distribution of OLS residuals is presented on the Fig. 4.10.

Table 4.13. Estimation results of the Model UK alternative specifications

Model		*Intercept*	*log(Routes)*	*log(Population100km)*	*Island*	σ_v	σ_u	ρ_Y	ρ_v	ρ_u
OLS	Estimate	9.098	1.380	0.047	-0.554	0.650				
	Std. Error	1.155	0.072	0.081	0.474					
	Sig.	0.000	$< 10^{-16}$	0.568	0.250					
	Likelihood									-39.373
SAR	Estimate	7.163	1.404	0.352	-0.192			-0.013		
	Std. Error	1.116	0.060	0.110	0.410			0.004		
	Sig.	0.000	$< 10^{-16}$	0.001	0.641			0.000		
	Likelihood									-33.988
SEM	Estimate	8.954	1.381	0.057	-0.505				0.010	
	Std. Error	1.123	0.068	0.078	0.448				0.030	
	Sig.	0.000	$< 10^{-16}$	0.470	0.260				0.745	
	Likelihood									-39.354
SF	Estimate	10.501	1.197	0.052	-1.059	0.001	1.153			
	Std. Error	0.112	0.005	0.007	0.038	0.002	0.126			
	Sig.	$< 10^{-16}$	$< 10^{-16}$	0.000	$< 10^{-16}$	0.748	$< 10^{-16}$			
	Likelihood									-36.508
SSF (1,0,0,0)	**Estimate**	**6.933**	**1.239**	**0.497**	**-0.506**	**0.005**	**1.020**	**-0.016**		
	Std. Error	**1.318**	**0.061**	**0.138**	**0.256**	**0.007**	**0.111**	**0.004**		
	Sig.	**0.000**	**$< 10^{-16}$**	**0.000**	**0.049**	**0.450**	**$< 10^{-16}$**	**0.000**		
	Likelihood									**-31.602**
SSF (0,0,1,0)	Estimate	10.023	1.266	0.054	-0.744	0.281	0.910		0.098	
	Std. Error	na	0.000	0.000	na	0.000	na		0.000	
	Sig.		0.000	0.000		0.000			0.000	
	Likelihood									-38.207
SSF (1,0,1,0)	Estimate	6.933	1.239	0.497	-0.506	0.005	1.020	-0.016	-0.043	
	Std. Error	0.032	0.014	0.010	0.009	0.000	0.004	0.001	na	
	Sig.	$< 10^{-16}$	$< 10^{-16}$	$< 10^{-16}$	$< 10^{-16}$	$< 10^{-16}$	$< 10^{-16}$	$< 10^{-16}$		
	Likelihood									-31.581
SSF (1,0,0,1)	Estimate	6.963	1.239	0.490	-0.482	0.005	1.021	-0.016		0.001
	Std. Error	na	na	na	0.000	na	na	0.000		0.000
	Sig.				$< 10^{-16}$			$< 10^{-16}$		0.002
	Likelihood									-31.644

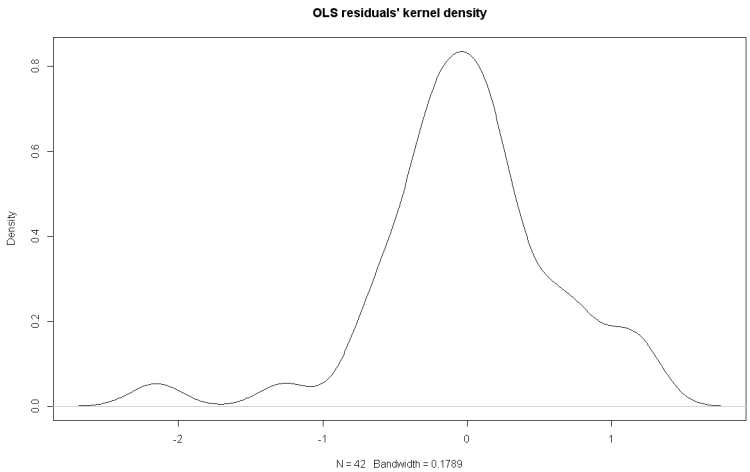


Fig. 4.10. Empirical kernel density of the Model UK OLS residuals

The plot is slightly left-skewed (sample skewness value is -0.672), which can be considered as an evidence for inefficiency in data. A hypothesis about inefficiency is supported by a

statistically significant estimate of inefficiency standard deviation σ_u (1.153), provided by the classical SF and different SSF model specifications.

Presence of spatial effects in data is not so obvious. The Table 4.14 contains results of formal statistical tests for spatial effects in OLS residuals.

Table 4.14. Results of spatial independence testing of the Model UK OLS residuals

Test statistic	*Value*	*p-value*	*Conclusion*
Moran's I	0.006	0.332	Insignificant spatial autocorrelation of OLS residuals
Lagrange multiplier test for spatial lags	9.041	0.003	Significant positive spatial lag in the OLS model residuals
Lagrange multiplier test for spatial errors	0.0167	0.898	Insignificant spatial errors in the OLS model residuals

Only spatial lags (endogenous spatial effects) are found significant in OLS residuals. The SAR model specification supports this conclusion: spatial lags are also found significant there. At the same time, the SEM model testifies against spatial heterogeneity in data. Note that spatial effects are tested separately and a more complicated spatial structure with different types of acting spatial effects can be not correctly recognised.

SSF models solve this problem and separately estimate every type of spatial effects. Two concurrent SSF model specifications SSF(1,0,0,0) and SSF(1,0,0,1) demonstrates similar goodness of fit and outperform other presented specifications. Difference between two mentioned models is not considered as significant (on the base of a formal likelihood ratio test), so the simpler model specification SSF(1,0,0,0) is preferred.

An interesting observation can be made comparing classical SF and SSF(1,0,0,0) models. The SF model states significant effects of all explanatory variables – a number of served routes, population in 100 km around an airport and a dummy for small island airports. Directions of these effects are expected – positive influence of number of routes and population and a negative effect for island airports. The SSF model gives the same direction of these effects, but their statistical significance is lower, especially for population and island variables, representing geographical environment. These effects are successfully replaced with a significant spatial lag. This result is very similar to the Box-Jenkins approach[223] to time series analysis, where the structure of the dependent variable is considered as a good replacement for influencing factors. Significant negative spatial lags, estimated by the SSF(1,0,0,0) model, can be explained with competition between UK airports on a local market.

Summarising executed spatial analysis of UK airports, we state that:

- Significant difference is observed between partial factor productivity of Spanish and UK airports. UK airports demonstrate higher average values of financial PFP indicators. This fact can be explained by a relatively higher level of de-monopolisation of the airport

industry in the UK and also on a stronger effect of world financial crisis on Spanish economics.

- Spatial effects are very weak for PFP indicators in the UK airports sample.
- Presence of inefficiency in data is strictly proven by the classical stochastic frontier model. This expected conclusion supports the hypothesis about different organisation of business in UK airports and a relatively competitive industry organisation.
- Stochastic frontier model with spatial lags SSF(1,0,0,0) outperforms other model specifications, which supports the hypothesis about significant endogenous spatial effects. The negative direction of spatial effects can be considered as a sign of spatial competition between UK airports.

4.5. Empirical analysis of Greek airports

4.5.1. Data set description

This data set contains cross-sectional information on traffic and infrastructure values in Greek airports in 2007. The data set is kindly provided by Dr. Tsekeris[215], who applied DEA methodology to analysis of Greek airports' efficiency. An original source of information is the Civil Aviation Authority of the Greek Ministry of Transport. Whereas there are significant seasonal demand variations in the Greek airport industry, data on passengers, cargos, flights and operating hours are separated into summer (between end of March and end of October) and winter (the rest of the year) periods. The Greek airport industry has its own peculiarities, related with a large number of islands and mountainous terrain, which make air transport indispensable for population. Nevertheless four major airports (in Athens, Thessaloniki, Heraklion, and Rhodes) concentrated about 72% of the total passenger traffic and 94% of the total amount of cargo in 2007. All Greek airports, except of the international airport of Athens, are state-owned and managed by the Civil Aviation Authority; the airport of Athens is operated as a private company. The Table 4.15 presents a technical description of the data set.

Summary statistics of the data set variables and a list of sample airports are presented in the Appendix 16. Spatial distribution of summer ATM, served by airports, is presented on the Fig. 4.11.

Table 4.15 Description of the Greek airports data set

Country	Greece		
Number of airports	42		
Years	2007		
Variables			
	Variable	*Description*	*Source*
	name	Airport title	DAFIF
	ICAO	Airport ICAO code	DAFIF
	lat	Airport latitude	DAFIF
	lon	Airport longitude	DAFIF
	APM_winter	A number of passengers carried during winter period	Tsekeris
	APM_summer	A number of passengers carried during summer period	
	APM	A number of passengers carried (winter + summer)	
	cargo_winter	A total volume of cargo served by an airport during winter period	
	cargo_summer	A total volume of cargo served by an airport during summer period	
	cargo	A number volume of cargo served by an airport (winter + summer)	
	ATM_winter	A number of air transport movements served by an airport during winter period	
	ATM_summer	A number of air transport movements served by an airport during summer period	
	ATM	A number of air transport movements served by an airport (winter + summer)	
	openning_hours_winter	A total number opening hours during winter period	
	openning_hours_summer	A total number opening hours during summer period	
	openning_hours	A total number opening hours (winter + summer)	
	runway_area	A total area of airport runways	
	terminal_area	A total area of airport terminal(s)	
	parking_area	A total area of airport parking area	
	island	1 if an airport is located on an island; 0 otherwise	
	international	1 if an airport is international; 0 otherwise	
	mixed_use	1 if an airport is in mixed use; 0 otherwise	
	WLU	A total volume of WLU served by an airport	
	NearestCity	A road network distance between an airport and its nearest city	

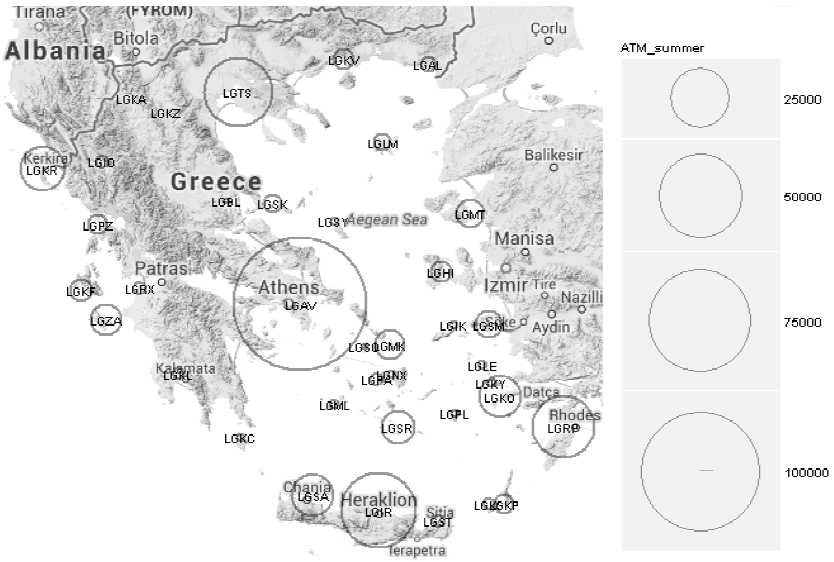


Fig. 4.11. Summer ATM in the Greek airports data set, 2007

A main feature of this data set is availability of data for winter and summer periods separately. This fact allows executing of seasonal comparison of research results. Another peculiarity of the data set is a high level of geographical isolation of Greek airports due to mountainous terrain and scattered islands.

4.5.2. Spatial analysis of airports' PFP indexes

The data set includes only physical characteristics of airports and traffic flows, so a list of study PFP indicators includes:

- ATM/WLU per Runway Area
- ATM/WLU per opening hour
- ATM/WLU per Terminal Area

Descriptive statistics of the PFP indicators are presented in the Appendix 17.

The indicators are studied separately for winter and summer periods. Passenger air traffic flows in Greece are significantly tourist-related, so values of the PFP indicators have strong seasonal differences. Box plots for WLU per runway area in summer and winter period are presented on the Fig. 4.12.

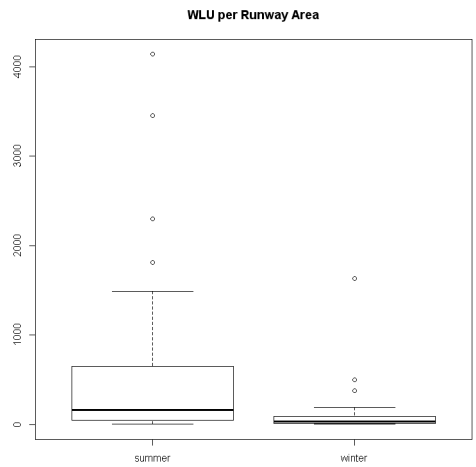


Fig. 4.12. Box plots of WLU per Runway Area of Greek airports (summer and winter)

Analysis of spatial dependencies in PFP indicators of Greek airport is quite limited. The Table 4.16 contains results of tests for spatial autocorrelation of PFP indicators' values for winter and summer periods.

Table 4.16. Results of spatial autocorrelation testing for PFP indicators of Greek airports

	Winter period			*Summer period*		
	Moran's I	Geary's C	Mantel	Moran's I	Geary's C	Mantel
WLU per Runway Area	-0.021 (0.861)	0.844 (0.299)	0.025 (0.327)	-0.056 (0.444)	0.935 (0.520)	0.085 (0.135)
WLU per opening hour	-0.017 (0.705)	0.799 (0.334)	0.038 (0.342)	-0.019 (0.818)	0.897 (0.434)	0.062 (0.233)
WLU per Terminal Area	0.024 (0.150)	0.860 (0.139)	0.067 (0.195)	-0.035 (0.809)	1.145 (0.161)	-0.013 (0.519)
ATM per Runway Area	-0.007 (0.566)	0.878 (0.251)	0.044 (0.276)	-0.063 (0.342)	0.981 (0.827)	0.032 (0.270)
ATM per opening hour	-0.026 (0.961)	0.831 (0.322)	0.043 (0.299)	-0.035 (0.823)	0.889 (0.397)	0.036 (0.290)
ATM per Terminal Area	0.028* (0.099)	1.091 (0.442)	0.025 (0.333)	0.025 (0.103)	1.174 (0.209)	-0.009 (0.514)

The general conclusion is a complete absence of statistically significant spatial effects both for winter and summer periods. This conclusion is quite expected subject to geographical separateness of Greek airports.

4.5.3. SSF analysis of Greek airports efficiency and spatial effects

A selected specification of the frontier is formulated as:

$$\log(WLU) = \beta_0 + \rho_Y W \log(WLU) + \beta_1 \log(OpenningHours) + \beta_2 \log(RunwayArea) + \\ + \beta_3 \log(TerminalArea) + \beta_4 Island + \beta_5 International \quad (4.7)$$

A list of selected inputs includes opening hours, runway and terminal areas (infrastructure inputs), and dummy variables for island and international airports. A general model with this frontier specification is referred as Model Greece.

Different specifications of the model, described in the paragraph 4.1.4, are estimated and analysed. All models are estimated separately for winter and summer periods. Parameter estimates and necessary statistics for selected model specification are summarised in the Table 4.17; the complete list of models and their detailed estimation results are presented in the Appendix 18.

Table 4.17. Estimation results of the Model Greece alternative specifications

Model		*Intercept*	*log(OpenningHours)*	*log(RunwayArea)*	*log(TerminalArea)*	*Island*	*International*	σ_v	σ_u	ρ_Y	ρ_v
Summer											
OLS	Estimate	0.926	2.178	-0.447	0.551	-0.545	0.209	0.628			
	Std. Error	2.875	0.270	0.286	0.111	0.327	0.326				
	Sig.	0.749	0.000	0.128	0.000	0.105	0.526				
	Likelihood										-33.945
SAR	Estimate	0.587	2.176	-0.432	0.546	-0.591	0.256			0.001	
	Std. Error	2.710	0.247	0.264	0.102	0.311	0.311			0.002	
	Sig.	0.828	$< 10^{-16}$	0.102	0.000	0.057	0.411			0.592	
	Likelihood										-33.801

Model		*Intercept*	*log(OpenningHours)*	*log(RunwayArea)*	*log(TerminalArea)*	*Island*	*International*	σ_v	σ_u	ρ_Y	ρ_v
SEM	Estimate	0.926	2.138	-0.433	0.561	-0.475	0.251				-0.038
	Std. Error	2.608	0.249	0.259	0.099	0.290	0.300				0.039
	Sig.	0.723	$< 10^{-16}$	0.095	0.000	0.101	0.403				0.333
	Likelihood										-33.656
SF	***Estimate***	***3.085***	***2.204***	***-0.558***	***0.512***	***-0.777***	***0.175***	***0.001***	***1.003***		
	Std. Error	***2.546***	***0.244***	***0.263***	***0.097***	***0.303***	***0.256***	***0.002***	***0.113***		
	Sig.	***0.226***	***$< 10^{-16}$***	***0.034***	***0.000***	***0.010***	***0.493***	***0.750***	***$<10^{-16}$***		
	Likelihood										***-28.456***
SSF (1,0,0,0)	Estimate	2.052	2.246	-0.420	0.418	-0.713	-0.006	0.005	1.005	0.001	
	Std. Error	1.1560	0.194	0.099	0.040	0.188	0.046	0.006	0.114	0.001	
	Sig.	0.076	$< 10^{-16}$	0.000	$< 10^{-16}$	0.001	0.898	0.433	$<10^{-16}$	0.750	
	Likelihood										-28.726
Winter											
OLS	Estimate	0.305	2.318	-0.425	0.359	-0.016	-0.074	1.166			
	Std. Error	5.375	0.453	0.522	0.194	0.566	0.646				
	Sig.	0.955	0.000	0.421	0.073	0.977	0.910				
	Likelihood										-58.059
SAR	Estimate	-1.117	2.258	-0.341	0.343	-0.239	0.181			0.005	
	Std. Error	4.921	0.408	0.471	0.174	0.530	0.606			0.003	
	Sig.	0.820	0.000	0.469	0.049	0.652	0.765			0.160	
	Likelihood										-57.072
SEM	Estimate	0.763	2.472	-0.537	0.321	0.052	-0.056				-0.034
	Std. Error	4.882	0.406	0.476	0.175	0.504	0.591				0.039
	Sig.	0.876	0.000	0.260	0.067	0.918	0.925				0.679
	Likelihood										-57.973
SF	***Estimate***	***0.083***	***2.691***	***-0.455***	***0.369***	***-0.441***	**-0.790**	**0.001**	**1.876**		
	Std. Error	***4.847***	***0.141***	***0.605***	***0.117***	***0.060***	**0.307**	**0.003**	**0.212**		
	Sig.	***0.986***	***$< 10^{-16}$***	***0.452***	***0.002***	***0.000***	**0.010**	**0.752**	**$<10^{-16}$**		
	Likelihood										***-52.882***
SSF (1,0,0,0)	Estimate	-2.121	2.595	-0.250	0.247	-0.547	-0.253	0.008	1.854	0.006	
	Std. Error	5.355	0.181	0.633	0.160	0.160	0.564	0.011	0.211	0.006	
	Sig.	0.692	$< 10^{-16}$	0.693	0.122	0.001	0.654	0.471	$<10^{-16}$	0.265	
	Likelihood										-52.569

We started with the simplest OLS model and enhanced it with inefficiency components and spatial effects, according to the model hierarchy presented on the Fig. 4.1. An empirical distribution of OLS residuals is presented on the Fig. 4.13.

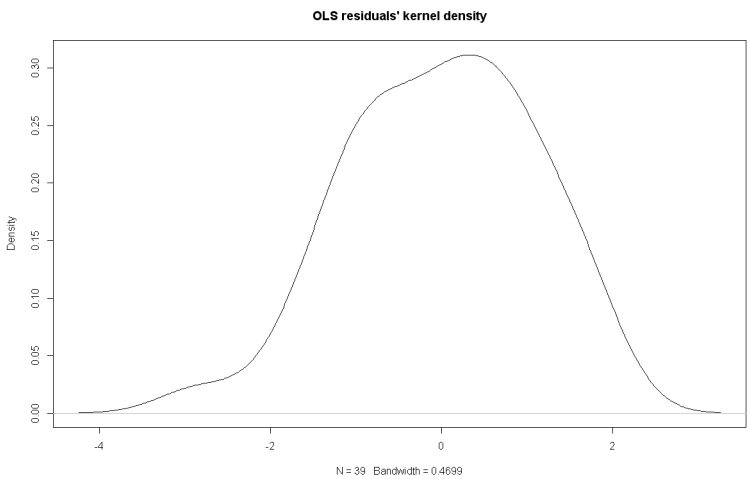


Fig. 4.13. Empirical kernel density of the Model Greece OLS residuals (summer season)

The plot is slightly left-skewed (sample skewness value is -0.309 for winter and -0.633 for summer season), which can be considered as an evidence for inefficiency in data. A hypothesis about inefficiency is supported by a statistically significant estimate of inefficiency standard deviation σ_u (1.003 and 1.876 for summer and winter seasons respectively), provided by the classical SF and SSF model specifications.

The Table 4.18 contains results of formal statistical tests for spatial effects in OLS residuals.

Table 4.18. Results of spatial independence testing of the Model Greece OLS residuals

Test statistic	*Value*	*p-value*	*Conclusion*
Moran's I	-0.026	0.995	Insignificant spatial autocorrelation of OLS model residuals
Lagrange multiplier test for spatial lags	0.289	0.592	Insignificant spatial lags in the OLS model residuals
Lagrange multiplier test for spatial errors	0.249	0.618	Insignificant spatial errors in the OLS model residuals

The general conclusion is a complete absence of spatial effects in Greek airports activity. This conclusion is supported by different approaches: tests for spatial autocorrelation between PFP indicators' values and between OLS and SF models' residuals and direct estimation of different types of spatial effects with SSF models. Under these conditions the classical SF model is a preferred specification (which is formally proven on the base of likelihood ratio tests).

Elasticity of inputs, estimated with the SF model, match our original expectations. Opening hours have a statistically significant positive effect with high absolute values (2.204 and 2.691 for summer and winter respectively). A terminal area is also considered as an important input for served traffic in both seasons. A runway area is estimated as insignificant resource in the winter season, but significant in the summer season, which can be explained by overall seasonal congestion of Greek airports. Location of an airport on a small island has an expected negative effect, consistent for both seasons. An international status of an airport appears as a significant negative factor for winter season only. This fact also can be explained by seasonal specifics of traffic in Greece airports, but require additional research.

Individual efficiency levels of Greek airports significantly differ for summer and winter seasons (mean efficiency, estimated with the classical SF model, is 0.588 for summer season and 0.335 for the winter season). This difference is expected, because infrastructure inputs (runway and terminal areas) are estimated as permanent resources, but a level of their utilisation is highly season-specific.

Summarising executed spatial analysis of Greek airports, we state that:

- Spatial effects are not discovered in efficiency of Greek airports. This result is obtained both for PFP indicators and SSF models and can be explained by geographical peculiarities – mountainous terrain and complexes of islands.
- Efficiency of Greek airports significantly varies for summer and winter seasons, which is related with tourist and other seasonal traffic flows.

4.6. Conclusions

This chapter is devoted to empirical analysis of spatial effects in four different European airports' data sets. We utilised financial and physical approaches to airport benchmarking and different airport inputs/outputs specifications.

Analysis of spatial effects includes testing of spatial autocorrelation between selected PFP indicators of airports and estimating of special types of spatial effects (spatial endogenous effects, spatially correlated random disturbances, and spatially related efficiency) using 11 alternative SSF model specifications. We used spatial specifications of the SF model, introduced in the chapter 3 of this thesis; a detailed hierarchy of model specifications can be found in the chapter. Parameters of all models were estimated using the derived MLE, implemented in the developed *spfrontier* package. We also calculated all necessary statistics for every model and estimated individual levels of inefficiency.

Research data sets include European airports data set (359 airports, 2008-2012), Spanish airports data set (38 airports, 2009-2010), UK airports data set (48 airports), and Greek airports data set (42 airports, 2007). Every data set has its own specifics, related with presence of inefficiency and spatial effects in data.

Conclusions for the European airports data set. Significant spatial autocorrelation is discovered for all considered PFP indicators – ATM/PAX/WLU per runway/per route and PAX per capita in a catchment area. We analysed two different specifications of the stochastic frontier – single-output (PAX) and multi-output (PAX and cargo) and obtained similar results. Both approaches support our initial assumption about significant spatial effects in data. The selected specification of the stochastic frontier model is SSF(1,0,1,0), which includes spatial endogenous effects and spatially correlated random disturbances. Thus we discovered statistically significant negative endogenous spatial effects, which are explained by spatial competition for passengers and cargo flows between neighbour airports, and spatially positively correlated random disturbances, which is a result of unobserved area-specific factors. Spatial effects between inefficiency values are not discovered for the data set.

Conclusions for the Spanish airports data set. The Spanish airport industry is fairly monopolistic; all 47 commercial airports in Spain are managed by AENA. Probably due to this fact we didn't discover significant inefficiency in the data set (in respect to the selected specification of the frontier) in this research. This result is clearly explained by comparative approach to inefficiency estimation of SFA and monopolistic structure of the Spanish airport industry. At the same time, we discovered significant spatial effects in this data set. Availability of financial information allows us utilising both physical and financial approaches to airport benchmarking. Positive spatial autocorrelation is found for partial factor productivity of Spanish airports, so the airports are geographically clustered in respect to considered PFP indicators (both physical and financial). Absence of comparative inefficiency in data allows utilising of standard spatial regression techniques, in particular SAR and SEM models. We discovered a statistical supremacy of the SEM model, which indicates spatial heterogeneity of Spanish airports.

Conclusions for the UK airports data set. After a set of airport sales and acquisitions, initiated by UK Competition Commission, UK airports are generally managed by of different operators. Different operators are supposed to act as competitors (including competition in spatial settings), enforcing economic efficiency of each other. Presence of inefficiency in data is strictly proven by the analysis. The stochastic frontier model with spatial lags, SSF(1,0,0,0), outperforms other model specifications, which supports the hypothesis about significant endogenous spatial effects. The negative direction of spatial effects can be considered as a sign of spatial competition between UK airports.

Conclusions for the Greek airports data set. Peculiarities of the Greek airport industry, related with a large number of islands and mountainous terrain, make spatial relationship less probable. Additionally, all Greek airports, except of the international airport of Athens, are state-owned and managed by the Civil Aviation Authority. As a result, spatial effects are not discovered in efficiency of Greek airports. This result is obtained both for PFP indicators and SSF models. Also our analysis demonstrates significant variation of Greek airports efficiency in summer and winter seasons, which is related with tourist and other seasonal traffic flows.

Detailed conclusions, made for every data set, are presented at the end of corresponding paragraphs.

Application of the SSF models to data sets in different spatial settings allowed practical examining the proposed methodology and supporting our main hypothesis about importance of spatial components in efficiency analysis. All data set and executed calculations are included into the *spfrontier* package, developed by the author and publicly available in the CRAN archive, to ensure research reproducibility.

CONCLUSIONS

Statement of the main research results

1. This research is devoted to enhancing of the methodology of statistical estimation of efficiency subject to presence of spatial effects. The work was focused on development of the spatial stochastic frontier model and its application to analysis of the European airport industry.
2. The critical review of existing airport benchmarking researches was performed. Actual methodologies of efficiency analysis were discussed and classified, and a wide range of their applications to the airport industry are reviewed. The review was focused on revealing spatial effects (spatial heterogeneity and spatial dependence). A theoretical background of spatial interactions between airports was reviewed and existing empirical evidences of presence of spatial effects in the European airport industry were presented.
3. Principles of stochastic frontier analysis and spatial econometrics were reviewed with a special attention to incorporating of spatial effects into stochastic frontier models. Despite the fact that the importance of spatial relationships for SFA is widely acknowledged in literature, number of researches, where spatial effects are included into consideration, is very limited. Mainly researchers ignore the presence of spatial effects or include them in an observed form only. Also we noted an absence of a general specification of the stochastic frontier model with spatial effects and, consequently, a lack of a unified software tool for estimation of such models.
4. Four possible types of spatial effects in SFA are identified. These effects include spatial exogenous effects, spatial endogenous effects, spatially correlated random disturbances, and spatially related efficiency. We presented reasoning for these spatial effects as phenomena in different branches of knowledge.
5. The spatial stochastic frontier model, incorporating spatial effects into the stochastic frontier analysis, was proposed. The SSF model was stated formally, in a reasonably general form, where spatial effects were included as first-order spatial lags. A number of practically effective private cases of the SSF model were also discussed. Specification of the SSF model is an important component of this research novelty.
6. A special attention is devoted to the problem of model parameter identification. Parameter identification is one of important issues, frequently noted both in spatial econometrics and stochastic frontier modelling literature. The SSF model as a combination of stochastic frontier and spatial regression models also suffer from weak parameter identification. In this research we presented an initial theoretical justification of

the parameter identification problem and illustrated it with real and simulated data examples.

7. One of the main practical results of this research is a derived maximum likelihood estimator for the SSF model parameters. A distribution law of the composed error term of the SSF model is derived and stated as a private case of the closed multivariate skew normal distribution. Using the derived distribution of the SSF model's error term, the likelihood function is specified and a related estimator is constructed. Individual inefficiency estimation is one of the main benefits of the classical stochastic frontier models, so we also derived formulas for estimates of individual inefficiency values in the SSF model.
8. The derived MLE for the SSF model parameters is implemented as a package for CRAN R software, called *spfrontier*. The package includes all derived algorithms for the SSF model estimation and accepted and published in the official CRAN archive. The package can be considered as a significant part of the practical value of this research.
9. The derived MLE and the developed package are validated using designed statistical simulation studies. We organised a set of simulation experiments, which allows investigating of the SSF model estimate properties for different specifications and sample sizes. According to the executed simulation experiments, the derived estimator provides statistically unbiased and consistent estimates and allows confidently distinguishing between different types of spatial effects. We also compared estimates of a private case of the SSF model with results of existing software that designed for classical stochastic frontier model and found them almost identical.
10. Empirical analysis of spatial effects in four different European airports' data sets is executed. Analysis consists of testing of spatial autocorrelation between airports' selected PFP indicators and estimating of alternative specifications of the SSF model. Research data sets include European airports data set (359 airports, 2008-2012), Spanish airports data set (38 airports, 2009-2010), UK airports data set (48 airports, 2011-2012), and Greek airports data set (42 airports, 2007). Conclusions were made separately for every data set.
    - *Conclusions for the European airports data set*. We discovered statistically significant negative endogenous spatial effects, which are explained by spatial competition for passengers and cargo flows between neighbour airports, and spatially positively correlated random disturbances, which is a result of unobserved area-specific factors.

- *Conclusions for the Spanish airports data set*. The Spanish airport industry is fairly monopolistic; thus in this research we didn't discover significant inefficiency in the data set. At the same time, we discovered significant spatial heterogeneity in this data set and applied methods of classical spatial econometrics for empirical analysis.
- *Conclusions for the UK airports data set*. Applying the SSF model, we discovered significant inefficiency and endogenous spatial effects for the UK airports sample. These finding supports our hypothesis about spatial competition in the relatively competitive UK airport industry.
- *Conclusions for the Greek airports data set*. Peculiarities of the Greek airport industry, related with a large number of islands and mountainous terrain, and common ownership of Greek airports make spatial relationships weaker. As a result, significant spatial effects were not discovered in efficiency of Greek airports. Also our analysis demonstrated significant variation of Greek airports efficiency in summer and winter seasons, which is related with tourist and other seasonal traffic flows.

Detailed conclusions on all research data sets are presented in the Chapter 4. Application of the SSF models to data sets in different spatial settings allowed practical examining the proposed methodology and supporting our main hypothesis about importance of spatial components in efficiency analysis.

Novelty of the research

The following results can be considered as a scientific novelty of the research:

1. The proposed SSF model, which aggregate principles of spatial econometrics and stochastic frontier analysis. The model allows estimation of the general production frontier and unit-specific inefficiency values, taking potential spatial effects into account. Four different types of spatial effects are explicitly incorporated into the model: endogenous spatial effects, exogenous spatial effects, spatially correlated random disturbances, and spatially related efficiency.
2. The derived estimator for the proposed SSF model. The estimator is based on maximum likelihood principles and allows estimating the SSF model parameters. A separate estimator is derived for unit-specific inefficiency values. The derived estimator is validated using designed simulation studies and real-world data sets.
3. The SSF model is applied to empirical investigation of spatial effects in the European airport industry. To the best of our knowledge, this thesis is the first systematic application of spatial econometrics to the airport industry. Developed model

specifications and obtained results present a novelty of this research for analysis of the airport industry and specifically for airport benchmarking.

Practical value of the research

The practical importance of the research consists of:

1. The developed software package *spfrontier*, implementing the derived estimator of the SSF model and a set of related utilities. The package is implemented as a module for the R environment and accepted in the official CRAN archive. The package includes functions for: estimation of the SSF model parameters; estimation of unit-specific inefficiency values; numerical calculation of the estimates' Hessian matrix; testing of parameter estimates' significance; and designed simulation studies for analysis of estimates' statistical properties. The package can be used for efficiency estimation in different application areas: transport economics, regional science, urban economics, housing, agriculture, ecology, and other areas, where spatial effects play an important role.
2. The results of application of spatial statistics techniques, including the developed SSF model, to the European airport industry. Four data sets, related to different economic and spatial environments, were separately investigated: Spanish airports, UK airports, Greek airports, and a joined sample of European airports. Using the developed SSF model, significant spatial effects were discovered and their analysis was executed. The obtained results can be utilised by the following stakeholders: airport management, airline management, municipalities, and policy makers.

Further research directions

There is a wide range of theme-related potential research directions. Among these directions, the following can be mentioned as the most important ones:

1. Further development of the SSF model. There are a number of possible improvements of the SSF models: usage of different spatial dependency forms, analysis of model parameters' identification, research of different spatial matrices specifications.
    a. Spatial effects are modelled in the SSF model using first-order spatial lags. Different approaches like spatial moving average or higher order spatial lags can be reasonably applied.
    b. The identification problem (whether the four types of spatial effects, considered in this thesis, can be distinguished from each other) is a well known curse of spatial models, and additional analysis of this problem should be executed for the proposed SSF model.

c. Importance of alternative spatial matrix specifications for the SSF model estimation is another point, which requires extensive research.

2. Enhancements of the derived MLE. Estimation, based on the derived MLE, is a multivariate optimisation task, which can be solved in different ways. This problem is especially significant since analytical gradient and Hessian matrixes are not derived within the scope of this research and numerical methods are used for optimisation. Obtaining of the analytical derivatives or application of modern optimisation techniques without analytical gradients is necessary for extended empirical applications of the SSF model. Another possible enhancement consists of usage of the expectation–maximization optimisation algorithm.
3. Development of other estimators for the SSF model. Estimation of the multivariate closed skew-normal distribution parameters, which plays a primary role in the SSF model, is another theoretical task, which attracts attention of scientific community. The possible set of methods includes, but is not limited with, generalised method of moments, generalised maximum entropy and Bayesian estimators.
4. Applications of the SSF model in different research areas. In this research we focused on application of the SSF model to analysis of the airport industry, but other application areas are queued up. Presence both of spatial effects and units' inefficiency is also a feature of regional science, urban economics, education economics, real estate economics and others. Application of the SSF model to these areas is a broad direction of further research.

BIBLIOGRAPHY

1. European Commission (1992). Council Regulation on licensing of air carriers. .
2. European Commission (1992). Council Regulation on access for Community air carriers to intra-Community air routes. .
3. European Commission (1992). Council Regulation on fares and rates for air services. .
4. Fu, X., Oum, T.H., Zhang, A. (2010). Air Transport Liberalization and Its Impacts on Airline Competition and Air Passenger Traffic, *Transportation Journal*, Vol. 49, No 4, pp. 24–41.
5. Competitive Interaction between Airports, Airlines and High-Speed Rail (2009). Joint Transport Research centre, Discussion Paper.
6. Oum, T.H., Adler, N., Yu, C. (2006). Privatization, corporatization, ownership forms and their effects on the performance of the world's major airports, *Journal of Air Transport Management*, Vol. 12, No 3, pp. 109–121.
7. Oum, T.H. (1992). Concepts, methods and purposes of productivity measurement in transportation, *Transportation Research Part A: Policy and Practice*, Vol. 26, No 6, pp. 493–505.
8. Scotti, D. (2011). Measuring Airports' Technical Efficiency: Evidence from Italy, PhD thesis, University of Bergamo, Italy.
9. Müller-Rostin, C., Niemeier, H.M., Ivanova, P., Müller, J., Hannak, I., Ehmer, H. (2010). Airport Entry and Exit: A European Analysis, in *Airport Competition: The European Experience*, England: Farnham: Ashgate Publishing Limited, pp. 27–46.
10. European Commission (2006). Commission Regulation laying down a common charging scheme for air navigation services Text with EEA relevance. .
11. Anselin, L. (1988). *Spatial econometrics: methods and models*. Dordrecht: Kluwer Academic Publishing, 304 p.
12. Doganis, R. (1992). *The airport business*. London: Routledge, 240 p.
13. Doganis, R., Graham, A. (1987). Airport Management: The Role of Performance Indicators, Polytechnic of Central London, London, UK, Transport Studies Group Research Report 13.
14. Holvad, T., Graham, A. (2000). Efficiency Measurement for Airports, presented at the Annual Transport Conference, Aalborg University, Denmark, pp. 331–343.
15. Graham, A. (2005). Airport benchmarking: a review of the current situation, *Benchmarking: An International Journal*, Vol. 12, pp. 99–111.

16. Graham, A., Vogel, H. (2006). A comparison of alternative airport performance measurement techniques: a European case study, *Journal of Airport Management*, No 1, pp. 59–74.
17. Gillen, D., Lall, A. (1997). Developing measures of airport productivity and performance: an application of data envelopment analysis, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 33, No 4, pp. 261–273.
18. Barros, C.P., Sampaio, A. (2004). Technical and allocative efficiency in airports, *International Journal of Transport Economics*, Vol. 31, No 3, pp. 355–378.
19. Barros, C.P. (2008). Technical efficiency of UK airports, *Journal of Air Transport Management*, Vol. 14, No 4, pp. 175–178.
20. Barros, C.P., Marques, R.C. (2008). Performance of European Airports: Regulation, Ownership and Managerial Efficiency, School of Economics and Management, Lisbon, Portugal, Working Paper 25/2008/DE/UECE.
21. Barros, C.P., Weber, W.L. (2009). Productivity growth and biased technological change in UK airports, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 45, No 4, pp. 642–653.
22. Assaf, A.G., Gillen, D., Barros, C. (2012). Performance assessment of UK airports: Evidence from a Bayesian dynamic frontier model, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 48, No 3, pp. 603–615.
23. Gitto, S., Mancuso, P. (2012). Bootstrapping the Malmquist indexes for Italian airports, *International Journal of Production Economics*, Vol. 135, No 1, pp. 403–411.
24. Gitto, S. (2008). The measurement of productivity and efficiency: theory and applications, PhD thesis, University of Rome "Tor Vergata," Rome, Italy.
25. Liebert, V., Niemeier, H.M. (2011). Benchmarking of Airports-A Critical Assessment, presented at the 12th World Conference on Transport Research, Lisbon, Portugal, p. 46.
26. Liebert, D.V.F.V.P. (2011). Airport benchmarking: an efficiency analysis of European airports from an economic and managerial perspective, University of Applied Sciences.
27. Adler, N., Liebert, V. (2011). Joint Impact of Competition, Ownership Form and Economic Regulation on Airport Performance, Jacobs University Bremen.
28. Oum, T.H., Yu, C., Choo, Y. (2011). ATRS Global Airport Performance Benchmarking Project, The Air Transport Research Society.
29. Merkert, R., Pagliari, R., Odeck, J., Brathen, S., Halpern, N., Husdal, J. (2010). Benchmarking Avinor's Efficiency, Møreforsking Molde AS, Molde, Norway, 1006.
30. Civil Aviation Authority (2000). The Use of Benchmarking in the Airport Reviews, Civil Aviation Authority, London, UK.

31. Kumbhakar, S.C., Lovell, C.A.K. (2003). *Stochastic frontier analysis*. Cambridge: Cambridge Univ. Press, 344 p.
32. Aigner, D., Lovell, C., Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models, *Journal of econometrics*, Vol. 6, No 1, pp. 21–37.
33. Meeusen, W., van der Broeck, J. (1977). Efficiency Estimation from Cobb-Douglas Production Function with Composed Error, *International Economic Review*, Vol. 8, pp. 435–444.
34. Pels, E., Nijkamp, P., Rietveld, P. (2003). Inefficiencies and scale economies of European airport operations, *Transportation Research Part E: Logistics and Transportation Review*, No 39, pp. 341–361.
35. Abrate, G., Erbetta, F. (2007). Investigating Returns to Scope and Operational Efficiency in Airport Business: an Input Distance Function Approach, *Hermes Working Papers*, Vol. 3.
36. Jing, X.Y. (2007). Benchmarking competitiveness of cargo airports, MSc thesis, National University of Singapore, Singapore.
37. Barros, C.P., Managi, S., Yoshida, Y. (2008). Technical Efficiency, Regulation, and Heterogeneity in Japanese Airports, School of Economics and Management, Lisbon, Portugal, Working Paper 43/2008/DE/UECE.
38. Voltes, A.J. (2008). Stochastic frontier estimation of airports' cost function, PhD thesis, University of Las Palmas de Gran Canaria, Spain.
39. Martín, J.C., Román, C., Voltes-Dorta, A. (2009). A stochastic frontier analysis to estimate the relative efficiency of Spanish airports, *Journal of Productivity Analysis*, Vol. 31, pp. 163–176.
40. Muller, J., Ulku, T., Zivanovic, J. (2009). Privatization, restructuring and its effects on performance: A comparison between German and British airports, German Airport Performance Project, Germany, Working Paper 16.
41. Malighetti, P., Martini, G., Paleari, S., Redondi, R. (2009). The Impacts of Airport Centrality in the EU Network and Inter-Airport Competition on Airport Efficiency, MPRA, Munich, Germany, Paper 17673.
42. Scotti, D., Malighetti, P., Martini, G., Volta, N. (2012). The impact of airport competition on technical efficiency: A stochastic frontier analysis applied to Italian airport, *Journal of Air Transport Management*, Vol. 22, pp. 9–15.
43. Borins, S., Advani, A. (2002). Managing airports: a test of the New Public Management, *International Public Management Journal*, Vol. 4, No 1, pp. 91–107.

44. Malighetti, P., Martini, G., Scotti, D., Volta, N. (2010). The impact of airport competition on technical efficiency: A Stochastic Frontier Analysis applied to Italian airports, MPRA, Munich, Germany, Paper 24648.
45. Adler, N., Liebert, V. (2011). Competition and regulation (when lacking the former) outrank ownership form in generating airport efficiency, presented at the GAP workshop "Benchmarking of Airports," Berlin, Germany, p. 27.
46. Anselin, L. (2010). Thirty years of spatial econometrics, *Papers in Regional Science*, Vol. 89, No 1, pp. 3–25.
47. Arbia, G. (2006). *Spatial econometrics: statistical foundations and applications to regional convergence*. Berlin: Springer, 219 p.
48. Ülkü, T., Jeleskovic, V., Müller, J. (2014). How scale and institutional setting explain the costs of small airports? An application of spatial regression analysis, Joint Discussion Paper Series in Economics.
49. Druska, V., Horrace, W.C. (2004). Generalized moments estimation for spatial panel data: Indonesian rice farming, *American Journal of Agricultural Economics*, Vol. 86, No 1, pp. 185–198.
50. Fahr, R., Sunde, U. (2005). Regional Dependencies in Job Creation: An Efficiency Analysis for Western Germany, IZA, Bonn, Germany, 1660.
51. Barrios, E.B. (2007). Spatial Effect in the Efficient Access of Rural Development, Asian Development Bank Institute, Japan, Discussion Paper 65.
52. Schettini, D., Igliori, D.C., Azzoni, C.R. (2007). Productive Efficiency across Regions in Brazil: a Spatial Stochastic Frontier Analysis, Nemesis, Brazil, Working Paper.
53. Affuso, E. (2010). Spatial autoregressive stochastic frontier analysis: An application to an impact evaluation study, Auburn University, USA, Working Paper.
54. Lin, J., Long, Z., Lin, K. (2010). Spatial Panel Stochastic Frontier Model and Technical Efficiency Estimation, *Journal of Business Economics*, Vol. 223, No 5, pp. 71–78.
55. Lin, J., Long, Z., Lin, K. (2010). Simulated maximum likelihood estimation of spatial stochastic frontier model and its application, in *Future Information Technology and Management Engineering (FITME), 2010 International Conference on*, Vol. 3, pp. 330–333.
56. Areal, F.J., Balcombe, K., Tiffin, R. (2010). Integrating spatial dependence into stochastic frontier analysis, University Library of Munich, Munich, Germany, Paper 24961.
57. Tonini, A., Pede, V. (2011). A Generalized Maximum Entropy Stochastic Frontier Measuring Productivity Accounting for Spatial Dependency, *Entropy*, Vol. 13, No 11, pp. 1916–1927.

58. Mastromarco, C., Serlenga, L., Shin, Y. (2012). Modelling Technical Efficiency in Cross Sectionally Dependent Panels, University of Salento and CESifo, Italy, Working Paper.
59. Glass, A., Kenjegalieva, K., Sickles, R.C. (2013). A Spatial Autoregressive Production Frontier Model for Panel Data: With an Application to European Countries, *SSRN Electronic Journal*.
60. Fusco, E., Vidoli, F. (2013). Spatial stochastic frontier models: controlling spatial global and local heterogeneity, *International Review of Applied Economics*, Vol. 27, No 5, pp. 679–694.
61. Pavlyuk, D. (2014). *CRAN R package "spfrontier"* .
62. O'Sullivan, A., Sheffrin, S.M. (2003). *Economics: principles in action*. Needham, Mass.: Prentice Hall, 592 p.
63. Zenglein, M.J., Müller, J. (2007). Non-Aviation Revenue in the Airport Business– Evaluating Performance Measurement for a Changing Value Proposition, Berlin School of Economics, Berlin, Germany.
64. Yu, M. (2004). Measuring physical efficiency of domestic airports in Taiwan with undesirable outputs and environmental factors, *Journal of Air Transport Management*, Vol. 10, No 5, pp. 295–303.
65. Pathomsiri, S., Haghani, A., Dresner, M., Windle, R.J. (2008). Impact of undesirable outputs on the productivity of US airports, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 44, No 2, pp. 235–259.
66. Hirschhausen, C., Cullman, A. (2005). Questions to Airport Benchmarkers - Some Theoretical and Practical Aspects Learned from Benchmarking Other Sectors, presented at the German Aviation Research Society Workshop, Vienna, Austria.
67. Abdesaken, G., Cullman, A. (2006). The Relative Efficiency of German Airports, presented at the GARS Amsterdam Student Workshop, Amsterdam.
68. Törnqvist, L. (1981). *Collected scientific papers of Leo Törnqvist.* Helsinki: Research Institute of the Finnish Economy, 423 p.
69. Caves, D.W., Christensen, L.R., Diewert, W.E. (1982). Multilateral Comparisons of Output, Input, and Productivity Using Superlative Index Numbers, *The Economic Journal*, Vol. 92, No 365, pp. 73–86.
70. Nyshadham, E.A., Rao, V.K. (2000). Assessing Efficiency of European Airports: A Total Factor Productivity Approach, *Public Works Management & Policy*, Vol. 5, No 2, pp. 106–114.
71. Charnes, A., Cooper, W.W., Rhodes, E. (1978). Measuring the efficiency of decision making units, *European journal of operational research*, Vol. 2, No 6, pp. 429–444.

72. Simar, L., Wilson, P.W. (2000). Statistical Inference in Nonparametric Frontier Models: The State of the Art, *Journal of Productivity Analysis*, Vol. 13, No 1, pp. 49–78.
73. Simar, L., Wilson, P.W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes, *Journal of econometrics*, Vol. 136, No 1, pp. 31–64.
74. Ulku, T. (2009). Efficiency of German Airports and Influencing Factors, MSc thesis, Humboldt University, Berlin, Germany.
75. Sarkis, J. (2000). An analysis of the operational efficiency of major airports in the United States, *Journal of Operations Management*, Vol. 18, No 3, pp. 335–352.
76. Tapiador, F.J., Mateos, A., Martí-Henneberg, J. (2008). The geographical efficiency of Spain's regional airports: A quantitative analysis, *Journal of Air Transport Management*, Vol. 14, No 4, pp. 205–212.
77. Martin, J.C., Roman, C. (2001). An application of DEA to measure the efficiency of Spanish airports prior to privatization, *Journal of Air Transport Management*, Vol. 7, No 3, pp. 149–157.
78. Psaraki-Kalouptsidi, V., Kalakou, S. (2011). Assessment of efficiency of Greek airports, *Journal of Airport Management*, Vol. 5, No 2, pp. 170–186.
79. Razali, S.R., Shah, M.Z. (2010). Performance Measurement of Malaysian Airports using DEA method, in *Proceeding of Malaysian Universities Transportation Research Forum and Conferences*, Universiti Tenaga Nasional, Malaysia, pp. 487–493.
80. Perelman, S., Serebrisky, T. (2010). Measuring the technical efficiency of airports in Latin America, The World Bank, Liege, Belgium, Policy Research Working Paper WPS5339.
81. Barros, C.P., Assef, A. (2009). Productivity change in USA airports: the Gillen ans Lall approach revisited, School of Economics and Management, Lisbon, Portugal, Working Paper 22/2009/DE/UECE.
82. Barros, C.P., Assaf, A., Lipovich, G.A. (2010). Productivity Analysis of Argentine Airports, School of Economics and Management, Lisbon, Portugal, Working Paper.
83. Barros, C.P., Peypoch, N. (2007). A comparative analysis of productivity change in Italian and Portuguese airports, School of Economics and Management, Lisbon, Portugal, Working Paper 006/2007/DE.
84. Barros, C.P., Peypoch, N., Villard, P. (2011). Productivity changes in canadian airports and technological change analysis, School of Economics and Management, Lisbon, Portugal, Working Paper 05/2011/DE/UECE.
85. Kamp, V., Niemeier, H., Schmidt, P. (2004). Benchmarking of German Airports - Some first Results, presented at the GARS Research Seminar "How to make Slot Markets Work," Bremen, Germany, p. 17.

86. Kamp, V. (2007). Airport Benchmarking - An Empirical Research on the Performance Measurement of German Airports with Data Envelopment Analysis, *Aerlines e-zine edition*, No 36, pp. 1–4.
87. Curi, C., Gitto, S., Mancuso, P. (2009). Managerial assessment of Italian airport efficiency: a statistical DEA approach, University of Tor Vergata, Rome, Italy, Discussion Paper.
88. Gitto, S., Mancuso, P. (2010). Airport efficiency: a DEA two stage analysis of the Italian commercial airports, MPRA, Munich, Germany, Paper 34366.
89. Curi, C., Gitto, S., Mancuso, P. (2011). New Evidence on the Efficiency of Italian Airports: A Bootstrapped DEA Analysis, *Socio-Economic Planning Sciences*, Vol. 45, No 2, pp. 84–93.
90. Barros, C.P., Dieke, P. (2007). Performance evaluation of Italian airports: A data envelopment analysis, *Journal of Air Transport Management*, Vol. 13, No 4, pp. 184–191.
91. Malighetti, P., Martini, G., Paleari, S., Redondi, R. (2007). Efficiency of Italian airports management: the implications for regulation, Italy, Working Paper.
92. Malighetti, P., Martini, G., Paleari, S., Redondi, R. (2008). The efficiency of European airports: Do the importance in the EU network and the intensity of competition matter?, University of Bergamo, Department of Economics and Technology Management, Bergamo, Italy, Working Paper 04-2008.
93. Suzuki, S., Nijkamp, P., Pels, E., Rietveld, P. (2009). Comparitive Performance Analysis of European Airports by Means of Extended Data Envelopment Analysis, Tinbergen Institute, The Netherlands, Discussion Paper 2009-024/3.
94. Deprins, D., Simar, L., Tulkens, H. (1984). Measuring labor-efficiency in post offices, in *The Performance of public enterprises: concepts and measurement*, Amsterdam: North-Holland, pp. 243–267.
95. Keeler, T. (1970). Airport Costs and Congestion, *The American Economist*, Vol. 14, No 1, pp. 47–53.
96. Doganis, R., Thompson, G. (1974). Establishing Airport Cost and Revenue Functions, *Aeronautical Journal*, Vol. 78, pp. 285–304.
97. Jeong, J. (2005). An investigation of operating cost of airports: focus on the effect of output scale, MSc thesis, The University of British Columbia, Canada.
98. Greene, W. (2005). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model, *Journal of Econometrics*, Vol. 126, No 2, pp. 269–303.
99. Pels, E. (2001). Relative efficiency of European airports, *Transport Policy*, Vol. 8, No 3, pp. 183–192.

100. Oum, T.H., Yan, J., Yu, C. (2008). Ownership forms matter for airport efficiency: A stochastic frontier investigation of worldwide airports, *Journal of Urban Economics*, Vol. 64, No 2, pp. 422–435.
101. Pavlyuk, D. (2009). Spatial Competition Pressure as a Factor of European Airports' Efficiency, *Transport and Telecommunication*, Vol. 10, No 4, pp. 8–17.
102. Pavlyuk, D. (2010). Multi-tier Spatial Stochastic Frontier Model for Competition and Cooperation of European Airports, *Transport and Telecommunication*, Vol. 11, No 3, pp. 57–66.
103. Forsyth, P.J., Niemeier, H.-M. (2011). The Economic Regulation of Airport Services, Submission to Productivity Commission Inquiry.
104. Assaf, A. (2009). Accounting for size in efficiency comparisons of airports, *Journal of Air Transport Management*, Vol. 15, No 5, pp. 256–258.
105. D'Alfonso, T., Daraio, C., Nastasi, A. (2013). Assesing the Impact of Competition on the Efficiency of Italian Airports.
106. Czerny, A.I. (2006). Price-cap Regulation of Airports: Single-till Versus Dual-till, *Journal of Regulatory Economics*, Vol. 30, No 1, pp. 85–97.
107. Bel, G., Fageda, X. (2009). Privatization, regulation and airport pricing: an empirical analysis for Europe, *Journal of Regulatory Economics*, Vol. 37, pp. 142–161.
108. Starkie, D. (2002). Airport regulation and competition, *Journal of Air Transport Management*, Vol. 8, pp. 63–72.
109. Hotelling, H. (1929). Stability in competition, *The economic journal*, Vol. 39, No 153, pp. 41–57.
110. D' Aspremont, C., Gabszewicz, J.J., Thisse, J.-F. (1979). On Hotelling's "Stability in Competition," *Econometrica*, Vol. 47, No 5, pp. 1145–1150.
111. Salop, S. (1979). Monopolistic competition with outside goods, *The Bell Journal of Economics*, Vol. 10, No 1, pp. 141–156.
112. Irmen, A., Thisse, J.-F. (1998). Competition in Multi-characteristics Spaces: Hotelling Was Almost Right, *Journal of Economic Theory*, Vol. 78, No 1, pp. 76–102.
113. Van Dender, K. (2005). Duopoly prices under congested access, *Journal of Regional Science*, Vol. 45, No 2, pp. 343–362.
114. Haskel, J., Iozzi, A., Valletti, T.M. (2011). Market structure, countervailing power and price discrimination: the case of airports, Imperial College, London, UK, Discussion Paper 2011/03.
115. Biscaia, R., Mota, I. (2011). Models of Spatial Competition: a Critical Review, Department of Economics, University of Porto, Portugal, Working Paper 411.

116. Regional and Small Airports Study (2004). Department of Transport, Canada, TP 14283B.
117. Starkie, D. (2009). The airport industry in a competitive environment: A United Kingdom Perspective, in *Competitive Interaction between Airports, Airlines and High-Speed Rail*, OECD Publishing, pp. 67–93.
118. Civil Aviation Authority (2011). Catchment area analysis, Civil Aviation Authority, UK, Working Paper.
119. Strobach, D. (2006). Competition between airports with an application to the state of Baden-Württemberg, University of Hohenheim, Germany, Working Paper 272/2006.
120. Malina, R. (2006). Competition and regulation in the German airport market, Institute of Transport Economics, Germany, Discussion Paper 10.
121. Hancioglu, B. (2008). The Market Power of Airports, Regulatory Issues and Competition between Airports, German Airport Performance Project, Germany, Working Paper.
122. Bel, G., Fageda, X. (2009). Factors explaining charges in European airports: Competition, market size, private ownership and regulation, Fundación de Estudios de Economía Aplicada, Madrid, Spain, Working Paper 2009-31.
123. Study on Competition between Airports and the Appiication of State Aid Rules (2002). Air Transport Group, UK, Final Report 2002/287.
124. Gillen, D., Niemeier, H.M. (2006). Airport Economics, Policy and Management: The European Union, presented at the Fundación Rafael del Pino workshop on Infrastructure economics: a comparative analyses of the main worldwide airports, Madrid, Spain, p. 53.
125. Black, J. (2009). *A dictionary of economics*, 3rd ed. Oxford ; New York: Oxford University Press, 505 p.
126. Fuss, M., McFadden, D. (1978). *Production Economics : a dual approach to theory and applications*. North-Holland.
127. Koopmans, T.C. (1951). *Activity analysis of production and allocation*. Wiley.
128. Debreu, G. (1951). *The Coefficient of Resource Utilization*. Cowles Commission for Research in Economics, University of Chicago.
129. Farrell, M.J. (1957). The Measurement of Productive Efficiency, *Journal of the Royal Statistical Society. Series A (General)*, Vol. 120, No 3, p. 253.
130. Shephard, R.W. (1970). *Theory of cost and production functions*. Princeton, N.J: Princeton University Press, 308 p.
131. Luenberger, D.G. (1992). New optimality principles for economic efficiency and equilibrium, *Journal of Optimization Theory and Applications*, Vol. 75, No 2, pp. 221–264.
132. Chambers, R.G., Chung, Y., Färe, R. (1996). Benefit and distance functions, *Journal of economic theory*, Vol. 70, No 2, pp. 407–419.

133. Primont, D., Fare, R. (2004). Directional Duality Theory.
134. Greene, W.H. (2012). *Econometric analysis*, 7 edition. Harlow; New York: Pearson Addison Wesley, 1232 p.
135. Coelli, T., Perelman, S. (1996). Efficiency measurement, multiple-output technologie and distance functions: With application to European Railways.
136. Fuentes, H.J., Grifell-Tatjé, E., Perelman, S. (2001). A Parametric Distance Function Approach for Malmquist Productivity Index Estimation, *Journal of Productivity Analysis*, Vol. 15, No 2, pp. 79–94.
137. Lothgren, M. (2000). Specification and estimation of stochastic multiple-output production and technical inefficiency, *Applied Economics*, Vol. 32, No 12, pp. 1533–1540.
138. Roibas, D., Arias, C. (2004). Endogeneity Problems in the Estimation of Multi-Output Technologies, University of Oviedo, Spain, 06/2004.
139. Kumbhakar, S.C. (2011). Estimation of multiple output production functions, Tech. rep., Department of Economics.
140. Van den Broeck, J., Koop, G., Osiewalski, J., Steel, M.F.. (1994). Stochastic frontier models: A Bayesian perspective, *Journal of Econometrics*, Vol. 61, No 2, pp. 273–303.
141. Koop, G., Steel, M.F.J. (2001). Bayesian Analysis of Stochastic Frontier Models, in *A Companion to Theoretical Econometrics*, B. H. Baltagi, Ed. Malden, MA, USA: Blackwell Publishing Ltd, pp. 520–537.
142. Olson, J.A., Schmidt, P., Waldman, D.M. (1980). A Monte Carlo study of estimators of stochastic frontier production functions, *Journal of Econometrics*, Vol. 13, No 1, pp. 67–82.
143. Greene, W.H. (2008). The Econometric Approach to Efficiency Analysis, in *The Measurement of Productive Efficiency and Productivity Change*, H. O. Fried, C. A. K. Lovell, and S. S. Schmidt, Eds. Oxford University Press, pp. 92–250.
144. Campbell, R., Rogers, K., Rezek, J. (2008). Efficient frontier estimation: a maximum entropy approach, *Journal of Productivity Analysis*, Vol. 30, No 3, pp. 213–221.
145. Stevenson, R. (1980). Likelihood Functions for Generalized Stochastic Frontier Functions, *Journal of Econometrics*, No 13, pp. 57–66.
146. Greene, W.H. (1990). A gamma-distributed stochastic frontier model, *Journal of econometrics*, Vol. 46, No 1–2, pp. 141–163.
147. Battese, G.E., Coelli, T. (1993). *A stochastic frontier production function incorporating a model for technical inefficiency effects*. University of New England. Department of Econometrics.
148. Azzalini, A. (1985). A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, Vol. 12, pp. 171–178.

149. Azzalini, A. (2005). The Skew-normal Distribution and Related Multivariate Families, *Scandinavian Journal of Statistics*, Vol. 32, No 2, pp. 159–188.
150. Domınguez-Molina, J.A., González-Farıas, G., Ramos-Quiroga, R. (2003). Skew-normality in stochastic frontier analysis, *Skew-elliptical distributions and their applications: A journey beyond normality*, pp. 223–241.
151. Jondrow, J., Knox Lovell, C.A., Materov, I.S., Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model, *Journal of econometrics*, Vol. 19, No 2, pp. 233–238.
152. Horrace, W.C., Schmidt, P. (1996). Confidence statements for efficiency estimates from stochastic frontier models, *Journal of Productivity Analysis*, Vol. 7, No 2–3, pp. 257–282.
153. Elhorst, J.P. (2009). Spatial panel models, in *Handbook of Applied Spatial Analysis*, New York: Springer: Berlin Heidelberg, pp. 377–407.
154. Bragg, L.A. (2005). Spatial dependence and omitted variable bias effects on efficiency analysis: A study of the Maine dairy industry, The University of Maine.
155. Igliori, D. (2005). Determinants of technical efficiency in agriculture and cattle ranching: A spatial analysis for the brazilian amazon, *University of Cambridge Land Economy Working Paper No. 09.2005*, p. 20.
156. Hadley, D. (2006). Patterns in Technical Efficiency and Technical Change at the Farm-level in England and Wales, 1982-2002, *Journal of Agricultural Economics*, Vol. 57, No 1, pp. 81–100.
157. Schmidt, P., Sickles, R.C. (1984). Production Frontiers and Panel Data, *Journal of Business & Economic Statistics*, Vol. 2, No 4, pp. 367–74.
158. Battese, G.E., Coelli, T.J. (1992). Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India, *Journal of Productivity Analysis*, Vol. 3, No 1–2, pp. 153–169.
159. Kumbhakar, S.C. (1991). Estimation of technical inefficiency in panel data models with firm- and time-specific effects, *Economics Letters*, Vol. 36, No 1, pp. 43–48.
160. Greene, W. (2005). Fixed and Random Effects in Stochastic Frontier Models, *Journal of Productivity Analysis*, Vol. 23, No 1, pp. 7–32.
161. Greene, W. (2003). Distinguishing Between Heterogeneity and Inefficiency: Stochastic Frontier Analysis of the World Health Organization's Panel Data on National Health Care Systems, Stern School of Business, New York University, New York.
162. Wang, H.-J., Ho, C.-W. (2010). Estimating fixed-effect panel stochastic frontier models by model transformation, *Journal of Econometrics*, Vol. 157, No 2, pp. 286–296.

163. Abrate, G., Piacenza, M., Fraquelli, G. (2001). Cost inefficiency or just heterogeneity? An application of stochastic frontier models to the Italian water industry, *Economia e politica industriale*, pp. 51–82.
164. Kopsakangas-Savolainen, M., Svento, R. (2011). Observed and unobserved heterogeneity in stochastic frontier models: An application to the electricity distribution industry, *Energy Economics*, Vol. 33, No 2, pp. 304–310.
165. Ahn, S.C., Sickles, R.C. (2000). Estimation of long-run inefficiency levels: a dynamic frontier approach, *Econometric Reviews*, Vol. 19, No 4, pp. 461–492.
166. Tsionas, E.G. (2006). Inference in dynamic stochastic frontier models, *Journal of Applied Econometrics*, Vol. 21, No 5, pp. 669–676.
167. Emvalomatis, G. (2012). Adjustment and unobserved heterogeneity in dynamic stochastic frontier models, *Journal of Productivity Analysis*, Vol. 37, No 1, pp. 7–16.
168. Schettini, D. (2010). Eficiência produtiva da indústria de transformação nas regiões brasileiras: uma análise de fronteiras estocásticas e cadeias espaciais de Markov, São Paulo.
169. Feng, X., Yuan, Q., Jia, P., Hayashi, Y. (2011). Effect of High-Speed Rail Development on the Progress of Regional Economy, *Journal of the Society for Transportation and Traffic Studies*, Vol. 2, No 1, pp. 46–55.
170. Misra, K. (2011). Does competition improve public school efficiency? A spatial analysis, Mississippi State University.
171. Tobler, W.R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region, *Economic Geography*, Vol. 46, p. 234.
172. Paelinck, J.H.P. (1979). *Spatial econometrics*. Farnborough, Eng: Saxon House, 211 p.
173. Anselin, L. (2001). Spatial Econometrics, in *A Companion to Theoretical Econometrics*, B. H. Baltagi, Ed. Malden, MA, USA: Blackwell Publishing Ltd, pp. 310–331.
174. Upton, G.J.G. (1985). *Spatial data analysis by example*. Chichester ; New York: Wiley, 2 p.
175. LeSage, J.P., Pace, R.K. (2009). *Introduction to spatial econometrics*. Boca Raton: CRC Press, 374 p.
176. LeSage, J.P. (1999). *Spatial econometrics*. Regional Research Institute, West Virginia University, 296 p.
177. Kelejian, H.H., Prucha, I.R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances, *The Journal of Real Estate Finance and Economics*, Vol. 17, No 1, pp. 99–121.
178. Lee, L. (2007). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models, *Journal of Econometrics*, Vol. 137, No 2, pp. 489–514.

179. Elhorst, J.P. (2010). Applied spatial econometrics: raising the bar, *Spatial Economic Analysis*, Vol. 5, No 1, pp. 9–28.
180. Ord, K. (1975). Estimation Methods for Models of Spatial Interaction, *Journal of the American Statistical Association*, Vol. 70, No 349, pp. 120–126.
181. Pace, R.K., Barry, R. (1997). Sparse Spatial Autoregressions, *Statistics & Probability Letters*, Vol. 33, No 3, pp. 291–297.
182. Pace, R.K., LeSage, J.P. (2004). Chebyshev approximation of log-determinants of spatial weight matrices, *Computational Statistics & Data Analysis*, Vol. 45, No 2, pp. 179–196.
183. Smirnov, O., Anselin, L. (2001). Fast maximum likelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach, *Computational Statistics & Data Analysis*, Vol. 35, No 3, pp. 301–319.
184. Manski, C.F. (1993). Identification of endogenous social effects: the reflection problem, *The review of economic studies*, Vol. 60, No 3, pp. 531–542.
185. Lee, L., Liu, X., Lin, X. (2010). Specification and estimation of social interaction models with network structures, *The Econometrics Journal*, Vol. 13, No 2, pp. 145–176.
186. Cochrane, D., Orcutt, G.H. (1949). Application of Least Squares Regression to Relationships Containing Auto- Correlated Error Terms, *Journal of the American Statistical Association*, Vol. 44, No 245, p. 32.
187. Barrios, E.B., Lavado, R.F. (2010). Spatial stochastic frontier models, Philippine Institute for Development Studies, The Philippines, Discussion Paper 2010-08.
188. Landagan, O., Barrios, E. (2007). An estimation procedure for a spatial–temporal model, *Statistics & Probability Letters*, Vol. 77, No 4, pp. 401–406.
189. Pavlyuk, D. (2012). Maximum Likelihood Estimator for Spatial Stochastic Frontier Models, in *Proceedings of the 12th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'12)*, Riga, Latvia, pp. 11–19.
190. Simwaka, K. (2012). Maximum likelihood estimation of a stochastic frontier model with residual covariance, MRPA, Germany, Paper 39726.
191. Battese, G.E., Coelli, T.J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data, *Empirical economics*, Vol. 20, No 2, pp. 325–332.
192. Schmidt, A.M., Moreira, A.R.B., Helfand, S.M., Fonseca, T.C.O. (2008). Spatial stochastic frontier models: accounting for unobserved local determinants of inefficiency, *Journal of Productivity Analysis*, Vol. 31, No 2, pp. 101–112.
193. Basar, T., Olsder, G.J. (1998). *Dynamic Noncooperative Game Theory, 2nd Edition*. Society for Industrial and Applied Mathematics.

194. Gaertner, W. (1974). A dynamic model of interdependent consumer behavior, *Journal of Economics*, Vol. 34, No 3–4, pp. 327–344.
195. Pavlyuk, D. (2013). Distinguishing Between Spatial Heterogeneity and Inefficiency: Spatial Stochastic Frontier Analysis of European Airports, *Transport and Telecommunication*, Vol. 14, No 1.
196. Anselin, L., Bera, A. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics., in *Handbook of applied economic statistics*, New York: Marcel Dekker.
197. González-Farías, G., Domínguez-Molina, J.A., Gupta, A. (2004). The closed skew-normal distribution, in *Skew-elliptical distributions and their applications: a journey beyond normality*, Chapman & Hall/CRC, Boca Raton, FL, pp. 25–42.
198. Horrace, W.C. (2005). Some results on the multivariate truncated normal distribution, *Journal of Multivariate Analysis*, Vol. 94, No 1, pp. 209–221.
199. Azzalini, A., Arellano-Valle, R.B. (2013). Maximum penalized likelihood estimation for skew-normal and skew-< i> t</i> distributions, *Journal of Statistical Planning and Inference*, Vol. 143, No 2, pp. 419–433.
200. Lin, T.I. (2009). Maximum likelihood estimation for multivariate skew normal mixture models, *Journal of Multivariate Analysis*, Vol. 100, No 2, pp. 257–265.
201. Flecher, C., Naveau, P., Allard, D. (2009). Estimating the closed skew-normal distribution parameters using weighted moments, *Statistics & Probability Letters*, Vol. 79, No 19, pp. 1977–1984.
202. Tallis, G.M. (1961). The moment generating function of the truncated multi-normal distribution, *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 23, No 1, pp. 223–229.
203. Greene, W.H. (2002). *LIMDEP 10 Econometric Modeling Guide*. New York: Econometric Software, 2500 p.
204. Stakhovych, S., Bijmolt, T.H.A. (2009). Specification of spatial models: a simulation study on weights matrices, *Papers in Regional Science*, Vol. 88, No 2, pp. 389–408.
205. Genz, A. (2004). Numerical computation of rectangular bivariate and trivariate normal and t probabilities, *Statistics and Computing*, Vol. 14, No 3, pp. 251–260.
206. CRAN (2014). *R software*. CRAN.
207. Olsen, R.J. (1978). Note on the Uniqueness of the Maximum Likelihood Estimator for the Tobit Model, *Econometrica*, Vol. 46, No 5, pp. 1211–1215.

208. Coelli, T. (1996). A guide to FRONTIER version 4.1: a computer programme for frontier production function estimation, Centre for Efficiency and Productivity Analysis, University of New England, Armidale, Australia, CEPA Working Paper 96/08.
209. Henningsen, A. (2013). *frontier: A Package for Stochastic Frontier Analysis*. .
210. Banker, R.D., Natarajan, R. (2008). Evaluating Contextual Variables Affecting Productivity Using Data Envelopment Analysis, *Operations Research*, Vol. 56, No 1, pp. 48–58.
211. Eurostat (2014). European Statistics Database, Statistical Office of the European Communities (Eurostat).
212. DAFIF (2014). The Digital Aeronautical Flight Information File Database.
213. OpenFlights (2014). The OpenFlights Airports, Airlines, and Routes Database.
214. CIESIN (2014). The Gridded Population of the World Database, Centre for International Earth Science Information Network.
215. Tsekeris, T. (2011). Greek airports: Efficiency measurement and analysis of determinants, *Journal of Air Transport Management*, Vol. 17, No 2, pp. 140–142.
216. Plumper, T., Neumayer, E. (2010). Model specification in the analysis of spatial dependence, *European Journal of Political Research*, Vol. 49, No 3, pp. 418–442.
217. O'Sullivan, D. (2010). *Geographic information analysis*, 2nd ed. Hoboken, N.J: John Wiley & Sons, 405 p.
218. Florax, R.J.G.., Folmer, H., Rey, S.J. (2003). Specification searches in spatial econometrics: the relevance of Hendry's methodology, *Regional Science and Urban Economics*, Vol. 33, No 5, pp. 557–579.
219. TDC (2012). Informe de fiscalización de la imputación por la entidad "aeropuertos españoles y navegación aérea" (AENA) a cada uno de los aeropuertos de los ingresos, gastos, e inversiones correspondientes a la actividad aeroportuaria, en los ejercicios 2009 y 2010., Tribunal de Cuentas, Spain, Doc 938.
220. Fageda, X., Voltes-Dorta, A. (2012). Efficiency and profitability of Spanish airports: a composite non-standard profit function approach, Universitat de Barcelona, Spain, Working Paper.
221. Kao, Y.-H., Bera, A.K. (2013). Spatial regression: the curious case of negative spatial dependence, in *VII World Conference of the Spatial Econometrics Association. Accessed online: www. spatialeconometricsassociation. org*.
222. CAA (2011). The Airport Charges Regulations, Civil Aviation Authority, UK, 2491.
223. Box, G.E.P., Jenkins, G.M., Reinsel, G.C. (2008). *Time series analysis: forecasting and control*, 4th ed. Hoboken, N.J: John Wiley, 746 p.

224. Abbott, M., Wu, S. (2002). Total Factor Productivity and Efficiency of Australian Airports, *The Australian Economic Review*, Vol. 35, No 3, pp. 244–260.
225. Ablanedo-Rosas, J.H., Gemoets, L.A. (2010). Measuring the efficiency of Mexican airports, *Journal of Air Transport Management*, Vol. 16, No 6, pp. 343–345.
226. Adler, N., Berechman, J. (2001). Measuring airport quality from the airlines' viewpoint: an application of data envelopment analysis, *Transport Policy*, Vol. 8, No 3, pp. 171–181.
227. Assaf, A., Gillen, D. (2012). Measuring the joint impact of governance form and economic regulation on airport efficiency, *European Journal of Operational Research*, Vol. 220, No 1, pp. 187–198.
228. Assaf, A. (2010). The cost efficiency of Australian airports post privatisation: A Bayesian methodology, *Tourism Management*, Vol. 31, No 2, pp. 267–273.
229. Assaf, A. (2011). Bootstrapped Malmquist indices of Australian airports, *The Service Industries Journal*, Vol. 31, No 5, pp. 829–846.
230. Barros, C.P., Managi, S. (2008). Productivity Change of UK Airports: 2000-2005, School of Economics and Management, Lisbon, Portugal, Working Paper 22/2008/DE/UECE.
231. Barros, C.P. (2012). Performance, heterogeneity and managerial efficiency of African airports: the Nigerian Case.
232. Barros, C.P. (2011). Cost efficiency of African airports using a finite mixture model, *Transport Policy*.
233. Bazargan, M. (2003). Size versus efficiency: a case study of US commercial airports, *Journal of Air Transport Management*, Vol. 9, No 3, pp. 187–193.
234. Beaudoin, J. (2006). An empirical investigation of the effects of managerial and ownership structure on the efficiency of North America's airports, The University of British Columbia, Vancouver, Canada.
235. Chi-Lok, A.Y., Zhang, A. (2009). Effects of competition and policy changes on Chinese airport productivity: An empirical investigation, *Journal of Air Transport Management*, Vol. 15, No 4, pp. 166–174.
236. Chow, C.K.W., Fung, M.K.Y. (2012). Estimating indices of airport productivity in Greater China, *Journal of Air Transport Management*, Vol. 24, pp. 12–17.
237. De Azevedo, A.S. (2011). Exploratory analysis on LCC potential to influence airport efficiency.
238. Dresner, M. (2006). Productive and Operational Efficiency of US Airports with Joint Consideration of both Desirable and Undesirable Outputs, p. 19.
239. Fernandes, E., Pacheco, R.R. (2002). Efficient use of airport capacity, *Transportation Research Part A: Policy and Practice*, Vol. 36, No 3, pp. 225–238.

240. Ferro, G., Garitta, F., A Romero, C., others (2010). Relative efficiency of Argentinean airports.
241. Fung, M.K.Y., Chow, C.K.W. (2011). Note on Productivity Convergence of Airports in China, *Pacific Economic Review*, Vol. 16, No 1, pp. 120–133.
242. Fung, M.K.Y., Wan, K.K.H., Hui, Y.V., Law, J.S. (2008). Productivity changes in Chinese airports 1995–2004, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 44, No 3, pp. 521–542.
243. Gillen, D., Lall, A. (2001). Non-parametric measures of efficiency of U.S. airports, *International Journal of Transport Economics*, Vol. 28, No 3, pp. 283–305.
244. Gök, U. (2012). Evaluating Turkish Airports Efficiencies Using Data Envelopment Analysis, Eastern Mediterranean University (EMU).
245. Hooper, P.., Hensher, D.. (1997). Measuring total factor productivity of airports— an index number approach, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 33, No 4, pp. 249–259.
246. Jardim, J.P.F. (2012). Airports Efficiency Evaluation Based on MCDA and DEA Multidimensional Tools, The University of Beira Interior, Covilhã, Portugal.
247. Kocak, H. (2011). Efficiency examination of Turkish airport with DEA approach, *International Business Research*, Vol. 4, No 2.
248. Lai, P.-L. (2013). A Study on the Relationship between Airport Privatisation and Airport Efficiency, Cardiff University.
249. Lam, S.W., Low, J.M.W., Tang, L.C. (2009). Operational efficiencies across Asia Pacific airports, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 45, No 4, pp. 654–665.
250. Liebert, V. (2011). Airport Benchmarking: An Efficiency Analysis of European Airports from an Economic and Managerial Perspective, Jacobs University, Bremen, Germany.
251. Lin, L.C., Hong, C.H. (2006). Operational performance evaluation of international major airports: An application of data envelopment analysis, *Journal of Air Transport Management*, Vol. 12, No 6, pp. 342–351.
252. Lozano, S., Gutiérrez, E. (2011). Efficiency Analysis and Target Setting of Spanish Airports, *Networks and Spatial Economics*, Vol. 11, No 1, pp. 139–157.
253. Marques, R.C., Simões, P. (2010). Measuring the influence of congestion on efficiency in worldwide airports, *Journal of Air Transport Management*, Vol. 16, No 6, pp. 334–336.
254. Martín, J.C., Román, C. (2006). A Benchmarking Analysis of Spanish Commercial Airports. A Comparison Between SMOP and DEA Ranking Methods, *Networks and Spatial Economics*, Vol. 6, No 2, pp. 111–134.

255. Martini, G., Scotti, D., Volta, N. (2011). The impact of local air pollution on airport efficiency assessment: evidence from Italy, in *EWEPA 2011. XII European Workshop on Efficiency and Productivity Analysis*.
256. Murillo-Melchor, C. (1999). An analysis of technical efficiency and productivity changes in Spanish airports using the Malmquist index, *International Journal of Transport Economics*, Vol. 26, No 2, pp. 271–292.
257. Oum, T.H., Yu, C., Fu, X. (2003). A comparative analysis of productivity performance of the world's major airports: summary report of the ATRS global airport benchmarking research report—2002, *Journal of Air Transport Management*, Vol. 9, No 5, pp. 285–297.
258. Pacheco, R.R., Fernandes, E. (2003). Managerial efficiency of Brazilian airports, *Transportation Research Part A: Policy and Practice*, Vol. 37, No 8, pp. 667–680.
259. Parker, D. (1999). The Performance of BAA before and after Privatisation: A DEA Study, *Journal of Transport Economics and Policy*, Vol. 33, No 2, pp. 133–145.
260. Sarkis, J., Talluri, S. (2004). Performance based clustering for benchmarking of US airports, *Transportation Research Part A: Policy and Practice*, Vol. 38, No 5, pp. 329–346.
261. Schaar, D., Sherry, L. (2008). Comparison of data envelopment analysis methods used in airport benchmarking, in *Proceedings of 3rd International Conference on Research in Air Transportation (ICRAT)*, pp. 339–346.
262. Suzuki, S., Nijkamp, P. (2013). A Stepwise Efficiency Improvement DEA Model for Airport Management with a Fixed Runway Capacity, Tinbergen Institute Discussion Paper.
263. Tovar, B., Martín-Cejas, R.R. (2010). Technical efficiency and productivity changes in Spanish airports: A parametric distance functions approach, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 46, No 2, pp. 249–260.
264. Vasigh, B., Gorjidooz, J. (2006). Productivity analysis of public and private airports: a casual investigation, *Journal of Air Transportation*, Vol. 11, No 3, pp. 144–163.
265. Wanke, P.F. (2012). Capacity shortfall and efficiency determinants in Brazilian airports: Evidence from bootstrapped DEA estimates, *Socio-Economic Planning Sciences*, Vol. 46, No 3, pp. 216–229.
266. Wanke, P.F. (2012). Efficiency of Brazil's airports: Evidences from bootstrapped DEA and FDH estimates, *Journal of Air Transport Management*, Vol. 23, No C, pp. 47–53.
267. Yang, H.-H. (2010). Measuring the efficiencies of Asia–Pacific international airports – Parametric and non-parametric evidence, *Computers & Industrial Engineering*, Vol. 59, No 4, pp. 697–702.
268. Yoshida, Y., Fujimoto, H. (2004). Japanese-airport benchmarking with the DEA and endogenous-weight TFP methods: testing the criticism of overinvestment in Japanese

regional airports, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 40, No 6, pp. 533–546.

269. Yoshida, Y. (2004). Endogenous-weight TFP measurement: methodology and its application to Japanese-airport benchmarking, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 40, No 2, pp. 151–182.
270. Yu, M.-M., Hsu, S.-H., Chang, C.-C., Lee, D.-H. (2008). Productivity growth of Taiwan's major domestic airports in the presence of aircraft noise, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 44, No 3, pp. 543–554.
271. Zhang, B., Wang, J., Liu, C., Zhao, Y. (2012). Evaluating the technical efficiency of Chinese airport airside activities, *Journal of Air Transport Management*, Vol. 20, pp. 23–27.

APPENDICES

Appendix 1. List of existing airport benchmarking studies

Table A1.1. Summary of used airport inputs, outputs and benchmarking methodologies in existing studies

Source	Inputs	Outputs	Methodology
Abbott & Wu 2002[224]	Employment (FTE) Capital (stock) Runways (length)	APM Cargo	DEA
Abdesaken & Cullman 2006[67]	Runways (number) Gates (number) Terminals (area) Employment (FTE) Baggage belts (number) Car parking (places)	WLU	DEA, PFP
Abdesaken & Cullman 2006[67]	Terminals (area) Runways (number) Runways (length) Employment (FTE)	ATM	DEA
Ablanedo-Rosas & Gemoets 2010[225]	Operations per hour Passengers per hour	ATM APM Cargo	DEA
Abrate & Erbetta 2007[35]	Employment (cost) Operational costs Terminals (area) Aircraft stands (number) Runways (length)	APM ATM Cargo Airport fees Aeronautical revenue Non-aeronautical revenue	SFA
Adler & Berechman 2001[226]	Terminals (number) Runways (number) Distance to city centres Minimum connecting time	Principal component, calculated from a questionnaire	DEA
Assaf & Gillen 2012[227]	Employment (FTE) Runways (number) Terminals (area) Operational costs	APM ATM Non-aeronautical revenue	SFA
Assaf 2010[228]	Employment (FTE) Terminals (area) Runways (number)	ATM APM Cargo	DEA
Assaf 2011[229]	Employment (FTE) Terminals (area) Operational costs	ATM APM Cargo	TFP
Assaf et al. 2012[22]	Employment (costs) Capital (costs) Operational costs	Aeronautical revenue Non-aeronautical revenue	SFA
Barros & Assef 2009[81]	Gates (number) Terminals (area) Runways (number) Employment (costs)	APM Cargo ATM (Air Carrier and Commuter Movements)	DEA
Barros & Dieke 2007[90]	Employment (costs) Capital (costs) Operational costs	ATM APM Cargo Handing revenues Aeronautical revenue Non-aeronautical revenue	DEA
Barros & Managi 2008[230]	Operational costs Employment (FTE) Capital (stock)	ATM APM Cargo	DEA, TFP

Source	Inputs	Outputs	Methodology
Barros & Marques 2008[20] Barros 2012[231]	Employment (costs) Operational costs Capital (investments)	ATM APM	SFA
Barros & Peypoch 2007[83]	Operational costs Capital (investments)	ATM APM Cargo Aeronautical revenue Handling revenue Non-aeronautical revenue	DEA
Barros & Sampaio 2004[18]	Employment (FTE) Capital (costs)	ATM APM Cargo Non-aeronautical revenue Aeronautical revenue	DEA
Barros & Weber 2009[21]	Operational costs Employment (FTE) Capital (stock)	ATM APM Cargo	DEA
Barros 2011[232]	Operational cost Employment (FTE) Capital (stock)	Capital (investments) ATM APM	SFA
Barros et al. 2008[37]	Terminals (area) Employment (FTE) Runways (area)	WLU	SFA
Barros et al. 2010[82]	Employment (FTE) Terminals (area) Runways (area) Car parking (places)	ATM APM Cargo	DEA, TFP
Barros et al. 2011[84]	Operational costs Runways (area) Capital (stock)	ATM APM Cargo	DEA
Bazargan 2003[233]	Operational costs Non-operating costs Runways (number) Gates (number) ATM APM	Aeronautical revenue Non-aeronautical revenue	DEA
Beaudoin 2006[234]	Employment (FTE) Runways (number) Runways (length) Terminals (area)	ATM APM Aeronautical revenues	DEA, SFA
Chi-Lok & Zhang 2009[235]	Runways (length) Terminals (area)	ATM APM Cargo	DEA
Chow & Fung 2012[236]	Runways (length) Terminals (area)	ATM APM Cargo	TFP
Curi et al. 2009[87]	Employment (FTE) Runways (number) Aircraft stands (area)	ATM APM Cargo	DEA
Curi et al. 2009[87]	Employment (FTE) Runways (number) Terminals (area)	Aeronautical revenue Non-aeronautical revenue	DEA
Curi et al. 2011[89]	Employment (FTE) Runways (number) Aircraft stands (area)	ATM APM Cargo	DEA

Source	Inputs	Outputs	Methodology
D'Alfonso et al. 2013[105]	Airport (area) Runways (number) Runways (area) Terminals (area) Terminals (number) Gates (number) Check-in desks (number)	ATM APM Cargo	DEA
de Azevedo Domingues 2011[237]	Aircraft stands (number) Gates (number) Runways (capacity) Check-in desks (number) Terminals (area) Baggage belts (number)	APM ATM	DEA
Dresner 2006[238]	Airport (area) Runways (number) Runways (area)	APM ATM (non-delayed) ATM (delayed) Cargo Time delays	DEA
Fernandes & Pacheco 2002[239]	Aircraft stands (area) Departure lounges (area Baggage belts (number) Check-in desks (number) Car parking (places) Curb frontage (length)	APM (domestic)	DEA
Ferro et al. 2010[240]	Runways (area) Employment (FTE) Ramp (area) Terminals (area)	ATM APM Cargo	DEA
Fung & Chow 2011[241]	Runways (length) Terminals (area)	ATM APM Cargo	TFP
Fung et al. 2008[242]	Runways (length) Terminals (area)	ATM APM Cargo	DEA, TFP
Gillen & Lall 1997[17]	Gates (number) Runways (number) Runways (area) Terminals (area) Employment (FTE) Baggage belts (number) Car parking (places) Airport (area)	ATM APM Cargo	DEA
Gillen & Lall 2001[243]	Gates (number) Runways (number) Employment (FTE) Baggage belts (number) Car parking (places)	APM Cargo	DEA
Gillen & Lall 2001[243]	Airport (area) Runways (number) Runways (area) Employment (FTE)	ATM (Air Carrier and Commuter Movements)	DEA
Gitto & Mancuso 2012[23]	Employment (costs) Capital (investments) Operational costs	ATM APM Cargo Aeronautical revenue Non-aeronautical revenue	TFP
Gitto 2008[24] Gitto & Mancuso 2010[88]	Airport (area) Runways (area) Employment (FTE)	ATM APM Cargo	DEA, FDH

Source	Inputs	Outputs	Methodology
Gitto 2008[24] Gitto & Mancuso 2010[88]	Employment (costs) Operational costs Capital (investments)	Aeronautical revenue Non-aeronautical revenue	DEA
Gök 2012[244]	Runways (length) Terminals (area)	ATM APM Cargo	DEA
Holvad & Graham 2000[14]	Employment (FTE) Capital (costs) Operational Costs	APM Cargo	DEA, FDH
Hooper & Hensher 1997[245]	Employment (costs) Capital (costs) Operational costs	Aeronautical revenue Non-aeronautical revenue	TFP
Jardim 2012[246]	Runways (number) Aircraft stands (number) Terminals (area) passenger and cargo	ATM APM Cargo	DEA
Kocak 2011[247]	Operational costs Employment (FTE) Runways (capacity) Potential passengers (number)	ATM APM Cargo	DEA
Lai 2013[248]	Employment (FTE) Gates (number) Runways (number) Terminals (area) Runways (length) Operational costs	ATM APM Cargo Total Revenue	DEA, AHP
Lam et al. 2009[249]	Employment (FTE) Capital (stock) Operational costs Trade value	ATM APM Cargo	DEA
Liebert 2011[250]	Employment (costs) Operational costs Runways (capacity)	ATM APM Cargo Non-aeronautical revenue	DEA
Lin & Hong 2006[251]	Employment (FTE) Check-in desks (number) Runways (number) Gates (number) Employment (FTE) Baggage belts (number) Aircraft stands (number) Terminals (area)	ATM APM Cargo	DEA
Lozano & Gutiérrez 2011[252]	Runways (area) Aircraft stands (area) Terminals (area) Check-in desks (number) Gates (number) Baggage belts (number)	ATM APM Cargo	DEA
Malighetti et al. 2007[91]	Airport (area) Runways (length) Aircraft stands (number)	ATM	DEA
Malighetti et al. 2008[92]	ATM Terminals (area) Check-in desks (number) Aircraft stands (number) Baggage belts (number)	APM	DEA
Malighetti et al. 2009[41]	Airport (area) Runways (length) Aircraft stands (number)	ATM	DEA

Source	Inputs	Outputs	Methodology
Malighetti et al. 2009[41]	Terminals (area) ATM Check-in desks (number) Aircraft stands (number) Baggage belts (number)	APM	DEA
Malighetti et al. 2010[44]	Runways (capacity) Aircraft stands (number) Terminals (area) Check-in desks (number) Baggage belts (number) Employment (FTE)	ATM WLU	SFA
Marques & Simões 2010[253]	Runways (number) Gates (number) Terminals (area) Employment (FTE)	ATM APM Cargo	DEA
Martin & Roman 2001[77]	Employment (costs) Capital (costs) Operational costs	ATM APM Cargo	DEA
Martín & Román 2006[254]	Employment (costs) Capital (costs) Operational costs	ATM APM Cargo Aeronautical revenue Non-aeronautical revenue	DEA
Martín et al. 2009[39]	Employment (FTE) Capital (costs) Operational costs	ATM WLU	SFA
Martini et al. 2011[255]	Terminals (area) Check-in desks (number) Employment (FTE) Runways (capacity) Aircraft stands (number)	ATM WLU Pollution index	SFA
Merkert et al. 2010[29]	Employment (FTE) Runways (number) Terminals (area) Gates (number)	ATM APM Cargo	DEA, PFP
Muller et al. 2009[40]	Terminals (area) Check-in desks (number) Gates (number)	APM	DEA, SFA, PFP
Murillo-Melchor 1999[256]	Employment (FTE) Capital (costs) Operational costs	APM	TFP
Oum et al. 2003[257]	Employment (FTE) Runways (number) Terminals (area) Gates (number)	ATM APM Cargo Non-aeronautical revenue	TFP
Oum et al. 2006[6]	Employment (FTE) Operational costs	ATM APM Non-aeronautical revenue	TFP
Oum et al. 2008[100]	Employment (costs) Operational costs Runways (number) Terminals (area)	ATM APM Non-aeronautical revenue	SFA
Oum et al. 2011[28]	Operational Costs Employment (costs)	ATM APM Cargo Non-aeronautical revenue	TFP

Source	Inputs	Outputs	Methodology
Pacheco & Fernandes 2003[258]	Aircraft stands (area) Departure lounges (area) Baggage belts (number) Check-in desks (number) Car parking (places) Curb frontage (length)	APM	DEA
Parker 1999[259]	Employment (FTE) Capital (stock) Operational costs	APM Cargo	DEA
Pathomsiri et al. 2008[65]	Airport (area) Runways (number) Runways (area)	ATM (not delayed) ATM (delayed) APM Cargo Time delays	TFP
Pels et al. 2003[34]	Check-in desks (number) Baggage belts (number) Terminals (area) Aircraft stands (area)	ATM	DEA, SFA
Pels et al. 2003[34]	Predicted ATM Check-in desks (number) Baggage belts (number)	APM	DEA, SFA
Perelman & Serebrisky 2010[80]	Employment (FTE) Runways (number) Gates (number)	ATM APM Cargo	DEA
Psaraki-Kalouptsidi & Kalakou 2011[78]	Terminals (area) Departure lounges (area) Arrival lounges (area) Check-in desks (area) Employment (FTE) Aircraft stands (area)	ATM APM	DEA
Sarkis & Talluri 2004[260]	Operational costs Employment (FTE) Runways (number) Gates (number)	ATM APM Cargo Total revenue	DEA
Sarkis 2000[75]	Operational costs Employment (FTE) Runways (number) Gates (number)	ATM APM Cargo Total revenue	DEA
Schaar & Sherry 2008[261]	Operational costs Total costs Runways (number) Gates (number)	Aeronautical revenue Non-aeronautical revenue ATM	DEA
Scotti 2011[8]	Runways (capacity) Aircraft stands (number) Terminals (area) Check-in desks (number) Baggage belts (number) Employment (FTE)	ATM APM Cargo	SFA
Scotti et al. 2012[42]	Runways (capacity) Aircraft stands (number) Terminals (area) Check-in desks (number) Baggage belts (number) Employment (FTE)	ATM APM Cargo	SFA
Suzuki & Nijkamp 2013[262]	Operational Costs Employment (costs) Runways (length)	Total revenue	DEA
Suzuki et al. 2009[93]	Runways (number) Terminals (area) Gates (number) Employment (FTE)	ATM APM	DEA

Source	Inputs	Outputs	Methodology
Tovar & Martín-Cejas 2010[263]	Gates (number) Employment (FTE) Airport (area)	ATM Average size of aircraft Aeronautical revenue Non-aeronautical revenue	SFA
Tsekeris 2011[215]	Runways (number) Operating hours Terminals (area) Aircraft stands (area)	ATM APM Cargo	DEA
Ulku 2009[74]	Employment (costs) Operational Costs Capital (stock)	ATM APM	DEA
Vasigh & Gorjidooz 2006[264]	Operational Costs Capital (stock) Runways (area)	ATM APM Aeronautical revenue Non-aeronautical revenue Landing fee	TFP
Voltes 2008[38]	Total costs Terminals (area) Runways (length) Warehouse (area) Gates (number) Baggage belts (number) Check-in desks (number) Employment (FTE) Total landed maximum takeoff weight	ATM APM Cargo Aeronautical revenue	SFA
Wanke 2012a [265]	Airport (area) Aircraft stands (area) Aircraft stands (number) Runways (number) Runways (length) Terminals (area) Car parking (places)	ATM APM Cargo	DEA, FDH
Wanke 2012b [266]	ATM	APM Cargo	DEA, FDH
Yang 2010[267]	Employment (FTE) Runways (number) Operational costs	Total revenue	DEA, SFA
Yoshida & Fujimoto 2004[268]	Runways (length) Terminals (area) Access cost Employment (FTE)	ATM APM Cargo	DEA, TFP
Yoshida 2004[269]	Terminals (area) Runways (length)	ATM APM Cargo	TFP
Yu 2004[64]	Runways (area) Terminals (area) Aircraft stands (area) Active route Population	ATM APM Aircraft noise	DEA
Yu et al. 2008[270]	Employment (FTE) Capital (stock) Operational costs	APM	DEA
Zhang et al. 2012 [271]	Take-off distance available Landing distance available Aircraft-parking position	ATM APM Cargo	DEA

Appendix 2. Source codes for sample DGP simulations

```
# Define DGP A parameters
params <- list(n=40, beta0=5, beta1=10, beta2=1,
                  sigmaV=0.5, sigmaU=2.5, rhoU=0.5)
parDef <- createParDef(selection = params,
            banker = list(loggingLevel="debug",inefficiency="half-normal",
            control=list(),
            parDef=createParDef(selection = params, banker=list())))
# Set random number generator seed for reproducibility
set.seed(3)
# Generate DGP A values
dgp <- evalFunctionOnParameterDef(parDef, spfrontier.dgp)
# Plot DGP A values and frontier
plot(dgp$x, dgp$y,pch=15,col="red", xlab="x", ylab="y")
xf<-seq(1,10,by=0.01)
yf<-5+10*log(xf)+log(xf)^2
lines(xf, yf,col="red")

# Define DGP B parameters
params$beta0 <- 2
params$rhoU <- -0.5
parDef <- createParDef(selection = params,
            banker = list(loggingLevel="debug",inefficiency="half-normal",
            control=list(),
            parDef=createParDef(selection = params, banker=list())))
# Set random number generator seed for reproducibility
set.seed(3)
# Generate DGP B values
dgp2 <- evalFunctionOnParameterDef(parDef, spfrontier.dgp)
# Plot DGP A values and frontier
points(dgp2$x, dgp2$y,pch=16, col="dark blue")
xf<-seq(1,10,by=0.01)
yf<-2+10*log(xf)+log(xf)^2
lines(xf, yf,col="dark blue",lty=2)

# Define DGP C parameters
params$rhoU <- 0
params$sigmaU <- 0.5
params$sigmaV <- 1.5
parDef <- createParDef(selection = params,
            banker = list(loggingLevel="debug",inefficiency="half-normal",
            control=list(),
            parDef=createParDef(selection = params, banker=list())))
# Set random number generator seed for reproducibility
```

```
set.seed(3)
# Generate DGP B values
dgp3 <- evalFunctionOnParameterDef(parDef, spfrontier.dgp)
# Plot DGP A values and frontier
points(dgp3$x, dgp3$y,pch=17, col="green")


# Plot the legend
legend("bottomright",
      legend = c("Process A true", "Processes B and C true frontier","Process
A data points", "Process B data points","Process C data points"),
      lty=c(1,2,NA,NA,NA),
      pch = c(NA,NA,15,16,17),
      col = c('red','dark blue', 'red','dark blue', 'green') ,merge = TRUE)

# Prepare a list of spatial weights
listw <- mat2listw(dgp$W_u,style="B")


# Test spatial autocorrelation for DGPs
moran.test(as.vector(dgp$y - 5-10*log(dgp$x)-log(dgp$x)^2),
     listw, alternative="two.sided")
moran.test(as.vector(dgp2$y - 2-10*log(dgp2$x)-log(dgp2$x)^2),
     listw, alternative="two.sided")
moran.test(as.vector(dgp3$y - 2-10*log(dgp3$x)-log(dgp3$x)^2),
     listw, alternative="two.sided")
```

Appendix 3. Official documentation of the *spfrontier* package

Package ‘spfrontier’

December 22, 2014

Type Package

Title Spatial Stochastic Frontier models estimation

Version 0.1.12

Date 2014-12-21

Author Dmitry Pavlyuk <Dmitry.V.Pavlyuk@gmail.com>

Maintainer Dmitry Pavlyuk <Dmitry.V.Pavlyuk@gmail.com>

Description A set of tools for estimation of various spatial specifications of stochastic frontier models

License GPL (>= 2)

Depends R (>= 3.0),moments,ezsim,tmvtnorm,mvtnorm,maxLik

Imports methods, parallel,spdep

ZipData no

Repository CRAN

Repository/R-Forge/Project spfrontier

Repository/R-Forge/Revision 45

Repository/R-Forge/DateTimeStamp 2014-12-21 16:00:12

Date/Publication 2014-12-21 18:05:06

NeedsCompilation no

R topics documented:

1

spfrontier-package Spatial Stochastic Frontier

Description

Spatial Stochastic Frontier

Details

A set of tools for estimation (MLE) of various spatial specifications of stochastic frontier models

Author(s)

Dmitry Pavlyuk <Dmitry.V.Pavlyuk@gmail.com>

airports European airports statistical data

Description

The spfrontier package includes the dataset airports, containing information about European airports infrastructure and traffic statistics in 2011.

Format

An unbalanced panel of 395 Euripean airports in 2008-2012 (1763 observations) on the following 31 variables.

ICAO Airport ICAO code

AirportName Airport official name

Country Airport's country name

longitude Airport longitude

latitude Airport latitude

Year Observation year

PAX A number of carried passengers

ATM A number of of air transport movements served by an airport

Cargo A total volume of cargo served by an airport

Population100km A number of inhabitants, living in 100 km around an airport

Population200km A number of inhabitants, living in 200 km around an airport

Island 1 if an airport is located on an island; 0 otherwise

GDPpc Gross domestic product per capita in airport's NUTS3 region

RevenueTotal Airport total revenue

RevenueAviation Airport aviation revenue

RevenueNonAviation Airport non-aviation revenue

RevenueHandling Airport revenue from handling services

RevenueParking Airport revenue from parking services

EBITDA Airport earnings before interest, taxes, depreciation, and amortization

NetProfit Airport net profit

DA Airport deprecation, and amortization

StaffCount A number of staff employed by an airport

StaffCost Airport staff cost

RunwayCount A number of airport runways

CheckinCount A number of airport check-iun facilities

GateCount A number of airport gates

TerminalCount A number of airport terminals

ParkingSpaces A number of airport parking spaces

RoutesDeparture A number of departure routes, served by an airport

RoutesArrival A number of arrival routes, served by an airport

Routes (RoutesDeparture + RoutesArrival)/2

Source

Eurostat (2013). European Statistics Database, Statistical Office of the European Communities (Eurostat)

Airports' statistical reports(2011)

Open Flights: Airport, airline and route data http://openflights.org/ (2013-05-31)

TDC (2012). Informe de fiscalizacion de la imputacion por la entidad "Aeropuertos Espanoles y Navegacion Aerea" (AENA) a cada uno de los aeropuertos de los ingresos, gastos, e inversiones correspondientes a la actividad aeroportuaria, en los ejercicios 2009 y 2010., Tribunal de Cuentas, Spain, Doc 938.

CIESIN, Columbia University. Gridded Population of the World: Future Estimates (GPWFE). (2005)

---

airports.greece Greece airports statistical data

Description

The spfrontier package includes the dataset airports, containing information about Greece air-ports infrastructure and traffic statistics in 2011.

Format

A dataframe with 39 observations on the following 24 variables.

name Airport title

ICAO_code Airport ICAO code lat Airport

latitude

lon Airport longitude

APM_winter A number of passengers carried during winter period

APM_summer A number of passengers carried during summer period

APM A number of passengers carried (winter + summer)

cargo_winter A total volume of cargo served by an airport during winter period

cargo_summer A total volume of cargo served by an airport during summer period cargo

A number volume of cargo served by an airport (winter + summer)

ATM_winter A number of air transport movements served by an airport during winter period

ATM_summer A number of air transport movements served by an airport during summer period

ATM A number of air transport movements served by an airport (winter + summer)

openning_hours_winter A total number openning hours during winter period

openning_hours_summer A total number openning hours during summer period

openning_hours A total number openning hours (winter + summer) runway_area A total

area of airport runways

terminal_area A total area of airport terminal(s) parking_area A

total area of airport parking area island 1 if an airpiort is located

on an island; 0 otherwise international 1 if an airpiort is

international; 0 otherwise mixed_use 1 if an airpiort is in mixed

use; 0 otherwise

WLU A total volume of work load units (WLU) served by an airport

NearestCity A road network distance between an airport and its nearest city

Source

"Airport efficiency and public investment in Greece" (2010) In Proceeding of the 2010 Interna-tional Kuhmo-Nectar Conference on Transport Economics, University of Valencia, Spain.

genW Standard spatial contiguity matrixes

Description

genW generates an spatial contiguity matrix (rook or queen)

rowStdrt standartizes spatial contiguity matrix by rows

constructW contructs a spatial contiguity matrix using object longitude and latitude coordinates

Usage

```
genW(n, type = "rook", seed = NULL)

rowStdrt(W)

constructW(coords, labels)
```

Arguments

n	a number of objects with spatial interaction to be arranged.See ’Details’ for objects arranging principle
type	an optional type of spatial interaction. Currently ’rook’ and ”queen’ values are supported, to produce Rook and Queen Contiguity matrix. See references for more info. By default set to rook.
seed	an optional random number generator seed for random matrices
W	a spatial contiguity matrix to be standatised
coords	a matrix of two columns, where every row is a longitude-latitude pair of object coordinates
labels	a vector of object lables to mark rows and columns of the resulting contiguity matrix

Details

To generate spatial interaction between n objects the function arranges them on a chess board. A number of columns is calculated as a square root of n, rounded to the top. The last row contains empty cells, if n is not quadratic

The function divides every element in an argument matrix by the sum of elements in its row. Some spatial estimation requires this standartisation (generally - for faster calculations)

The function contructs a spatial contiguity matrix using object longitude and latitude coordinates. Eucledean distance is currently used.

References

Anselin, L. (1988). Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Examples

```
# Completely filled 10x10 rook contiguity matrix rookW <- genW(100)
rookW

# Partly filled 10x10 rook contiguity matrix rookW <- genW(90)
rookW

# Completely filled 10x10 queen contiguity matrix queenW <- genW(100, type="queen")
queenW
# Completely filled 10x10 queen contiguity matrix queenW <- genW(100, type="queen")
queenW

# Standartisation
stQueenW <- rowStdrt(queenW) stQueenW
data(airports)

W <- constructW(cbind(airports$lon, airports$lat),airports$ICAO_code)
```

---

logLikelihood	Calculation of the log likelihood function for the spatial stochastic frontier model

---

Description

logLikelihood returns a value of the log likelihood function for the spatial stochastic frontier model

Usage

```
logLikelihood(formula, data, W_y = NULL, W_v = NULL, W_u = NULL, inefficiency =
   "half-normal", values, logging = c("quiet", "info", "debug"), costFrontier = F)
```

Arguments

formula	an object of class "formula"
data	data frame, containing the variables in the model
W_y	a spatial weight matrix for spatial lag of the dependent variable
W_v	a spatial weight matrix for spatial lag of the symmetric error term
W_u	a spatial weight matrix for spatial lag of the inefficiency error term
inefficiency	sets the distribution for inefficiency error component. Possible values are 'half-normal' (for half-normal distribution) and 'truncated' (for truncated normal distribution). By default set to 'half-normal'.

Details

This function is exported from the package for testing and presentation purposes A list of arguments of the function exactly matches the corresponding list of the spfrontier function

---

ModelEstimates-class Model Estimation Results

Description

ModelEstimates stores information about MLE estimates of a spatial stochastic frontier model

Method status returns estimation status

Method resultParams returns raw estimated coefficients Method hessian

returns Hessian matrix for estimated coefficients

Method stdErrors returns standard errors of estimated coefficients Method

efficiencies returns efficiency estimates

Method show prints estimated coefficients

Method coefficients returns estimated coefficients Method

fitted returns model fitted values

Method residuals returns residuals

Method summary prints summary of the estimated model

Usage

```
status(object)

resultParams(object)

hessian(object)

stdErrors(object)

efficiencies(object)

## S4 method for signature 'ModelEstimates' show(object)

## S4 method for signature 'ModelEstimates'
```

coefficients(object)

S4 method for signature 'ModelEstimates'
resultParams(object)

S4 method for signature 'ModelEstimates' fitted(object)

S4 method for signature 'ModelEstimates'
efficiencies(object)

S4 method for signature 'ModelEstimates'
residuals(object)

S4 method for signature 'ModelEstimates'
stdErrors(object)

S4 method for signature 'ModelEstimates' hessian(object)

S4 method for signature 'ModelEstimates' status(object)

S4 method for signature 'ModelEstimates'
summary(object)

Arguments

object an object of ModelEstimates class

Details

ModelEstimates stores all parameter estimates and additional statistics, available after estimation of a spatial stochastic frontier model.

Slots

coefficients estimated values of model parameters

resultParams raw estimated values

status model estimation status: 0 -
Success
1 - Failed; convergence is not achieved
1000 - Failed; unexpected exception
1001 - Failed; Initial values for MLE cannot be estimated
1002 - Failed; Maximum likelihood function is infinite

logL value of the log-likelihood function

logLcalls information abour a number of log-likelihood function and its gradient function calls

hessian Hessian matrix for estimated coefficients

stdErrors standard errors of estimated coefficients

residuals model residuals

fitted model fitted values

efficiencies estimates of efficiency values for sample observations

---

spfrontier Spatial stochastic frontier model

Description

spfrontier estimates spatial specifications of the stochastic frontier model.

Usage

```
spfrontier(formula, data, W_y = NULL, W_v = NULL, W_u = NULL, inefficiency =
  "half-normal", initialValues = "errorsarlm",
  logging = c("quiet", "info", "debug"), control = NULL, onlyCoef = F, costFrontier = F)
```

Arguments

formula	an object of class "formula": a symbolic description of the model to be fitted. The details of model specification are given under 'Details'.
data	data frame, containing the variables in the model
W_y	a spatial weight matrix for spatial lag of the dependent variable
W_v	a spatial weight matrix for spatial lag of the symmetric error term
W_u	a spatial weight matrix for spatial lag of the inefficiency error term
inefficiency	sets the distribution for inefficiency error component. Possible values are 'half-normal' (for half-normal distribution) and 'truncated' (for truncated normal distribution). By default set to 'half-normal'. See references for explanations
initialValues	an optional vector of initial values, used by maximum likelihood estimator. If not defined, estimator-specific method of initial values estimation is used.
logging	an optional level of logging. Possible values are 'quiet','warn','info','debug'. By default set to quiet.
control	an optional list of control parameters, passed to optim estimator from the 'stats package
onlyCoef	allows calculating only estimates for coefficients (with inefficiencies and other additional statistics). Developed generally for testing, to speed up the process.
costFrontier	is designed for selection of cost or production frontier

Details

Models for estimation are specified symbolically, but without any spatial components. Spatial components are included implicitly on the base of the model argument.

References

Kumbhakar, S.C. and Lovell, C.A.K (2000), Stochastic Frontier Analysis, Cambridge University Press, U.K.

Examples

```
data( airports )
airports2011 <- subset(airports, Year==2011)
W <- constructW(cbind(airports2011$longitude, airports2011$latitude),airports2011$ICAO) formula <- log(PAX)
~ log(Population100km) + log(Routes) + log(GDPpc)
ols <- lm(formula , data=airports2011) summary(ols )
plot(density(stats::residuals(ols)))
skewness(stats::residuals(ols))

# Takes >5 sec, see demo for more examples
# model <- spfrontier(formula , data=airports2011)
# summary(model )

# model <- spfrontier(formula , data=airports2011, W_y=W)
# summary(model )
```

---

spfrontier.true.value True value for simulation

Description

spfrontier.true.value returns true parameter values for a simulation process

ezsimspfrontier tests estimators of a spatial stochastic frontier model with different parameters

Usage

```
spfrontier.true.value()

ezsimspfrontier(runs, params, inefficiency = "half-normal", logging = "info",
   control = list())
```

Arguments

runs	a number of simulated samples
params	a set with parameters to be used in simulation.

inefficiency sets the distribution for inefficiency error component. Possible values are 'half-normal' (for half-normal distribution) and 'truncated' (for truncated normal distribution). By default set to 'half-normal'. See references for explanations

logging an optional level of logging. Possible values are 'quiet','warn','info','debug'. By default set to quiet.

control an optional list of control parameters for simulation process. Currently the procedure supports: ignoreWy (TRUE/FALSE) - the spatial contiguity matrix for a dependent variable is not provided to spfrontier estimator (but used in DGP) ignoreWv (TRUE/FALSE) - the spatial contiguity matrix for a symmetric error term is not provided to spfrontier estimator (but used in DGP) ignoreWu (TRUE/FALSE) - the spatial contiguity matrix for a inefficiency error term is not provided to spfrontier estimator (but used in DGP) parallel (TRUE/FALSE) - whether to use parallel computer seed - a state for random number generation in R. If NULL (default), the initial state is random. See set.seed for details. auto_save - saves intermediate results to files. See ezsim for details.

Details

The spfrontier.true.value function should notbe used directly, it is exported for supporting ezsim

The ezsimspfrontier function executes multiple calls of the spfrontier estimator on a simulated data set, generated on the base of provided parameters. The resulting estimates can be analysed for biasedness, efficiency, etc.

See Also

ezsim

Examples

```
params000 <- list(n=c(50, 100),beta0=5, beta1=10,
                  beta2=1,
                  sigmaV=0.5,
                  sigmaU=2.5) ctrl <-
list(seed=999, cores=1)
res000 <- ezsimspfrontier(2, params = params000, inefficiency =
                  "half-normal", logging = "info", control=ctrl)

summary(res000)
```

Appendix 4. R source codes for simulation study of the *spfrontier* package

```
# Define DGP parameters
params000 <- list(n=c(50,100,200,300),
                  beta0=5,
                  beta1=2,
                  beta2=3,
                  sigmaV=0.1,
                  sigmaU=0.5)

# Define alternative DGPs

params000T <- c(params000, list(mu=1))
params100 <- c(params000, list(rhoY=0.2))
params100T <- c(params000T, list(rhoY=0.2))

params110 <- c(params100, list(rhoV=0.4))
params101 <- c(params100, list(rhoU=0.4))
params111 <- c(params110, list(rhoU=0.4))

params010 <- params110
params010$rhoY <- NULL

params011 <- params111
params011$rhoY <- NULL

params001 <- params011
params001$rhoV <- NULL

# Set up control parameters

ctrl <- list(true.initial=F, seed=999, cores=detectCores())

# Run simulation study for classic SF with half-normal inefficiency

res000 <- ezsimspfrontier(100, params = params000,  inefficiency = "half-
normal",logging = "info", control=ctrl)
save(res000, file="res000.rData")

# Run simulation study for classic SF with truncated normal inefficiency

res000T <- ezsimspfrontier(100, params = params000T, inefficiency =
"truncated",logging = "info", control=ctrl)
save(res000T, file="res000T.rData")

# Run simulation study for SSF(1,0,0,0) with half-normal inefficiency

res100 <- ezsimspfrontier(100, params = params100,  inefficiency = "half-
normal",logging = "info", control=ctrl)
save(res100, file="res100.rData")

# Run simulation study for SSF(1,0,0,0), ignoring spatial lags (biasedness is
expected)

res100A <- ezsimspfrontier(100, params = params100,  inefficiency = "half-
normal",logging = "info", control=c(ctrl,list(ignoreWy=T)))
save(res100A, file="res100A.rData")

# Run simulation study for SSF(1,0,0,0) with truncated normal inefficiency

res100T <- ezsimspfrontier(100, params = params100T, inefficiency =
"truncated",logging = "info", control=ctrl)
```

```
save(res100T, file="res100T.rData")

# Run simulation study for SSF(1,0,0,0) without spatial lags in DGP

params001A <- c(params001, list(rhoY=0))
res001A <- ezsimspfrontier(100, params = params001A, inefficiency = "half-
normal",logging = "info", control=c(ctrl,list(ignoreWu=T,replaceWyWu=T)))
save(res001A, file="res001A.rData")


# Run simulation study for SSF(1,0,0,0), replacing spatial lags spatial
errors in DGP

params010A <- c(params010, list(rhoY=0))
res010A <- ezsimspfrontier(1, params = params010A, inefficiency = "half-
normal",logging = "info", control=c(ctrl,list(ignoreWv=T,replaceWyWv=T)))
save(res010A, file="res010A.rData")


# Set up multithreaded settings

ctrl <- list(true.initial=TRUE, seed=0, cores=detectCores()-1)

# Run simulation study for SSF(0,0,0,1)
# Takes ~20 hrs on 8 cores
params001$n <- c(50,100,200,300)
res001<- ezsimspfrontier(100, params = params001, inefficiency = "half-
normal",logging = "info",control=ctrl)
save(res001, file="res001.rData")

# Run simulation study for SSF(0,0,1,0)
res010 <- ezsimspfrontier(100, params = params010, inefficiency = "half-
normal",logging = "info",control=ctrl)

save(res010, file="res010.rData")


# Run simulation study for SSF(1,0,1,0)
res110 <- ezsimspfrontier(100, params = params110, inefficiency = "half-
normal",logging = "info",control=ctrl)

save(res110, file="SimE07_res110.rData")


# Run simulation study for SSF(1,0,0,1)
res101 <- ezsimspfrontier(100, params = params101, inefficiency = "half-
normal",logging = "info",control=ctrl)
save(res101, file="SimE08_res101.rData")


# Run simulation study for SSF(0,0,1,1)
res011 <- ezsimspfrontier(100, params = params011, inefficiency = "half-
normal",logging = "info",control=ctrl)
save(res011, file="SimE09_res011.rData")

# Run simulation study for SSF(1,0,1,1)

res111 <- ezsimspfrontier(100, params = params111, inefficiency = "half-
normal",logging = "info",control=ctrl)
save(res111, file="SimE10_res111.rData")
```

Appendix 5. Computing environment used for simulation experiments

Computing cluster

Amazon EC2 Instance: c3.2xlarge

Amazon EC2 Instance description: A compute-optimized instance based on 8 Intel Xeon E5-2680 v2 (Ivy Bridge) processors

Total number of cores: 8

Number of cores in a cluster: 8

EC2 compute units: 28

Available memory: 15 GB

Software environment

Amazon Machine Image: Bioconductor AMI

Bioconductor version: 2.14

R version: 3.1.0

Main R packages

Package 'spfrontier': 0.1.8 (2014-06-26)

Package 'ezsim': 0.5.5 (2014-06-26)

Additional R packages

Package 'tmvtnorm': 1.4-9 (2014-03-04)

Package 'sandwich': 2.3-0 (2013-10-05)

Package 'moments': 0.13 (2012-01-24)

Appendix 6. Results of simulation studies.

***Simulation Experiment*: SimE1**

DGP: $\mu^* = 0, \rho_Y^* = 0, \rho_v^* = 0, \rho_u^* = 0$

Estimator: SSF(0,0,0,0), half-normal inefficiency

Sample size: 50, 100, 200, 300

Simulation runs: 100

Execution time: 13.9 mins

Main conclusions:

- unbiased estimates for frontier and inefficiency parameters;
- consistent estimates both for frontier and inefficiency parameters

Table A6.1.1. SimE1 simulation results

n	*Parameter*	*True Values*	*Mean*	*Bias*	*Bias, %*	*SD*	*RMSD*
50	β_0	5	4.9281	-0.0719	0.818	0.8211	-0.0144
	β_1	2	2.0187	0.0187	0.3415	0.342	0.0094
	β_2	3	2.9487	-0.0513	0.372	0.3756	-0.0171
	σ_v	0.5	0.3499	-0.1501	0.3546	0.385	-0.3001
	σ_u	2.5	2.3219	-0.1781	0.5844	0.6109	-0.0712
100	β_0	5	4.956	-0.044	0.5983	0.5999	-0.0088
	β_1	2	2.022	0.022	0.1974	0.1986	0.011
	β_2	3	3.0198	0.0198	0.2409	0.2417	0.0066
	σ_v	0.5	0.3705	-0.1295	0.2672	0.2969	-0.259
	σ_u	2.5	2.4805	-0.0195	0.2979	0.2985	-0.0078
200	β_0	5	4.9824	-0.0176	0.3465	0.3469	-0.0035
	β_1	2	1.9934	-0.0066	0.1568	0.1569	-0.0033
	β_2	3	3.0105	0.0105	0.1499	0.1503	0.0035
	σ_v	0.5	0.4722	-0.0278	0.1291	0.1321	-0.0556
	σ_u	2.5	2.4827	-0.0173	0.1907	0.1915	-0.0069
300	β_0	5	4.9819	-0.0181	0.3065	0.307	-0.0036
	β_1	2	2.0098	0.0098	0.1334	0.1337	0.0049
	β_2	3	2.9972	-0.0028	0.1205	0.1205	-0.0009
	σ_v	0.5	0.4894	-0.0106	0.1005	0.101	-0.0212
	σ_u	2.5	2.4846	-0.0154	0.1517	0.1525	-0.0062

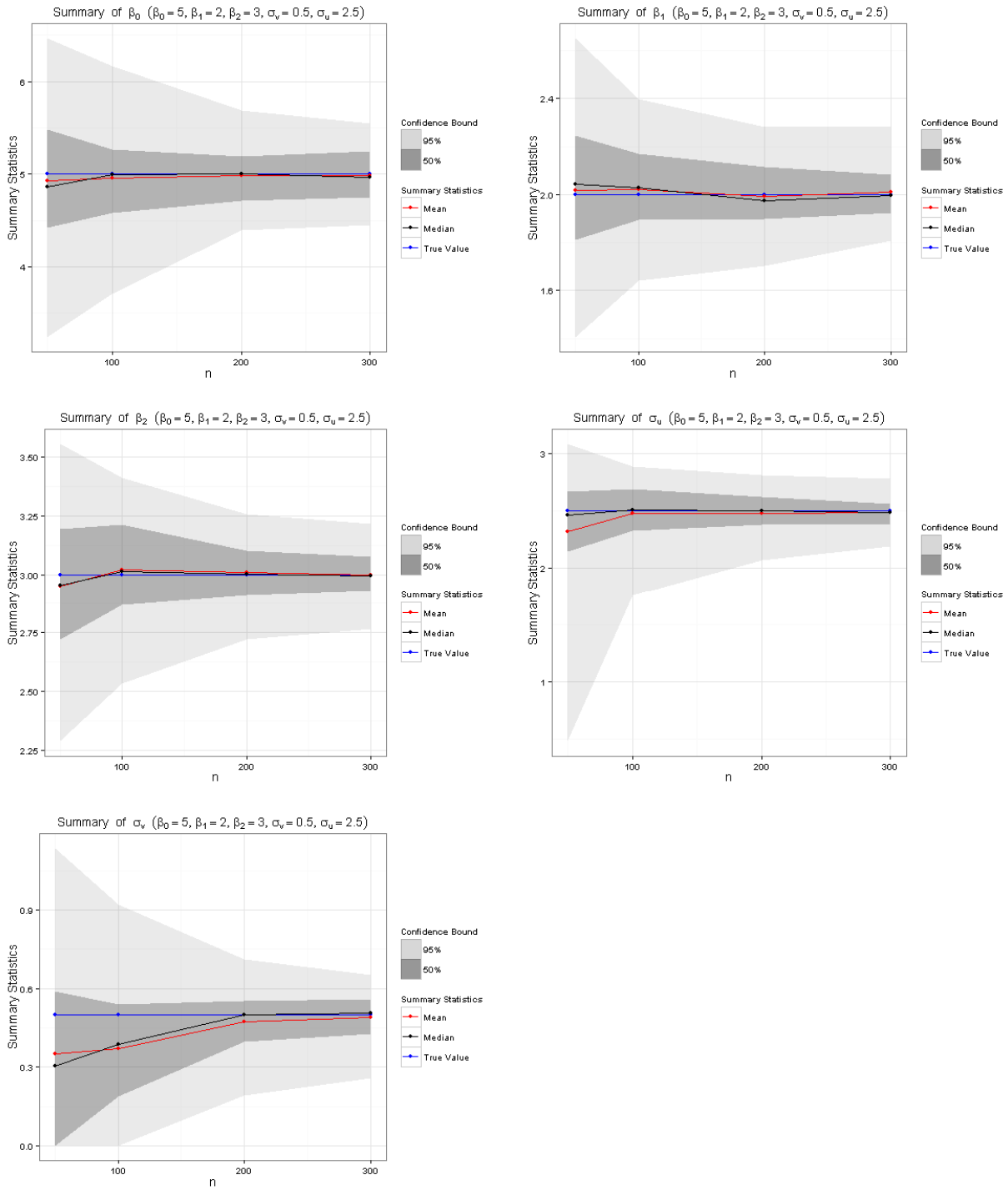


Fig. A6.1.1. Summary of SimE1 estimates

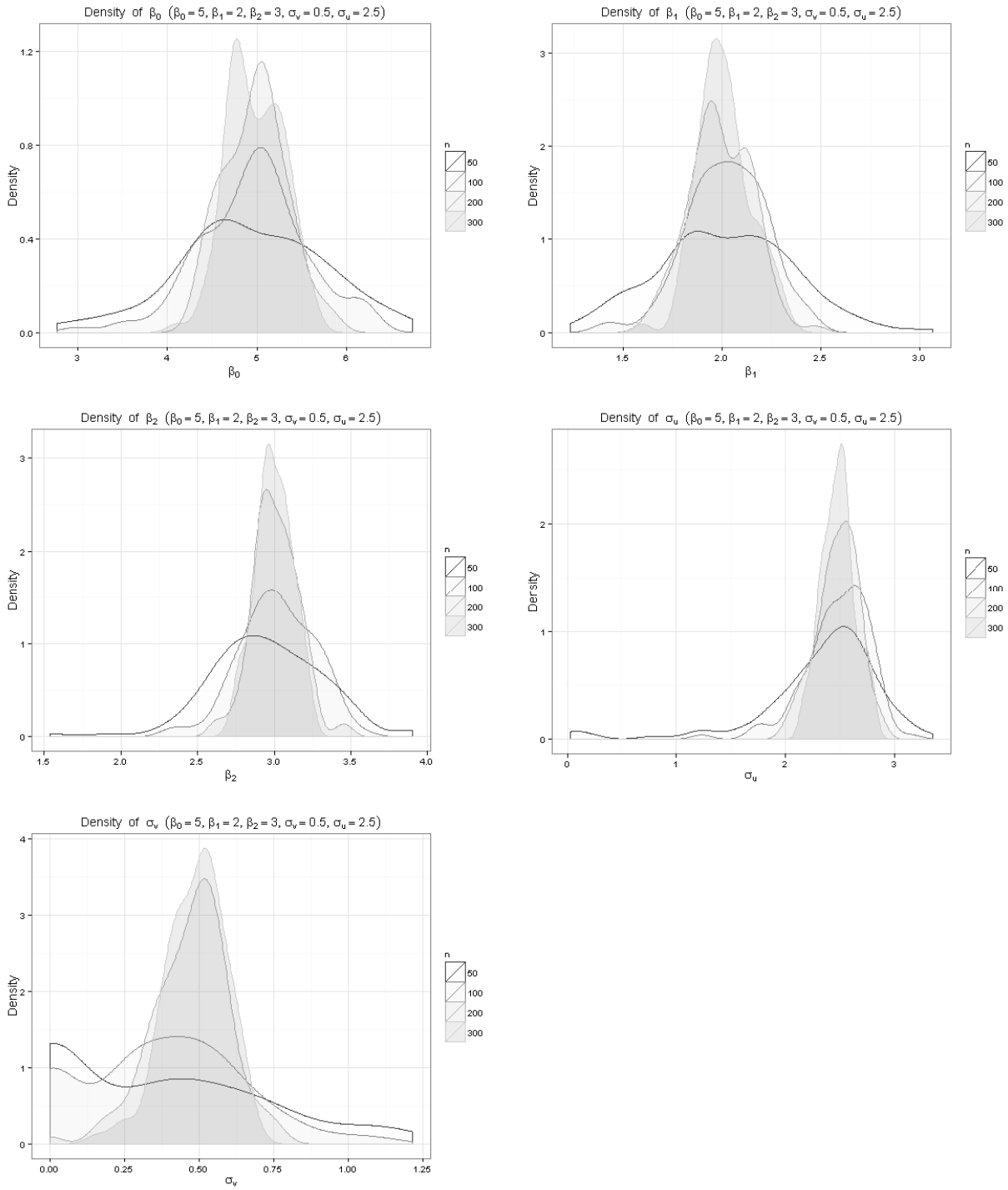


Fig. A6.1.2. Empirical kernel densities of SimE1 estimates

***Simulation Experiment*: SimE2**

DGP: $\mu^* = 1, \rho_Y^* = 0, \rho_v^* = 0, \rho_u^* = 0$

Estimator: SSF(0,0,0,0), truncated normal inefficiency

Sample size: 50, 100, 200, 300

Simulation Runs: 100

Execution time: 17.9 mins

Main conclusions:

- unbiased estimates for frontier and inefficiency parameters;
- consistent estimates both for frontier and inefficiency parameters;
- weak identification of σ_u and μ, especially for small samples.

Table A6.2.1. SimE2 simulation results

n	*Parameter*	*True Values*	*Mean*	*Bias*	*Bias, %*	*SD*	*RMSD*
50	β_0	5	4.7301	-0.2699	1.01	1.0454	-0.054
	β_1	2	2.0369	0.0369	0.4025	0.4042	0.0184
	β_2	3	2.931	-0.069	0.385	0.3911	-0.023
	σ_v	0.5	0.3962	-0.1038	0.4326	0.4449	-0.2077
	σ_u	2.5	4.2846	1.7846	5.2064	5.5038	0.7138
	μ	1	-28.4411	-29.4411	106.0615	110.0718	-29.4411
100	β_0	5	4.8551	-0.1449	0.729	0.7432	-0.029
	β_1	2	2.0362	0.0362	0.2444	0.2471	0.0181
	β_2	3	3.0317	0.0317	0.2759	0.2777	0.0106
	σ_v	0.5	0.3526	-0.1474	0.3124	0.3454	-0.2948
	σ_u	2.5	3.0215	0.5215	1.981	2.0485	0.2086
	μ	1	-3.3969	-4.3969	19.917	20.3966	-4.3969
200	β_0	5	4.9921	-0.0079	0.5017	0.5017	-0.0016
	β_1	2	1.9963	-0.0037	0.1848	0.1849	-0.0018
	β_2	3	3.0088	0.0088	0.1812	0.1815	0.0029
	σ_v	0.5	0.4286	-0.0714	0.2263	0.2372	-0.1427
	σ_u	2.5	2.5864	0.0864	0.4432	0.4515	0.0346
	μ	1	0.5848	-0.4152	1.6166	1.6691	-0.4152
300	β_0	5	5.0325	0.0325	0.4115	0.4128	0.0065
	β_1	2	2.0152	0.0152	0.1568	0.1576	0.0076
	β_2	3	2.9927	-0.0073	0.1362	0.1364	-0.0024
	σ_v	0.5	0.4449	-0.0551	0.174	0.1825	-0.1102
	σ_u	2.5	2.5272	0.0272	0.6441	0.6447	0.0109
	μ	1	0.6424	-0.3576	4.0231	4.0389	-0.3576

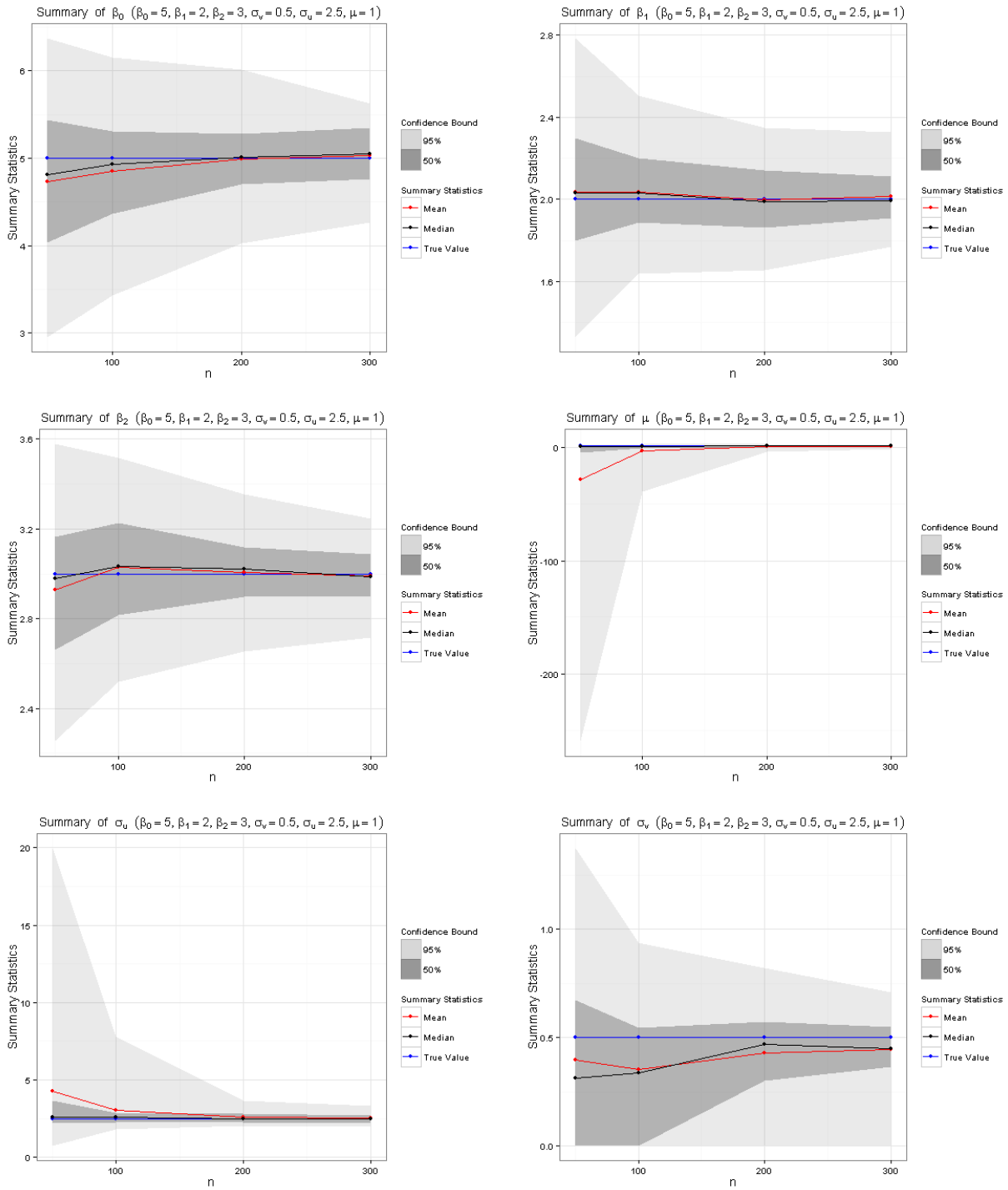


Fig. A6.2.1. Summary of SimE2 estimates

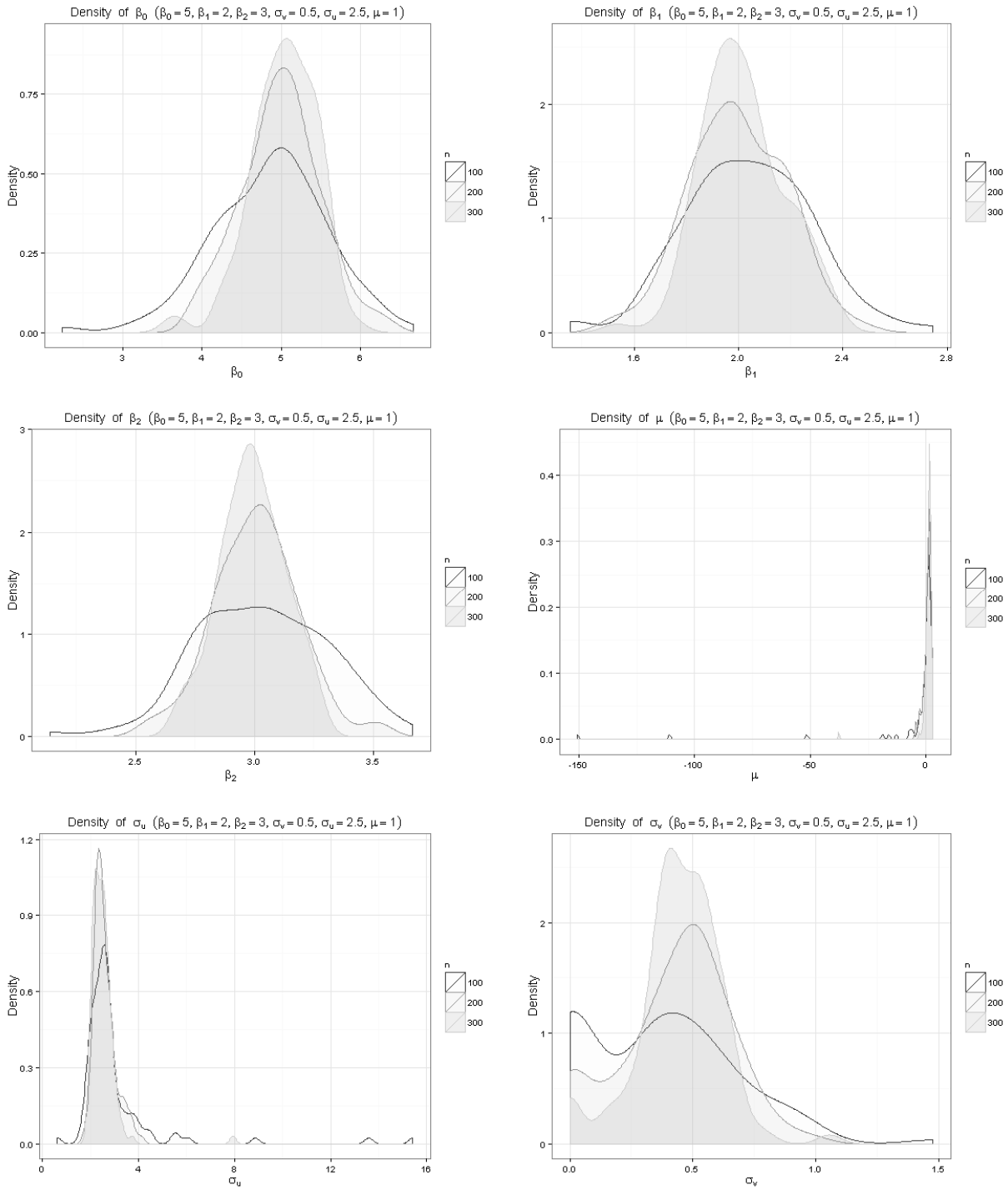


Fig. A6.2.2. Empirical kernel densities of SimE2 estimates

***Simulation Experiment*: SimE3**

DGP: $\mu^* = 0, \rho_Y^* = 0.2, \rho_v^* = 0, \rho_u^* = 0$

Estimator: SSF(1,0,0,0), half-normal inefficiency

Sample size: 50, 100, 200, 300

Simulation runs: 100

Execution time: 50.0 mins

Main conclusions:

- unbiased estimates for frontier and inefficiency parameters;
- consistent estimates both for frontier and inefficiency parameters;
- unbiased and consistent estimates for endogenous spatial effects parameter ρ_Y.

Table A6.3.1. SimE3 simulation results

n	*Parameter*	*True Values*	*Mean*	*Bias*	*Bias, %*	*SD*	*RMSD*
50	β_0	5	4.6	-0.4	2.9182	2.9455	-0.08
	β_1	2	2.0071	0.0071	0.3386	0.3387	0.0035
	β_2	3	2.9513	-0.0487	0.3927	0.3957	-0.0162
	ρ_Y	0.2	0.2243	0.0243	0.2241	0.2254	0.1216
	σ_v	0.5	0.285	-0.215	0.3791	0.4358	-0.43
	σ_u	2.5	2.2895	-0.2105	0.6252	0.6597	-0.0842
100	β_0	5	4.5107	-0.4893	2.2939	2.3455	-0.0979
	β_1	2	2.0163	0.0163	0.2261	0.2267	0.0082
	β_2	3	3.0059	0.0059	0.2637	0.2638	0.002
	ρ_Y	0.2	0.2358	0.0358	0.1643	0.1682	0.1791
	σ_v	0.5	0.3165	-0.1835	0.2847	0.3387	-0.367
	σ_u	2.5	2.4984	-0.0016	0.3211	0.3212	-0.0006
200	β_0	5	4.6669	-0.3331	1.2795	1.3221	-0.0666
	β_1	2	1.9919	-0.0081	0.1705	0.1706	-0.004
	β_2	3	3.0062	0.0062	0.1563	0.1564	0.0021
	ρ_Y	0.2	0.2252	0.0252	0.0942	0.0975	0.1259
	σ_v	0.5	0.446	-0.054	0.1711	0.1794	-0.108
	σ_u	2.5	2.4956	-0.0044	0.2267	0.2268	-0.0018
300	β_0	5	4.6545	-0.3455	0.9837	1.0426	-0.0691
	β_1	2	2.0117	0.0117	0.1381	0.1386	0.0058
	β_2	3	2.9941	-0.0059	0.1219	0.1221	-0.002
	ρ_Y	0.2	0.2245	0.0245	0.0682	0.0724	0.1224
	σ_v	0.5	0.4804	-0.0196	0.1214	0.1229	-0.0392
	σ_u	2.5	2.4884	-0.0116	0.1732	0.1736	-0.0046

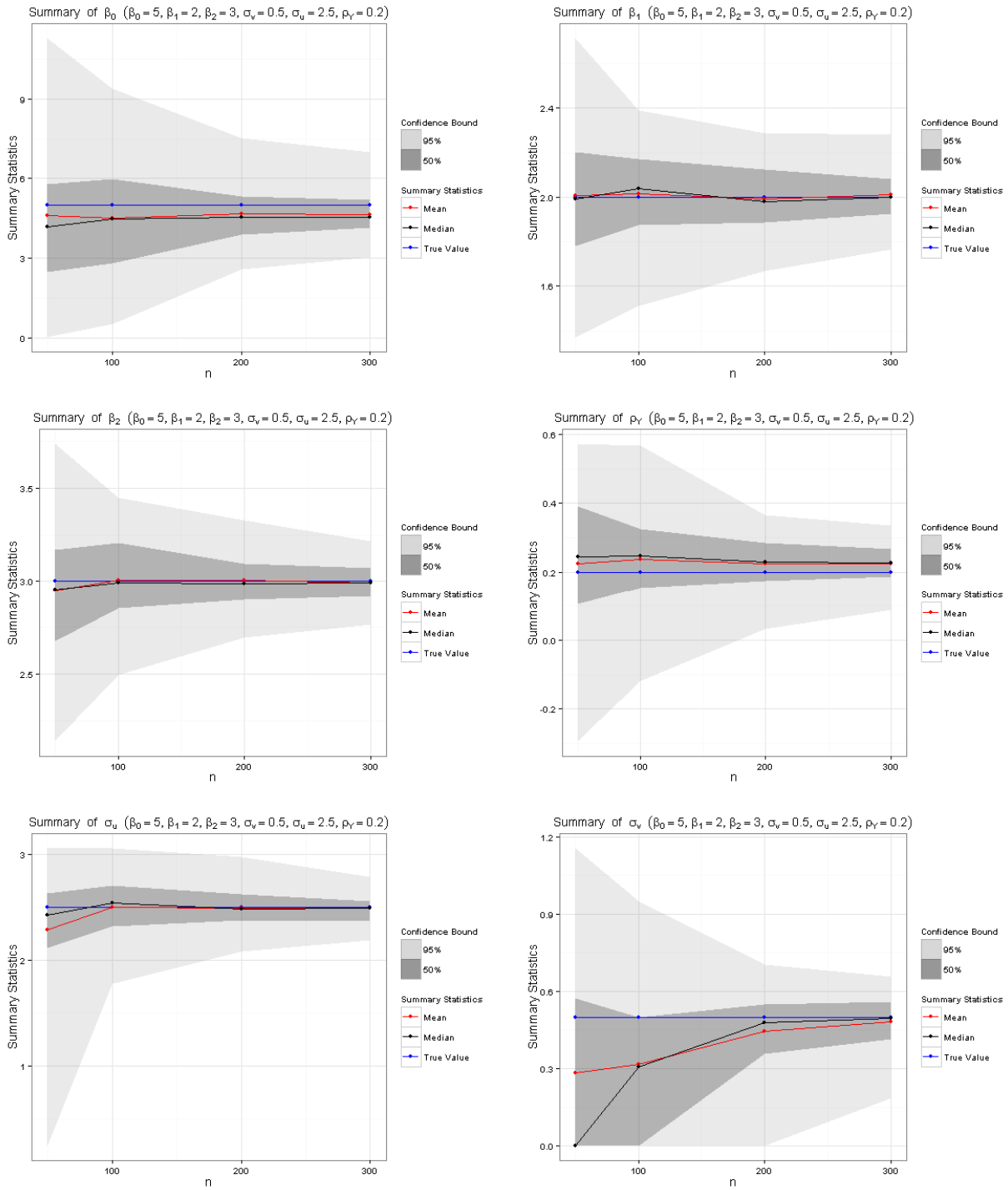


Fig. A6.3.1. Summary of SimE3 estimates

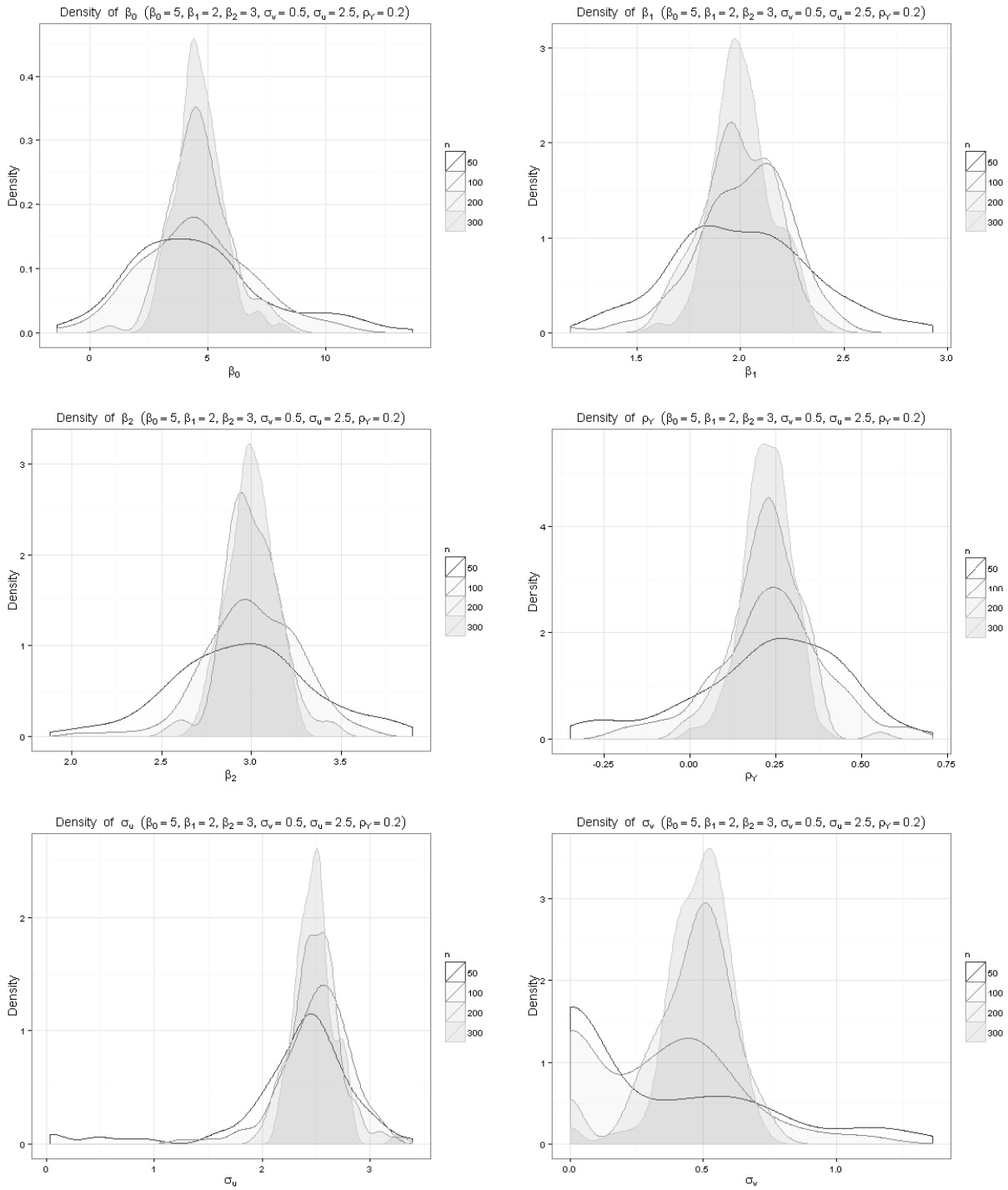


Fig. A6.3.2. Empirical kernel densities of SimE3 estimates

***Simulation Experiment*: SimE3b**

DGP: $\mu^* = 0, \rho_Y^* = 0.2, \rho_v^* = 0, \rho_u^* = 0$

Estimator: SSF(0,0,0,0), half-normal inefficiency

Sample size: 50, 100, 200, 300

Simulation Runs: 100

Execution time: 12.0 mins

Main conclusions:

- biased and inconsistent estimates for frontier intercept and random disturbances' standard deviations (as expected due to missed endogenous spatial effects in the estimator).

Table A6.3b.1 SimE3b simulation results

n	*Parameter*	*True Values*	*Mean*	*Bias*	*Bias, %*	*SD*	*RMSD*
50	β_0	5	7.5966	2.5966	0.9426	2.7624	0.5193
	β_1	2	2.017	0.017	0.3688	0.3692	0.0085
	β_2	3	2.997	-0.003	0.3877	0.3877	-0.001
	σ_v	0.5	0.3608	-0.1392	0.3792	0.404	-0.2784
	σ_u	2.5	2.3705	-0.1295	0.6467	0.6596	-0.0518
100	β_0	5	7.6533	2.6533	0.6395	2.7293	0.5307
	β_1	2	2.0473	0.0473	0.2166	0.2217	0.0236
	β_2	3	3.0313	0.0313	0.2489	0.2508	0.0104
	σ_v	0.5	0.397	-0.103	0.3011	0.3183	-0.206
	σ_u	2.5	2.5211	0.0211	0.349	0.3497	0.0084
200	β_0	5	7.6459	2.6459	0.3517	2.6692	0.5292
	β_1	2	1.9974	-0.0026	0.1575	0.1575	-0.0013
	β_2	3	3.0349	0.0349	0.1541	0.158	0.0116
	σ_v	0.5	0.5133	0.0133	0.1111	0.1119	0.0266
	σ_u	2.5	2.5006	0.0006	0.1831	0.1831	0.0002
300	β_0	5	7.6623	2.6623	0.3452	2.6846	0.5325
	β_1	2	2.0175	0.0175	0.1363	0.1374	0.0087
	β_2	3	3.0093	0.0093	0.1288	0.1292	0.0031
	σ_v	0.5	0.5388	0.0388	0.0991	0.1064	0.0776
	σ_u	2.5	2.4967	-0.0033	0.1571	0.1572	-0.0013

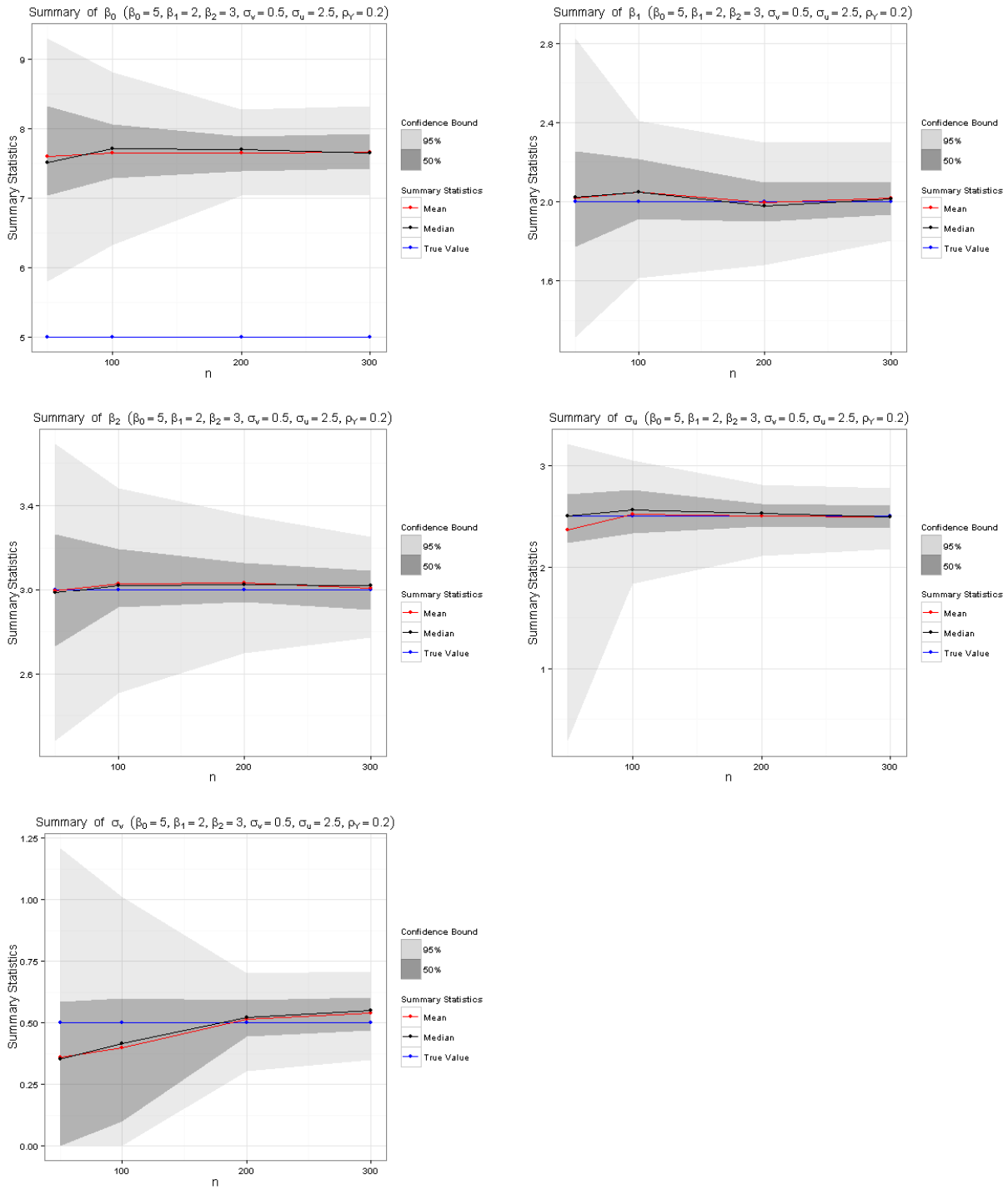


Fig. A6.3b.1. Summary of SimE3b estimates

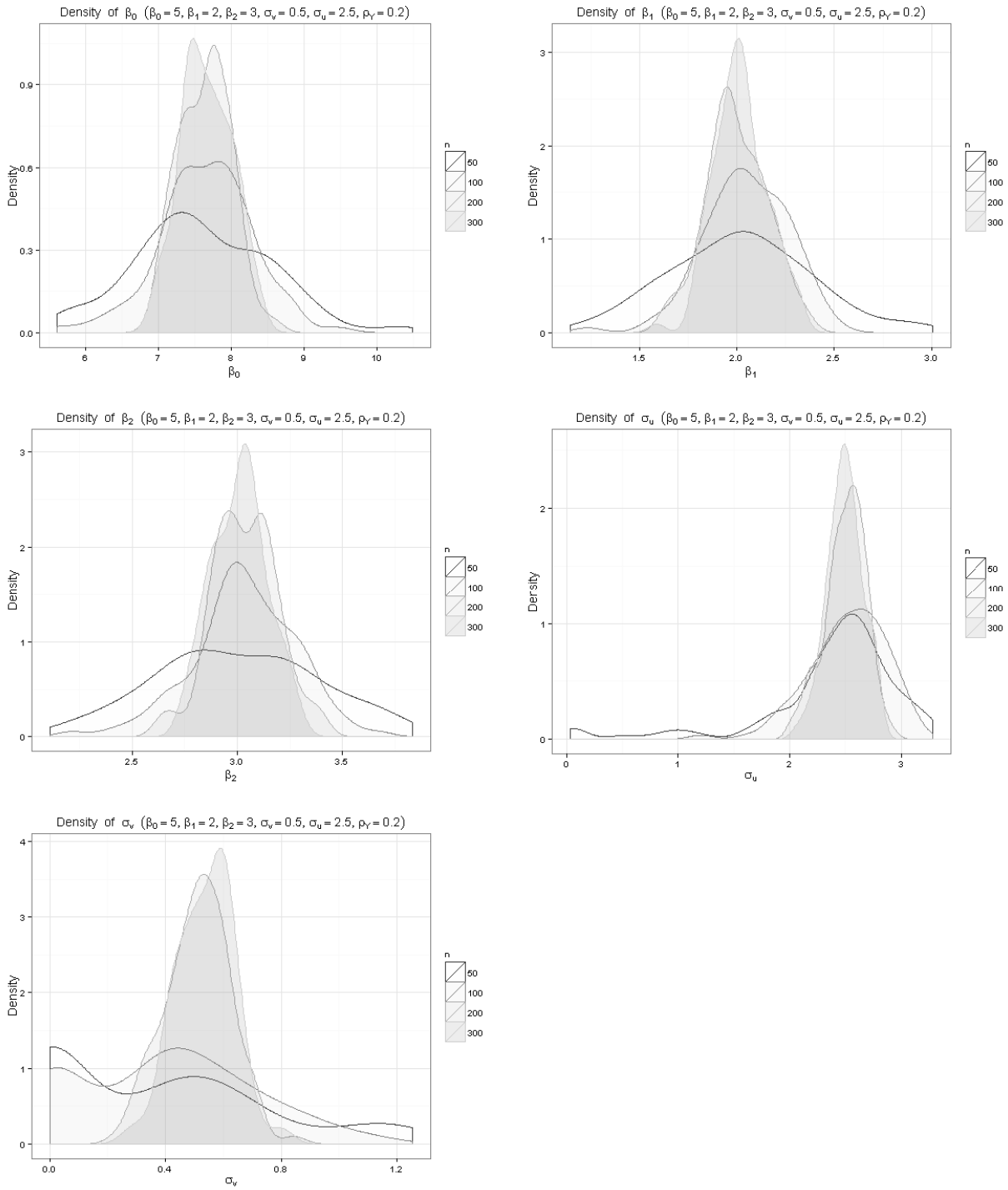


Fig. A6.3b.2. Empirical kernel densities of SimE3b estimates

***Simulation Experiment*: SimE4**

DGP: $\mu^* = 1, \rho_Y^* = 0.2, \rho_v^* = 0, \rho_u^* = 0$

Estimator: SSF(1,0,0,0), truncated normal inefficiency

Sample size: 50, 100, 200, 300

Simulation Runs: 100

Execution time: 32.6 mins

Main conclusions:

- unbiased estimates for frontier and inefficiency parameters;
- consistent estimates both for frontier and inefficiency parameters;
- unbiased and consistent estimates for endogenous spatial effects parameter ρ_Y;
- weak identification of σ_u and μ.

Table A6.4.1. SimE4 simulation results

n	*Parameter*	*True Values*	*Mean*	*Bias*	*Bias, %*	*SD*	*RMSD*
50	β_0	5	4.4505	-0.5495	1.2547	1.3697	-0.1099
	β_1	2	2.2493	0.2493	0.433	0.4996	0.1246
	β_2	3	3.33	0.33	0.4181	0.5326	0.11
	ρ_Y	0.2	0.1631	-0.0369	0.1529	0.1573	-0.1844
	σ_v	0.5	0.1082	-0.3918	0.2369	0.4579	-0.7836
	σ_u	2.5	2.7005	0.2005	0.1762	0.2669	0.0802
	μ	1	1.9198	0.9198	0.6823	1.1452	0.9198
100	β_0	5	5.8619	0.8619	2.6641	2.8	0.1724
	β_1	2	2.1594	0.1594	0.2764	0.3191	0.0797
	β_2	3	2.9437	-0.0563	0.4719	0.4753	-0.0188
	ρ_Y	0.2	0.1375	-0.0625	0.1847	0.1949	-0.3124
	σ_v	0.5	0.2684	-0.2316	0.3991	0.4614	-0.4633
	σ_u	2.5	3.0375	0.5375	0.3962	0.6677	0.215
	μ	1	2.4486	1.4486	0.4071	1.5047	1.4486
200	β_0	5	4.0887	-0.9113	1.2955	1.5839	-0.1823
	β_1	2	2.1493	0.1493	0.1421	0.2061	0.0746
	β_2	3	3.1082	0.1082	0.1645	0.1969	0.0361
	ρ_Y	0.2	0.2621	0.0621	0.0935	0.1122	0.3104
	σ_v	0.5	0.5426	0.0426	0.1268	0.1337	0.0851
	σ_u	2.5	2.8508	0.3508	0.1679	0.3889	0.1403
	μ	1	2.1796	1.1796	0.1407	1.1879	1.1796
300	β_0	5	5.2658	0.2658	1.2423	1.2704	0.0532
	β_1	2	2.1315	0.1315	0.1819	0.2245	0.0658
	β_2	3	3.0684	0.0684	0.0837	0.1081	0.0228
	ρ_Y	0.2	0.1748	-0.0252	0.0858	0.0894	-0.1261
	σ_v	0.5	0.5346	0.0346	0.0965	0.1026	0.0692
	σ_u	2.5	2.8097	0.3097	0.1725	0.3545	0.1239
	μ	1	1.9757	0.9757	0.742	1.2258	0.9757

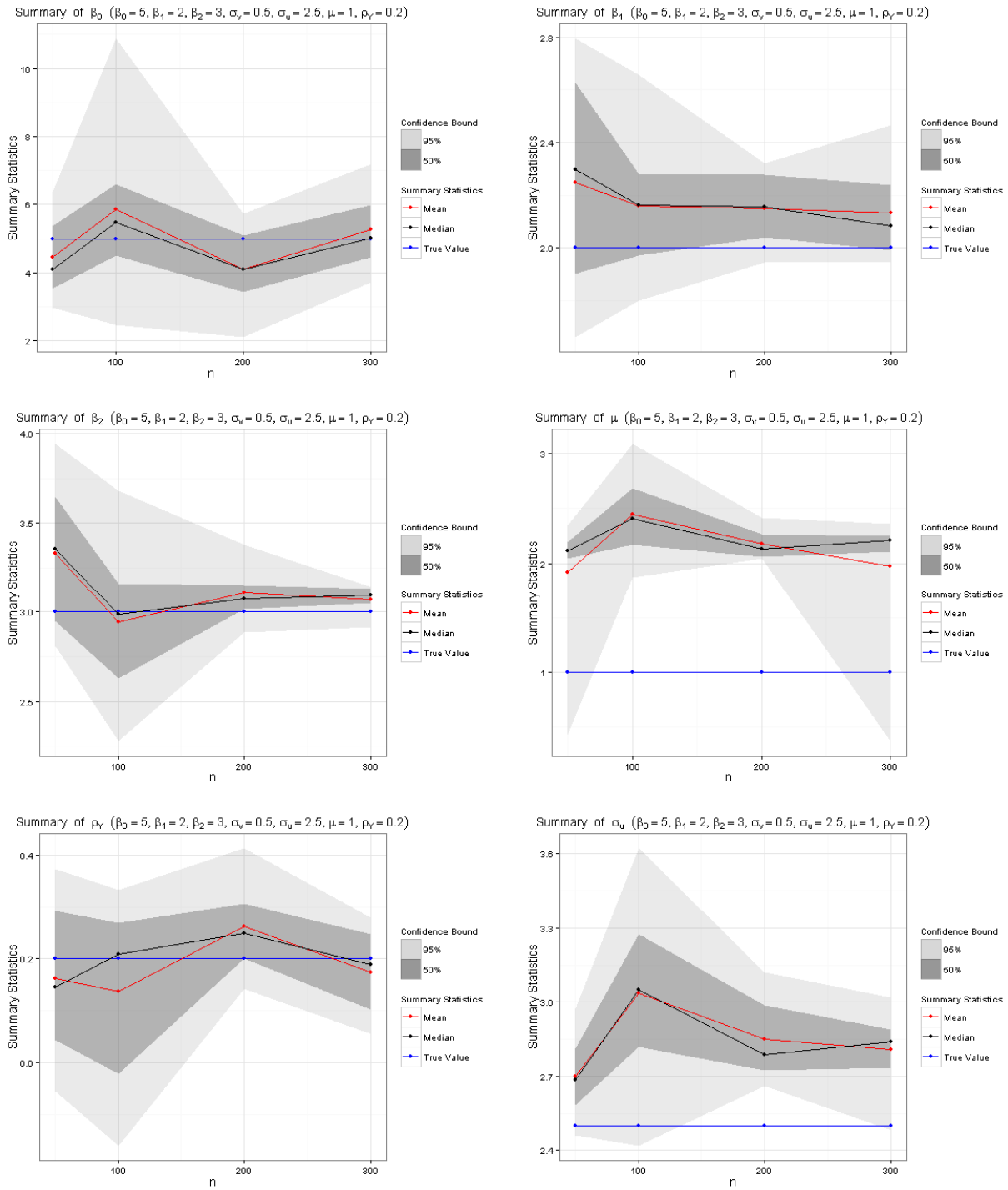


Fig. A6.4.1. Summary of SimE4 estimates

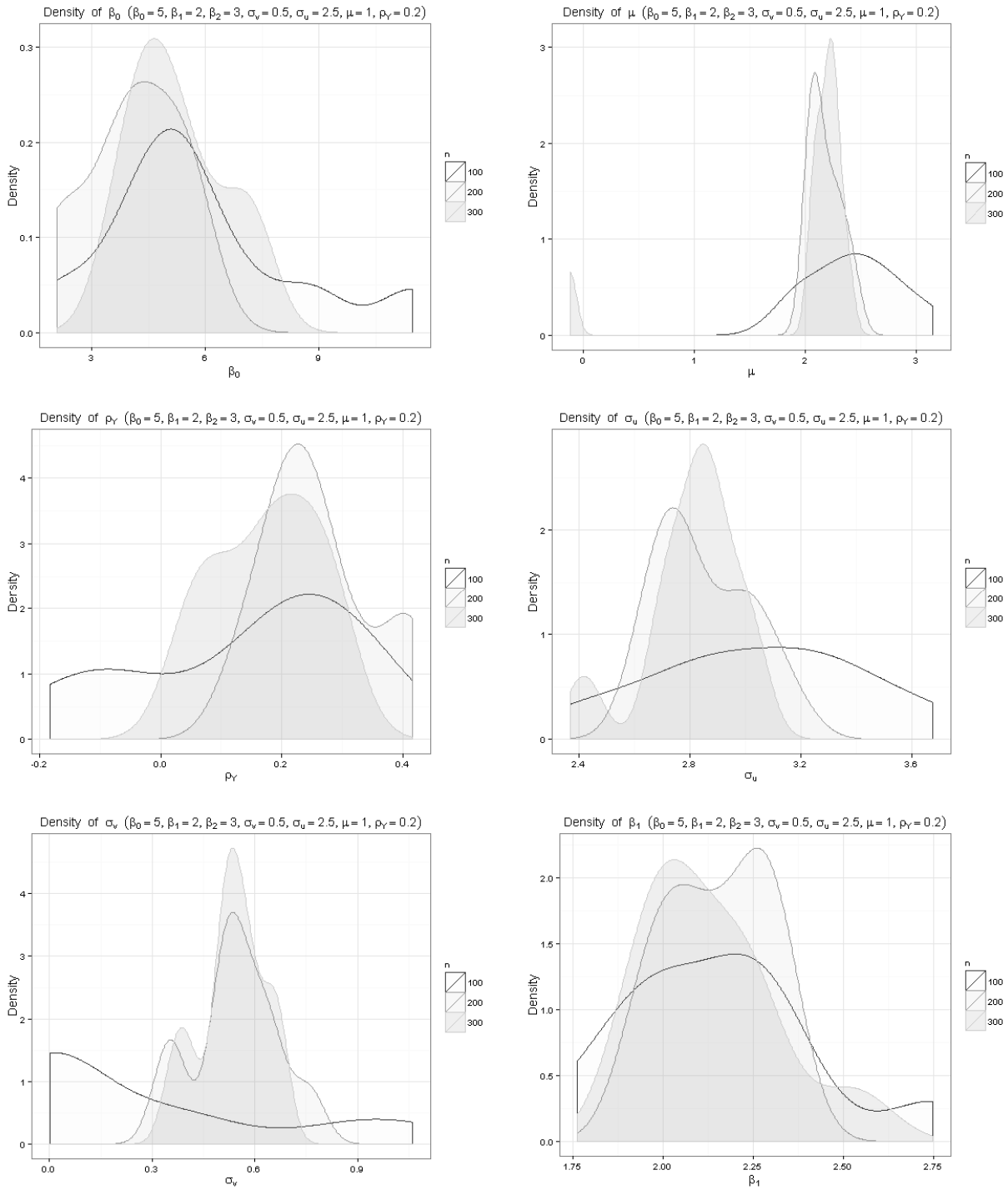


Fig. A6.4.2. Empirical kernel densities of SimE4 estimates

***Simulation Experiment*: SimE5**

DGP: $\mu^* = 0, \rho_Y^* = 0, \rho_v^* = 0.4, \rho_u^* = 0$

Estimator: SSF(0,0,1,0), half-normal inefficiency

Sample size: 50, 100, 200, 300

Simulation Runs: 100

Execution time: ~20 hrs

Main conclusions:

- unbiased and consistent estimates for frontier parameters, except σ_v and σ_u;
- consistent estimates for the spatially correlated random disturbances parameter ρ_v;
- large sample variance of the spatially correlated random disturbances parameter ρ_v and inefficiency standard deviation σ_u for small samples. So this is not recommended to apply MLE estimator of the SSF model for small samples;
- estimation of the model for samples of 1000 or more objects is impossible in the specified environment due to double-precision floating-point limits;
- model estimation takes a long time in a relatively powerful environment.

Table A6.5.1. SimE5 simulation results

n	*Parameter*	*True Values*	*Mean*	*Bias*	*Bias, %*	*SD*	*RMSD*
50	β_0	5	4.9224	-0.0776	0.4425	0.4492	-0.0155
	β_1	2	1.9933	-0.0067	0.083	0.0833	-0.0033
	β_2	3	2.9938	-0.0062	0.0692	0.0695	-0.0021
	σ_v	0.1	0.1444	0.0444	0.0503	0.0671	0.4443
	σ_u	0.5	0.3409	-0.1591	0.1916	0.249	-0.3181
	ρ_v	0.4	-0.0413	-0.4413	0.3985	0.5946	-1.1032
100	β_0	5	4.9665	-0.0335	0.5179	0.519	-0.0067
	β_1	2	2.001	0.001	0.0713	0.0713	0.0005
	β_2	3	3.0103	0.0103	0.0666	0.0674	0.0034
	σ_v	0.1	0.1434	0.0434	0.0529	0.0685	0.4341
	σ_u	0.5	0.3828	-0.1172	0.1566	0.1956	-0.2344
	ρ_v	0.4	0.1681	-0.2319	0.2917	0.3727	-0.5798
200	β_0	5	4.908	-0.092	0.2919	0.306	-0.0184
	β_1	2	2.0063	0.0063	0.049	0.0494	0.0032
	β_2	3	2.9972	-0.0028	0.0385	0.0386	-0.0009
	σ_v	0.1	0.142	0.042	0.0493	0.0648	0.4196
	σ_u	0.5	0.392	-0.108	0.1629	0.1954	-0.2159
	ρ_v	0.4	0.1658	-0.2342	0.2211	0.3221	-0.5856
300	β_0	5	4.9755	-0.0245	0.2549	0.2561	-0.0049
	β_1	2	1.9975	-0.0025	0.0333	0.0334	-0.0013
	β_2	3	2.9923	-0.0077	0.0533	0.0539	-0.0026
	σ_v	0.1	0.1463	0.0463	0.0413	0.062	0.463
	σ_u	0.5	0.428	-0.072	0.1294	0.1481	-0.1439
	ρ_v	0.4	0.2021	-0.1979	0.1822	0.269	-0.4947

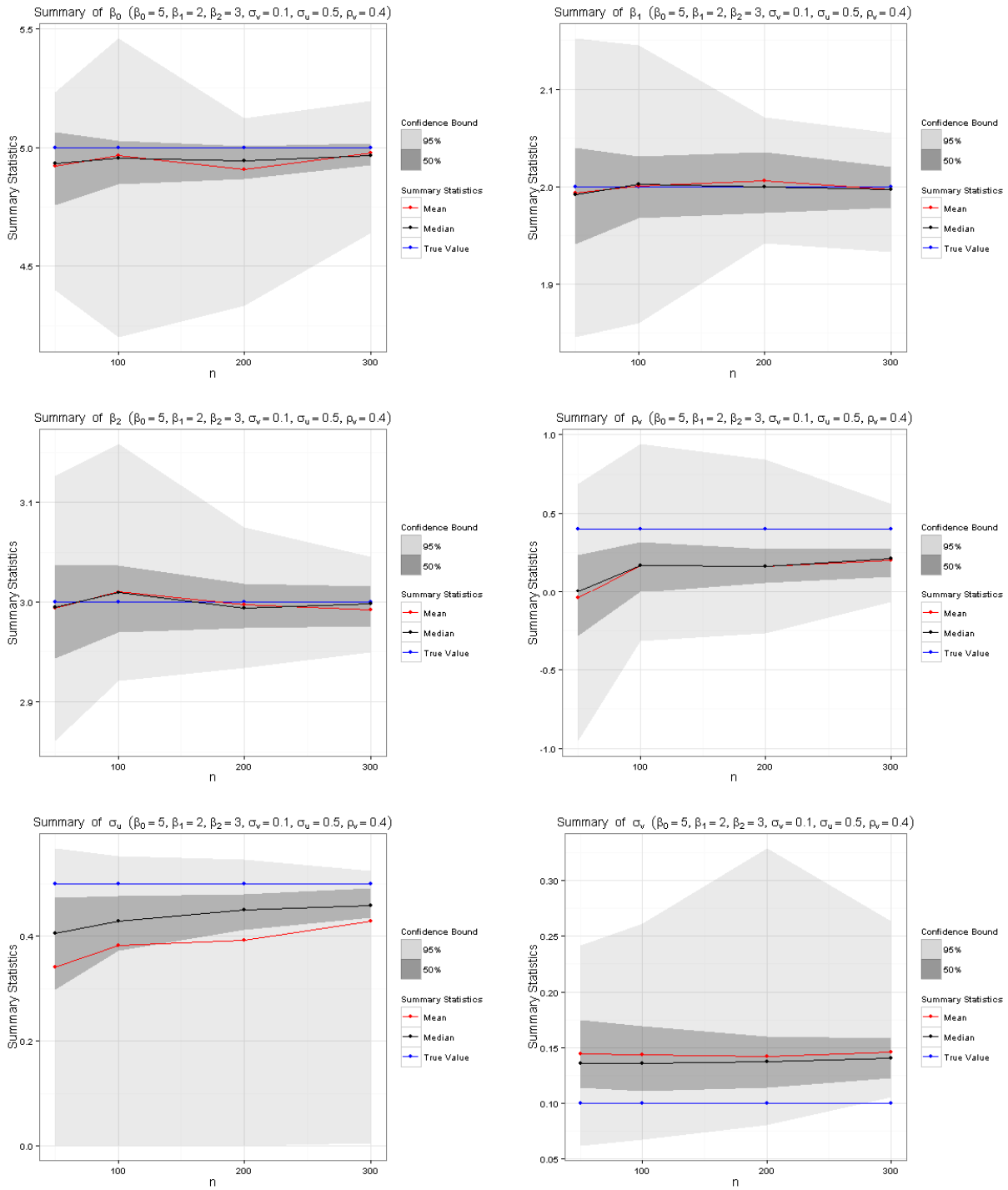


Fig. A6.5.1. Summary of SimE5 estimates

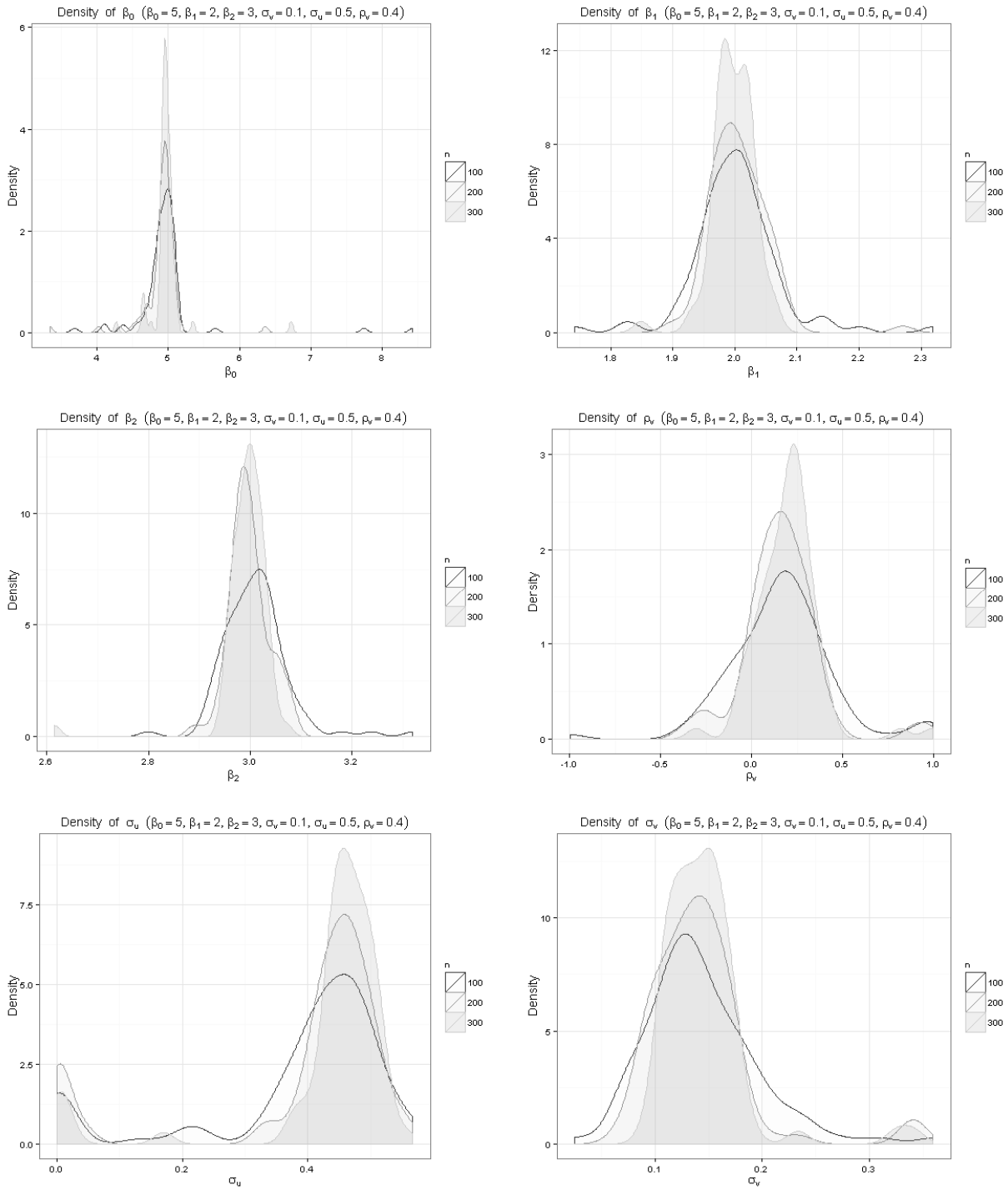


Fig.A6.5.1. Empirical kernel densities of SimE5 estimates

***Simulation Experiment*: SimE5b**

DGP: $\mu^* = 0, \rho_Y^* = 0, \rho_v^* = 0.4, \rho_u^* = 0$

Estimator: SSF(1,0,0,0), half-normal inefficiency

Sample size: 50, 100, 200, 300

Simulations: 100

Execution time: 21.4 mins

Main conclusions:

- unbiased and consistent estimates for frontier parameters, except σ_v and σ_u;
- unbiased and consistent estimates for absent endogenous spatial effects parameter ρ_Y, so there is no replacement of spatially correlated random disturbances with endogenous spatial effects.

Table A6.5b.1. SimE5b simulation results

n	*Parameter*	*True Values*	*Mean*	*Bias*	*Bias, %*	*SD*	*RMSD*
50	β_0	5	4.8803	-0.1197	2.6218	2.6246	-0.0239
	β_1	2	1.9854	-0.0146	0.3249	0.3252	-0.0073
	β_2	3	2.9324	-0.0676	0.3801	0.386	-0.0225
	ρ_Y	0	0.0091	0.0091	0.2459	0.2461	
	σ_v	0.5	0.2913	-0.2087	0.3862	0.439	-0.4174
	σ_u	2.5	2.2792	-0.2208	0.6383	0.6754	-0.0883
	ρ_v	0.4					
100	β_0	5	4.7642	-0.2358	1.8761	1.8909	-0.0472
	β_1	2	2.0269	0.0269	0.2119	0.2136	0.0135
	β_2	3	3.0142	0.0142	0.2515	0.2519	0.0047
	ρ_Y	0	0.0183	0.0183	0.1564	0.1575	
	σ_v	0.5	0.338	-0.162	0.2783	0.322	-0.3239
	σ_u	2.5	2.488	-0.012	0.3221	0.3224	-0.0048
	ρ_v	0.4					
200	β_0	5	4.9612	-0.0388	1.1652	1.1658	-0.0078
	β_1	2	1.9887	-0.0113	0.1661	0.1664	-0.0057
	β_2	3	3.0056	0.0056	0.1515	0.1516	0.0019
	ρ_Y	0	0.0042	0.0042	0.1035	0.1036	
	σ_v	0.5	0.4585	-0.0415	0.1566	0.1621	-0.0831
	σ_u	2.5	2.4888	-0.0112	0.2153	0.2156	-0.0045
	ρ_v	0.4					
300	β_0	5	4.9396	-0.0604	0.8454	0.8476	-0.0121
	β_1	2	2.0101	0.0101	0.1352	0.1356	0.0051
	β_2	3	2.9965	-0.0035	0.1218	0.1219	-0.0012
	ρ_Y	0	0.0034	0.0034	0.0739	0.074	
	σ_v	0.5	0.4921	-0.0079	0.1043	0.1045	-0.0157
	σ_u	2.5	2.4794	-0.0206	0.1533	0.1546	-0.0082
	ρ_v	0.4					

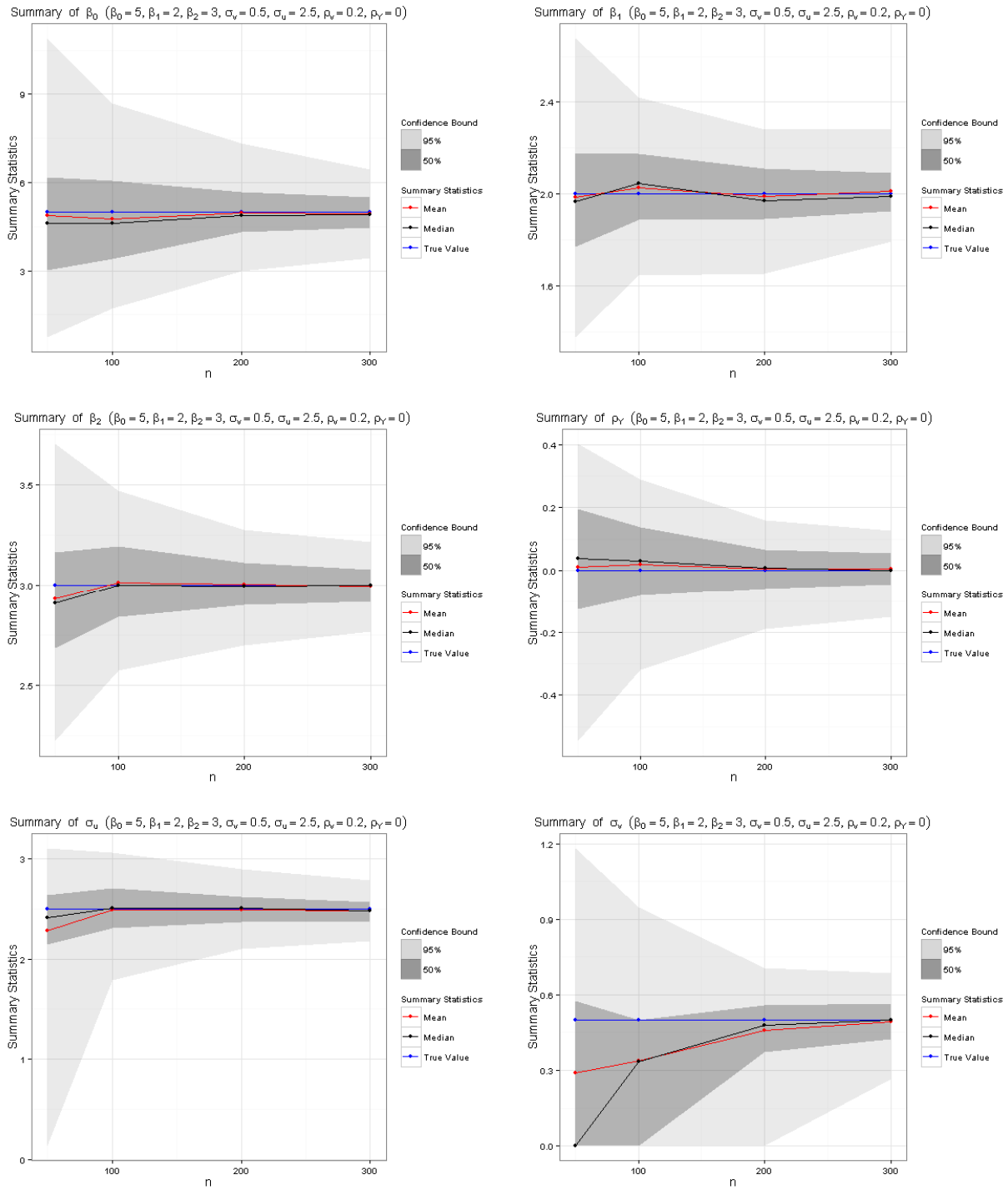


Fig. A6.5b.1. Summary of SimE5b estimates

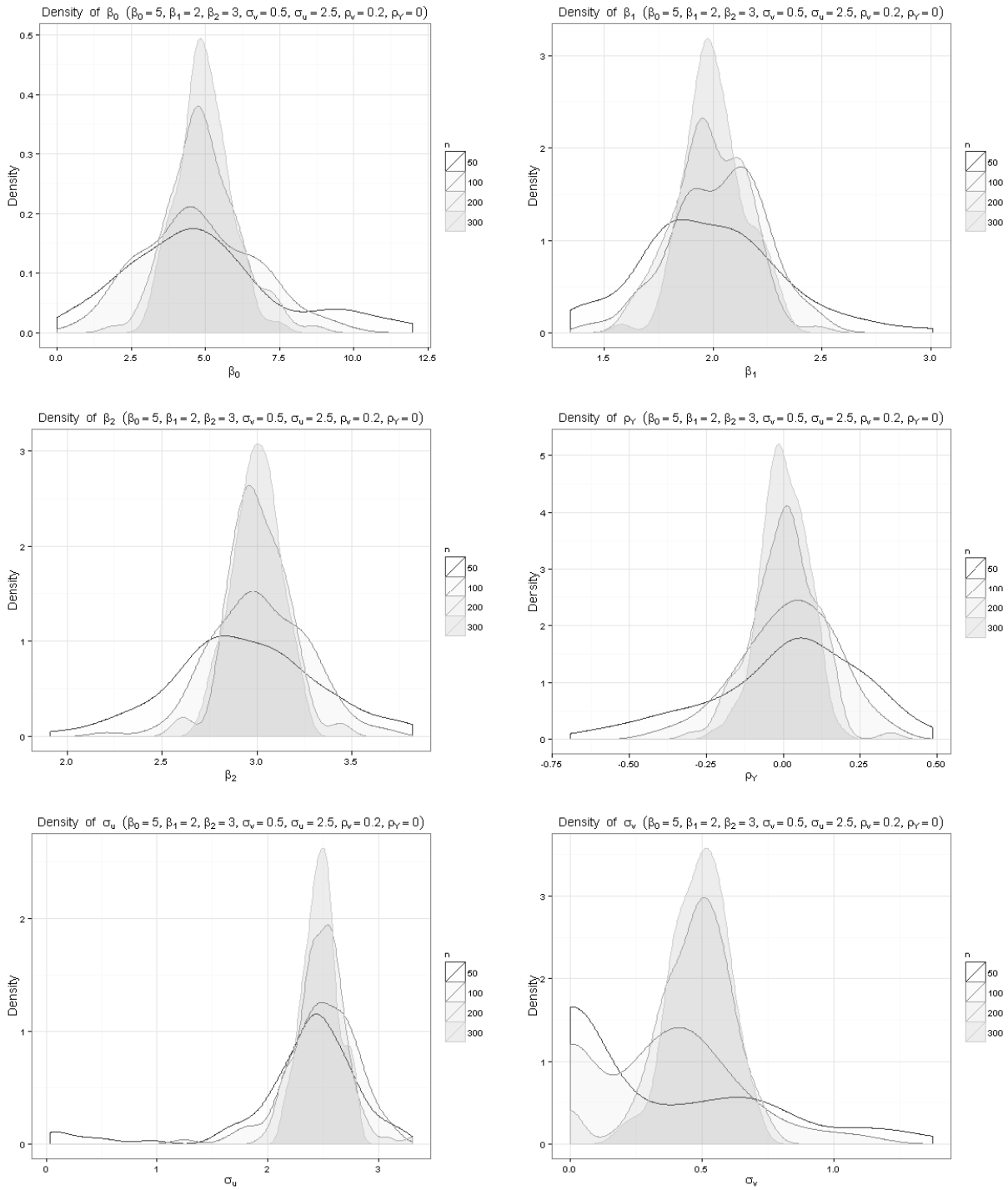


Fig.A6.5b.1. Empirical kernel densities of SimE5b estimates

***Simulation Experiment*: SimE6**

DGP: $\mu^* = 0, \rho_Y^* = 0, \rho_v^* = 0, \rho_u^* = 0.4$

Estimator: SSF(0,0,0,1), half-normal inefficiency

Sample size: 50, 100, 200, 300

Simulations: 100

Execution time: ~20 hrs

Main conclusions:

- unbiased and consistent estimates for frontier parameters;
- consistent estimates for the spatially related efficiency parameter ρ_u;
- large sample variance of the spatially related efficiency parameter ρ_u and inefficiency standard deviation σ_u for a small sample of 100 objects. So this is not recommended to apply MLE estimator of the SSF model for small samples;
- estimation of the model for samples of 1000 or more objects is impossible in the specified environment due to double-precision floating-point limits;
- model estimation takes a long time in a relatively powerful environment.

Table A6.6.1. SimE6 simulation results

n	*Parameter*	*True Values*	*Mean*	*Bias*	*Bias, %*	*SD*	*RMSD*
50	β_0	5	5.5095	0.5095	3.5904	3.6263	0.1019
	β_1	2	1.9907	-0.0093	0.0755	0.0761	-0.0047
	β_2	3	3.0006	0.0006	0.0673	0.0673	0.0002
	σ_v	0.1	0.1317	0.0317	0.0518	0.0608	0.3171
	σ_u	0.5	0.3706	-0.1294	0.1559	0.2026	-0.2587
	ρ_u	0.4	0.3297	-0.0703	0.3409	0.3481	-0.1759
300	β_0	5	5.6143	0.6143	4.6745	4.7147	0.1229
	β_1	2	2.0028	0.0028	0.0514	0.0514	0.0014
	β_2	3	3.0082	0.0082	0.0599	0.0605	0.0027
	σ_v	0.1	0.1392	0.0392	0.0505	0.0639	0.3922
	σ_u	0.5	0.4014	-0.0986	0.1575	0.1858	-0.1972
	ρ_u	0.4	0.2641	-0.1359	0.2882	0.3186	-0.3397
200	β_0	5	4.9783	-0.0217	1.4523	1.4525	-0.0043
	β_1	2	2.0074	0.0074	0.0624	0.0628	0.0037
	β_2	3	2.995	-0.005	0.0533	0.0535	-0.0017
	σ_v	0.1	0.1459	0.0459	0.0369	0.0589	0.4592
	σ_u	0.5	0.4284	-0.0716	0.1319	0.1501	-0.1432
	ρ_u	0.4	0.244	-0.156	0.2597	0.3029	-0.3901
300	β_0	5	5.2245	0.2245	2.3552	2.3659	0.0449
	β_1	2	1.9964	-0.0036	0.0371	0.0372	-0.0018
	β_2	3	3.0018	0.0018	0.0359	0.0359	0.0006
	σ_v	0.1	0.1418	0.0418	0.0326	0.053	0.4176
	σ_u	0.5	0.4418	-0.0582	0.1115	0.1258	-0.1164
	ρ_u	0.4	0.2713	-0.1287	0.2718	0.3007	-0.3217

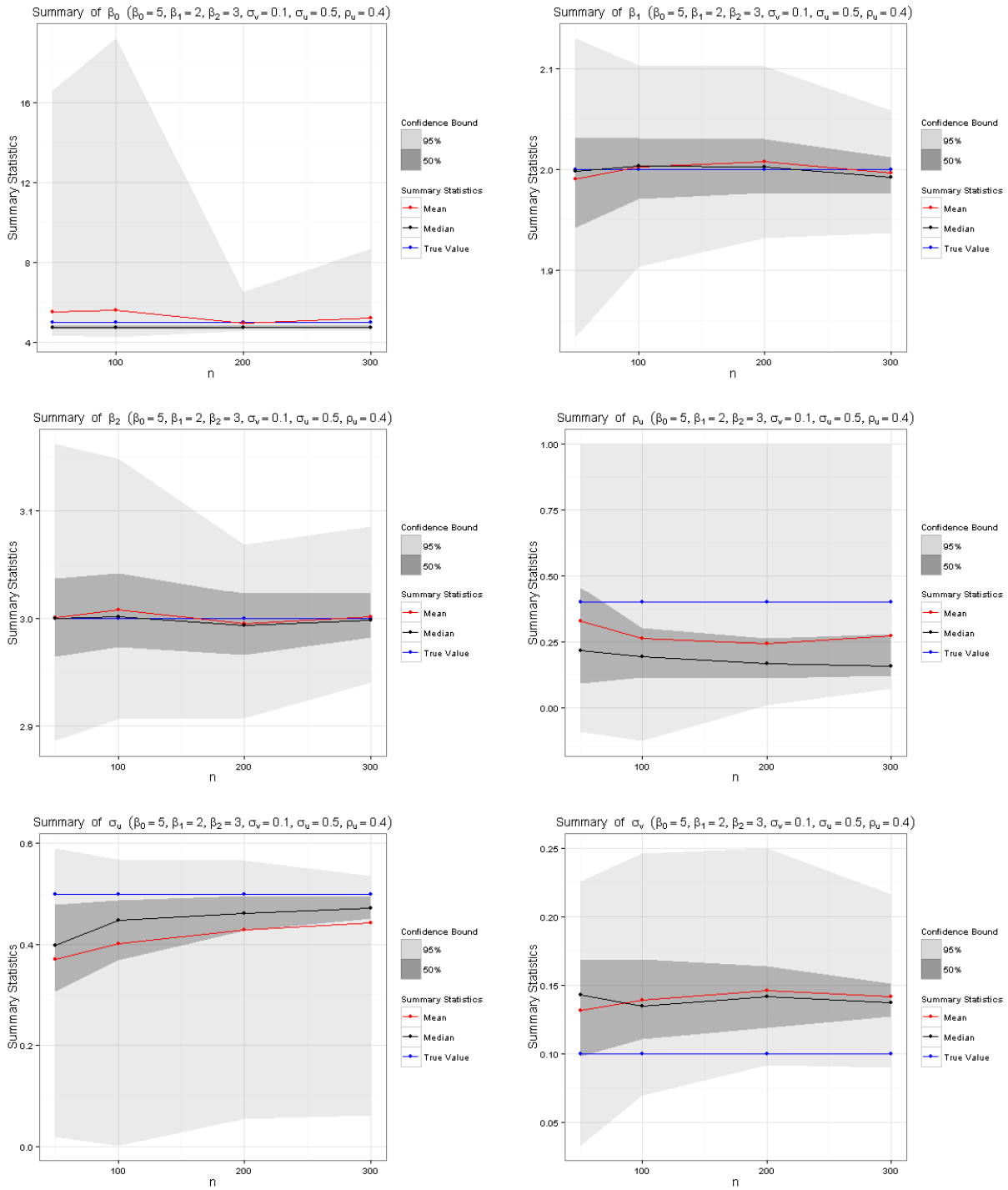


Fig. A6.6.1. Summary of SimE6 estimates

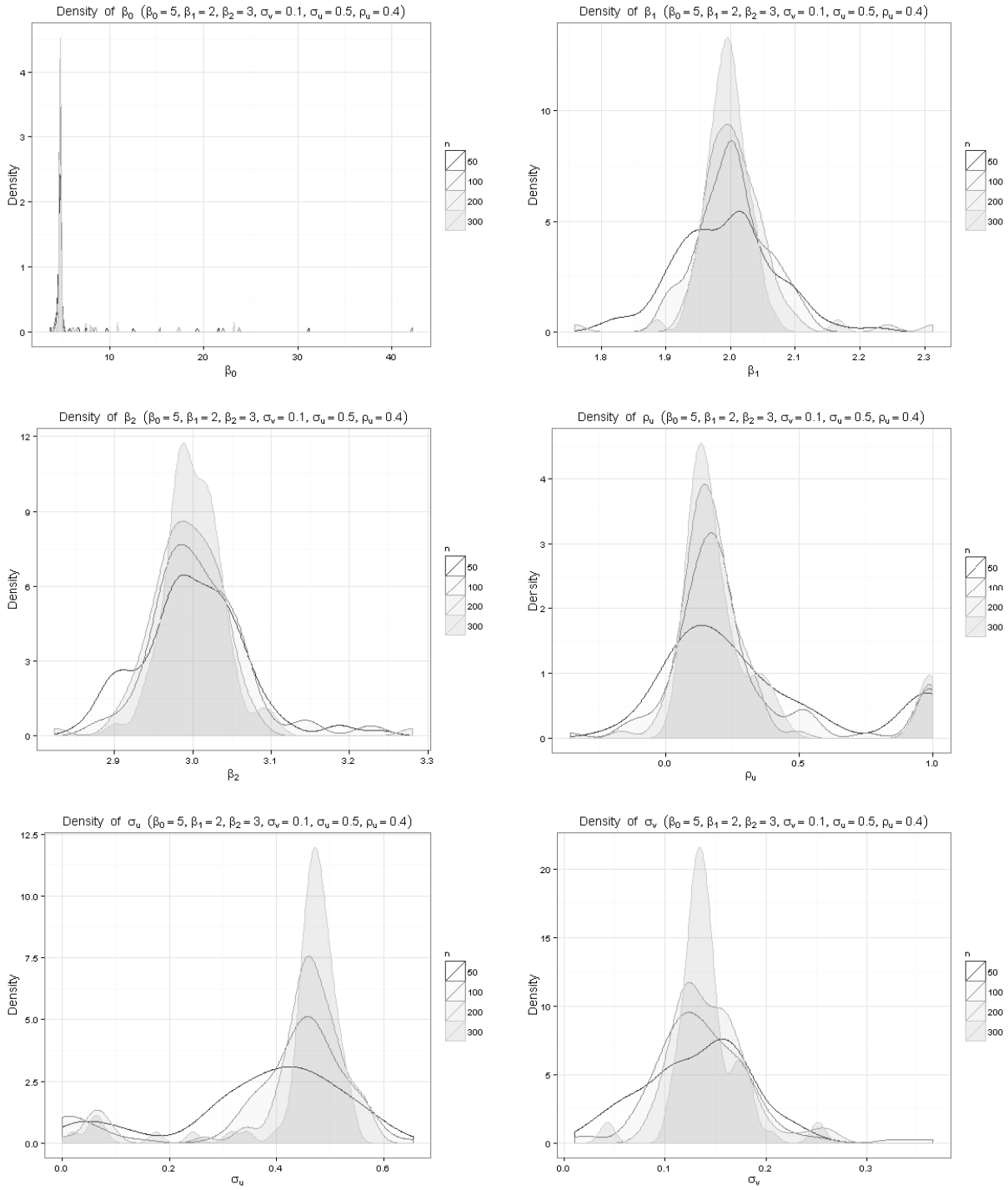


Fig. A6.6.2. Empirical kernel densities of SimE6 estimates

***Simulation Experiment*: SimE6b**

DGP: $\mu^* = 0,\ \rho_Y^* = 0,\ \rho_v^* = 0,\ \rho_u^* = 0.4$

Estimator: SSF(1,0,0,0), half-normal inefficiency

Sample size: 50, 100, 200, 300

Simulations: 100

Execution time: 21.3 mins

Main conclusions:

- unbiased and consistent estimates for frontier parameters, except σ_v and σ_u;
- unbiased and consistent estimates for absent endogenous spatial effects parameter ρ_Y, so there is no replacement of spatially related inefficiencies with endogenous spatial effects.

Table A6.6b.1. SimE6b Simulation results

n	Parameter	True Values	Mean	Bias	Bias, %	SD	RMSD
50	β_0	5	3.3136	-1.6864	2.6104	3.1078	-0.3373
	β_1	2	2.0086	0.0086	0.3499	0.35	0.0043
	β_2	3	2.9516	-0.0484	0.394	0.3969	-0.0161
	ρ_Y	0	0.1074	0.1074	0.2538	0.2755	
	σ_v	0.5	0.3122	-0.1878	0.4038	0.4453	-0.3756
	σ_u	2.5	2.2667	-0.2333	0.7089	0.7463	-0.0933
	ρ_u	0.4					
300	β_0	5	3.5275	-1.4725	1.8782	2.3866	-0.2945
	β_1	2	2.0225	0.0225	0.2194	0.2205	0.0113
	β_2	3	3.0001	0.0001	0.2587	0.2587	0
	ρ_Y	0	0.092	0.092	0.1615	0.1859	
	σ_v	0.5	0.3579	-0.1421	0.2768	0.3111	-0.2843
	σ_u	2.5	2.4752	-0.0248	0.3082	0.3092	-0.0099
	ρ_u	0.4					
200	β_0	5	3.6254	-1.3746	1.1315	1.7804	-0.2749
	β_1	2	1.9998	-0.0002	0.1669	0.1669	-0.0001
	β_2	3	3.0048	0.0048	0.1504	0.1505	0.0016
	ρ_Y	0	0.0858	0.0858	0.0996	0.1315	
	σ_v	0.5	0.4512	-0.0488	0.1782	0.1847	-0.0975
	σ_u	2.5	2.5035	0.0035	0.2146	0.2146	0.0014
	ρ_u	0.4					
300	β_0	5	3.4677	-1.5323	0.8081	1.7324	-0.3065
	β_1	2	2.0099	0.0099	0.134	0.1344	0.005
	β_2	3	2.999	-0.001	0.1233	0.1233	-0.0003
	ρ_Y	0	0.0984	0.0984	0.0718	0.1218	
	σ_v	0.5	0.4942	-0.0058	0.1031	0.1033	-0.0115
	σ_u	2.5	2.4929	-0.0071	0.157	0.1571	-0.0029
	ρ_u	0.4					

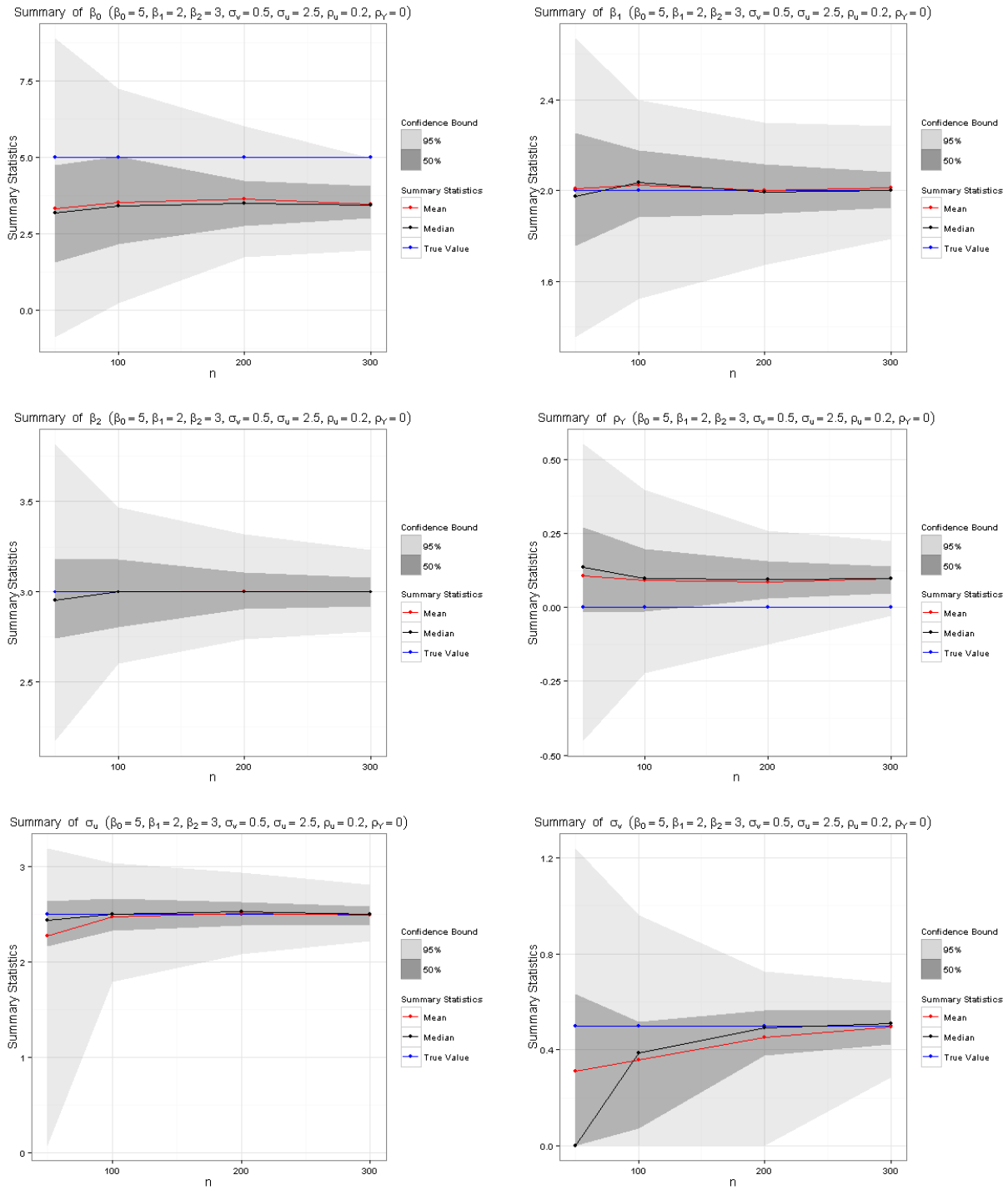


Fig. A6.6b.1. Summary of SimE6b estimates

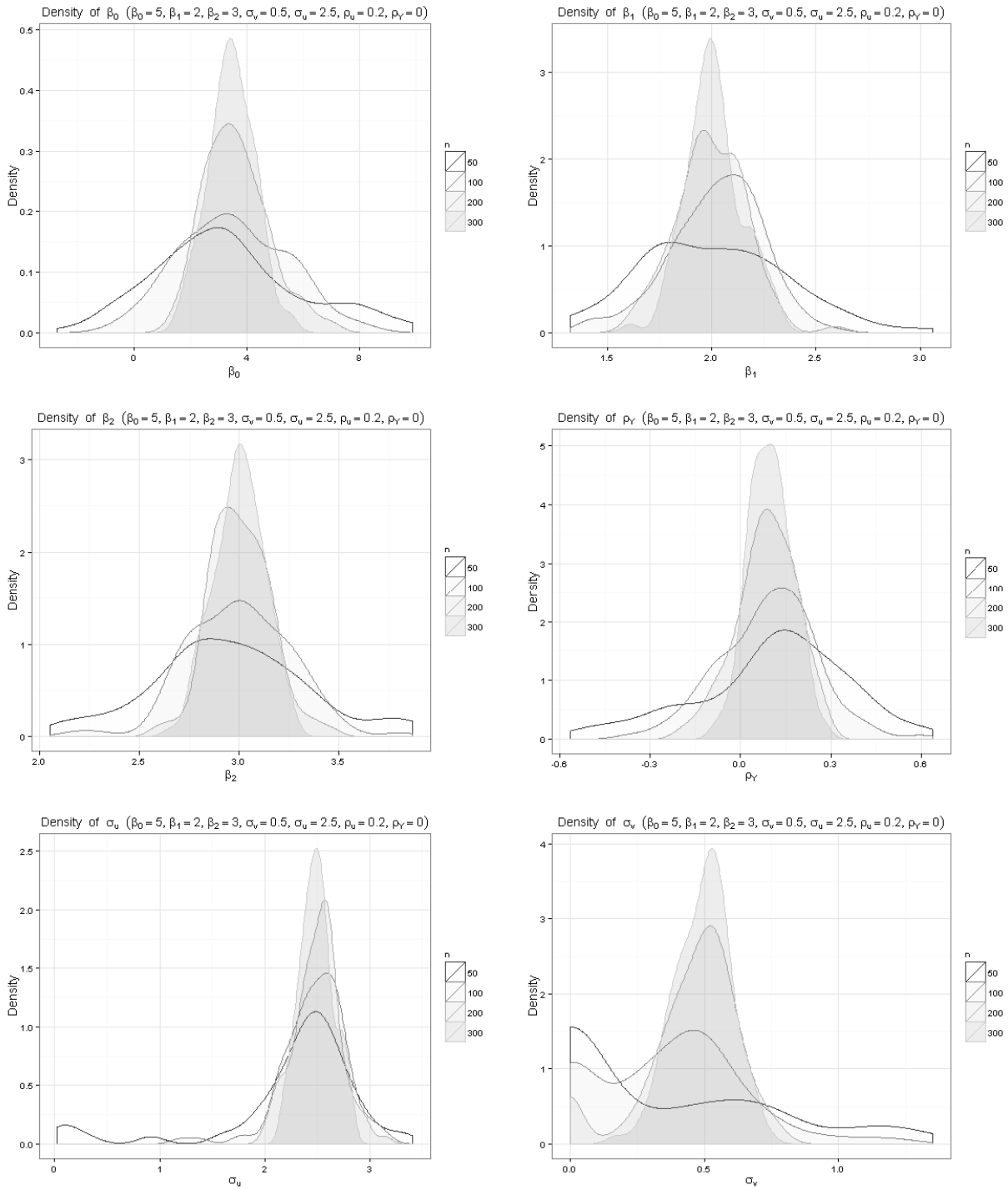
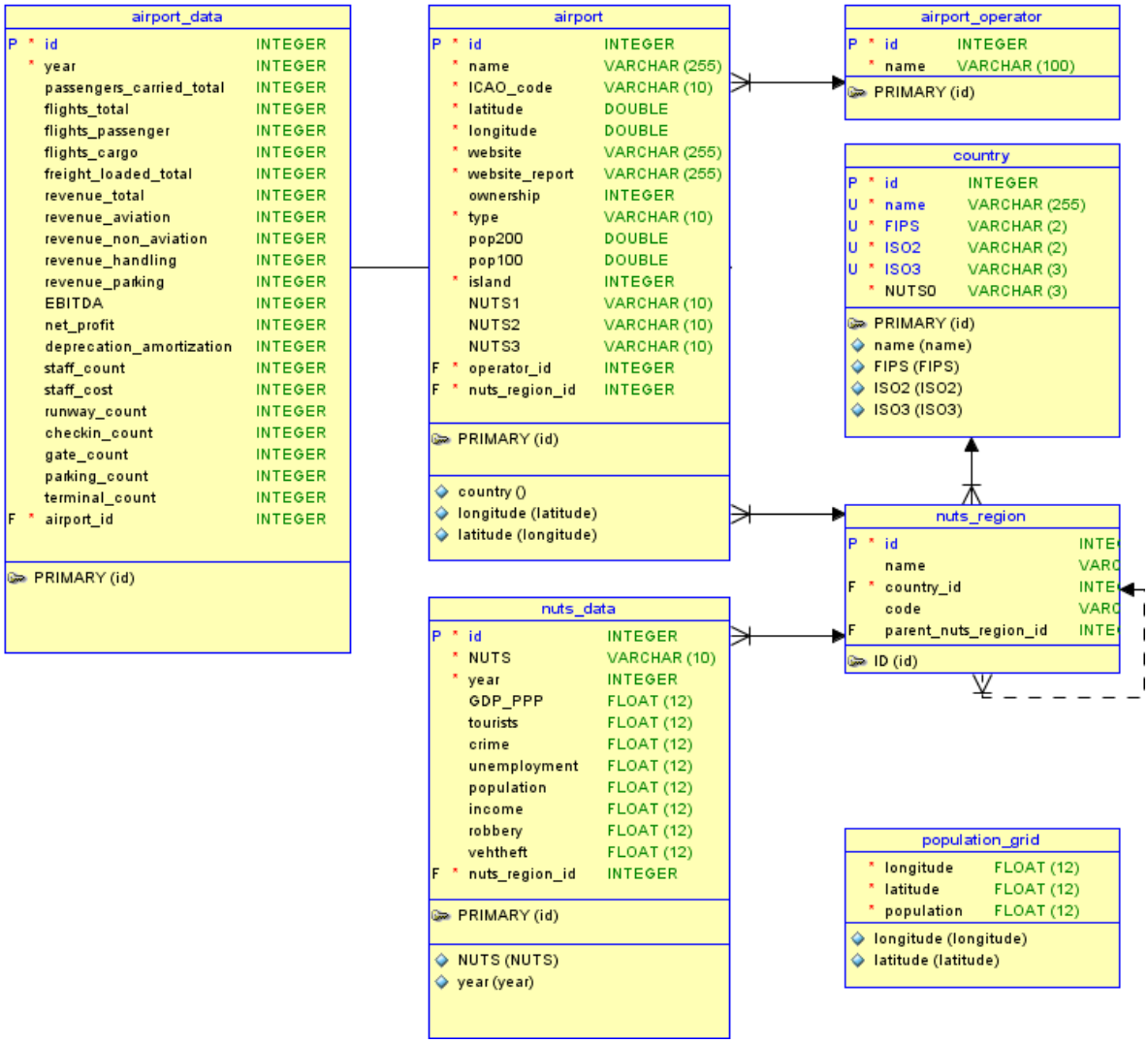


Fig. A6.6b.2. Empirical kernel densities of SimE6b estimates

Appendix 7. Entity-Relationship diagram of the research database



Appendix 8. Summary statistics of the data set of European airports

Table A8.1. Summary statistics of the data set of European airports, 2011

Variable	Units	Minimum	Maximum	Median	Mean	Standard deviation	Zero values	Not available
PAX	number	1500	69388110	892080	3817089.00	8615464.0	0	0
ATM	number	44	507398	11264	39090.93	75101.02	0	0
Cargo	tonnes	0	2214648	338	41512.25	199132.90	62	0
Population100km	number	5698	19281820	2142411	3467474.00	4144592	0	0
Population200km	number	59016	48377190	7011503	11030450.0	10959160	0	0
Island	logical	0	1	0	0.12	0.33	316	0
GDPpc	EUR	5900	153400	21200	22454.23	10386.80	0	0
RunwayCount	number	1	5	2	1.67	0.79	0	224
CheckinCount	number	4	600	42	76.32	96.48	0	228
GateCount	number	2	350	21	39.91	51.15	0	228
ParkingSpaces	number	89	27500	3600	5232.85	5296.25	0	262
RoutesDeparture	number	1	457	9	37.16	69.03	0	0
RoutesArrival	number	1	458	9	37.10	68.88	0	0

Appendix 9. Correlations of infrastructure indicators

Table A9.1. Pearson's correlation coefficient values for infrastructure indicators of European airports, 2011

	RunwayCount	*GateCount*	*ParkingSpaces*	*CheckinCount*	*Routes*
RunwayCount	1	0.68	0.57	0.64	0.69
GateCount	0.68	1	0.78	0.9	0.83
ParkingSpaces	0.57	0.78	1	0.69	0.69
CheckinCount	0.64	0.9	0.69	1	0.82
Routes	0.69	0.83	0.69	0.82	1

All correlation values are highly significant (sig. $< 10^{-16}$)

Appendix 10. Descriptive statistics of PFP indicators' values of European airports

Table A10.1. Descriptive statistics of PFP indicators' values of European airports, 2011

	Min.	1st Qu	Median	Mean	3rd Qu	Max.	NA's
ATM per Runway	630	22292	43420	49361	68160	199720	224
WLU per Runway	2120000	191700000	370000000	487300000	610500000	2313000000	224
PAX per Runway	21200	1916896	3700108	4872811	6104481	23129368	224
ATM per Route	11	393	544	699	763	5957	0
WLU per Route	37500	2385622	4015204	4712943	5577856	30736270	0
PAX per Route	375	23856	40151	47126	55777	307362	0
PAX per capita in 100 km	0.00008	0.14182	0.52922	2.57148	2.01826	60.14876	0

Appendix 11. Model Europe1 estimates of airport efficiency levels

Table A11.1. European airports' efficiency levels, estimated using the Model Europe1

	Country	ICAO	AirportName	PAX	SFA efficiency values	SSFA(1,1,0) efficiency values
1	Greece	LGKO	Kos	1926223	0.818	0.840
2	France	LFTH	Le Palyvestre	574122	0.837	0.835
3	Greece	LGSR	Santorini	785547	0.812	0.828
4	Greece	LGIR	Nikos Kazantzakis	5247007	0.828	0.828
5	Greece	LGKR	Ioannis Kapodistrias Intl	1844173	0.842	0.827
6	Croatia	LDPL	Pula	344640	0.791	0.816
7	Spain	LEIB	Ibiza	5612913	0.814	0.816
8	Bulgaria	LBBG	Burgas	2227430	0.817	0.807
9	Greece	LGSA	Souda	1774708	0.813	0.806
10	Denmark	EKKA	Karup	292972	0.790	0.800
11	Spain	LEMH	Menorca	2561368	0.801	0.800
12	Greece	LGRP	Rhodes Diagoras	4148386	0.794	0.776
13	Finland	EFOU	Oulu	973127	0.760	0.770
14	Greece	LGMK	Mikonos	482809	0.760	0.770
15	France	LFBT	Lourdes	446347	0.739	0.763
16	Czech Republic	LKMT	Mosnov	245596	0.756	0.760
17	Sweden	ESPA	Kallax	1066702	0.742	0.757
18	Norway	ENDU	Bardufoss	196980	0.748	0.756
19	Germany	EDXW	Westerland Sylt	195438	0.731	0.753
20	Greece	LGKV	Megas Alexandros Intl	252307	0.755	0.750
21	Greece	LGSK	Alexandros Papadiamantis	246658	0.759	0.747
22	France	LFRG	St Gatien	119804	0.636	0.738
23	France	LFMP	Rivesaltes	367726	0.709	0.735
24	Greece	LGZA	Dionysios Solomos	920701	0.774	0.735
25	Germany	EDDG	Munster Osnabruck	1293315	0.713	0.726
26	Italy	LIPR	Rimini	913190	0.717	0.717
27	Spain	LELC	Murcia San Javier	1262534	0.729	0.712
28	Czech Republic	LKTB	Turany	525432	0.695	0.702
29	Finland	EFKU	Kuopio	284098	0.679	0.696
30	Sweden	ESOW	Vasteras	144434	0.692	0.696
31	Spain	LERS	Reus	1347890	0.716	0.694
32	Sweden	ESNZ	Ostersund Airport	377976	0.665	0.684
33	Italy	LIEO	Olbia Costa Smeralda	1825580	0.681	0.684
34	Germany	EDDE	Erfurt	264726	0.664	0.682
35	Greece	LGPZ	Aktio	294156	0.724	0.677
36	Finland	EFJO	Joensuu	117109	0.656	0.677
37	Finland	EFRO	Rovaniemi	395473	0.666	0.675
38	United Kingdom	EGKK	Gatwick	33638323	0.615	0.672
39	United Kingdom	EGHH	Bournemouth	612499	0.588	0.671
40	France	LFRQ	Pluguffan	110804	0.622	0.668
41	Spain	LESO	San Sebastian	240767	0.659	0.666
42	Lithuania	EYKA	Kaunas Intl	870801	0.676	0.665
43	United Kingdom	EGAC	Belfast City	2392382	0.623	0.662
44	United Kingdom	EGLL	Heathrow	69388105	0.597	0.660
45	Spain	GCXO	Tenerife Norte	4118009	0.657	0.660
46	Spain	LEPA	Son Sant Joan	22702799	0.698	0.660
47	United Kingdom	EGAA	Belfast Intl	4101907	0.622	0.659
48	United Kingdom	EGSS	Stansted	18043407	0.594	0.659
49	Croatia	LDDU	Dubrovnik	1326250	0.681	0.658
50	Sweden	ESMS	Sturup	1945895	0.645	0.649
51	Spain	LEPP	Pamplona	230401	0.643	0.647

52	France	LFKB	Poretta	1023566	0.627	0.647
53	France	LFRB	Guipavas	968695	0.610	0.644
54	Ireland	EINN	Shannon	1342753	0.631	0.644
55	Italy	LIML	Linate	9061749	0.642	0.643
56	France	LFPO	Orly	27099908	0.609	0.640
57	Sweden	ESNU	Umea	956255	0.635	0.639
58	United Kingdom	EGGW	Luton	9509911	0.568	0.637
59	Italy	LICA	Lamezia Terme	2293744	0.677	0.636
60	Slovakia	LZIB	M R Stefanik	1577414	0.650	0.632
61	Germany	EDDK	Koln Bonn	9599976	0.614	0.631
62	Croatia	LDZD	Zadar	265982	0.640	0.631
63	Norway	ENGM	Gardermoen	21102984	0.648	0.629
64	Spain	LEVD	Valladolid	452919	0.625	0.628
65	Romania	LRBC	Bacau	293965	0.660	0.627
66	Croatia	LDSP	Split	1271202	0.655	0.627
67	Denmark	EKYT	Aalborg	1377821	0.635	0.625
68	United Kingdom	EGGP	Liverpool	5246414	0.567	0.625
69	United Kingdom	EGAE	City of Derry	405568	0.582	0.625
70	Spain	LEGE	Girona	2991065	0.648	0.620
71	Spain	GCHI	Hierro	169314	0.585	0.618
72	Spain	LECO	A Coruna	1006727	0.621	0.617
73	Bulgaria	LBWN	Varna	1164805	0.673	0.615
74	United Kingdom	EGPF	Glasgow	6858264	0.565	0.610
75	Germany	EDDC	Dresden	1902662	0.605	0.610
76	Germany	EDSB	Baden Airpark	1107496	0.559	0.607
77	Germany	EDDF	Frankfurt Main	56276006	0.602	0.602
78	France	LFTW	Garons	192494	0.574	0.602
79	Italy	LIEE	Elmas	3681944	0.631	0.601
80	Ireland	EIDW	Dublin	18719711	0.588	0.597
81	Italy	LICD	Lampedusa	161291	0.559	0.595
82	United Kingdom	EGBB	Birmingham	8606497	0.499	0.594
83	Sweden	ESGG	Landvetter	4906329	0.593	0.593
84	Spain	LEMD	Barajas	49531687	0.643	0.591
85	Spain	LEGR	Granada	862621	0.617	0.588
86	Italy	LIRA	Ciampino	4741287	0.625	0.588
87	France	LFPG	Charles De Gaulle	60742357	0.559	0.584
88	Norway	ENCN	Kjevik	945316	0.575	0.583
89	Netherlands	EHAM	Schiphol	49690392	0.562	0.582
90	Italy	LIMJ	Genova Sestri	1393985	0.573	0.581
91	Spain	LEAM	Almeria	763762	0.602	0.579
92	Norway	ENBR	Flesland	5184549	0.577	0.578
93	Austria	LOWL	Linz	659787	0.571	0.577
94	Germany	EDLP	Paderborn Lippstadt	954095	0.574	0.576
95	France	LFBZ	Anglet	1031474	0.594	0.576
96	Belgium	EBCI	Brussels South	5883173	0.525	0.576
97	France	LFBE	Roumaniere	290020	0.492	0.575
98	United Kingdom	EGPH	Edinburgh	9383242	0.539	0.574
99	France	LFBH	La Rochelle-Ile de Re	227841	0.511	0.573
100	Italy	LICJ	Palermo	4966162	0.616	0.573
101	Denmark	EKEB	Esbjerg	87196	0.585	0.573
102	Greece	LGAL	Dimokritos	238265	0.615	0.572
103	Italy	LIMP	Parma	268369	0.569	0.571
104	Germany	EDDN	Nurnberg	3933626	0.568	0.571
105	France	LFMT	Mediterranee	1308195	0.555	0.571
106	Norway	ENZV	Sola	3881453	0.565	0.570
107	United Kingdom	EGPK	Prestwick	1295512	0.517	0.569
108	United Kingdom	EGGD	Bristol	5767628	0.495	0.568
109	Belgium	EBAW	Deurne	114681	0.495	0.568
110	United Kingdom	EGPN	Dundee	61629	0.498	0.564

111	Germany	EDLW	Dortmund	1808956	0.568	0.563
112	France	LFRD	Pleurtuit	131004	0.479	0.561
113	Italy	LIMF	Torino	3700108	0.534	0.561
114	Sweden	ESDF	Ronneby	227497	0.536	0.560
115	Germany	EDDH	Hamburg	13528395	0.581	0.558
116	Norway	ENVA	Vaernes	3901645	0.573	0.558
117	Norway	ENAL	Vigra	892080	0.547	0.557
118	Germany	EDFH	Frankfurt Hahn	2829589	0.515	0.554
119	United Kingdom	EGCC	Manchester	18803819	0.500	0.554
120	United Kingdom	EGNJ	Humberside	273096	0.476	0.550
121	Germany	EDNY	Friedrichshafen	539376	0.498	0.550
122	Italy	LIPQ	Ronchi Dei Legionari	854252	0.570	0.548
123	Italy	LIME	Bergamo Orio Al Serio	8410684	0.562	0.548
124	Italy	LIRN	Capodichino	5728402	0.609	0.546
125	United Kingdom	EGPD	Dyce	3082575	0.498	0.545
126	Switzerland	LSZH	Zurich	24313250	0.523	0.543
127	Norway	ENHD	Karmoy	597053	0.531	0.542
128	Germany	EDDM	Franz Josef Strauss	37593829	0.554	0.541
129	Norway	ENEV	Evenes	582338	0.552	0.541
130	Norway	ENML	Aro	434844	0.529	0.540
131	Greece	LGKF	Kefallinia	346397	0.624	0.539
132	Germany	EDLV	Niederrhein	2410060	0.512	0.538
133	France	LFBO	Blagnac	6936702	0.520	0.538
134	Denmark	EKAH	Aarhus	587559	0.550	0.538
135	Italy	LIRP	Pisa	4509561	0.562	0.538
136	Greece	LGKP	Karpathos	181084	0.482	0.538
137	France	LFLC	Auvergne	388909	0.457	0.536
138	Ireland	EIKY	Kerry	310905	0.529	0.536
139	Greece	LGSM	Samos	408720	0.577	0.535
140	Italy	LIPK	Forli	344168	0.563	0.535
141	United Kingdom	EGNX	Nottingham East Midlands	4208260	0.455	0.534
142	Germany	EDDS	Stuttgart	9536000	0.526	0.534
143	France	LFBP	Pau Pyrenees	639020	0.544	0.533
144	Italy	LIBR	Casale	2049754	0.599	0.531
145	Italy	LIMC	Malpensa	19087098	0.533	0.531
146	Finland	EFVA	Vaasa	338140	0.527	0.529
147	Germany	EDDB	Schonefeld	7098842	0.533	0.526
148	United Kingdom	EGNT	Newcastle	4336304	0.467	0.525
149	Germany	EDDP	Leipzig Halle	1834406	0.530	0.525
150	France	LFSD	Longvic	44538	0.391	0.524
151	France	LFMK	Salvaza	368003	0.492	0.524
152	Spain	LEST	Santiago	2446685	0.552	0.523
153	France	LFKJ	Campo Dell Oro	1171198	0.516	0.521
154	Sweden	ESSA	Arlanda	19058651	0.542	0.519
155	Sweden	ESKN	Skavsta	2581305	0.546	0.518
156	United Kingdom	EGFF	Cardiff	1208202	0.452	0.517
157	United Kingdom	EGCN	Robin Hood Doncaster Sheffield Airport	821603	0.441	0.516
158	France	LFLP	Meythet	42875	0.377	0.515
159	France	LFMN	Cote D'Azur	10405876	0.530	0.515
160	Italy	LIPZ	Venezia Tessera	8553639	0.519	0.514
161	Norway	ENBO	Bodo	1544353	0.523	0.514
162	France	LFMH	Boutheon	108648	0.426	0.514
163	Spain	LEJR	Jerez	953631	0.551	0.514
164	United Kingdom	EGLC	City	2941781	0.449	0.513
165	Spain	LEBB	Bilbao	4036722	0.521	0.513
166	Spain	LEXJ	Santander	1115200	0.516	0.513
167	Germany	EDDT	Tegel	16892424	0.527	0.513
168	Spain	LEVX	Vigo	975982	0.538	0.512

169	Ireland	EICK	Cork	2350843	0.509	0.511
170	France	LFOK	Vatry	50817	0.425	0.510
171	Belgium	EBBR	Brussels Natl	18613386	0.470	0.510
172	Spain	LEAS	Asturias	1333656	0.520	0.510
173	Italy	LIPY	Falconara	597099	0.549	0.510
174	Ireland	EIKN	Ireland West Knock	653237	0.493	0.509
175	Denmark	EKCH	Kastrup	22606904	0.538	0.508
176	Slovakia	LZKZ	Kosice	263502	0.529	0.508
177	Spain	LEZL	Sevilla	4940062	0.550	0.506
178	France	LFOB	Tille	3677236	0.472	0.504
179	Italy	LIPE	Bologna	5820813	0.535	0.504
180	Finland	EFTP	Tampere Pirkkala	657368	0.509	0.502
181	Italy	LIEA	Alghero	1511472	0.523	0.501
182	Germany	EDDW	Neuenland	2552632	0.517	0.499
183	France	LFRH	Lann Bihoue	181524	0.451	0.498
184	Finland	EFHK	Helsinki Vantaa	14871299	0.534	0.498
185	Cyprus	LCLK	Larnaca	5431272	0.557	0.498
186	Iceland	BIKF	Keflavik International Airport	2462894	0.536	0.495
187	France	LFBD	Merignac	4020670	0.475	0.494
188	Austria	LOWK	Woerthersee International Airport	377974	0.493	0.493
189	Cyprus	LCLK	Larnaca	5431272	0.551	0.491
190	Germany	EDHL	Lubeck Blankensee	329159	0.500	0.490
191	Switzerland	LSGG	Geneve Cointrin	13003611	0.399	0.490
192	France	LFLL	Saint Exupery	8318143	0.405	0.490
193	United Kingdom	EGNH	Blackpool	235669	0.431	0.490
194	Germany	EDDV	Hannover	5302487	0.508	0.489
195	Italy	LIRF	Fiumicino	37404513	0.560	0.489
196	Finland	EFPO	Pori	53619	0.475	0.489
197	Czech Republic	LKPR	Ruzyne	11724179	0.503	0.485
198	France	LFMU	Vias	193702	0.467	0.484
199	Greece	LGIO	Ioannina	88597	0.570	0.482
200	Spain	LEBL	Barcelona	34314376	0.555	0.482
201	United Kingdom	EGNM	Leeds Bradford	2909527	0.419	0.482
202	Norway	ENKB	Kvernberget	283591	0.470	0.482
203	Sweden	ESNQ	Kiruna	164208	0.469	0.481
204	France	LFRK	Carpiquet	99169	0.405	0.481
205	Portugal	LPFR	Faro	5617688	0.516	0.481
206	Norway	ENTC	Langnes	1698357	0.514	0.480
207	Italy	LICC	Catania Fontanarossa	6771238	0.554	0.480
208	Germany	EDDL	Dusseldorf	20298970	0.480	0.478
209	Italy	LICG	Pantelleria	132487	0.489	0.478
210	Finland	EFLP	Lappeenranta	116369	0.474	0.475
211	Hungary	LHBP	Ferihegy	8884837	0.528	0.475
212	Germany	EDJA	Allgau	755458	0.457	0.474
213	Poland	EPKT	Pyrzowice	2513417	0.509	0.472
214	Sweden	ESTA	Angelholm-Helsingborg Airport	396720	0.480	0.469
215	Norway	ENRY	Moss	1666446	0.484	0.468
216	Italy	LICR	Reggio Calabria	519454	0.529	0.467
217	Sweden	ESNN	Sundsvall Harnosand	282291	0.465	0.465
218	Greece	LGHI	Chios	229500	0.514	0.463
219	France	LFJL	Metz Nancy Lorraine	249884	0.388	0.463
220	United Kingdom	EGHI	Southampton	1761961	0.395	0.462
221	Spain	GCLA	La Palma	1048911	0.478	0.462
222	Denmark	EKBI	Billund	2638612	0.519	0.457
223	Sweden	ESGP	Save	772858	0.468	0.456
224	Spain	GEML	Melilla	277717	0.496	0.455
225	Italy	LIPX	Villafranca	3348933	0.478	0.454

226	Sweden	ESSV	Visby	338688	0.415	0.450
227	Italy	LIBD	Bari	3700248	0.523	0.448
228	United Kingdom	EGSH	Norwich	413837	0.385	0.446
229	France	LFRS	Nantes Atlantique	3158378	0.393	0.444
230	Sweden	ESSB	Bromma	2182992	0.463	0.443
231	France	LFKF	Sud Corse	445286	0.455	0.443
232	Spain	LEMG	Malaga	12759548	0.501	0.441
233	Spain	LEZG	Zaragoza Ab	750527	0.461	0.440
234	Finland	EFKS	Kuusamo	91696	0.427	0.439
235	Spain	LEVC	Valencia	4967230	0.499	0.435
236	Switzerland	LSZA	Lugano	165054	0.415	0.435
237	Netherlands	EHRD	Rotterdam	1081841	0.403	0.434
238	Spain	LEAL	Alicante	9892302	0.516	0.430
239	Switzerland	LSZR	St Gallen Altenrhein	94834	0.386	0.429
240	Greece	LGAV	Eleftherios Venizelos Intl	14325505	0.535	0.428
241	United Kingdom	EGPE	Inverness	579123	0.364	0.425
242	Italy	LICT	Trapani Birgi	1468041	0.479	0.423
243	Italy	LIRQ	Firenze	1893238	0.459	0.420
244	Greece	LGTS	Makedonia	3958475	0.511	0.419
245	Norway	ENAT	Alta	346366	0.413	0.418
246	Greece	LGMT	Mitilini	469380	0.486	0.414
247	Spain	GCFV	Fuerteventura	4895403	0.424	0.412
248	Cyprus	LCPH	Pafos Intl	1759292	0.471	0.411
249	Austria	LOWW	Schwechat	21106426	0.469	0.410
250	Finland	EFIV	Ivalo	125318	0.393	0.409
251	France	LFOT	Val De Loire	120485	0.344	0.408
252	Finland	EFTU	Turku	376767	0.414	0.407
253	Finland	EFKE	Kemi Tornio	93753	0.420	0.406
254	Sweden	ESMX	Kronoberg	180692	0.400	0.405
255	Cyprus	LCPH	Pafos Intl	1759292	0.464	0.404
256	Netherlands	EHEH	Eindhoven	2670269	0.375	0.401
257	Netherlands	EHBK	Maastricht	337347	0.360	0.400
258	Portugal	LPPR	Porto	6004500	0.454	0.400
259	Italy	LIBP	Pescara	545099	0.446	0.400
260	France	LFRN	St Jacques	431698	0.341	0.399
261	Poland	EPSC	Goleniow	244663	0.396	0.397
262	Poland	EPWA	Okecie	9352979	0.444	0.397
263	France	LFML	Provence	7223736	0.430	0.397
264	Austria	LOWG	Graz	959463	0.424	0.396
265	Finland	EFJY	Jyvaskyla	88823	0.396	0.393
266	France	LFQQ	Lesquin	1143242	0.356	0.389
267	France	LFRO	Lannion	35492	0.335	0.388
268	Italy	LIMZ	Levaldigi	220958	0.378	0.388
269	Croatia	LDZA	Zagreb	2262627	0.431	0.382
270	Portugal	LPPT	Lisboa	14806537	0.445	0.380
271	Spain	GCLP	Gran Canaria	10339466	0.436	0.380
272	Poland	EPLL	Reymont Airport	384063	0.399	0.380
273	Estonia	EETN	Tallinn	1907569	0.409	0.378
274	France	LFCR	Marcillac	139387	0.310	0.378
275	Norway	ENFL	Floro	122030	0.356	0.374
276	Luxembourg	ELLX	Luxembourg	1836920	0.325	0.373
277	Finland	EFKI	Kajaani	78071	0.372	0.372
278	Poland	EPRZ	Jasionka	491173	0.404	0.369
279	France	LFBA	La Garenne	34875	0.329	0.369
280	Poland	EPPO	Lawica	1416685	0.382	0.368
281	Poland	EPKK	Balice	2994359	0.411	0.363
282	Norway	ENTO	Torp	1338616	0.384	0.361
283	Romania	LRCL	Cluj Napoca	1004946	0.400	0.357
284	United Kingdom	EGPI	Islay	25784	0.280	0.356

285	Poland	EPBY	Bydgoszcz Ignacy Jan Paderewski Airport	276705	0.362	0.354
286	France	LFBL	Bellegarde	335111	0.286	0.352
287	Latvia	EVRA	Riga Intl	5098360	0.386	0.346
288	France	LFLW	Aurillac	26407	0.271	0.346
289	Poland	EPWR	Strachowice	1609014	0.367	0.343
290	United Kingdom	EGPB	Sumburgh	142612	0.284	0.342
291	Germany	EDDR	Saarbrucken	411473	0.298	0.341
292	France	LFBI	Biard	90713	0.281	0.340
293	Greece	LGKL	Kalamata	99451	0.402	0.339
294	Finland	EFKT	Kittila	237999	0.327	0.339
295	Spain	GCTS	Tenerife Sur	8507260	0.376	0.338
296	United Kingdom	EGNV	Durham Tees Valley Airport	190284	0.281	0.336
297	Slovenia	LJLJ	Ljubljana	1358792	0.366	0.335
298	Spain	GCRR	Lanzarote	5440041	0.356	0.334
299	Poland	EPGD	Lech Walesa	2456025	0.354	0.332
300	Italy	LIRZ	Perugia	171071	0.372	0.331
301	Norway	ENKR	Hoybuktmoen	295214	0.336	0.329
302	Portugal	LPPS	Porto Santo	108383	0.334	0.329
303	France	LFST	Entzheim	1065568	0.297	0.328
304	Lithuania	EYPA	Palanga Intl	111788	0.332	0.327
305	United Kingdom	EGPO	Stornoway	122439	0.252	0.318
306	France	LFKC	Saint Catherine	294352	0.316	0.317
307	Greece	LGPA	Paros	36271	0.361	0.316
308	Portugal	LPMA	Madeira	2312923	0.334	0.312
309	France	LFOH	Octeville	26956	0.272	0.308
310	Finland	EFKK	Kruunupyy	94555	0.313	0.307
311	Germany	ETNL	Laage	164227	0.299	0.299
312	Austria	LOWI	Innsbruck	995848	0.297	0.294
313	Netherlands	EHGG	Eelde	115553	0.283	0.290
314	Italy	LIPH	Treviso	1075319	0.308	0.289
315	United Kingdom	EGTE	Exeter	707414	0.254	0.287
316	Sweden	ESOK	Karlstad Airport	108784	0.284	0.283
317	Bulgaria	LBSF	Sofia	3465823	0.355	0.281
318	Greece	LGML	Milos	30351	0.319	0.279
319	Romania	LROP	Henri Coanda	5028201	0.358	0.276
320	Austria	LOWS	Salzburg	1695962	0.267	0.253
321	Spain	LEVT	Vitoria	23109	0.255	0.247
322	Lithuania	EYVI	Vilnius Intl	1709406	0.282	0.241
323	United Kingdom	EGBJ	Gloucestershire	14737	0.182	0.238
324	Greece	LGNX	Naxos	25783	0.282	0.231
325	Romania	LRTR	Traian Vuia	1227943	0.270	0.230
326	Greece	LGLM	Limnos	92952	0.252	0.225
327	United Kingdom	EGPC	Wick	24262	0.167	0.216
328	France	LFCK	Mazamet	37792	0.199	0.215
329	United Kingdom	EGPL	Benbecula	34240	0.143	0.201
330	France	LFRZ	Montoir	14112	0.165	0.196
331	France	LFLS	Saint Geoirs	335060	0.159	0.195
332	Czech Republic	LKPD	Pardubice	59034	0.203	0.191
333	United Kingdom	EGPA	Kirkwall	133930	0.138	0.175
334	Romania	LRSB	Sibiu	176876	0.204	0.175
335	United Kingdom	EGPU	Tiree	8310	0.121	0.171
336	Finland	EFMA	Mariehamn	53562	0.162	0.168
337	France	LFLB	Aix Les Bains	233420	0.131	0.165
338	United Kingdom	EGEC	Campbeltown Airport	9201	0.124	0.144
339	Czech Republic	LKKV	Karlovy Vary	96291	0.147	0.142
340	Greece	LGKC	Kithira	27391	0.160	0.138
341	Greece	LGIK	Ikaria	37534	0.166	0.136
342	Switzerland	LSZB	Bern Belp	169288	0.124	0.135

343	Finland	EFET	Enontekio	18238	0.128	0.135
344	Greece	LGST	Sitia	39604	0.157	0.126
345	France	LFHP	Loudes	7859	0.105	0.126
346	United Kingdom	EGPR	Barra Airport	10482	0.084	0.120
347	Greece	LGSO	Syros Airport	9872	0.134	0.108
348	Greece	LGLE	Leros	30906	0.123	0.102
349	Finland	EFSA	Savonlinna	14113	0.103	0.097
350	Greece	LGKJ	Kastelorizo	8723	0.114	0.094
351	Greece	LGPL	Astypalaia	13350	0.092	0.078
352	Greece	LGKY	Kalymnos Island	24249	0.102	0.071
353	Finland	EFVR	Varkaus	8671	0.070	0.066
354	Greece	LGSY	Skiros	4488	0.076	0.064
355	United Kingdom	EGMC	Southend	42401	0.058	0.063
356	France	LFBX	Bassillac	5942	0.047	0.056
357	France	LFGJ	Tavaux	2387	0.041	0.053
358	Greece	LGKS	Kasos	4866	0.043	0.039
359	United Kingdom	EGTK	Kidlington	1500	0.017	0.020

Appendix 12. Descriptive statistics of the Spanish airports data set

Table A12.1. Descriptive statistics of the Spanish airports data set, 2009-2010

Variable	Units	Minimum	Maximum	Median	Mean	Standard deviation	Zero values	Not available
2009								
PAX	number	34605	47943510	1309685	4968699.00	9272604.00	0	0
ATM	number	1917	427168	14435	47308.62	82494.08	0	0
Cargo	tonnes	0	330161	1784	16121.57	55616.94	2	0
Population100km	number	10162	6104502	1553328	1838295.00	1619292.00	0	0
Population200km	number	806708.9	9206386	5079862	4989138.00	2413455.00	0	0
Island	logical	0	1	0	0.32	0.47	25	0
RevenueTotal	thsd. EUR	359	590369	11897	50867.95	108750.80	0	0
EBITDA	thsd. EUR	-6190	219501	1808	18598.86	42358.91	0	0
NetProfit	thsd. EUR	-213087	28955	-4184	-7690.87	36617.98	0	0
DA	thsd. EUR	802	288864	5292	17510.49	49749.90	0	0
StaffCost	thsd. EUR	1045	52176	5221	9241.46	10713.84	0	0
RunwayCount	number	1	4	1	1.38	0.68	0	0
TerminalCount	number	1	4	1	1.19	0.57	0	0
RoutesDeparture	number	1	294	12	48.00	73.66	0	0
RoutesArrival	number	1	293	12	47.81	73.68	0	0
2010								
PAX	number	24527	49797630	1347612	5118363.00	9657971.00	0	0
ATM	number	1776	426941	12750	47211.27	82468.65	0	0
Cargo	tonnes	0	400477	1675	18459.24	67235.76	2	0
Population100km	number	10162	6104502	1553328	1838295.00	1619292.00	0	0
Population200km	number	806708.9	9206386	5079862	4989138.00	2413455.00	0	0
Island	logical	0	1	0	0.32	0.47	25	0
RevenueTotal	thsd. EUR	330	614076	11042	52462.59	113616.20	0	0
EBITDA	thsd. EUR	-6411	247171	1263	19401.19	45634.68	0	0
NetProfit	thsd. EUR	-127873	32117	-4642	-6063.27	26036.36	0	0
DA	thsd. EUR	791	275771	4745	19353.24	50320.56	0	0
StaffCost	thsd. EUR	1140	52810	5046	9418.84	10843.18	0	0
RunwayCount	number	1	4	1	1.35	0.68	0	0
TerminalCount	number	1	4	1	1.22	0.63	0	0
RoutesDeparture	number	1	294	12	48.00	73.66	0	0
RoutesArrival	number	1	293	12	47.81	73.68	0	0

Table A12.2. List of Spanish airports in the data set

ICAO	name	ICAO	name	ICAO	name	ICAO	name
GCFV	Fuerteventura	LEAM	Almeria	LELN	Leon Airport	LEVC	Valencia
GCGM	La Gomera Airport	LEAS	Asturias	LELO	Logrono-Agoncillo Airport	LEVD	Valladolid
GCHI	Hierro	LEBB	Bilbao	LEMD	Barajas	LEVT	Vitoria
GCLA	La Palma	LEBL	Barcelona	LEMG	Malaga	LEVX	Vigo
GCLP	Gran Canaria	LECO	A Coruna	LEMH	Menorca	LEXJ	Santander
GCRR	Lanzarote	LEGE	Girona	LEPA	Son Sant Joan	LEZG	Zaragoza Ab
GCTS	Tenerife Sur	LEGR	Granada	LEPP	Pamplona	LEZL	Sevilla
GCXO	Tenerife Norte	LEIB	Ibiza	LERS	Reus		
GEML	Melilla	LEJR	Jerez	LESA	Salamanca		
LEAL	Alicante	LELC	Murcia San Javier	LESO	San Sebastian		

Appendix 13. Descriptive statistics of PFP indicators' values of Spanish airports

Table A13.1. Descriptive statistics of PFP indicators' values of Spanish airports, 2010

	Min	Max	Median	Mean	Standard deviation
ATM per Route	293.5506	12244.0000	1396.2860	1978.6160	2109.4220
WLU per Route	2452700.0000	40074170.0000	8956421.0000	11527320.0000	8290138.0000
WLU per StaffCost	574.0373	94303.4300	27692.7700	32987.5600	24681.4400
Revenue per Route	330.0000	2957.0000	925.1259	1111.8190	703.2425
Revenue per PAX	0.0058	0.0782	0.0094	0.0115	0.0115
EBITDA per Route	-6411.0000	1238.8800	126.3000	-333.4339	1389.4790
Revenue per WLU	0.0001	0.0008	0.0001	0.0001	0.0001
EBITDA per WLU	-0.0017	0.0001	0.0000	-0.0001	0.0003
EBITDA per Revenue	-7.9030	0.5577	0.1087	-0.5421	1.7175
WLU per Population100km	1.1680	5084.8440	134.5388	869.5215	1447.3220
Revenue per Population100km	0.0002	0.5304	0.0124	0.0740	0.1253
EBITDA per Population100km	-0.2645	0.2658	0.0004	0.0158	0.0787

Appendix 14. Descriptive statistics of the UK airports data set

Table A14. 1. Descriptive statistics of the data set of UK airports, 2011-2012

Variable	Units	Minimum	Maximum	Median	Mean	Standard deviation	Zero values	Not available
2011								
PAX	number	1500	69388110	764508.5	5198920	12041540	0	0
ATM	number	70	476293	12387	47523.14	84572.38	0	0
Cargo	tonnes	0	1569303	269.5	59227.33	246116.6	7	0
Population100km	number	17988.54	19281820	4676217	6953842	6503658	0	0
Population200km	number	59015.81	45006490	18819340	18654350	14423620	0	0
Island	logical	0	1	0	0.12	0.33	37	0
RevenueTotal	thsd. EUR	2895	2456000	63559	224348.3	531260.70	0	21
RevenueAviation	thsd. EUR	5457	1379000	37186	133355.2	311857.50	0	23
RevenueNonAviation	thsd. EUR	0	1077000	30639	112482.2	245414.00	2	23
EBITDA	thsd. EUR	332	1207000	26376	98071.29	261649.90	0	21
DA	thsd. EUR	293	575000	9395	46891.19	124478.30	0	21
StaffCost	thsd. EUR	2735	314000	14169	44977.39	78298.67	0	24
StaffCount	number	77	5265	308	767.95	1232.04	0	22
RunwayCount	number	1	3	1	1.50	0.60	0	20
TerminalCount	number	1	4	1	1.38	0.80	0	21
RoutesDeparture	number	1	457	14	47.12	83.31	0	0
RoutesArrival	number	1	458	14	47.12	83.65	0	0
2012								
PAX	number	5903	69983470	694041	5236092	12163740	0	0
ATM	number	643	471452	12259.5	46926.83	83785.17	0	0
Cargo	tonnes	0	1555992	279.5	59324.90	244302.40	10	0
Population100km	number	17988.54	19281820	4676217	6953842	6503658	0	0
Population200km	number	59015.81	45006490	18819340	18654350	14423620	0	0
Island	logical	0	1	0	0.12	0.33	37	0
RevenueTotal	thsd. EUR	3128	2718000	84063.5	276823.3	630750.20	0	24
RevenueAviation	thsd. EUR	5503	1564000	51320	161938.4	371923.70	0	25
RevenueNonAviation	thsd. EUR	0	1154000	44602	129452.6	273782.80	1	25
EBITDA	thsd. EUR	-733	1237000	28197	113266.6	288813.40	0	24
DA	thsd. EUR	327	574000	11106.5	54210.94	133834.50	0	24
StaffCost	thsd. EUR	4151	356000	23224	61550.93	97296.57	0	28
StaffCount	number	121	5278	465.5	936.13	1336.20	0	26
RunwayCount	number	1	3	1	1.50	0.62	0	24
TerminalCount	number	1	4	1	1.44	0.86	0	24
RoutesDeparture	number	1	457	14	47.12	83.31	0	0
RoutesArrival	number	1	458	14	47.12	83.65	0	0

Table A14. 2. List of UK airports in the data set

ICAO	*name*	*ICAO*	*name*	*ICAO*	*name*	*ICAO*	*name*
EGAA	Belfast Intl	EGGW	Luton	EGNV	Durham Tees Valley Airport	EGPL	Benbecula
EGAC	Belfast City	EGHH	Bournemouth	EGNX	Nottingham East Midlands	EGPN	Dundee
EGAE	City of Derry	EGHI	Southampton	EGPA	Kirkwall	EGPO	Stornoway
EGBB	Birmingham	EGKK	Gatwick	EGPB	Sumburgh	EGPR	Barra Airport
EGBJ	Gloucestershire	EGLC	City	EGPC	Wick	EGPU	Tiree
EGCC	Manchester	EGLL	Heathrow	EGPD	Dyce	EGSH	Norwich
EGCN	Doncaster Sheffield Airport	EGMC	Southend	EGPE	Inverness	EGSS	Stansted
EGEC	Campbeltown Airport	EGNH	Blackpool	EGPF	Glasgow	EGTE	Exeter
EGFF	Cardiff	EGNJ	Humberside	EGPH	Edinburgh	EGTK	Kidlington
EGGD	Bristol	EGNM	Leeds Bradford	EGPI	Islay		
EGGP	Liverpool	EGNT	Newcastle	EGPK	Prestwick		

Appendix 15. Descriptive statistics of PFP indicators' values of UK airports

Table A15.1. Descriptive statistics of PFP indicators' values of UK airports, 2012

	Min	Max	Median	Mean	Standard deviation
ATM per Runway	3634	157151	51045	59291	44064
WLU per Runway	30848700	2333301000	426452900	644495800	657795600
PAX per Runway	308487	23327820	4262981	6444262	6576946
ATM per Route	322	3110	941	1117	620
WLU per Route	295150	15300340	5859828	6480708	4319772
PAX per Route	2952	152969	58598	64803	43193
PAX per Population100km	0.000	6.679	0.305	1.019	1.410
WLU per Population100km	0.032	667.958	30.465	101.887	141.035

Appendix 16. Descriptive statistics of the data set of Greek airports

Table A16.1. Descriptive statistics of the data set of Greek airports, 2007

Variable	Units	Minimum	Maximum	Median	Mean	Standard deviation	Zero values	Not available
APM	number	4139	16632800	145586	1079417.0	2845883	0	0
APM_winter	number	232	5230591	10994	215195.90	851852.70	0	0
APM_summer	number	2895	11402210	136755	864221.40	2047626	0	0
cargo	kg	0	121782100	43299	3797959.0	19474180	7	0
cargo_winter	kg	0	49524120	7114	1528956.0	7920955	10	0
cargo_summer	kg	0	72257990	24323	2269003.0	11553410	7	0
ATM	number	198	193123	2526	11671.64	31972.90	0	0
ATM_winter	number	10	68376	508	3166.21	11091.30	0	0
ATM_summer	number	114	124747	2070	8505.44	21129.58	0	0
openning_hours	hrs	1231.96	8760	2330.5	3605.27	2660.38	0	0
openning_hours_winter	hrs	549.46	3720	686.43	1334.85	1167.27	0	0
openning_hours_summer	hrs	682.5	5040	1626.29	2270.42	1574.32	0	0
runway_area	sq.m.	17750	351000	91350	93653.72	66025.22	0	0
terminal_area	sq.m.	100	238000	2150	13336.95	38890.03	0	0
parking_area	sq.m.	1500	900000	22100	60014.69	144894.70	0	0
island	logical	0	1	1	0.72	0.46	11	0
international	logical	0	1	0	0.38	0.49	24	0
mixed_use	logical	0	1	0	0.31	0.47	27	0
WLU	WLU	4139	17850620	150080	1117397	3027194	0	0
NearestCity	km	1	45	8	11.18	10.00	0	0

Table A16.2. List of Greek airports in the data set

ICAO	*name*	*ICAO*	*name*	*ICAO*	*name*	*ICAO*	*name*
LGBL	Aghialos	LGIK	Ikaria	LGKO	Kos	LGRP	Rhodes
LGPZ	Aktio	LGIO	Ioannina	LGKZ	Kozani	LGSM	Samos
LGAL	Alexandroupoli	LGKL	Kalamata	LGKC	Kythira	LGSR	Santorini
LGRX	Araxos	LGKY	Kalymnos	LGLE	Leros	LGST	Sitia
LGPL	Astypalaia	LGKP	Karpathos	LGLM	Limnos	LGSK	Skiathos
LGAV	Athens	LGKS	Kasos	LGML	Milos	LGSY	Skyros
LGSA	Chania	LGKJ	Kastelorizo	LGMK	Mykonos	LGSO	Syros
LGHI	Chios	LGKA	Kastoria	LGMT	Mytilini	LGTS	Thessaloniki
LGKR	Corfu	LGKV	Kavala	LGNX	Naxos	LGZA	Zakynthos
LGIR	Heraklion	LGKF	Kefalonia	LGPA	Paros		

Appendix 17. Descriptive statistics of PFP indicators' values of Greek airports

Table A17.1. Descriptive statistics of PFP indicators' values of Greek airports, 2007

		Min	*Max*	*Median*	*Mean*	*Standard deviation*
WLU per Runway Area	summer	2.5862	4140.9910	163.0222	591.5872	948.8032
WLU per Runway Area	winter	0.1724	1631.2910	29.1950	105.6822	270.4834
WLU per Hour	summer	0.8813	659.0991	22.6117	57.5533	114.7941
WLU per Hour	winter	0.1044	421.6992	4.4632	20.7435	67.9296
WLU per Terminal Area	summer	273.038	45201.280	4915.427	7454.086	8678.674
WLU per Terminal Area	winter	11.600	21446.290	782.568	1789.136	3515.624
ATM per Runway Area	summer	0.0009	0.3554	0.0275	0.0607	0.0796
ATM per Runway Area	winter	0.0001	0.1948	0.0096	0.0186	0.0324
ATM per Hour	summer	0.0004	0.0678	0.0033	0.0060	0.0111
ATM per Hour	winter	0.0000	0.0504	0.0019	0.0034	0.0080
ATM per Terminal Area	summer	0.0991	10.8800	0.6181	1.3785	2.1665
ATM per Terminal Area	winter	0.0050	5.0800	0.2271	0.5654	1.0854

Appendix 18. Model Greece estimation results

Table A18.1. Estimation results of the Model Greece specifications (summer)

Model		Intercept	log(OpeningHours)	log(RunwayArea)	log(TerminalArea)	Island	International	σ_v	σ_u	ρ_Y	ρ_v	ρ_u
OLS	Estimate	0.926	2.178	-0.447	0.551	-0.545	0.209	0.628				
	Std. Error	2.875	0.270	0.286	0.111	0.327	0.326					
	Sig.	0.749	0.000	0.128	0.000	0.105	0.526					
	Likelihood	-33.945										
SAR	Estimate	0.587	2.176	-0.432	0.546	-0.591	0.256			0.001		
	Std. Error	2.710	0.247	0.264	0.102	0.311	0.311			0.002		
	Sig.	0.828	$< 10^{-16}$	0.102	0.000	0.057	0.411			0.592		
	Likelihood	-33.801										
SEM	Estimate	0.926	2.138	-0.433	0.561	-0.475	0.251				-0.038	
	Std. Error	2.608	0.249	0.259	0.099	0.290	0.300				0.039	
	Sig.	0.723	$< 10^{-16}$	0.095	0.000	0.101	0.403				0.333	
	Likelihood	-33.656										
SF	Estimate	3.085	2.204	-0.558	0.512	-0.777	0.175	0.001	1.003			
	Std. Error	2.546	0.244	0.263	0.097	0.303	0.256	0.002	0.113			
	Sig.	0.226	$< 10^{-16}$	0.034	0.000	0.010	0.493	0.750	$<10^{-16}$			
	Likelihood	-28.456										
SSF (1,0,0,0)	Estimate	2.052	2.246	-0.420	0.418	-0.713	-0.006	0.005	1.005	0.001		
	Std. Error	1.1560	0.194	0.099	0.040	0.188	0.046	0.006	0.114	0.001		
	Sig.	0.076	$< 10^{-16}$	0.000	$< 10^{-16}$	0.001	0.898	0.433	$<10^{-16}$	0.750		
	Likelihood	-28.726										
SSF (0,0,1,0)	Estimate	0.999	2.232	-0.387	0.492	-0.547	0.077	0.299	0.835		-0.035	
	Std. Error	0.000	0.000	na	na	0.000	ns	0.000	0.000		na	
	Sig.	$< 10^{-16}$	$< 10^{-16}$			$<10^{-16}$		$<10^{-16}$	$<10^{-16}$			
	Likelihood	-32.716										
SSF (0,0,0,1)	Estimate	1.062	2.314	-0.426	0.465	-0.694	0.262	0.285	0.815			-0.005
	Std. Error	na	0.000	0.000	0.000	na	na	na	0.000			na
	Sig.		$< 10^{-16}$	$< 10^{-16}$	$< 10^{-16}$				$<10^{-16}$			
	Likelihood	-33.410										

Table A18.2. Estimation results of the Model Greece specifications (winter)

Model		Intercept	log(OpeningHours)	log(RunwayArea)	log(TerminalArea)	Island	International	σ_v	σ_u	ρ_Y	ρ_v	ρ_u
OLS	Estimate	0.305	2.318	-0.425	0.359	-0.016	-0.074	1.166				
	Std. Error	5.375	0.453	0.522	0.194	0.566	0.646					
	Sig.	0.955	0.000	0.421	0.073	0.977	0.910					
	Likelihood	-58.059										
SAR	Estimate	-1.117	2.258	-0.341	0.343	-0.239	0.181			0.005		
	Std. Error	4.921	0.408	0.471	0.174	0.530	0.606			0.003		
	Sig.	0.820	0.000	0.469	0.049	0.652	0.765			0.160		
	Likelihood	-57.072										
SEM	Estimate	0.763	2.472	-0.537	0.321	0.052	-0.056				-0.034	
	Std. Error	4.882	0.406	0.476	0.175	0.504	0.591				0.039	
	Sig.	0.876	0.000	0.260	0.067	0.918	0.925				0.679	
	Likelihood	-57.973										
SF	Estimate	0.083	2.691	-0.455	0.369	-0.441	-0.790	0.001	1.876			
	Std. Error	4.847	0.141	0.605	0.117	0.060	0.307	0.003	0.212			
	Sig.	0.986	$< 10^{-16}$	0.452	0.002	0.000	0.010	0.752	$<10^{-16}$			
	Likelihood	-52.882										
SSF (1,0,0,0)	Estimate	-2.121	2.595	-0.250	0.247	-0.547	-0.253	0.008	1.854	0.006		
	Std. Error	5.355	0.181	0.633	0.160	0.160	0.564	0.011	0.211	0.006		
	Sig.	0.692	$< 10^{-16}$	0.693	0.122	0.001	0.654	0.471	$<10^{-16}$	0.265		
	Likelihood	-52.569										
SSF (0,0,1,0)	Estimate	0.394	2.378	-0.391	0.361	0.031	-0.055	0.802	1.202		-0.004	
	Std. Error	0.001	na	na	0.000	na	0.000	na	na		0.000	
	Sig.	$< 10^{-16}$			$< 10^{-16}$		$<10^{-16}$				$<10^{-16}$	
	Likelihood	-57.737										
SSF (0,0,0,1)	Estimate	0.359	2.350	-0.390	0.373	0.032	-0.064	0.792	1.212			-0.026
	Std. Error	na	0.000	0.000	0.000	0.000	0.000	0.000	0.000			0.000
	Sig.		$<10^{-16}$	$<10^{-16}$	$<10^{-16}$	$<10^{-16}$	$<10^{-16}$	$<10^{-16}$	$<10^{-16}$			$<10^{-16}$
	Likelihood	-57.569										