TRANSPORT AND TELECOMMUNICATION INSTITUTE

Dmitry Pavlyuk

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

Summary of the promotion work
to obtain the scientific degree Doctor of Science in Engineering (Dr.sc.ing.)

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

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

RIGA – 2015

UDK 519.2:656

P 34

Transport and Telecommunication Institute

Pavlyuk D.

P 34 Study of European airports' efficiency on the basis of spatial stochastic frontier analysis: Summary of the promotion work. Riga: Transport and Telecommunication Institute, 2015.

In the submission kept the author's style and design.

ISBN 978-9984-818-75-7

THE PROMOTION WORK PRESENTED TO THE TRANSPORT AND TELECOMMUNICATION INSTITUTE TO OBTAIN THE SCIENTIFIC DEGREE – DOCTOR OF SCIENCE IN ENGINEERING (Dr.Sc.Ing.)

OFFICIAL REVIEWERS:

____
____
____

____
____
____

____
____
____

The defence of the thesis will be delivered in ________, 2015 at ___:__ o'clock at the Promotion Council of Transport and Telecommunication Institute on award of a scientific degree – doctor of science in engineering: 1, Lomonosova Street, conference hall 130, Riga, Latvia, tel. +371 67100594, fax: +371 67100535.

CONFIRMATION

I confirm that I developed the promotion work that is presented to the promotion council of Transport and Telecommunication Institute to obtain the scientific degree of Doctor of Science in Engineering. The promotion work has not ever been presented to any other promotional council to the scientific degree.

__________, 2015 D. Pavlyuk

The promotion work is written in English, it contains an introduction, 4 chapters, conclusions, 23 figures, 27 tables, 156 pages, and 18 appendixes. Bibliography contains 271 sources.

Contents

ANNOTATION 5
1. THE RELEVANCE OF THE PROBLEM AND MOTIVATION OF THE RESEARCH 6
2. THE GOAL AND TASKS OF THE RESEARCH 7
3. THE OBJECT AND SUBJECT OF THE RESEARCH 7
4. DEGREE OF THE THEME RESEARCH 8
5. STATEMENT OF THE PROBLEM 10
6. THE METHODOLOGY AND THE METHODS OF THE RESEARCH 10
7. SCIENTIFIC NOVELTY OF THE WORK 10
8. PRACTICAL VALUE AND REALIZATION OF THE WORK 11
9. APPROBATION OF THE RESEARCH 11
10. STRUCTURE OF THE THESIS 11
11. THESES WHICH ARE SUBMITTED FOR DEFENCE 12
12. SUMMARY OF THESIS CHAPTERS 13
12.1. Airport Benchmarking Methodologies and their Empirical Applications in Spatial Settings 13
12.2. Stochastic Frontier Analysis and a Problem of Spatial Effects Incorporation 18
12.3. Spatial Stochastic Frontier Model and its Parameters Estimation 23
12.4. Empirical Study of the European Airport Industry 29
13. CONCLUSION 38
14. PUBLICATIONS WITH AUTHOR'S PARTICIPATION 41

ANNOTATION

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.

1. THE RELEVANCE OF THE PROBLEM AND MOTIVATION OF THE RESEARCH

The legislative liberalisation process of the European air transportation market was completed in 1997. 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, 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.

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 to a private sector company, many European airports have become partly or completely private. 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.

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

- 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.

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, 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, 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 can be used for enhancing airport benchmarking procedures.

2. 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.

3. 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.

4. DEGREE OF THE THEME RESEARCH

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.

Since 1980-ties great efforts have been made in developing of performance measurements in the airport industry. 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, Gillen and Lall, Barros, Gitto and Mancuso, and Liebert, among others. A number of valuable reports in this area are also published: the Global Airport Performance Benchmarking Reports (2003-2011), the Airport Performance Indicators and Review of Airport Charges reports (2011-2013), the Airport Service Quality programme (2006-2015). Some local authorities, which control the airport sector, also provide their own benchmarking reports.

Different methodologies are utilised in literature for airport performance measurement: partial factor productivity (PFP) indicators, data envelopment analysis (DEA), stochastic frontier analysis (SFA). SFA, a methodological base of this research, is a popular frontier-based econometric approach to efficiency estimation. 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 for airport benchmarking, includes works of Pels (2003), Abrate and Erbetta (2007), Jing (2007), Barros (2008), Martin and Voltes (2009), Muller (2009), Malighetti, Martini(2009), and Scotti (2012).

Despite a large number of airport benchmarking studies, spatial effects are rarely included into consideration. Researches, conducted by Borins and Advani (2002), Jing (2007), Malighetti (2010) and Scotti (2011), and Adler and Liebert (2011), 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 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. Nevertheless, to the best of our knowledge, Ulku (2014) conducted the only application of these methods to analysis of airport productivity and efficiency.

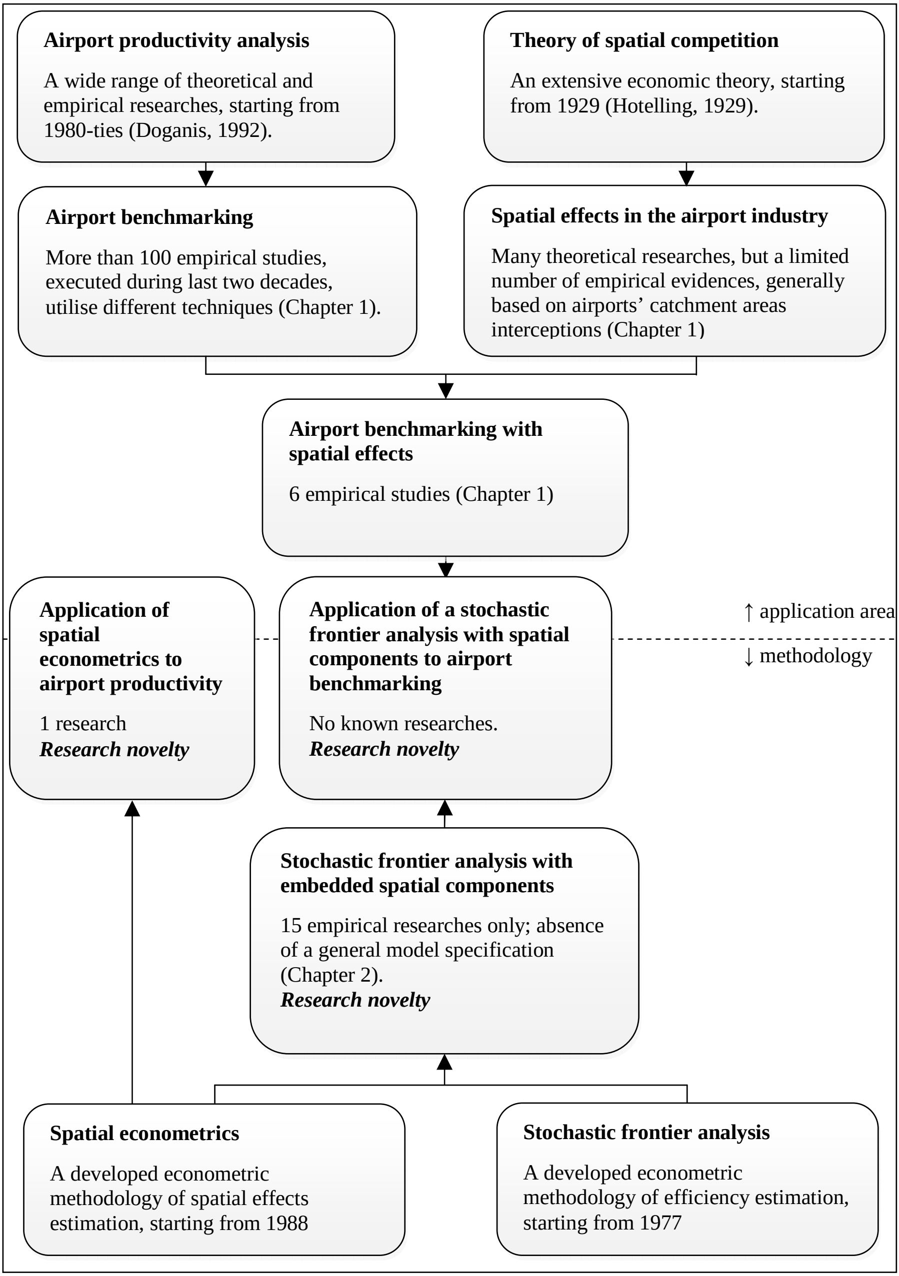

Fig. 1. Place of the study in a hierarchy of research areas

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 (2004), Fahr and Sunde (2005), Barrios (2007), Schettini (2007), Affuso (2010), Lin(2010), Areal(2010), Tonini and Pede (2011), Mastromarco (2012), Glass (2013), and Fusco and Vidoli (2013). All presented studies consider one particular type of possible spatial effects, so a general model specification seems to be necessary.

5. 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.

6. THE METHODOLOGY AND THE 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 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.

7. SCIENTIFIC NOVELTY OF THE WORK

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, spatial heterogeneity, 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. Statistical analysis 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.

8. PRACTICAL VALUE AND REALIZATION OF THE WORK

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.

9. APPROBATION OF THE RESEARCH

The results of research were reported at 8 scientific and research conferences in Latvia, Poland, and Russia. Also feedback on the developed *spfrontier* package has been received from the scientific community.

10. 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, 2011-2012), 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.

11. THESES WHICH ARE SUBMITTED FOR DEFENCE

The following points 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. 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.

12. SUMMARY OF THESIS CHAPTERS

12.1. Airport Benchmarking Methodologies and their Empirical Applications in Spatial Settings

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

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

Considering airports as commercial organisations, their activity result can be defined as a 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. 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. 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) for simpler comparison of airport productivity.

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 to make the resources comparable. 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.

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 most popular inputs and outputs of models, used in applied studies, in the Table 1 (a full list of analysed researches with inputs and outputs used can be found in the thesis).

Table 1.

Airport inputs and outputs, most frequently used in existing studies

Inputs		*Outputs*	
Indicator	*Number of researches (of 96)*	*Indicator*	*Number of researches (of 96)*
Employment	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

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.

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. If 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. 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 data sets with missed items is very under-researched area, so a complete data set becomes critically important.

Methodology of efficiency estimation

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.

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 partial factor productivity (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:

- 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 a 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, 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 2.

Table 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 (2011) and Hirschhausen and Culman (2005)

Methodologies, based on averaging of values, consider a relationship between weighted airport outputs and inputs. Total factor productivity 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. 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.

The most popular statistical model is a classical regression, which estimates a relationship between an expected value of a dependent variable and a set of explanatory variables. 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 statistical approach to frontier construction and efficiency estimation brought to development of a set of models. Stochastic frontier analysis (SFA) is one of the most popular approaches. The main strength of SFA is a statistical approach both to 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. SFA, rarely used for airports efficiency analysis before, recently became quite popular. Considering empirical applications, we note a growing academic interest to usage of the SFA approach to airports efficiency estimation, but at the same time a lack of studies with a heterogeneous frontier, which supposed to be a right choice for variegated environment of the airport industry.

Spatial heterogeneity in the airport industry

The majority of airport benchmarking methodologies are based on comparison between airports in a sample. Effective utilisation of such 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 local), 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. Endogenous heterogeneity in practice is frequently noticed as inefficiency, when exogenous is stated as a benchmarking difficulty. Discussing exogenous heterogeneity, Forsyth and Niemeier (2011) state that "a central problem of benchmarking is the heterogeneity of airports, which must be taken account". The importance of heterogeneity in airport benchmarking is widely acknowledged in literature.

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 (2008) and Liebert (2011) separately note the importance of unobserved heterogeneity for airport benchmarking.

In this research we focus on factors of spatial heterogeneity, related with airports' geographical locations. Spatial heterogeneity is based on uneven distribution of efficiency-related factors within a geographic area. Spatial heterogeneity can be partly represented in models by observed factors, but also latent accounting of unobserved factors is also technically possible. The main premise, which allows indirect including of airport heterogeneity into a model specification, is a similarity of unobserved spatial effects for neighbour airports.

Sources of spatial heterogeneity in the airport industry include:

- Natural sources: climate, landscape.
- Origin sources: demographic, economic and social conditions, labour market, population habits and local peculiarities.

- Destination sources: nearby touristic attractors, logistic centres, ports, and other cargo distribution units, local transport infrastructure, residential population of airport catchment area.
- Administrative and historical sources: common ownership, legislative environment, economic regulation.

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.

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. The theory of spatial competition is well established and there are a significant number of its applications in different economic areas. Empirical estimation of spatial competition among airports is weakly covered by researches.

Chapter conclusions

In this chapter we reviewed existing approaches to airport benchmarking. 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.

Finally, we note that a relationship between spatial effects and efficiency of airports is a weakly researched area.

12.2. Stochastic Frontier Analysis and a Problem of Spatial Effects Incorporation

Theoretical background of stochastic frontier analysis

A process of production in classical economics is defined as the usage of material and immaterial resources for making goods and services. 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 $x=(x_1, x_2, \ldots, x_K)$, to produce M outputs $y=(y_1, y_2, \ldots, y_M)$. 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. This set of input and output pairs is well known as a production possibility set and we will denote it by $PPS = \{x, y: x \text{ can produce } y\}$. The set of feasible outputs for an input vector can be defined as:

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

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. 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 (Debreu-Farrell definition):

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

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:

$$y = f(x) \cdot TE(x, y) = f(x) \cdot \exp(-u), u \ge 0\,, \quad (2)$$

where

$f(x)$ is a production function;

u is an inverse to the technical efficiency value, so it is frequently noticed as an inefficiency term.

Presence of random disturbances in practice is widely acknowledged and considered as a background for econometric analysis. Introducing the random disturbances v into the formula (2), we obtain a stochastic frontier (SF) model:

$$y = f(x) \cdot \exp(v) \cdot \exp(-u) \quad (3)$$

For econometric estimation of this model parameters 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:

$$y_i = f(x_i, \beta) \cdot \exp(v_i) \cdot \exp(-u_i) \tag{4}$$

In this form the model can be estimated using standard econometric techniques. A wide range of statistical methods can be applied to estimate parameters of the production frontier and inefficiency terms of the model (4). If distribution laws of v and u are 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.

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. Taking advantages and shortcoming of different distribution specifications, in this research we concentrate on the specification with the truncated normal distribution.

The composed error term of the SF model (4) is constructed as a difference of random variables with normal and truncated normal distributions: $\varepsilon_i = v_i - u_i$.

The density distribution function of a sum of normal and truncated normal distribution is well known. This distribution is known as an extended skew normal distribution, introduced by Azzalini (1985).

The log-likelihood function for the SF model with the truncated normal distribution of u and a sample of n observation is well known. Note that both random disturbances and inefficiency terms are supposed to be independent one from each other and for different sample observations.

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. 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:

- 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*).
- Exogenous effects represent a relationship between outputs of a given company and input of neighbour companies.
- Activity of a company can also be affected by area-specific factors. Many influencing factors are unevenly distributed over the space (spatial heterogeneity).

Potential problems of spatial effects in SFA were noted in early frontier researches. Farrell (1957) constructed a production frontier for 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. 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.

One of the most natural and theoretically well-grounded forms of this assumption is based on a known 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. 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.

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 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. 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 $\boldsymbol{W}$ are conventionally put to zeros to exclude self-dependency. An aggregate spatial influence of neighbour producers can be presented as a weighted sum of neighbour parameter values (a spatial lag):

$$[Wy]_i = \sum_{j=1}^{n} w_{ij} Y_j \tag{5}$$

A general spatial regression model is expressed in linear form as:

$$\begin{aligned} Y &= \rho_Y W_Y Y + X\beta + W_X X\beta^{(s)} + v, \\ v &= \rho_v W_v v + \tilde{v}, \end{aligned} \tag{6}$$

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

According to the general spatial model specification (6), output of producers $\boldsymbol{y}$ is influenced by its own inputs, spatially weighted outputs and inputs of neighbour producers (endogenous and exogenous spatial effects accordingly), and spatially related random disturbance (spatial heterogeneity).

Incorporating of spatial econometric principles into SFA is covered by a very limited number of researches; a review of these researches is presented in the body of thesis. 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.

Chapter 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. 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:

- 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.

- Theories of stochastic frontier analysis and spatial econometrics are very well developed, but there are almost no systematic researches on merging their principles. There is no general formulation of the stochastic frontier model with different types of spatial effects. This leads a significant number of private-case models, formulated and estimated by different researchers.
- 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.

12.3. Spatial Stochastic Frontier Model and its Parameters Estimation

Formal specification of the proposed SSF model

In the context of the SF model we specify a hypothesis about existence of the following four types of spatial effects:

Type 1. Endogenous spatial effects: a relationship between outputs (or, more generally, decisions) of an airport and outputs of its neighbours (spatial dependency).

Type 2. Exogenous spatial effects: a relationship between an output of an airport and inputs (resources) of its neighbours

Type 3. Spatially correlated random disturbances: uneven distribution of unobserved influencing factors over the space (spatial heterogeneity)

Type 4. Spatially related efficiency: a relationship between efficiency of neighbour airports.

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

$$Y_i = \rho_Y \sum_{j=1}^{n} w_{Y,ij} Y_j + \sum_{k=1}^{K} X_{ki}\beta_k + \sum_{k=1}^{K}\left(\beta_k^{(s)} \sum_{j=1}^{n} w_{X,ij} X_{kj}\right) + v_i - u_i, \tag{7}$$

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

where

i,j are company indexes, $i = 1,\ldots, n$, $j = 1,\ldots, n$,

Y_i is an output of a company i,

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

$w_{Y,ij}$, $w_{X,ij}$, $w_{v,ij}$, and $w_{u,ij}$ are spatial weights for spatial endogenous, exogenous effects, spatially correlated random disturbances and spatially related efficiency between companies *i* and *j* accordingly,
v_i and u_i are a random disturbance and inefficiency terms,
β_k are unknown coefficients, representing direct effects of inputs,
ρ_Y, $\beta^{(s)}{}_k$, ρ_v, and ρ_u are unknown coefficients for spatial effects,
$\tilde{v}_i$ and $\tilde{u}_i$ are IID random disturbances and inefficiency levels.

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

$$Y = \rho_Y W_Y Y + X\beta + W_X X\beta^{(s)} + v - u, \qquad (8)$$
$$v = \rho_v W_v v + \tilde{v}, u = \rho_u W_u u + \tilde{u}.$$

Considering possible ways of model generalisation, the model (8) 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 parameters in brackets represent orders of spatial autoregressive terms in a dependent variable, explanatory variables, random disturbances, and inefficiency terms respectively.

Classical estimators 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 assumptions about normal distribution of random disturbances and truncated normal distribution of inefficiency values. The distribution of the composed term of the SSF model 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 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 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:

$$\varepsilon \sim CSN_{n,n}(\mu', \Sigma', \Gamma', \nu', \Delta'), \tag{9}$$

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:

$$f_\varepsilon(\varepsilon) = [\Phi_n(0, -\mu, \Sigma_u)]^{-1} \Phi_n\left(-\Sigma_u(\Sigma_v + \Sigma_u)^{-1}(\varepsilon + \mu), -\mu, \left(\Sigma_v^{-1} + \Sigma_u^{-1}\right)^{-1}\right) \times \\ \times \varphi_n(\varepsilon, -\mu, \Sigma_v + \Sigma_u), \tag{10}$$

where φ_n and Φ_n are standard multivariate normal probability density and cumulative distribution functions.

Proof of the Theorem 1 can be found in the body of the thesis.

The Theorem 1 states that the composed error term of the SSF model has the closed skew normal distribution $CSN_{n,n}$ with the specified parameters. 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:

$$\begin{aligned} &\ln L\left(\beta, \beta^{(s)}, \sigma^2_{\tilde{v}}, \sigma^2_{\tilde{u}}, \mu, \rho_Y, \rho_v, \rho_u\right) = -\ln \Phi_n(0, -\mu, \Sigma_u) + \\ &+ \ln \Phi_n\left(-\Sigma_u(\Sigma_v + \Sigma_u)^{-1}(e + \mu), -\mu, \left(\Sigma_v^{-1} + \Sigma_u^{-1}\right)^{-1}\right) + \\ &+ \ln \varphi_n(e, -\mu, \Sigma_v + \Sigma_u), \\ &e = Y - \rho_Y W_Y Y - X\beta - W_X X\beta^{(s)}, \\ &\Sigma_v = \sigma^2_{\tilde{v}} \left((I_n - \rho_v W_v)^{-1}\right)^T (I_n - \rho_v W_v)^{-1}, \\ &\Sigma_u = \sigma^2_{\tilde{u}}, \left((I_n - \rho_u W_u)^{-1}\right)^T (I_n - \rho_u W_u)^{-1}. \end{aligned} \tag{11}$$

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

The second estimation goal 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. The conditional probability density function, derived in the thesis, matches the multivariate truncated normal probability density function, so

$$u|\varepsilon \sim MVTN_{0,+\infty}\left(\mu_{u|\varepsilon}, \Sigma_{u|\varepsilon}\right), \tag{12}$$

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

Given the conditional distribution of *u*, a vector of point estimates $\hat{u}$ can be found as a conditional expected value. Given distribution parameters, theoretical moments of the multivariate truncated normal distribution are well-known.

Implementation of the MLE of the SSF model parameters

Implementation of the proposed MLE of the SSF model parameters requires a set of functions (multivariate normal and truncated normal density and distribution functions, calculation of moments for multivariate truncated normal random variables, optimisation algorithms for MLE implementation), which are well-known in theory, but computationally hard. R is a freely available environment for statistical computing, where all of the required core algorithms are implemented. 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. 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; 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 online. The package is also enhanced with demo files and simulation tests.

There are several critical aspects of the MLE implementation, described in the chapter:

- Selection of the initial parameter values, which is extremely important for numeric maximisation of the non-convex likelihood function.
- Re-parameterisation of the likelihood function requires for making it smoother and computationally easier.
- Estimation of parameter estimates' variance-covariance matrix, which is required for statistical hypothesis testing. Complications, related with necessary backward re-parameterisation, are resolved using the sandwich estimator of the variance-covariance matrix.

Validation of the proposed MLE for the SSF model

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. Estimates, calculated by *frontier* and *spfrontier* packages for a model with half-normal and truncated normal inefficiencies, are matched perfectly.

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:

$$
\begin{aligned}
&DGP\left(\rho_Y^*, \rho_v^*, \rho_u^*, \mu^*\right): \\
&Y^* = \left(I_n - \rho_Y^* W_Y\right)^{-1} \times \\
&\qquad \times \left(5 + 10\log\left(X^*\right) + \log\left(X^*\right)^2 + \left(I_n - \rho_v^* W_v\right)^{-1}\tilde{v} - \left(I_n - \rho_u^* W_u\right)^{-1}\tilde{u}\right), \\
&X^* \sim U(1,10), \tilde{v} \sim N\left(0, 0.5^2\right), \tilde{u} \sim N\left(\mu^*, 2.5^2\right).
\end{aligned} \tag{13}
$$

The simulated production function is monotonic increasing and concave over the specified range of the input. Artificial rook- and queen-style spatial weight matrixes are used for spatial weights W_v W_Y and W_u specification.

A list of executed simulation experiments:

- SimE1: DGP(0,0,0,0), estimator SSF(0,0,0,0), half-normal
- SimE2: DGP(1,0,0,0), estimator SSF(0,0,0,0), truncated normal
- SimE3: DGP(0,0.2,0,0), estimator SSF(1,0,0,0), half-normal
- SimE3b: DGP(0,0.2,0,0), estimator SSF(0,0,0,0), half-normal
- SimE4: DGP(1,0.2,0,0), estimator SSF(1,0,0,0), truncated normal
- SimE5: DGP(0,0,0.4,0), estimator SSF(0,0,1,0), half-normal
- SimE5b: DGP(0,0,0.4,0), estimator SSF(1,0,0,0), half-normal
- SimE6: DGP(0,0,0,0.4), estimator SSF(0,0,0,1), half-normal
- SimE6b: DGP(0,0,0,0.4), estimator SSF(1,0,0,0), half-normal

Every experiment was executed 100 times for samples of 50, 100, 200, and 300 objects.

The estimator validity is measured using absolute and relative bias of estimates; standard deviation and root-mean-square deviation of estimates; estimate's confidence intervals to test estimate convergence to parameter's true value for larger samples; kernel density estimation of estimates' empirical probability density functions.

Spatially related efficiency is one of the key components of the introduced SSF model, so we 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. Estimates of the DGP parameters are unbiased 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.

Empirical kernel density of ρ_u and σ_u estimates is presented on the Fig. 2:

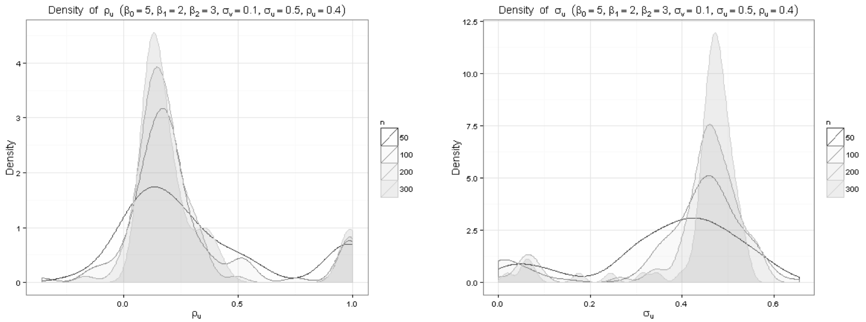

Fig. 2. Empirical kernel density plots for ρ_u and σ_u parameter estimates in SimE6

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

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

- The developed estimator provides unbiased and consistent estimates for classical non-spatial specifications of the stochastic frontier model (experiments SimE1 and SimE2).
- 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).
- Some parameters of the spatial stochastic frontier models are weakly identified and can be distinguished from each other (experiments SimE2, SimE4, SimE5b, and SimE6b).
- 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.

Chapter 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.

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*. This 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.

12.4. Empirical Study of the European Airport Industry

Taking features of airports data, discussed in the chapter 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.
- Geographical completeness of a dataset, so all neighbour airports are presented in the dataset.
- Availability of individual (disaggregated) airport data.

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. Collected data is received from public data sources only; no private information is used.

Four research 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.

All collected data sets are publicly available as a part of the *spfrontier* package, developed by the author.

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.
- Economic PFP indexes: WLU per employee cost; Revenue per WLU/ATM; EBITDA per WLU/ATM; EBIDTA per revenue.

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

The general specification of the SSF model, proposed in the chapter 3, was utilised for frontier estimation. A set of analysed private cases of the SSF model includes classical regression model (OLS), spatial autoregressive (SAR) and spatial error (SEM) models, non-spatial stochastic frontier (SF) model, and a set of private cases of SSF(1,0,1,1) model with selected spatial effects.

An inheritance diagram of the evaluated models is presented on the Fig. 3.

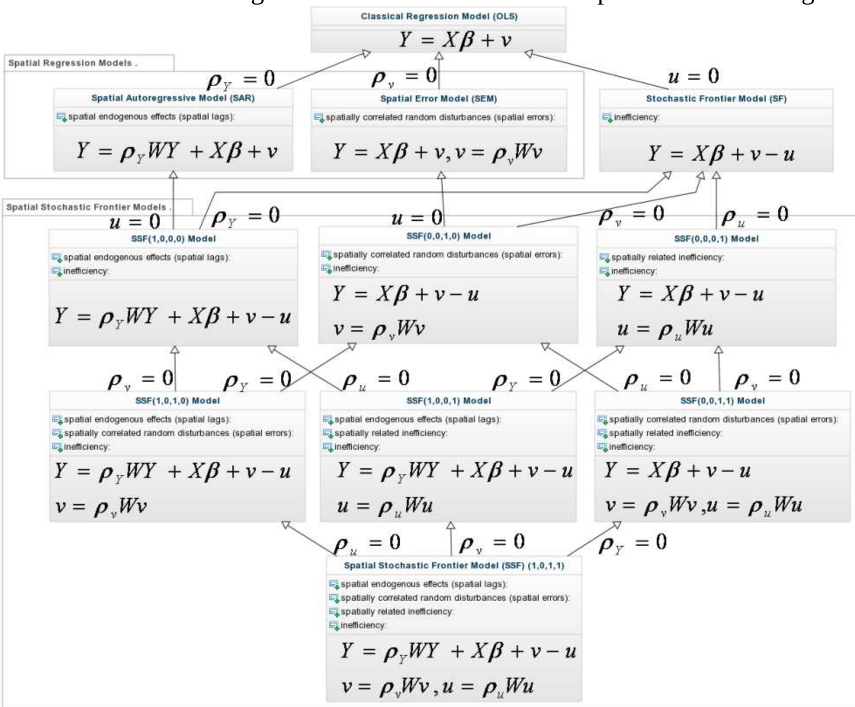

Fig. 3. Inheritance diagram of the research models

Empirical analysis of European airports data set

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.

As financial information is not available in the data set, this research is limited with the intermediary approach to airport activity. 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. Within this research, we decided to use a total number of routes (both arrival and departure) as a proxy for all infrastructure units.

The primary goal of this research is to discover possible spatial patterns in airport benchmarking. The Table 3 contains results of testing for spatial autocorrelation between all considered PFP indicators.

Table 3.

Spatial autocorrelation of PFP indicators of European airports

	Moran's I		*Geary's C*		*Mantel*	
	Coef.	*p-value*	*Coef.*	*p-value*	*Coef.*	*p-value*
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

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. Discovered spatial correlation of PAX per capita in 100 km is the most expected as population is unevenly distributed over Europe.

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, allowed getting over this problem, separately identifying different types of spatial effects and enhancing the results.

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

- Single-output (PAX) frontier (Model Europe1).
- Multi-output frontier with two outputs: PAX and Cargo (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:

$$\log(PAX) = \beta_0 + \rho_Y W \log(PAX) + \beta_1 \log(Routes) + \\ + \beta_2 \log(Population100km) + \beta_3 \log(GDPpc) \quad (14)$$

Ten different specifications of the model, presented on the Fig. 3, are estimated and analysed. 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.

Presence of inefficiency in data is supported by negatively skewed OLS residuals (skewness equals to –0.659) and statistically significant σ_u of the estimated classical SF model. Also spatial effects between OLS residuals are discovered using Moran's I and Lagrange multiplier tests.

Both classical SAR and SEM models provided statistically significant estimates of their specific types of spatial effects.

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. Selection of a model specification, which optimally fits the data, is based on the calculated values of a log-likelihood function. We selected the SSF(1,0,1,0) as the best model specification, and used this for further analysis. This model estimates are presented in the Table 4.

Table 4.

Estimation results of the Model Europe1 parameters

Model		*Intercept*	*log(Population100km)*	*log(Routes)*	*log(GDPpc)*	σ_v	σ_u	ρ_Y	ρ_v
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

All estimated SSF model specifications with spatial endogenous effects 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.

Significant spatial correlations of random disturbances are also discovered in all corresponding model specifications. 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 spatial heterogeneity.

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

Frontier parameter estimates (which are values of elasticity for 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*), also match the 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 conclude a significant level of inefficiency in data: a sample mean of efficiency is 0.479, sample median is 0.502. An important point of this research in context of the SSF model development is comparison of efficiency estimates, provide by classical SF and SSF models. The SSF model provided lower efficiency values for relatively isolated airports (Greek, Italian), and higher values for French and UK airports.

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 Model Europe2 is formulated as:

$$\begin{aligned} -\log(PAX) = \beta_0 + \rho_Y W \log(PAX) + \beta_1 \log(Cargo/PAX) + \\ + \beta_2 \log(Routes) + \beta_3 \log(Population100km) + \\ + \beta_4 \log(GDPpc) \end{aligned} \tag{15}$$

Note that 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. The chapter 2 contains a detailed description of the multi-output frontier specification.

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. Exclusion of the *log(Cargo/PAX)* reduces Model Europe2 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.

Empirical analysis of Spanish airports

The data set includes traffic, infrastructure and financial information about Spanish airports in 2009-2010. The Spanish airport industry is fairly monopolistic; all 48 commercial airports in Spain are managed by a public company AENA, dependent on the Ministry of Transports. 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. Financial data is available for 38 airports and is supplemented with traffic and infrastructure data, collected from the Eurostat and Open Flights databases.

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.

A final frontier specification, which was selected for presentation in this thesis, is formalised using the Cobb-Douglass function and has the following appearance (Model Spain):

$$\log(Revenue) = \beta_0 + \rho_Y W \log(Revenue) + \beta_1 \log(PAX) + \\ + \beta_2 \log(TerminalCount) + \beta_3 \log(Population100km) \tag{16}$$

Despite the initial assumption about presence of inefficiency in data, supported by previous researches, the simple OLS specification is almost perfect for the Model Spain: $R^2_{adj} = 0.9602$. 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 and likely use similar principles for traffic handling and revenue generation, so 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. 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 demonstrates a slightly better statistical performance and better matches our expectations, so the SEM model is selected as the best specification. The SEM model estimation results are presented in the Table 5.

Table 5.

Estimation results of the SEM Model Spain

Model		*Intercept*	*log(PAX)*	*log(TerminalCount)*	*log(Population100km)*	ρ_v
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

Results of the SEM model are generally expected. All three inputs have positive significant elasticity values (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.

Empirical analysis of UK airports

The UK data set includes traffic, infrastructure and financial information about 48 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.

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.

Although spatial autocorrelation is significant for some indicators, generally spatial effects are weak. The only indicator with highly significant spatial autocorrelation is WLU per route, representing efficiency of infrastructure usage.

The selected specification of the frontier has a standard Cobb-Douglass functional form (Model UK):

$$\begin{aligned}\log(PAX) = \beta_0 + \rho_Y W \log(PAX) + \beta_1 \log(Routes) + \\ + \beta_2 \log(Population100km) + \beta_3 Island\end{aligned} \tag{17}$$

The research hypothesis about inefficiency is supported by a statistically significant estimate of inefficiency standard deviation σ_u, provided by the classical SF and different SSF model specifications.

Presence of spatial effects in data is not so obvious. 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. The best SSF model specification is SSF(1,0,0,0); its estimates are presented in the Table 6.

Table 6.

Estimation results of the SSF(1,0,0,0) Model UK

Model		*Intercept*	*log(Routes)*	*log(Population100km)*	*Island*	σ_v	σ_u	ρ_Y
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

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.

Empirical analysis of Greek airports

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, who applied DEA methodology to analysis of Greek airports' efficiency. Whereas there are significant seasonal demand variations in the Greek airport industry, data on passengers, cargos, flights and operating hours are separated into summer and winter 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. All Greek airports, except of the international airport of Athens, are state-owned.

PFP index values 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. The general conclusion is a complete absence of statistically significant spatial effects both for winter and summer PFP index values.

A selected specification of the frontier is formulated as (Model Greece):

$$\begin{aligned}\log(WLU) = {} & \beta_0 + \rho_Y W \log(WLU) + \beta_1 \log(OpenningHours) + \\ & + \beta_2 \log(RunwayArea) + \beta_3 \log(TerminalArea) + \\ & + \beta_4 Island + \beta_5 International\end{aligned} \tag{18}$$

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 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 (Table 7).

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.

Table 7.

Estimation results of the SF Model Greece

Model		*Intercept*	*log(OpeningHours)*	*log(RunwayArea)*	*log(TerminalArea)*	*Island*	*International*	σ_v	σ_u
Summer									
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
Winter									
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

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.

Chapter 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 using alternative SSF model specifications. 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, Spanish airports data set, UK airports data set, and Greek airports data set. Every data set has its own specifics, affecting the presence of inefficiency and spatial effects in data.

General conclusions on the data sets are:

- ***European airports***: presence of significant inefficiency, spatial endogenous effects, and spatially correlated random disturbances.
- ***Spanish airports***: presence of spatial heterogeneity and absence of inefficiency components (explained by the frontier methodology limitation) and other types of spatial effects.

- ***UK airports***: presence of significant endogenous spatial effects (spatial competition) and inefficiency in data.
- ***Greek airports:*** presence of inefficiency and absence of all types of spatial effects in data.

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.

13. CONCLUSION

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); 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.

Further research directions can be mentioned:

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.
   - 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.
   - 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.
   - 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.

14. PUBLICATIONS WITH AUTHOR'S PARTICIPATION

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, Transport and Telecommunication, 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.