DISSERTATIONES RERUM OECONOMICARUM UNIVERSITATIS TARTUENSIS
59

TARMO PUOLOKAINEN

Public Agencies' Performance Benchmarking in the Case of Demand Uncertainty with an Application to Estonian, Finnish and Swedish Fire and Rescue Services

DISSERTATIONES RERUM OECONOMICARUM
UNIVERSITATIS TARTUENSIS

59

DISSERTATIONES RERUM OECONOMICARUM
UNIVERSITATIS TARTUENSIS
59

TARMO PUOLOKAINEN

Public Agencies' Performance Benchmarking in the Case of Demand Uncertainty with an Application to Estonian, Finnish and Swedish Fire and Rescue Services

UNIVERSITY OF TARTU
Press

School of Economics and Business Administration, University of Tartu, Estonia

The dissertation is accepted for the defence of the degree of Doctor Philosophiae (in Economics) on January 11th, 2018 by the Council of the School of Economics and Business Administration, University of Tartu.

Supervisors: Professor *emeritus* Janno Reiljan (PhD), University of Tartu, Estonia
Professor Dr. Dr. h.c. (*emeritus*) Peter Friedrich (PhD), University of Tartu, Estonia
Professor Christopher J. O'Donnell (PhD), the University of Queensland, Australia

Opponents: Professor Geert Bouckaert (PhD), Katholieke Universiteit Leuven, Belgium
Ants-Hannes Viira (PhD), Estonian University of of Life Sciences, Estonia

The public defence of the dissertation is on February 15th, 2018 at 11.00 in room 124, J. Liivi 4, University of Tartu

The publication of this dissertation is granted by the Doctoral School of Economics and Innovation of University of Tartu created under the auspices of European Social Fund and by University of Tartu School of Economics and Business Administration.

ISSN 1406-1309
ISBN 978-9949-77-666-5 (print)
ISBN 978-9949-77-667-2 (pdf)

Tartu University Press
www.tyk.ee

Contents

List of Figures

List of Tables

List of Abbreviations

BSC – Balanced Scorecard
CE – Cost efficiency
CES – Constant elasticity of substitution
CET – Constant elasticity of transformation
COLS – Corrected ordinary least squares
CPA – Comprehensive Performance Assessment
DEA – Data envelopment analysis
DFA – Deterministic frontier analysis
DMU – Decision-making unit
EMS – Emergency medical service
EU model – Expected utility model
FDH – Free disposal hull
FRS – Fire and rescue services
GIS – Geographic information system
IDF – Input distance function
ITE – Input-oriented technical efficiency
KPI – Key Performance Indicator
MBO – Management by objectives
ML – Maximum likelihood
NGO – Non-governmental organisation
NLS – Non-linear least squares
NPM – New Public Management
ODF – Output distance function
OPA – 'Old' Public Administration
OTE – Output-oriented technical efficiency
OTME – Output-oriented technical and mix efficiency
PMS – Performance measurement system
SFA – Stochastic frontier analysis
SUR – Seemingly unrelated regression

List of Author's Publications

I Articles related to the thesis

1. **Puolokainen, T.** (2017). Reforming fire and rescue services: a comparative study of Estonia and Georgia. *International Journal of Public Sector Management*, 30(3): 227–240.

2. **Puolokainen, T.**, Jaansoo, A. and Klaos, M. (Under review, 2018). Outsourcing rescue services in remote areas of Estonia.

3. Reiljan, J., **Puolokainen, T.**, and Ülper, A. (2015a). Complex benchmarking and efficiency measurement of public agencies: the case of Estonian rescue service brigades. *Public Finance and Management*, 15(2):130–152.

II Other publications

1. Espenberg, K. and **Puolokainen, T.** (2013). Etniline sooline palgalõhe Eestis (in Estonian; Ethnical gender wage gap in Estonia). *Ariadne Lõng*, 1/2:19–34.

2. **Puolokainen, T.** and Jaansoo, A. (2018). Bureaucracy and Service Delivery. *Global Encyclopedia of Public Administration, Public Policy, and Governance*, (Farazmand, A. (ed.)), pp. 1–6.

3. Reiljan, J., Paltser, I., and **Puolokainen, T.** (2016). The influence of research & development policy on business-sector performance: the Baltic states in the EU. *Public Finance and Management*, 16(3):192–225.

4. Reiljan, J. and **Puolokainen, T.** (2016). Riigiasutuse töö tõhususe hindamine: Keeruline, väga keeruline! (in Estonian; Evaluating the efficiency of a public agency: difficult, very difficult!). *Akadeemia*, 2016(10–11).

5. Reiljan, J., Ülper, A., and **Puolokainen, T.** (2012). Demand side factors of local public services in Estonia. *Discussions on Estonian Economic Policy: Theory and Practice of Economic Policy*, 20(1):147–164.

6. Reiljan, J., Ülper, A., and **Puolokainen, T.** (2015b). The impact of demand factors on the supply of local public services in Estonian municipalities. *Public Finance and Management*, 15(1):24–46.

INTRODUCTION

Motivation for the research

Public administrative units (public agencies) attempt to achieve public goals by providing goods to other economic units (e.g. private households, private firms, public firms, public administrative units). They have a public owner, they possess the long-term stock of production factors and their management is competent regarding the essential decisions related to production and delivery. They comprise legally dependent institutions (gross public offices) fully integrated into the budget planning (e.g. ministries, directories, courts, parliaments) but also public enterprises (Eichhorn and Friedrich 1976).

Considering the public administrative units in the wide meaning, the main task for them would usually be to design the goal achieving supply (characterised by production functions) which would balance the demands of society (welfare functions, goal achievement). The volume, quality, structure as well as the potential to provide the service has to be taken into account, while achieving the highest possible performance.

Thornhill (2006) emphasised the importance of public agencies. First, the public agencies are big employers. Second, the public agencies combined are a major provider of services in the society (and therefore affect the cost of inputs). Third, the public agencies are consumers of tax resources. Fourth, the public agencies (and governments) are mainly responsible for the survival of society. Altogether, the changes in public agencies have a significant influence on the economy and hence are objects of analysing how to improve their performance.

The issue of performance of the public agencies is very topical in Estonia as well as elsewhere. As the expectations of citizens are rising but the increase in resources for the public sector cannot keep up the same pace, public sector performance management can be seen as a possible solution (Van Dooren *et al.* 2015). Many countries face the new tendencies of developments within society – the problem of rapid ageing of population, the shrinking supply of labour force and ever scarcer public sector budget income sources (taxation, fees, public debts and donations). This all restricts available policy alternatives, which makes it more difficult to raise or even keep the current living standard. In these conditions a need to improve performance of public agencies will be a key economic policy objective. Economic growth and rising living standards depend on better performance (see e.g. Linna *et al.* 2010). Performance is closely related to terms productivity, efficiency and effectiveness. Productivity ratio can be defined as a ratio of outputs to inputs, where larger values of this ratio can be associated with better performance (Coelli *et al.* 2005). Based on Neely *et al.* (1995), effectiveness refers to the extent to which citizens requirements are met, while efficiency is a measure of how economically the public agencies resources are utilised when providing a given level of services. Production theory has developed a clearly defined concept of efficiency. According to Rasmussen (2013: 61), efficiency can be defined as the achieved compared to what can be achieved. One implication of this concept is that a unit which is more efficient than another is able to produce more outputs with a given level of inputs in the same technology setting (or produce the same level of outputs with lesser inputs). This implication makes the concept of efficiency very appealing for decision makers and politicians, because improving efficiency allows increasing societal gains without extra funds (Schubert 2009).

Discussions on what are public agencies' goals and outputs have taken place since the 17th century. However, the interest in the measurement and evaluation of the performance of the public agencies has intensified notably by the spread of the New Public Management (NPM) approach in the practice since the 1980s (for a description of the approach, see Hood 1991; Osborne and Gaebler 1993; Dunleavy and Hood 1994). The predomination of oversimplistic business methods, mainly caused by not considering the peculiarities of public agencies, resulted in some setbacks and cooling of the first euphoria of the emergence of the NPM paradigm. The recent developments in the field and additional available data have given an expectation that performance measurement can be a useful tool to strengthen the public agencies endeavours towards the improvement of performance. Measuring the performance in the public agencies is an increasingly important tool for public sector management.

The performance measurement has the goal to assess the public management's contribution to achieving the targets of the agency. It is therefore very important to assess the public service provision and whether the public agency copes with the tasks assigned to them optimally, i.e. balancing the performance with set constraints. The socially-acceptable performance can be regarded as doing the right things in the right way (Fried *et al.* 2008: vii). A good performance also indicates that the public goals are achieved: in the case of the Estonian Rescue Board the goal of their activity is to shape and maintain a safe environment, prevent risks and assist in case of emergency operatively and professionally (Rescue Act (Päästeseadus): §2).

Unlike the private sector with main objective to maximise the revenue or profit, the difficulty for public agency lies in shaping the performance goal. Using the finances allocated by the ministry via tax and other public budget income, the public agency has to achieve the optimal volume, structure, and quality of services provided. This is a complex problem because in many cases (e.g. the fire and rescue services) the demand of other economic units (e.g. households, private firms, public agencies) and individuals is uncertain at the time of planning. Thus it has to be indirectly determined by principals (e.g. politicians or government officials decide the capability of service provision through resource allocation). The consumers who receive the services do not shape the demand directly through willingness to pay for services. Often, they are not charged for the service as they are merit services (Musgrave 1959). A supply-demand balance is not achieved through market processes. Therefore, it is important to develop and to strengthen the theoretical basis of defining and assessing the optimality in supply of public services while considering the demand uncertainty.

It should be taken into account that the tasks of public sector institutions like public agencies are diverse. They differ dominantly from those of the private sector (firms and households) (Dixit 2002). Two aspects should be stressed here: a public agency has many stakeholders (e.g. taxpayers, consumers, politicians, other economic units) influencing public agencies' goals and hence in comparison to private economic units qualitatively different and comprehensive expectations to its outputs. Apart from the market-driven price signals, society needs other methods to indicate the performance of public agencies (such as welfare theory and social accounting, see Eerma and Friedrich 2016). One option is to develop a performance indicator system which characterises the public services provision processes parameters and outcomes, and assist finding out where it is necessary to make changes to achieve the desired goals. However, different stakeholders have various interests concerning the processes and outcomes, which might be conflicting. This in turn will greatly complicate the assessment of the performance of the public agency compared to profit-oriented companies.

Generally, all kinds of performance indicators are used to evaluate the performance of a public agency in case there is no market price that characterises the demand and the change in demand. On the choice of performance indicators (Schacter 1999; Bird *et al.* 2005), however, it is crucial to understand what kind of performance aspect the indicator characterises and whose interests will primarily be served through the selected indicators. No less important are the stakeholders, who choose the performance indicators, as the results of choice represent their interests. Choosing and measuring the performance indicators is the basis of performance assessment and chosen indicators have to cover all aspects of a public agency's performance. Thus, much effort is needed to create a comprehensive, complete and balanced system of performance indicators. One has to consider the volume, quality and structure of the public service provision; the time of production and delivery[1]; the effectiveness of service provision (whether the readiness and activities are in accord with the demands of society), the input-output ratios (to assess productivity and efficiency). Four key aspects, on which performance measurement focuses on are: what to measure, how to measure, how to interpret the measured results (depends on the evaluator), how to use performance measurements to improve decision-making (which depends on

[1]There are different solutions for public agencies: the recipients receive a service at the location of the public agency; the service is brought to the recipient or the recipient is served at his location.

the user of the results)? In the theoretical literature (see discussion in Van Thiel and Leeuw 2002) it is a common belief that the public agencies performance can be measured, and the measurement results can be applied to improve the performance management. However, there are only a few empirical analyses that confirm (or reject) this statement (Boyne *et al.* 2006: 4), meaning there are not many studies, that analyse the impact of actually introducing performance measurement to the decision-making. The current dissertation discusses the usage of performance measurement as a management tool.

The performance assessment has to support the performance management (for different management concepts the performance management is of different importance). According to different goals, the assessment results in different systems of indicators. When performance indicators are relevant, they should be determined quantitatively and qualitatively for inputs, outputs, processes and results in the past for one economic unit. Performance management, as the next step, is the implementation of the results of performance assessments on making management decisions to improve some aspect of the performance. The management of public agency has to take into account the interests of all stakeholders who determine the kind of goals to be achieved. Introduction of adequate assessments (complete and balanced indicators' system) in the public management has to be co-ordinated with political processes forming a comprehensive environment (legal, financial, etc.) for the public agency's management. The solution to these problems requires the deepening of theoretical basics for forming an adequate indicator system and is one research task of this doctoral thesis.

If an incomplete or unbalanced indicator system is used, an inadequate assessment is conducted and the results may influence the behaviour of the public agency to an undesired direction (the goals and their indicators can also be used for negotiation, compensation, threats, etc.; see Eichhorn and Friedrich 1976; Smith 1990; Smith and Goddard 2002). It is difficult to explore the interest and performance aspects of all stakeholders concerning public agencies and how to balance these interests. Problems related to completing and balancing the performance assessment of public agencies have not yet been fully addressed, so therefore has this thesis a general theoretical significance.

There are different ways to identify the relevant performance indicators. Some methods are used to identify desired goals of different stakeholders by considering laws, statutes, parliamentary decisions, political programmes, administrative acts or arrangements, etc. There are also methods to identify the priorities and preferences of desired results by brainstorming, voting, etc. Methods of investment accounting, accounting, statistical methods, etc. are directed to the assessment of performance indicators, but also characterising inputs and other conditions which are shaping performance. If the indicators are qualitatively formulated and quantitatively measured, the attainment of desired results of public agencies can be investigated and benchmarked in their working conditions (taking into account the inputs and other factors).

Benchmarking (Bogetoft and Otto 2010), or relative performance evaluation, is an assessment tool opening new possibilities for improving the performance especially in public agencies because of the attainability of comparable information about the performance of the public agencies' subunits. Theoretically, it is important to analyse to which extent benchmarking can be used to substitute the non-existence of competition in the public sector, as well as the possibilities and limits (opportunities and threats) of benchmarking in society. Benchmarking limits the improvement of the best subunits because other subunits are evaluated compared to the best. Hence, it is necessary to introduce incentives for the best to improve. Alternatively, such issue could be tackled by analysing the productivity growth of the subunits. In the thesis these aspects will be considered.

Much of the performance measurement and thus management is based on historical data of the economic unit, projecting it to the future. Often the assumptions made follow that the future is known with high degree of certainty (Otley 2014). This, however, is a simplified approach. In reality, uncertainty is much greater and may result in unexpected events and unforeseen consequences. Uncertainty can be defined as a lack of ability to predict what the future might hold (Otley 2014). Uncertainty may result from external environmental factors which are basically known or just slowly changing in time (such as the socio-economic level of public agencies' service area) or unknown (e.g. weather conditions, different shocks). This adds another layer of difficulties in creating a useful performance indicator system.

In the case of public agencies, the demand of public services is often unknown when planning the amount and quality of supply of the services. Some models have taken the unknown output of production into account (e.g. Chambers and Quiggin 2000; O'Donnell *et al.* 2010). The current thesis develops models, which consider the issues of demand uncertainty in production.

Several techniques have been developed to analyse performance from different perspectives as adequately as possible, including single performance indicators, ratio analysis, unit cost analysis, risk adjustment, Four Quadrant model, composite index analysis, cluster analysis, Balanced Scorecard, regression analysis, multilevel analysis, and seemingly unrelated regression (SUR) methods. In addition, frontier analysis methods have gained popularity in recent decades; they aim to evaluate the cost (production) frontier of the work of public agencies in the considered sample and find the distance of each agency from this frontier. The most common methods of frontier analysis are data envelopment analysis (DEA), which in its basic form is a deterministic and non-parametric method, and stochastic frontier analysis (SFA), which is a stochastic and parametric method. The current thesis discusses the possibilities of using these methods on public agencies' performance measurement when taking uncertainty into account.

The proposed concept is implemented using the case of fire and rescue service (FRS) subunits, which is a novel subject in the area of performance studies. Typically, fixed resources (funds, staff and equipment) are allocated to the subunits but the outcome is unknown since the amount of operations they provide varies. Therefore, they have to maintain a certain level of readiness at all times to be capable to respond in case of an emergency. Sustaining the readiness is the most expensive component in the budget, so it would be important to allocate the resources without much waste, e.g. minimising the level of readiness to provide an optimal service. To reach this goal, production functions with an uncertain amount of output should be developed. By evaluating the efficiency of the subunits, possible shortcomings of budget allocation can be ascertained.

As noted, the performance indicator systems' goal is to qualitatively determine the essence of subjects under investigation and quantitatively measure the parameters for all comparable objects in the same way. In the case of FRS subunits, the characterised phenomena are:

- The expected results of FRS subunits performance from the perspective of different stakeholders. Based on decision theory, this cannot be stated as the society's expected results, as there is no integral leading subject, which would formulate the absolute expectations. Therefore, the indicators are fixed to comprehensively take into account the different stakeholders' expectations.

- The performance of FRS subunits, which is characterised through the proposed performance indicators system and which follows the different stakeholders' expectations.

- The indicators that can be measured quantitatively can be used to measure the distance between the real and expected result. Different indicators can be prioritised differently.

- The real and expected results are dependent on allocated inputs and environmental factors, which should be discussed and taken into account.

- The performance measurement of FRS subunits gives a comprehensive overview of different possibilities of improving the performance of FRS brigades, adjusting the results due to differences in input allocation and environmental factors.

- The performance measurement of FRS subunits is a basis for evaluating the performance of the rescue authorities as the principal of FRS subunits.

These phenomena are analysed and discussed in-depth in the thesis.

To illustrate how to apply the proposed concept, the examples include the FRS from three countries, Estonia, Finland and Sweden. Being neighbouring countries, they have some similarities in terms of climate (with a colder climate, much effort goes on heating the houses, which is a potential fire threat) and population density (large sparsely populated areas, which are difficult and costly to serve). In addition, the developments of all three countries signal that achieving cost-efficiency (CE) is important. The last reform in Estonia (in 2012) resulted in closing of nine FRS brigades. The need to cut back on costs has also been prevalent in Sweden, where different types of cooperation with other entities as well as changes in process management are targeted at cost-efficiency (see e.g. Weinholt 2015; Holmgren and Weinholt 2016). Nonetheless, the socio-economic development of these countries is different, as Estonia, regaining its independence in 1991 has to do catching up to reach the development level of the Nordic countries.

However, these countries have qualitatively different management systems, as the degree of centralisation varies. In Estonia, the service is managed by a central agency (Estonian Rescue Board), in Finland, there are 22 fire departments with autonomous decision-making, and in Sweden the services are provided at a municipal level (which, in turn, might cooperate with each other). Such diverse management approaches would allow to test whether the concept would be suitable independent of the management approach. In addition, the Estonian Rescue Board has stated as its vision in its strategy for 2025 that it has reduced accidents and losses to the level seen in the Nordic countries. Analysing the three countries comparatively would open the opportunities for a discussion on how such levels could be reached.

To be of interest to a wider audience, one can consider that the FRS agencies deliver similar services and are thus comparable. Nonetheless, the services provided by the FRS can also be to an extent similar to the services provided by ambulance, police and other internal security agencies, where a quick response is needed, and the direct demand is unknown beforehand. In a wider sense, also public agencies which have localised branches, can be of relative similarity in the public sector performance measurement context. In the planning situation, private firms which deliver goods also have a few similarities (such as a fast food chain delivering pizzas – the demand for services is unknown and they have to respond quickly in case of an order).

The aim and research tasks of the thesis

The aim of the doctoral thesis is to develop the theoretical concept and application to measure the performance of public agencies in the case of demand uncertainty. The suggested models would be the basis for planning resource allocation improvement in public agencies. The models are implemented using the example of the Estonian, Finnish and Swedish FRS.

The following research tasks are set:

Theoretical framework:

- To analyse the theoretical basis of defining and measuring the performance in public agencies;
- To systematise the potential uses of performance measurements in public agencies, while assessing the possibilities and limitations of measurement;
- To analyse the impact of demand uncertainty on public agencies in the provision of services;

Methods:

- To systematise the methods implemented on the measurement of the public agencies' performance;
- To create a performance measurement methodology that would be able to incorporate demand uncertainty;

Application:

- To illustrate the proposed concept in a real situation of service provision in FRS (Estonia, Finland and Sweden), relying on previous research experience, legislation and development plans;
- To compare the characteristics of different measurement methods for the assessment of public agencies' subunits performance, to identify reasons for differences in performance assessments;
- To assess empirically the performance of Estonian, Finnish and Swedish FRS;
- To assess the possibilities of using the proposed concept incorporating demand uncertainty as a tool for improving decision-making.

The logical progression of research tasks is also illustrated in the Figure 1. The thesis is divided into three chapters that follow the proposed division.

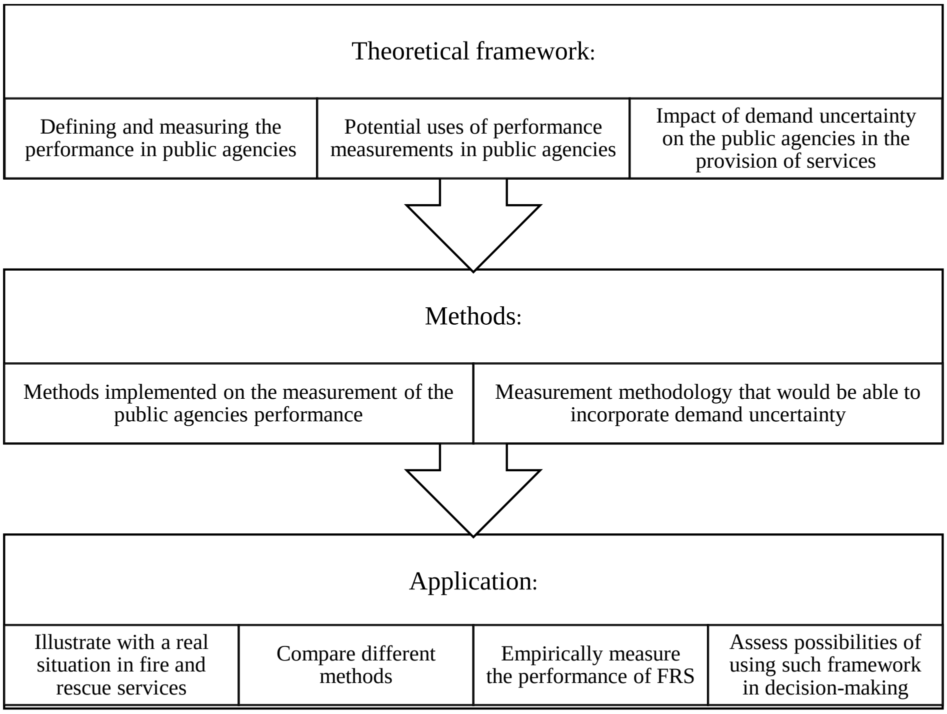


Figure 1: Research tasks (Source: Author's compilation)

Novelties of the study

The thesis contributes in three ways. First, it closes the gap between theoretical microeconomic production theory and more practical public sector management and administration fields in the approach to demand uncertainty in a public agencies' performance. Although the terms 'efficiency' and 'effectiveness' have a specific meaning in production theory, in the 'real world' they are used loosely, vaguely and inconsistently by policymakers (Pollitt and Bouckaert 2011).

Second, it introduces the concept of demand uncertainty into the public agencies performance measurement and management through frontier analysis methods.

Third, although a quite vast trade press covers the topics of FRS, this is a novel and understudied application in the scientific field of production theory in uncertainty.

In addition, this is the first attempt to analyse systematically the performance of FRS in three countries, Estonia, Finland and Sweden. This opens opportunities for a discussion on how to reach the levels of Scandinavian FRS quality, which is the goal of the Estonian Rescue Board, as stated in its strategy.

Research methodology

The research methods used to accomplish the dissertation aims are the following:

- Literature search, document analysis, law interpretation, background research and consultations with experts of FRS to gain thorough knowledge about the subject analysed (what to measure and to benchmark it);

- Literature search and document analysis for theoretical background and different approaches of performance measurement and assessment in public agencies (how to measure it);

- Empirical analysis of Estonian, Finnish and Swedish subunits using various frontier analysis methods (DEA, FDH, DFA) (how to interpret the results).

The delineation and limitations of the study

The theoretical part of the thesis is embedded in the Anglo-Saxon approach of public administration, thus in the context of performance measurement in the public sector. The works regarded in the paradigm of the New Public Management (NPM) are of high importance and receive much attention. This, however, means that the continental approaches and public management models (such as Harzburg Model (Höhn 1967), where the performance measurement is of less importance) are not dealt with to a great extent. Thus, the thesis advocates towards performance measurement. In addition, different alternative streams of literature, such as welfare measurement of public services and actions (Dupuit 1844), leading to cost-benefit analysis and utility analysis; measurement of preferences of citizens for public services in public finance (social choice, Arrow 1951); and public value in public management (Moore 1995), are not extensively covered in the thesis. The literature discussed is limited to the works published up to 2016.

As the functions of public agencies vary largely, a balance between generalisation and specification must be found. 'Public agency' in the theoretical treatment would be considered as a general term, without making any restrictions on the specific form (e.g. a public office, a public enterprise, a quango, a sector of public administration). Thus, the effects of the form of the public agencies on the performance are not analysed.

The thesis focuses on one type of publicly provided services, fire and rescue services, which in an organisational context can be organised and planned quite differently. Therefore, the analysis focuses on three countries. In the case of Estonia, the FRS are provided by a centralised public agency, in Finland by fire departments, and in the case of Sweden it is a duty for municipalities. For simplification, the upper level of management is considered to be the central agency (in Finnish and Swedish case, this could be considered more as industry-level analysis) and lower level as subunits (brigades for Estonia, fire departments for Finland and municipalities for Sweden). As the structure of providing the services differs, it also hinders the possibilities for comparison.

As the FRS are often provided in cooperation with other entities (such as volunteers), the entities discussed are restricted. In the case of Estonia, only the national FRS brigades are considered, mainly due to the lack of data for longer time periods about voluntary FRS brigades. The voluntary FRS is becoming more and more popular in Estonia and deserves further attention. As this is not the focus of this thesis, the issue has been analysed in Puolokainen *et al.* (2018). In the case of Finland and Sweden, the voluntary FRS provision is merged into the system for a longer time period and thus is reflected in the data. Other possible entities related to the service provision are not the focus of this thesis.

The thesis is restricted to economic issues and excludes the political problems, such as goal changes due to political conditions (voting situation). It is assumed that the public agency is pursuing rationality and benevolence. Thus, the influence of financing (how the budgets are formed for the central agency/industry), the influence of cooperation between different subunits, the influence of other stakeholders and private actors, and influence of principal-agent situations on the decision-making and thus to the performance of the public agency are not considered in the empirical case.

The case of uncertainty is focussed on the demand uncertainty, meaning that the amount of services demanded by the subunits is unknown beforehand. However, it should be noted that a public agency faces many uncertainties in its activities. For example, there is uncertainty with respect to resources (and how big is the allocated budget for the central agency/industry) and environmental factors (effect of socio-economic development, hazardous behaviour like smoking, changing climate, etc.).

The empirical part of the thesis is devoted to the verification of performance indicators and the implementation of measurement models. As the number of possible indicators is large, the analysis has to be restricted. Therefore, the thesis concentrates on performance in one type of public agency. The thesis concerns the indicators for inputs, processes, outputs and desired outcomes of FRS. The analysis is also restricted to the data that FRS managements are already collecting, which means that if the data is unavailable, or available only in a limited way (short time period or no disaggregation), it has to be excluded from the model. This, however, gives many opportunities to extend the discussed models in future research, when such data will become available. This is mostly true for every kind of prevention activities provided by fire authorities.

As the microeconomic approach requires consistent data for practical applicability, it will be complicated to use the proposed methods for a longer time period without much extra effort. In addition, evaluating the effect of different reforms (mergers of organisations, changes in assignments, etc.) have to be taken into account systematically.

Structure of the study

This thesis has three chapters. The first chapter gives a theoretical background of performance measurement in the public sector, the second compares different measurement methods and the third integrates the proposed model and applies it to FRS subunits.

The first chapter defines the performance in the public agency from the public administration perspective and discusses the reasons why it varies across different units. The classification of reasons and purposes why one might be interested in measuring the performance in public agencies is highlighted and analysed, while possible setbacks are also considered. As the performance of a public agency is complex, the dimensions of performance are considered in the input-output model framework and applied to the fire and rescue services. The performance is measured using different evaluation criteria, the differences and controversies are compared. To be more concrete, the performance measurement is analysed through the production theory framework and the efficiency and effectiveness are defined to be the basis for the following measurement. In the final section, the theoretical case of demand uncertainty is elaborated and the model that will be applied to the empirical case is analysed.

The second chapter analyses and compares the possible methods for performance measurement in public agencies. An emphasis is given to the frontier analysis methods which are the basis for estimating the models in the next chapter. The case of uncertainty in the productivity analysis is considered and a literature review on articles that use frontier analysis methods in the FRS field is given.

The last chapter focuses on the empirical case of FRS and the performance of the service providers in Estonia, Finland and Sweden. The countries are firstly analysed independently, estimating the cost-efficiency, under-resourcing and aggregate output efficiency using data envelopment analysis, free disposal hull and deterministic frontier analysis. The methods are used on two different models, one that takes demand uncertainty into account and one that does not (naïve model). Lastly, the results of the countries are compared and conclusions and possible policy implications are discussed.

Acknowledgements

Firstly, I would like to express my gratitude to all of my supervisors, Professors Janno Reiljan, Peter Friedrich and Chris O'Donnell, who have been an immense help in my development towards a decent researcher. Professor Reiljan has been a great support already since my bachelors' studies and is always willing to provide his opinion and backing on my concerns. Professor Friedrich has widened my perspective on many topics – all of the delineation of my work is based on his extensive knowledge. Special thanks go to Professor O'Donnell, who hosted me for a year at the University of Queensland, Australia and who patiently and repeatedly explained many concepts, so even I would eventually understand. Most of my thesis was developed and written there.

Besides the supervisors, many people throughout the years have provided me valuable feedback, insightful comments and encouragement on writing the thesis as well as on wider topics. Professors Tiiu Paas, Maaja Vadi, Raul Eamets, Urmas Varblane, Toomas Haldma and Kadri Ukrainski, senior fellows Jaan Masso, Priit Vahter and many other (often also anonymous) reviewers, to name but a few. And Andrew Rozeik for editing this thesis into nice English.

I would like to thank all of my supporters throughout the years, as I have never needed to pay for my education directly – Estonian taxpayers, European Union, Archimedes Foundation (Kristjan Jaak Scholarship, DoRA, DoRA+), the University of Tartu Foundation, Estonian American Fund (Andreas and Elmerice Traks Scholarship), Estonian Students Society (Udo Mällo Scholarship) and the Doctoral School in Economics and Innovation – without whose funding I would have never had the opportunity to gain such a wide perspective on world's issues.

The School of Economics and Business Administration in University of Tartu has been most helpful. For that, I have Anneli Kütt, Janek Saluse, Ülle Maidla and Katrin Tamm among others to thank. One learns to appreciate the neat Estonian bureaucracy a lot more, once thrown into a system that simply does not function.

My gratitude goes to the Estonian Rescue Board, especially to Helina Uku and Tarvi Ojala, for gathering all the necessary data for me. To Esa Kokki from the Finnish Emergency Services College for granting me access to the PRONTO database. From the Swedish MSB, I have Joakim Ekberg, Morgan Asp, Colin McIntyre and Mikael Malmqvist to thank for gathering the information on Swedish FRS and providing additional comments.

The thesis would not have been possible without the understanding of my bosses, Uku Varblane and Kerly Espenberg, who should know on their own, what writing a thesis while working means. Also, I have marvellous colleagues in the Centre for Applied Social Sciences.

The journey would have been a lot less colourful without the fellow students. I thank Tõnis Tänav and Kristian Pentus for the fruitful discussions and endless innovating ideas that took us to Shanghai and Astana among closer destinations. I would like to thank Annika Jaansoo, with whom we won a prize in Georgia. My friends have given me a push in pursuing higher values. I would like to thank everyone who has had a great impact on me in becoming the real me.

I would like to thank my parents for keeping me and not putting me up for adoption and providing me with the necessary secret fertiliser in order to succeed. And of course, Astrid, who has been by my side for all these years and tirelessly double-checking most of my correspondence with the outer World, so it would be well in line with the society's expectations. And lastly Kribu, our cat, who has provided a decent stress-relief service with sharp claws.

1. PERFORMANCE MEASUREMENT IN THE PUBLIC SECTOR

I was shown several empty wards, several administrative offices that were veritable hives of activity, and finally a huge deserted dusty operating theatre suite. I enquired about the cost of it. Mrs Rogers informed me that, together with Radiotherapy and Intensive Care, it cost two and a quarter million pounds.

I asked her if she was not horrified that the place was not in use.

'No,' she said cheerfully. 'Very good thing in some ways. Prolongs its life. Cuts down running costs.'

'But there are no patients,' I reminded her.

She agreed. 'Nonetheless,' she added, 'the essential work of the hospital has to go on.'

'I thought the patients were the essential work of the hospital.'

'Running an organisation of five hundred people is a big job, Minister,' said Mrs Rogers, beginning to sound impatient with me.

— J. Lynn & A. Jay, Yes Minister, The Compassionate Society

1.1. Performance, reasons for its variation and intended purposes of measuring it in the public sector

1.1.1. Performance measurement in the public administration context

Performance measurement in the public sector has been an interest to multiple fields of scientific research, such as public sector management, organisational theory, public administration, decision theory and microeconomics (production theory). This has led to a plethora of possible explanations and definitions for the phenomena discussed in the current thesis. This chapter analyses the different approaches as follows: first, the background, intended purposes and practical issues of performance measurement are discussed, and then a more clearly defined concept of efficiency from a microeconomic framework is introduced. The section focuses on the main goals, why one should measure the performance of a public agency and how the results can be applied to improve the performance through management. In addition, possible setbacks of performance measurement are discussed. This section mainly follows the interests of the public sector management field, as it has a more practical position.

The public sector management has in theory seen roughly three main paradigms and ideal settings: 'Old' Public Administration (OPA), New Public Management (NPM) and post-NPM paradigms (such as New Public Governance (NPG) and New Public Service (NPS)). OPA has the following characteristics (Denhardt and Denhardt 2000): politically neutral, focussing on direct delivery of services. According to this approach, the best organisational structure would be a centralised bureaucracy, with top-down control mechanisms, limiting discretion as much as possible. Bureaucracies target to be closed systems to the extent possible, and so minimise the possible participation of citizens. Public agencies value efficiency and rationality. Public administrators' are only in the background in policymaking and governance; they are rather seen as efficient implementers of public objectives. Public administrators tasks are described by Gulick's POSDCORB (Gulick 1937: 91). In the case of OPA, the results (and to an extent the performance) of public agency are fixed by laws and accounting techniques (information about costs), making the comparison of different subunits relatively easy.

The OPA approach was conquered with the methods of business-like management, labelled under NPM and popularised by Hood (1991), Osborne and Gaebler (1993), etc. in the literature. The NPM approach took off in the 1980s and have influenced public reforms to date. The NPM opposed the traditional OPA approach, as Hood (1991) points out in its main doctrines: hands-on professional management, explicit standards and measures of performance, greater emphasis on output controls, shift to disaggregation of units in the public sector, shift to greater competition in the public sector, stress on private-sector styles of management practice and on greater discipline and parsimony in resource use (efficiency improvement).

Although one of the NPM initial ideas has been the decentralisation and horizontal specialisation in public agencies, it has been evident that there is an increasing need for coordination and coherence in public policy, which has resulted in the re-centralisation and restoration of the hierarchy (Christensen 2012). This process has been defined as post-NPM reform and has been the reaction to the 'pillarisation' of the public sector (Pollitt 2003; Gregory 2006) – the structural devolution encouraged by NPM ideas had ignored the need for horizontal planning and coordination (Fimreite and Laegreid 2009). Bouckaert (2009) describes the zig-zag action-reaction character of public reforms, meaning that the reforms follow a sequence of finding a solution to a problem created by the last reform. This sequence is never-ending. The last decade, however, has not seen a dominant model of public management (Pollitt and Bouckaert 2011).

Following NPM, the post-NPM has the key concepts of coordination, centralisation, governance, networks and partnerships (Christensen 2012). He also argues that both approaches overlap and are not exclusive when discussing reform tools. The performance measurement as a management tool was not that important in the context of OPA, but has been in the spotlight of NPM as well as post-NPM approaches.

For an alternative to performance measurement as a basis of public management, Van Dooren *et al.* (2015) suggest Max Weber's three types of authority. These authority-types do not focus on the results, so it can be considered distinctive from a wide approach of performance management. In reality, the proposed idealisations do not exist in its pure form, but are a blend of each one. The charismatic authority assumes that the personal qualities of the manager also determine the legitimacy. The manager demands obedience and creates a sense of mission. The traditional authority follows the assumption that the position of the manager also determines its legitimacy, which follows the respect for traditions and routines. The rational-legal authority is based on following a code of rational rules and regulations. If a manager succeeds in this, it is considered legitimate. From this follows the bureaucratic management (OPA), which is about coordinating and directing within a set of rules. The rules are considered a political end.

Osborne and Gaebler (1993), as the early introducers of NPM approach, gave the following reasons to measure the performance in the public sector:

- If the performance is not measured, one cannot distinguish success from failure (the correct execution of law);
- If the success cannot be distinguished, one cannot appreciate and reward it;
- If the success is not appreciated and rewarded, the failed agency gains from it;
- If the success cannot be distinguished, one cannot learn from it;
- If the failure is not distinguished, one cannot fix it;
- If the (good) performance can be shown, the public sector might gain public support for its progress.

Therefore, performance measurement and evaluation of public agencies is necessary to plan better the activities and resource allocation of the institution to find the best possible way to balance the capacity and quality of service offered to the consumer. Through the performance measurement, it is possible to report to the stakeholders on the efficiency of the resource use for providing the necessary services. Different management styles, applied mainly in the Anglo-Saxon countries, such as management-by-objective (MBO), zero-based budgeting

(ZBB), planning-programming-budgeting system (PPBS), strategic management, performance budgeting, managing for results, results-based management and entrepreneurial budgeting, take such features into account, so tied to this, in these management settings the performance information is provided and used for decision-making.

Performance measurement is the process of quantifying action, where measurement is the process of quantification and action leads to performance (Neely *et al.* 1995: 1228). Performance measurement in the public sector has a long and strong theoretical tradition in public administration and organisation theory. This might be the reason why public sector performance measurement has been relatively more influenced by economics focussing on output and welfare, rather than financial aspects typical for private sector institutions. Public management has been influenced by political science and sociology in managing politics and institutionalisation. Thus, public management has focussed more on performance measurement issues, such as uncertainties and ambiguity (Johnsen 2000).

Meyer and Zucker (1989: 111) argue that: *'Generally, performance will be defined narrowly to the extent that (a) elites dominate an organisation, (b) a high degree of professionalisation exists, and (c) the organisation performs a technical function, outputs of which are measurable. Performance will be construed much more broadly, by contrast, to the extent that (a) the norm of participative democratic governance operates, sometimes in the formal structure or rules of an organisation, (b) the interests of multiple constituencies are given recognition, and (c) the organisation's function is non-technical and outputs elude measurement.'*

Identifying the performance and outcomes of public agencies – the essence of goods and services provided by public sector – is usually a difficult task, because different interest groups might understand the prospective performance qualitatively differently. For what reason (market failure, increasing equality) is a good or service provided by the public sector and what are its quantitative essence and characteristics? What goals has the society set on the quality and quantity of the publicly provided good or service and how should they be measured (see also Jaldell 2002: 27)? The next subsection considers the possible reasons, why the performance differs across public agencies.

1.1.2. Differences in performance and performance measurement in public agencies

There are many different aspects, when identifying the essence of performance in public agencies. Following the social constructionist approach, one might argue that the performance of a public agency depends on the evaluators' point of view. In addition, in the post-positivist approach, the actual performance of public agencies might (and will) differ from the measured performance. Smith (1990), Smith and Street (2005) and Smith (2006) state possible reasons for differences in performance, some of which are directly under the control of agencies in interest, some of which only indirectly:

- They have different objectives, which might be a result of differences in laws, and devote different levels of resources to services: even though the budgeting can be centralised, it might still result in different priorities leading to differences in services provided.
- They have different needs: the socio-demographic and economic background (environment) of regions are significantly different, which is why the demand for services varies across regions. This leads to variations in the feasible levels of performance.
- They have different ways of service provision: unlike the private sector, where the provider can freely choose the most suitable location, the location in the public sector is usually determined, so the agency has to conform accordingly (such as differences in wage levels and factor prices, urban and rural areas, quality of resources used, etc.).
- They have different related public agencies and other organisations, which have an impact on the pursued outcomes.
- They have differences in efficiency: different subunits have to provide a similar service in different locations, which is why different managerial competences are needed. This is one of the priorities of benchmarking – to decrease the variation in performance due to managerial skills. If the unit is fully efficient, it can improve the performance of the indicator only at the expense of worsening some other indicator.
- They have differences in accounting, reporting and measuring methodologies: it is impossible to ensure that all the subunits are following exactly the provided guidelines, so differences in interpretation of regulation is possible.
- Random (or idiosyncratic) fluctuation. Different unpredictable shocks and uncertainty about the future.

Multiple reasons for performance measurement (discussed in the next subsection) give rise to differences in the framework of measurement. Measuring and managing performance matters to all stakeholders – public organisations themselves and their principals (mostly for internal use), other public agencies/private firms and the public/taxpayers (mostly for external use). The results of the measurement also depend on the benchmark to which the entity is compared, and how the results could be applied. Therefore, the performance measurement is dependent on:

- The framework;
- The stakeholders;
- The benchmark.

The framework

Different performance measurement styles evaluate multiple aspects to different depths. On the one hand, there are performance measurement systems (PMS) which evaluate the organisation in detail – auditing processes (such as controlling all financial and operational aspects on-site, analysing the daily activities through documentation, etc.), for example inspections of schools or full auditing of public agencies. On the other hand, general indicators are provided, such as cross-country comparison of one certain indicator, for example the average test result of the school-leavers for some year in some country. Between these two extremes, there are specialised targeted performance measurement and evaluation reports (Propper and Wilson 2003: 254).

Some authors (e.g. McDavid and Hawthorn 2006; Van Dooren *et al.* 2015) distinguish the performance measurement and performance evaluation, but it is not a common practice. According to them, the PMS are continuous and routinised, while addressing general issues. The resources for measurement are part of the organisational infrastructure and thus the attribution is taken for granted. Following that, the managers have a key role in the measurement. The PMS develops and matures in time. On the other hand, the evaluation takes place only once, has a targeted issue and the measures are customised for each evaluation. The resources for evaluation are project-based and the attribution is a central issue. The evaluators are usually vaguely connected to the evaluation and have no personal interest in the results. The use of performance evaluation is usually negotiated beforehand. The current thesis follows the performance evaluation line.

The targeted performance measurement reports usually try to answer the following questions (Bolton 2003: 21):

- How effective is the public agency at fulfilling its mission? How should it be measured?
- How efficient is the public agency at fulfilling its mission?
- How should the public agency's performance be assessed with respect to other public agencies?

- How does the public agency inform the public (taxpayers) of its performance?
- How does the public agency take into account the received feedback from the public (taxpayers) on its performance?

Due to the multitude of priorities in performance measurement, the questions are of different importance. The first question is the most difficult to answer because of the lack of market competition in the public sector. The tasks for a public agency are long-term and might be politicised, which causes the risk of not reaching a desired result. Consequently, there are typically two main possibilities to measure the performance in the public sector (Steers 1975; Bolton 2003; Jung 2011). Firstly, one can compare the real results with the set goal – the performance is considered successful when the agency fulfils the goals and tasks in a certain time frame. However, it is difficult to assess the level of the initially set goal (do the initially set goals need highly efficient performance, are the goals only dependent on the performance of a public agency, is the result measurable in the same framework as the set goal?). Alternatively, one could measure and compare the dynamics of the performance – the growth of performance indicators value across different public agencies. This also has multiple setbacks – what caused the change, was it solely the public agency? For example, the increase in the number of crimes may be caused by changes in the society and influenced by the performance of many institutions (public and private), and not just the poor performance of the police or other public security agency. Comparing different public agencies' performance raises the question of the level of basis. Based on catch-up theory, from a relatively low level of basis the public agency may result in faster improvement (converges towards the relatively well performing agency) of performance than from a relatively high level of basis.

The stakeholders

The performance measurement depends on to whom the results targeted and thus has different goals. In a typical public administration literature context, one distinguishes the 'middlemen' (project and programme managers, senior officials in public agencies, other stakeholders, who are the users and suppliers of specific services) and 'end-users' (ministers, members of parliament and citizens) (Pollitt 2006).

Different stakeholders have different interests of the performance of public agency (Brignall and Modell 2000; Van Dooren *et al.* 2015). The performance indicators provide the possibility to follow and evaluate the performance of a public agency for all interested groups of society. In the case of the so-called 'middlemen', the performance measurement is of interest to learn and improve/innovate the service provision, raise quality and to give an account of the performance (economic, effective and efficient service provision). The 'end-users', however are more interested in the performance measurement as a guidance for reforms, control and evaluation (ministers), allocating and approving resources, financial results and resource utilisation (funding bodies, members of parliament) and being aware of and receiving an equal and good-quality service (to citizens, among others). Published results can create a pseudo-competition between different units of public agencies, where evaluation results may be used for resource reallocation purposes (Propper and Wilson 2003: 253). All this, however, assumes that the performance measurement in its essence is possible and rational (Van Thiel and Leeuw 2002: 268).

Pollitt (2013) goes more into detail and distinguishes and characterises seven different key groups who can be involved in public agency and its performance measurement. The motivations and levels of understanding of technical issues of different groups will be different. In the previously discussed 'middlemen' group, one would find: a) top officials, b) technocrats concerned with running performance management systems (including consultants), c) operational staff concerned with the delivery of public services; and in the 'end-user' group: d) ministers, e) members of legislatures, f) the mass media, and g) citizens. In addition, on could consider h) private firms as another key group that can be added.

In the times of the post-NPM paradigm (described above), the public policies and programmes are administered through increasingly complex governance structures, including networks, collaborations, and partnerships among public, non-profit and for-profit organisations (Lynn *et al.* 2000). Heinrich (2012) states: *'In accounting for performance across "diverse and dispersed" administrative entities and service units, many of which may operate in varying social, political and fiscal contexts, public managers need to achieve a tenable balance between demands for analytical rigor and accuracy in performance measurement and political and practical limitations on what is feasible to measure in complex governing systems.'*

It is dependent on the influence and power of these stakeholders, which targets are pursued and how it is reflected in performance measurement. Thus, the performance measurement allocates various weights to stakeholders and considering the dominant stakeholder or equally-balanced system of measurement is relevant. For further discussion on relationships between different stakeholders and how they can pressure the public agency and its performance, see Brignall and Modell (2000).

Carlsson *et al.* (2012) conducted a survey to find out whether the public administrators and citizens have differences in risk reduction priorities. The results were mostly consistent, as only minor differences prevailed between reducing the risk of many small or one large accident. The citizens preferred to avoid many small accidents instead of one large, among the administrators the views are split almost equally between these options. This gives a reason to assume that the priorities of public administrators and citizens are similar, when not considering the constraints (financial as well as environmental). However, other studies have reached a contrary conclusion. Ho (2008), which was used as an example by Van Dooren *et al.* (2015: 147–148), found that citizens have a different perspective than managers. Often the citizens value less the input and output measures and more outcomes and the citizen perception of service quality, responsiveness, customer services, intra-jurisdictional equity, transparency and effectiveness in public communication. This can be due to the fact that the citizens often relax different constraints (such as limited budget).

As such multiple and conflicting interests typically prevail, there is a need to balance these by trade-offs. This is more difficult when a public agency depends on a particular stakeholder more than others. Such dominant stakeholder can dictate the objectives and limit the public agency's ability to fulfil other stakeholder's objectives. In the case of NPM, much effort goes on financial and measurable indicators, leaving non-financial and probably also important aspects out of scope. One could argue that in this setting, the 'end-users' like ministers and funding bodies (members of parliament) and not the citizens dominate the public agency's goal-setting. They, however are accountable in turn to citizens, when they seek re-election to their position. Thus, the system is not clear-cut and has overlapping and complex relations between different stakeholders and their objectives.

The NPM approach draws on two economic theory branches affecting the stakeholders of performance measurement: property rights (Demsetz 1967) and principal-agent theory (Jensen and Meckling 1976; Holmström 1979; Zeckhauser and Pratt 1985). The property rights approach declares the impossibility to transfer ownership rights among the individuals in the public sector. As a result, there is less pressure on public agency, which does not have a competitive market environment. In addition, a public agency is typically a single provider of the service, and inefficiencies occur because of the monopoly position or due to X-inefficiency (Leibenstein 1966).

The principal-agent theory states that if the principal cannot observe the agent's effort level, the agent tends to work too little or has different goals to pursue from the principal. In the current case, the principal-agent relations are manifold. At the 'end-user' and 'middlemen' level, one can consider the ministers, MPs, citizens as principals and the public agency as the 'middleman'. The stakeholders are unable to specify incentives to the public agency, resulting in biased expectations. In addition, the citizens are principals to the legislative and executive bodies (members of parliament, funding bodies and ministers). On the other hand, one can consider the central public agency as the principal who makes planning decisions, and its subunits as agents who support the achievement of the central agencies objectives. In the current thesis, the main focus is on the assumption that the public agency will receive clear goals from its principals, and acts as a principal to its subunits, which are considered as agents. Therefore, it is assumed that the principal is rational and benevolent and motivated toward the maximisation of the social welfare by means of making the most equitable-effective and efficient use of central resources. The subunits, or agents, are only motivated to cover the local demand for their services.

The benchmark

The measurement can be divided mainly into two distinct categories: 1) in case one compares different (sub)units, it is not the 'absolute value' that matters, but rather the relative value which allows the benchmark; 2) in case one compares the same units' value in different years for example, it is the rate of change that matters, so it is important to use the same methodology across years to ensure its comparability.

Benchmarking as a term originated from the land survey as a mark, relative to which the survey was carried out. The term was taken over by the management theories, where it describes the systematic comparison of an organisation to the best practices (Bogetoft and Otto 2010). The goal of benchmarking is to identify the best performing (sub)units and find ways to implement the same strategies more widely on worse performing (sub)units. The benchmarking can be used to compare different units and organisations, as well as the dynamics by using time-series or panel data (Bogetoft and Otto 2010: 2).

Often a benchmark can be a set standard, performance target or result achieved by the best performing counterpart. If the performance record is generally consistent and does not fluctuate in high amplitude, it reassures that the operation is on track. However, if substantial gaps or fluctuation are apparent, the organisation in interest might conduct further analyses to gather possible solutions and make changes in the operations (Ammons 2008). This can be a result of management behaviour.

This subsection considered the main reasons why performance might differ in a public agency and how the performance measurement has to comply with alternatives. The next subsection analyses different purposes why one should conduct a performance measurement in the public sector.

1.1.3. Different purposes for measuring the performance in the public sector

As the public sector has to create additional value, the need for performance measurement and evaluation has to be justified. The reasons according to Pidd's (2012) classification, which in turn is based on Bird *et al.* (2005), Behn (2003) and Poister (2008), are divided into six (see Appendix 1). The main reasons are explained subsequently. Planning and improving is of most value for internal purposes; evaluation and benchmarking, on the other hand, is mostly targeted at external audiences who, in turn, can influence the operations of subunits. Van Dooren *et al.* (2015) divide the planning and improving in turn into two parts, namely following the learning and steering/controlling practices. The purpose of learning is focussed on the future, the steering/controlling on the present and evaluating/benchmarking on the past.

Planning and improving

Similar to the private sector, it is important to plan the activities in the public sector. Following the three-step planning, where the most important long-term decisions are made in the process of strategical planning, medium-term decisions are subject to the tactical planning and daily decisions are results of operative management. This applies to the public agencies as well. In the strategic planning the private sector firms follow the vision and mission of the firm – the equivalent to the public agencies is typically provided through political process (development plans, future development strategies, etc.), which means that it is crucial to think through what is needed to achieve that (Pidd 2012: 73). Mintzberg (1994) states that prediction of the future is impossible, therefore, the prognosis and analysis cannot be the essence of strategical planning. The goal of strategical planning should rather be mapping the possible scenarios and clarification, what these scenarios could result in.

The planning period also dictates the suitable performance indicators: indicators that characterise the added value of the public agency are used in the strategical planning. In case the goals or tasks of the public agency change, the performance indicator system used also needs updating (Pidd 2012: 76).

Performance measurement allows to identify subunits of public agency with best performance and so generates the learning effect: the subunits with worse performance evaluation can mimic the best (Koopmans 1951: 123–125). The results of performance measurement can also be the basis for coordination – one can identify shortages and surpluses in resource (budgeting and tasks) allocation. One could argue, therefore, that extra funds have been earned (as a result of showing good performance) or extra funds are needed for improvement (to catch up with better performing subunits). In addition, the results can have a punitive use (Pollitt 2013).

Evaluation and benchmarking

It must be stressed that evaluation/benchmarking is a formalised analysis process and not just observation of different (sub)units, to see ‘what others are doing’. In addition, it is not assumed that the best practices can be taken over directly, but rather to get insights on what might work to improve the performance (Pidd 2012: 113).

Most public sector organisations exist as local subunits, whether they offer services (such as unemployment insurance fund or road administration) or enforce measures (such as tax collection or the police). The subunits are assumed to function in a general framework, following justice and equality principles.

Since the framework is similar irrespective of the location, it is typical to compare the subunits in hope of finding the best practices. Much attention has to be put into the evaluation to ensure its validity, especially when the results are published. The subunits, whose performance is evaluated highly, will receive public approval. The opposite is true for the subunits whose performance is evaluated low, resulting in extra problems.

Benchmarking can be used widely on subunits which are meaningfully comparable, irrespective of the service provided (schools, hospitals, FRS brigades). Since the public agencies are accountable for providing the services at an optimal level, it is reasonable to compare their relative level of performance. One of the examples includes the test results of high school graduates. From the average test scores, some schools are considered better than others, which in turn results in a better or worse reputation. A bad reputation might entail a downward spiral: the competition for better reputation schools increases and the rest have to deal with the leftovers, thus resulting in worse test results in the future (Leckie and Goldstein 2009; MacLeod and Urquiola 2009).

Other purposes for measuring performance

Mannion and Goddard (2001) find that the reasons of performance measurement have changed in time. Firstly, it was used as internal management tool (planning and improving), later on it has become a tool to increase accountability of a public agency and an opportunity to compare different organisations (evaluation and benchmarking). As a result, the public critique directs the development of organisations. Performance indicators are also used for performance-based bonuses systems or alternatively providing greater independence of public agencies from central government (for example, the highly performing hospitals have more independence in resource allocations than their worse performing counterparts). In addition, there are some other distinct reasons to measure the performance of public agencies.

Monitoring and control – Every organisation has some sort of control mechanism, as executives and managers need the feedback on fulfilling the organisations mission. Typically, the public agencies are subject to internal monitoring as well as external scrutiny, like national audits. Thus, performance reviews are one option for providing this.

Accountability – Public agencies are mainly funded through taxpayers, so it would be important to indicate how well a public agency is performing. Accountability is stressed in democracies, so measuring the performance with respect to accountability is another opportunity.

Financial budgeting and planning – The performance measurement can be considered an extension to accountancy, and is often used to support budgeting decisions.

Individual performance measurement – Individual performance in public agencies can also be measured and used, for example, as a basis for a performance-based salary system. This, however, is the focus of human resources topics.

Symbolic or ritual use – Pollitt (2013) notes that there can be occasions when performance is measured just to show that it is done ('look, we have a modern PMS!').

This subsection analysed the performance measurement and its main intended purposes in the public sector. Most of the focus was put on planning, improving, evaluating and benchmarking, as this thesis concentrates on these topics. The next subsection analyses the shortcomings of conducting a performance measurement in the public sector framework.

1.1.4. Undesirable effects of performance measurement

As was stated in the introduction of the thesis, it is commonly assumed that public agencies performance can be measured and it is meaningful. Boyne *et al.* (2006: 4) pointed out that only a few empirical studies actually analyse this statement. Performance measurement is related to a number of challenges and deficiencies. First of all, performance measurement can be quite expensive and dysfunctional if it is done poorly.

When developing a PMS, one must take into account also the reaction of the organisation targeted. The goal of performance indicators should be to develop a stimulus for a subunit to act in a certain way or direction (to improve performance). The stimuli are knowingly designed management tools, connected with a certain goal or target, towards to which the performance is measured. Through bonuses and penalties the accomplishment towards goals are regulated. The stimuli are defined *ex ante* and are traceable for all parties. Sometimes, the stimuli are formulated badly (vaguely) or are dysfunctional. Typically, the issues of performance measurement and evaluation are classified into three main groups (Fryer *et al.* 2009), which can be divided further. Smith (1990), Smith and Goddard (2002) and Pollitt (2013) have made more detailed classifications and emphasise multiple reasons why performance indicators might result in an unwilling behaviour. These issues have been divided between the three main groups. One should notice that discussed issues are broader and might be overlapping across different groups.

1. **Technical issues** – related to the choice, collection, analysis, and interpretation of performance indicators (Dyson *et al.* 2001; Van Thiel and Leeuw 2002; Van de Walle 2008).

Attention should be paid to who is constructing and choosing the performance indicators. If it is done by the organisations themselves, they might try to selfishly collect, analyse and present the evaluation on their own terms. When the system is provided by the principals, the principals will receive only the data that they ask for, thus resulting in a shallow understanding of the real performance. It is important to ensure that all the requirements for reporting are fulfilled (that all activities are covered). In addition, one should take into account, how independent an organisation is in its activities – when the procedures are strictly regulated, it is more difficult to evaluate the performance (Van Thiel and Leeuw 2002: 276).

A PMS should focus on many aspects: in the case of indicators, the amount and suitability to measure every goal and activity has to be taken into account. When only a few indicators reflect a small part of public agencies' performance or there are big gaps in activities done but not measured, it might result in a performance paradox (Van Thiel and Leeuw 2002: 271). This is a situation when the values of performance indicators and real performance are only slightly correlated. It is important to realise that the performance paradox is not caused by the performance itself, but from measuring it (an invalid evaluation). Contrary to expectations, in the case of a performance paradox, the evaluation does not give a full overview of the performance. Thus, the real performance might be worse than that evaluated (overestimation) or better (underestimation). In the latter case, the paradox is harmless unless the measurement is used as basis for sanctions.

Definitional drift is another (to some extent) technical issue, which quite often emerges (Pollitt 2013). As many performance indicators have some sort of grouping (such as 'class A emergency'), the public agency might want to 'bend' or 'stretch' these categories in their favour. On the one hand, it is a technical flaw of the PMS and on the other hand it influences the behaviour of the public agency (involvement issue).

Symbolic use of information is yet another issue frequently seen in performance measurement. The performance information is gathered just for 'show' without a real intent to use it. *'Visible possession of performance data helps show that one is* "modern", "accountable", "in control" *or* "acting dynamically".' (Pollitt 2013: 354).

Smith and Goddard (2002) add the misinterpretation possibilities – making wrong conclusion in the performance evaluation, which is caused by not taking into account all sorts of potential effects.

2. Systemic or conceptual issues – related to the so-called 'big picture' problems, namely how to integrate the assessment as part of the management, the lack of strategic focus, goal ambiguity, disagreement about definitions, etc. (Fountain 2001; Ingraham 2005; Adcroft and Willis 2005; Van de Walle 2008).

Smith and Goddard (2002) name the following issues:

- Creating a tunnel-vision with too narrow PMS – as a result only some aspects of organisations activities are measured and improved (leaving many important aspects ignored). A similar concept is described by Pollitt (2013) as *synecdoche*. The more complex and multi-faceted the goal of a public agency is, the more likely such issue prevails.

- Over-prioritising measurable indicators – to show the improvement of measurable indicators might not adequately reflect the wished result.

- Sub-optimising – performance indicators allow local goals (of a subunit) to be reached, at the expense of the organisation as a whole.

- Myopia – performance indicators allow to concentrate on short-term and easily reachable targets at the expense of the long-term targets.

3. Involvement issues – related to human and management issues, while implementing performance management (Boyne *et al.* 2005; Pollitt 2006).

One cause of the performance paradox discussed above is the dilution of indicators over time, meaning that after some time the indicators do not adequately distinguish good performance from bad. The dilution of indicators can be a result of four processes, which can be classified as involvement issues (Van Thiel and Leeuw 2002: 271):

- Positive learning – the performance improves to the extent that the sensitivity of an indicator vanishes (according to this indicator, all the subunits have reached a similar level).

- Negative learning – subunits have learned which indicators are measured and which are not, so the undertaken actions are manipulated to receive the best possible performance evaluation.

- Selectivity – poor performers are substituted with better ones, which decreases the variation between different subunits ('the stronger will survive').

- Denial – it is easier to deny and ignore the differences in performance than to improve.

In addition to dilution of indicators, other involvement issues can be highlighted (Smith and Goddard 2002):

- Self-satisfaction – performance indicators do not promote ambition to over-achieve mediocracy, the further effort does not result in adequate recognition. Pollitt (2013) illustrates a *threshold effect*, as a situation when a minimum target is set and when reached, no more motivation remains to pursue higher targets.

- Cheating – performance indicators allow the intentional manipulation of the data, such as 'creative' accounting and fraud, which lead to overestimation of performance. This is most common in the case of satisfaction surveys, when negative responses are not taken into account.
- Gaming – a change in behaviour to gain a strategical advantage to get a better evaluation (most commonly on indicators which assess the year-to-year growth). Pollitt (2013) notes the *rachet effect*, which might occur, when next year's target is based on last year's performance. In such case, the managers are tempted to just hit (but not exceed) the target, so it would not be more difficult to achieve next year's target.
- Stagnation – organisational paralysis, caused by extensive measurement, which in turn might lag behind newest practices. For Pollitt (2013), such phenomenon is described as *logic of escalation*, that when one starts to measure a public agency's performance quantitatively, it is likely to develop into a massive (and overwhelming) system of measurement.

A common thread of the previous lists is the finding that uniform understanding of performance indicators and their use as behaviour altering tool in the management is crucial. It should be noted that these issues can and will occur simultaneously. Much research focuses on assessing the performance indicators, but not much attention goes to the assessment of the system itself and whether it is necessary and compatible for the measurement at all.

This section pointed out some possible setbacks related to performance measurement in the public sector management. As performance measurement is costly, the expenditure has to be justified. It should be stressed that in undertaking the measurement, one has a trade-off between comprehensiveness and usefulness – if the system is too complex, it will result in a biased result and if too frugal, the result would be useless and shallow. The next section identifies the different dimensions of performance which could be the building blocks for a comprehensive PMS.

1.2. The dimensions of performance

1.2.1. Main stages in the process of providing a service by a public agency

An overview of providing a good or service by public sector and possible performance measurement criteria of the provision are given in Figure 1.1. The following logic model approach is based on a simple input-output model (Schacter

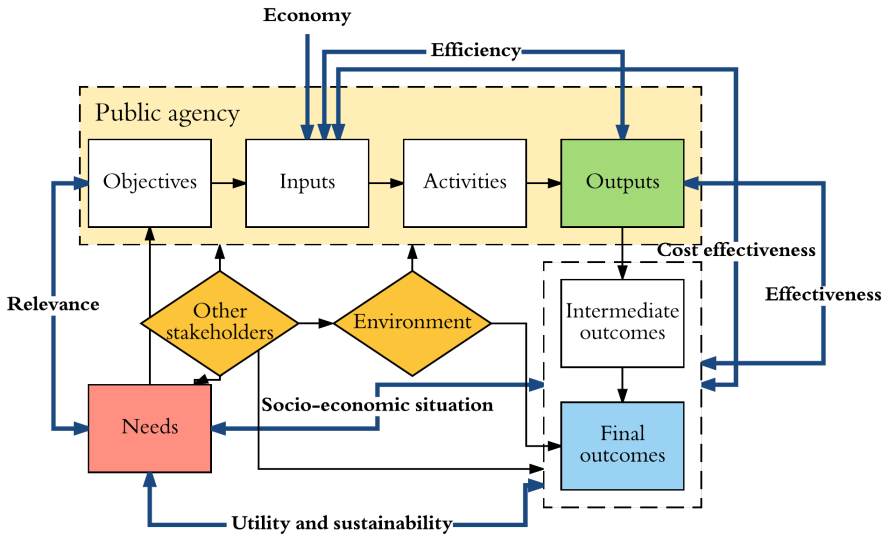


Figure 1.1: The conceptional model of performance (based on Pollitt and Bouckaert 2011; Van Dooren *et al.* 2015)

1999; Pidd 2012; Van Dooren *et al.* 2015). The society – in a very broad term – has some needs, based on which objectives are introduced to a public agency for it to fulfil these objectives and satisfy the needs. To do that, the public agency needs inputs (resources such as capital, time, knowledge and other assets). These are used in an organised way (activities) to create output (new value), which is illustrated as the mechanism of 'public agency' in the figure.

Inputs can be divided to two: ones that will end up as outputs (materials in public agencies, which are turned into tangible end-products like a passport or driving licence) and others, that are needed during the processes or providing services (such as labour, information, deliveries from other public agencies and office supplies).

Activities are necessary steps taken to transform the inputs into outputs and to achieve the objectives that are set to a public agency – therefore there is an opportunity to evaluate the process efficiency. Activities (processes) are distinguished as ones that are directly under the control of public agencies, and others that are affected only indirectly (such as involving voluntary rescuers).

As a result of activities, outputs (such as extinguished fires) are produced, which reflect the long-term goals of increasing the welfare of society. If the public agency is effective, then the outputs have an impact on outcomes. The provision approach sees the outputs as products or services, which are produced but not necessarily consumed (Van Dooren *et al.* 2015). Therefore, instead of extinguished fires, one might consider the working hours of rescuers on standby as an output measure. Such approach is common in public administration, allowing the people or entities to be held accountable. Public agencies are typically providing services of which they have no direct impact on the level of consumption. For example, in the FRS case, if less emergencies occur, the FRS brigades cannot be accountable for the low level of the consumption of their services.

Generally, defining and measuring the outcomes is more complex, because usually the phenomenon of the outcome is rather vague and ambiguous (such as safe living environment), thus it can be measured only indirectly (e.g. by public surveys). Since the outcome is dependent on the demand (needs), a situation may arise where the aforementioned phases result in a bias. Therefore, the volume and quality of goods and services provided by the public sector is questionable (Putnam *et al.* 1994; Van de Walle 2008). As the public sector is very diverse, and there is no way to directly measure the gains, implementing a variety of performance indicators is suggested (Propper and Wilson 2003; Modell 2003; Bird *et al.* 2005). In Figure 1.1, some indicators are more easily taken into account, such as inputs, because there is a long tradition of measuring them (accounting, budgeting).

The figure also illustrates the possibilities to measure the performance. It is assumed that the public agency acknowledges its objectives (the desirable outcome is unambiguous and clear) and acts towards achieving them (it is known how much resources are needed and which activities lead to intended results). The performance improves, when one uses the same amount of inputs but produces more outputs or produces same amount of outputs with less inputs.

Measuring the dimensions of performance (Pidd 2012: 16) answers three questions: How much was done? How well was it done? As a result, did someone's (mostly consumer's or taxpayer's) welfare increase? Measuring only inputs does not give an adequate answer, which is why activities, outputs and outcomes should also be measured.

Measuring activities answers the question how well the agency did, when providing goods or services. Process measurement represents also the evaluation of potential of provision. With respect to performance, the potential of provision can be taken as a capacity to create a certain base level (potential) to provide services, which is realised according to society's demand.

Measuring outputs answers the question of how many goods or services are produced by a public agency (such as number of departures). Outputs are mostly easy to define, to collect their data and understand their essence. Comparing the outputs to inputs, (cost-)efficiency is evaluated.

Probably the most fundamental objection to use outputs as performance indicators is the uncertainty of how and to what extent are the outputs and outcomes connected. There might be situations where creating more outputs result in worse outcomes (Bouckaert 1992). Another objection when using outputs as indicators is the fact that it does not take into account the quality of services. The quality dimension cannot be implied due to the non-existence of market price for goods provided by public sector, therefore it is impossible to use it as a feedback for quality. Quality is a demand side problem (Hjalmarsson 1991, cited in Jaldell 2002: 38), while efficiency is a supply side issue. In many cases it is impossible to evaluate, whether it is a 'good' quality (e.g. what characterises a fire well extinguished – quickly, safely, with low resources or minimising the losses?). It has been criticised that outputs only have a weak link with end-consumers, which is also true in the private sector (Burkhead and Hennigan 1978: 37).

Measuring the service quality gives a subjective answer to the question on how well the service is provided. In the case of FRS brigades, it might be the public opinion and satisfaction with the brigades. Although it is just a subjective opinion, it matters nonetheless. These opinions can be collected through public surveys. Fiszbein *et al.* (2011: 33) highlights three shortages of measuring service quality: 1) individuals may have different objectives for service provision than society as a whole does; 2) individuals may have been conditioned to have low expectations, leading them to report high satisfaction despite poor service; 3) information asymmetries; the providers have technical information and expertise that consumers lack (such as health services – patients' perception of service might be based on factors not relevant to technical quality – friendly doctor versus right prescription).

Measuring outcomes (or consequences) answers the question whether providing the service increased someone's welfare. In the case of FRS brigades, the outcome might be saved people or assets. In a wider sense, outcomes are also prevented fires and saved people and assets due to that. Measuring outcomes is substantially more difficult, because two situations cannot exist simultaneously. It cannot be evaluated how many fires would occur, if prevention activities would have not been carried out.

Outcomes or consequences are dependent on direct outputs as well as environmental factors, including other stakeholders (other public agencies, NGOs, firms, etc.). Stakeholders might influence the public agencies objective setting, or create independently outputs (like volunteers) that affect the outcomes. Using outcomes as performance measures are criticised due to the fact that public agencies options to directly influence them are low (Burkhead and Hennigan 1978). Similarly, the outcomes might not be a direct result of the work of public agency (maybe the pupils would have passed the tests without schools, patients would have recovered without the help of medicine workers, etc.). If outcomes are measured with respect to some certain standards (such as a relative level of outcomes without a public agency), the impact can be evaluated. The outcomes do not reflect the environmental conditions directly (regional specificities), which is why they should be adjusted accordingly (Propper and Wilson 2003: 255).

Outcomes or consequences are dependent on direct outputs as well as environmental factors. The distinction of outputs and outcomes is important because inputs trend might be quite different (cost-effectiveness studies versus efficiency studies; the first explores outcomes in relation to inputs and the second outputs to inputs). Outcomes are not only created as public sector supply but often also have some characteristics of public goods/services, such as non-excludability (as the consumers cannot be excluded from consuming the product if it is already created) and non-rivalry (as the consumption of one recipient does not affect the consumption of others). Due to these aspects, the environmental factors are important in performance measurement (Jaldell 2002).

The distinction of outputs and outcomes in the private sector is not problematic – to maximise the profit, one has to produce something for which there is demand (fulfilling the criteria of effectiveness) and do it without waste (fulfilling the criteria of efficiency). Therefore, the performance measurement is simplified in the private sector (Jaldell 2002: 38). In practice, there is seldom a choice whether to use outputs or outcomes: one uses the indicators, which are possible to define, measure and evaluate. If the goal is to measure the productivity, one should prefer outputs. If the line between outputs and outcomes is blurry, one should prefer indicators that are more useful for the decision makers (based on the objective of the public agency).

On top of such logic model one can also add the management perspective. Following the PDCA-cycle (plan-do-check-act), popularised by Deming (Deming 1952), the 'planning' step can take into account the needs, construct the objectives and formulate a plan for creating outputs. The 'do' step involves implementing the plan, or in other words turning inputs into outputs through necessary activities. The 'checking' step can involve the comparison of outputs and targeted outcomes. The 'act' step involves re-assessing the needs and adjusting the procedure to meet the new targets. From there, the cycle goes on. In the next subsection, such input-output model is explained using the current case of fire and rescue services.

1.2.2. The case of fire and rescue services in the public sector performance measurement

To conduct a performance measurement in a specific public agency, one must understand the essence and goals of the object under discussion. As public agencies vary in terms of provided services and structure, a tailor-made approach should be introduced. FRS, as the main interest of this thesis, are one of the emergency management services, which typically deal with reducing the adverse effects of diverse emergencies and disasters in the community. Dependent on the system, this might include the full ambulance service, but in most cases, it does not, as the FRS only provides first aid until the arrival of an ambulance.

The purposes why one should measure the FRS' performance is not widely discussed in the literature. Carvalho *et al.* (2006) conducted interviews with senior fire officers in Portugal, who agreed that performance indicators would be a useful management tool, which potentially increases the quality of services and allows comparisons between FRS brigades. In addition, the indicators can support the planning and development of budgets, contributing to the accountability of services, increasing the motivation of rescuers through incentives, rewards and sanctions, and stimulating the public's interest in the FRS as a public service. These purposes are in line with the general performance measurement purposes discussed previously.

The main goal of FRS is to provide a good, which can be called the feeling of safety (Jaldell 2002). The goal can be achieved mainly through two activities – prevention (including monitoring of safety) and reaction to the rescue events (see Figure 1.2). Therefore, the desired fire protection outcomes – which are not easily measured – include fires prevented or suppressed, and ultimately the human life and property preserved (Hatry *et al.* 2007).

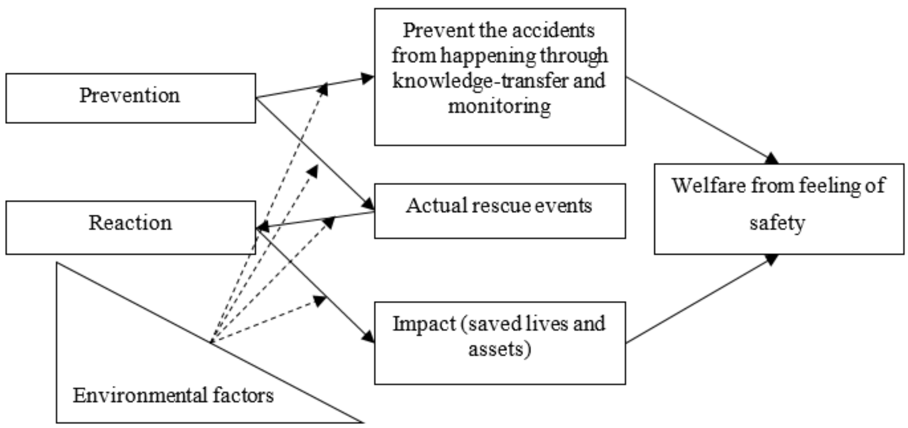


Figure 1.2: The main levers to reach the goal of a rescue service (based on Jaldell 2002: 45)

The main activity for a FRS agency is to prevent fires and other accidents (like traffic, water, chemical accidents, etc.) from happening. To improve the prevention, one can introduce building codes (non-flammable materials, good electrical fitting, centralised heating, etc.), fire inspections and prevention activities such as schooling and consulting. As outputs in that case, one could consider the number of people educated, inspections done, and so on. The number of fires prevented is not directly measurable. Another (directly unmeasurable) output from education is the peoples' knowledge how to react in case of an emergency, resulting in lower costs of an accident (saved lives and property).

If an accident still occurs, it might have been because of the poor performance of a FRS agency or due to factors which are outside the control of the agency (like environmental factors of FRS brigades, individuals and other organisations' activities). No matter how many resources are allocated to the prevention, it is highly unlikely to prevent all of the accidents. (Un)controllable environmental factors also have an impact on the reaction of a rescue event – the reaction depends on the service provider (the speed of providing the service; the number, skills and quality of service providers) as well as other factors like the socio-economic situation of the society, preparedness of other public agencies, public awareness, etc. The fire victims (as consumers of FRS) in the Deng *et al.* (2001) study put much emphasis on the response speed of FRS – the fire spreads rapidly and causes damage in a very short time period. Thus, following indicators are important – the identification and notification of fire (or other accident), the time between the call and departure from the fire department, the time to arrive at the scene, the time to get the fire under control. In addition, Jaldell (2017) has estimated a positive effect of longer response time to fatalities in residential home fires. He applied it using Swedish fire incident reports, and concluded that the decrease of median response time by one minute results in two additional lives saved annually. Therefore, many rescuers are constantly on standby, against the possibility of unexpected events, some of them highly unlikely but extremely large in scale when happening (ODPM 2005).

From a theoretical point of view, the FRS agency has many stakeholders they are accountable to – namely, politicians who set the direction of service provision, taxpayers who finance the services, and consumers who evaluate the quality of the service. Considering the stakeholders, the field of internal security in many countries relies (also) on volunteers (see Figure 1.3), who contribute to the solving of a concrete event as well as safekeeping the neighbourhood through prevention. Rescue services also cooperate intensively with other emergency services such as the ambulance service as well as police. The collaboration effects of FRS have not yet been studied to a large extent (for examples see Weinholt 2015; Puolokainen *et al.* 2018).

One can argue that the feeling of safety as well as willingness to contribute to safekeeping is stronger in regions where the community is stronger. According to a study on voluntary work, about a half of Estonians described themselves as an actor in their own right, one-third participates in local activities and only 6% are active in the activities or organise them (Turu-uuringute AS 2014). The main reason not to become a volunteer is the lack of time – yet for now the volunteers mainly contribute to the reaction in case of emergency events, which presumes special training and equipment. According to the Estonian Development Plan for Internal Security 2015–2020, it is important to increase the possibility to participate in the field that is familiar to the volunteer, especially conducting the prevention tasks (Siseministeerium 2015b: 15). The development plan also dictates that the willingness to get involved has to increase and this can be achieved by diverse civil initiative forms of participation (Siseministeerium 2015b: 27).

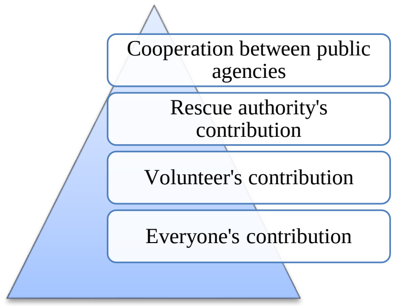


Figure 1.3: The pyramid of contributions to achieve the goals of internal safety (based on Siseministeerium 2015b: 13)

Given that most communities have decided not to exclude individuals from accessing FRS and one’s individual right does not diminish the rights for other consumers, it has elements of a public good. When a rescue service is provided, it is also difficult to exclude somebody – when a fire is extinguished, all of the neighbouring households (buildings) benefit. When conducting preventive tasks, it is theoretically possible to include only those who are the taxpayers (eligible for rescue services), but it might not be desirable, since prevented accident is beneficial for all of the nearby residents. Inversely, Ahlbrandt (1973) states that the FRS do not fit the description of a true public good, because the service is localised and targeted to a large extent to the individual consumer. The spillover effect is dependent on the community (density of population, housing type, etc.). Therefore, he suggests considering FRS to be somewhere between private and a quasi-public good. Due to the non-excludability, the individuals are able to consume the FRS, so there should be no incentive to provide it privately (as individuals would be rational and act as free riders), leading the public sector to undertake the provision of the service.

Rivalry is usually absent in case of prevention, in reaction there might be some rivalry if multiple accidents happen simultaneously and the rescue team has to choose, how to react (Jaldell 2002: 37). Therefore, one can conclude that the FRS has the main characteristics of a public service. Brueckner (1981) has evaluated the congestion of the FRS. A Samuelsonian pure public good is insensitive to the size of the consuming group; the congestion effect states that the individual public good consumption falls, holding public sector resources fixed, as the size of the consuming group increases. As a result, Brueckner (1981) stated that the FRS act rather as a public good in the sense of congestion effect (increasing the population in an area does not reduce the level of FRS). Another result from the study is the exhibit of increasing returns to scale in consumption of FRS. This means that a given level of FRS can be provided at a lower per capita cost in a larger community (area with higher population density).

The outcome of a FRS should be the higher feeling of security in the population. This would be difficult to measure, as peoples' psychology overvalues the closeness of a FRS brigade (Espenberg *et al.* 2013). The closeness of the FRS brigade does not prevent the accident from happening, so one could argue that people would be more careful and risk averse when the FRS brigade is actually further away. It would also result in higher independence in reacting to an accident (higher willingness to learn how to react in case of an emergency).

According to Jaldell (2002), the FRS might not be fully efficient in a competitive market sense due to three reasons: the public sector has also other objectives, public choice and property rights reasons, as well as monopolisation. By other objectives than efficiency, equity is often referred to (to provide a standardised level of service, independent of location); the objectives might also be vague and complex, which results in higher costs. According to the law of diminishing marginal utility, one should also take into account that when the goal is to reduce the number of fire deaths, saving additional lives is evermore expensive (and in some sense results in inefficiency as the social costs might at some point exceed the social benefits).

As discussed in Deng *et al.* (2001), people (as taxpayers and consumers of the service) would expect the best efficiency with the least resources – this expectation has obvious shortcomings and controversy. The FRS also has to control for casualties and property loss, which results in using many resources to save just a couple of lives. From the productivity evaluation point of view, it results a low efficiency, in case the judgement does not take the emphasis of human lives and public safety into account.

Deng *et al.* (2001) proposed building a standard concept for measuring the results of the FRS. As a novelty, they tried to take into account also people's feelings as beneficiaries of the public service. To achieve the goal, they used the Delphi method to reach consensus between fire victims, scholars who studied FRS and officers of FRS. As a result, six dimensional factors were mentioned to identify the quality of FRS: 1) timely firefighting response; 2) the correct judge of fire ground situations; 3) the control of the firefighting resources (achieve best results with constrained resources, such as manpower, materials and organisational culture); 4) the proper action of fire ground operations; 5) reducing life and property loss of fire; and 6) eliminating miscellaneous fire hazard. Many of these factors are also present in other previously discussed studies.

Review of fire and rescue services performance indicators based on performance management literature

Performance measurement and management has been widely introduced in the context of NPM. FRS, as mainly offered by public agencies, have also been a target of performance measurement and quantification of goals. In the frontline of performance measurement, typically the UK is referenced (Hood 1991; 1995; Kloot 2009). Countries vary from one extreme to another in using performance indicators for FRS – an article indicating the both ends compared Portugal and UK (Carvalho *et al.* 2006) for that matter. Portugal did not have any publicly available performance information, but it was encouraged by interviewed senior officers of fire and rescue services. The UK on the other hand, reported 51 indicators in 2001/02, which demonstrates the possibility of over-auditing, prescribing and regulating (as pointed out in Power 1997). Indicators themselves do not achieve better services. Bouckaert (2009) has noted that when a problem in the public sector management occured, the solution in reality was to have more of the same. A problem with indicators, objectives or audits resulted in more of the same. Boyne (2002) brought to attention that only 76% of performance indicators used in the UK were relevant, and the evidence that these indicators have had an effect on changing the FRS in England and Wales, is lacking.

The systematic conceptualisation of FRS' outputs and outcomes as performance information dates firstly in the 1970s. Examples from the United States include one of the first works from National Fire Protection Association (e.g. Schaenman and Swartz 1974; Schaenman 1977), which according to its website is a global non-profit organisation, established in 1896. Its functions are to deliver information and knowledge through more than 300 consensus codes and standards, research, training, education, outreach and advocacy; and by partnering with others who share an interest in furthering their mission (NFPA 2016). Therefore, the first initiative came from the organisation focussing on standards and knowledge. The most thorough literature review on quantification of FRS output measures with an empirical application is so far conducted by Jaldell (2002), without duplicating the results of these articles in the current thesis.

The FRS performance can be measured at two levels (Jaldell 2002: 26) – at an aggregated macro or national level, giving an assessment to the FRS as a whole, or within-unit level, as in the case of FRS brigades.

As pointed out by multiple articles (e.g. Brignall and Modell 2000; Kloot 2006; 2009), the result of NPM reforms relating performance indicators is focussing too much on financial aspects and not enough on operational aspects. This might be caused by the fact that NPM is introduced mainly from the accounting field and not (public) management. As a result, this is too one-sided and will lead to a biased result, when considering a multi-dimensional performance measurement.

Another issue, pointed out by Carvalho *et al.* (2006) and Espenberg *et al.* (2013), is that the FRS are mainly provided as municipal level services. Therefore, the relationships between national and municipal managers have some built-in difficulties – the national managers do not have a comprehensive set of information about the local level to coordinate the activities and municipal managers would like to have as much independence as possible. Thus, issues of accountability arise.

The following section analyses and concludes collected performance indicators implemented for FRS, based on articles which have systematically analysed countries or regions performance indicators. The complete table of indicators is given in the Appendix 2. The table includes indicators from Australia, the United Kingdom (England, Wales and Scotland) and the United States as some of the frontrunners of performance management possibilities.

Carvalho *et al.* (2006) compared the PMS of the UK and Portugal on FRS using mixed methods of document analysis and interviews with senior fire officers. As a result, they concludedf that the UK might overuse the monitoring which does not have enough effect on the services and Portugal lacks the indicators overall. Indicators collected by the Office for National Statistics in the UK in 2006 are given in the appendices, and the report also stated many shortcomings (ODPM 2005). Andrews (2010) evaluated the impact of modernisation on FRS performance, based on data for 46 fire authorities in England for 2001–2006. As a result, he found that the performance had increased compared to the pre-modernisation period. As the key dimensions of FRS performance, fire prevention (accidental fires), community safety (false fire alarms), resilience and emergencies (non-fire incidents), equality and diversity (the CRE Standard Score – the requirements of the duty placed on local authorities), staff safety (rate of firefighter injuries) and well-being (ill-health retirements) were used. Murphy and Greenhalgh (2013) review the performance management regimes in the UK (England and Wales) and propose a new, more efficient and effective regime.

The indicators used by Scotland were sharply defined, focussing on service delivery (Carvalho *et al.* 2006). This study was followed by Kloot (2009), who used content analysis on publicly available annual reports of FRS to determine performance indicators used for Australian state of Victoria. As a result, she points out the focus on merely financial indicators and not many operational indicators. The lack of demonstration of accountability to external stakeholders has been pointed out. Kloot showed her surprise in the results, as NPM has been implemented for many years in Australia. Then again, internal reports are conducted for managerial and professional stakeholders, which are not publicly available.

Scotland merged the eight FRS into the national Scottish Fire and Rescue Services (SFRS) in 2013. Audit Scotland (2012) provided guidelines for successful mergers, which have been mostly followed by SFRS (Audit Scotland 2015). Important to the current thesis, the importance of developing a PMS while planning the merger was stressed. As learning lessons from former mergers in Scotland, the absence or underdevelopment of PMS was pointed out. In addition, information about the baseline was lacking (the first year of operation of each body – unit costs, staffing levels and quality of services). It is difficult to demonstrate the impact of changes without proper data and it also hinders the ability of stakeholders to scrutinise the performance and expected benefits of the centralisation. Despite the budget constraints and staff cuts, the merged organisations in Scotland reported that they continue to deliver business as usual – this prevails mainly due to lack of performance information on service provision quality.

A report by Flynn (2009) identifies and presents potential performance indicators based on the United States, and discusses the possibilities and threats of the proposed indicators (for the full list, see the appendix). Hall *et al.* (2008) analysed the possibilities to evaluate the effectiveness of code compliance (supervision conducted by FRS). This, however, is not in the scope of the current thesis, as the supervision is usually not directly conducted by FRS brigades.

One reviewed issue on performance is attributed to representative bureaucracy theory, which notes that the public agency should consist of people like the population they serve. Andrews *et al.* (2014) analysed the relationship of FRS brigades' performance in the UK with respect to the gender and minority ethnic representation. They found some evidence that the more representative the fire authorities, the more effective the organisation, especially in the non-core tasks, which require more provider-client relationships to deal with. Therefore, one could argue for adding the gender and minority ethnic representation of public officials as a performance indicator illustrating the versatility of the workforce.

Krasuski *et al.* (2012) proposed a method to evaluate the commanding efficiency of FRS officers, based on incident data reports. The method was based on case-based reasoning, which is an artificial intelligence technique – it retrieves solutions which were used in the past to solve similar problems and then adapted to the current problem, thus introducing self-learning. As a result, the method used concluded that the shorter the duration of the action, the better would be the commanding officer.

The literature review on environmental factors determining the fire risk in urban residential fires is comprehensively undertaken by Jennings (1999; 2013). Most of the underlying seminal works assessing the effects of environmental factors on fires are exploratory in essence and are not based on a firm theoretical approach. From the very first studies, it was found that the occurrence of fires is not uniform or randomly distributed, but there is a systematic variation in the nature and severity of the fire problem across urbanised areas. One of the first relationships found between fire occurrence were associated with poverty and housing quality (for a list of articles confirming this relationship, see Jennings 2013; this is also confirmed in Estonia, based on interviews with senior fire chiefs, see Espenberg *et al.* 2013). The introduction of geographic information systems (GIS) has brought many new opportunities to analyse the relationships between fire risks and environmental factors. As a result of the literature review, Jennings (2013) comes up with a conceptual framework of fire risks. The framework consists of characteristics affecting the risk, which can be grouped as:

- dwelling characteristics (dwelling materials, structure, age, equipment, electrical fittings and content);
- physical environment (setting, vegetation cover and topography);
- neighbourhood characteristics (demography, cultural practices and socio-economic status);
- weather conditions (rainfall/snowfall, wind speed and temperature);
- behaviour (perception/attitude, values and beliefs, socio-economic status and cultural background);
- group behaviour (family lifecycle, household size and household composition);
- calendar events.

As one can conclude, there is no consensus in characterising the system of FRS provision, which complicates the further analysis. The FRS have been influenced by NPM-related reforms with respect to performance measurement, but as a complex service influenced by many external characteristics, it has not produced clear results. The literature analysed is lacking a systemic approach, as most of it is based on practical or exploratory grounds. As a next step, the evaluation criteria for performance measurement is analysed.

1.2.3. Evaluation criteria for performance measurement

The evaluation criteria can be considered as different relationships between the dimensions of performance. At the most basic level, according to Ross and Burkhead (1974) and Jaldell (2002), there are five possibilities for relating the outputs to inputs:

- Using work measures, which are defined similarly as are process indicators in this thesis.
- The measurement of outputs by inputs, e.g. the costs of service also indicate the result.

- The determinants approach uses costs as a dependent variable in multiple regression with all factors influencing the costs as independent variables. Two weaknesses are mentioned: in that case only supply side of services is considered, and there is no underlying behavioural theory, so the chosen factors have no theoretical justification.
- Using changes in outcomes or effects related to inputs.
- Using changes in the quantity of direct outputs related to inputs (production function approach), which is described more thoroughly in the following text.

Various dimensions of performance can be assessed in a number of different evaluation criteria (Bouckaert 1992; Boyne 2002; Mayne and Zapico-Goñi 2007; Pidd 2012). Taking into account the dimensions of supplying goods/services by the public sector, an overview of the main evaluation criteria for public agency's performance measurement is given. The general conception of performance criteria is to willingly or unwillingly evaluate the transformation (from inputs to outputs) and they have some similarities. The most well-known are economy, efficiency and effectiveness (Pidd 2012: 24).

- Economy – focuses only on the costs of services and therefore indicates only the use of inputs and describes little about how well a public agency is meeting its aims.
- Efficiency – can be defined as a comparison between the actual and optimal inputs-outputs sets (Koopmans 1951; Farrell 1957; Daraio and Simar 2007; Cooper *et al.* 2011). The comparison may be made in two forms: the optimal set of inputs in case the outputs are fixed and the optimal set of outputs in case the inputs are fixed.
- Effectiveness – will be used in case there is an objective function and therefore goal attainment can be used (Bogetoft and Otto 2010). In other words, effectiveness relates to the social objectives of the agency and is therefore a measure of how well it is meeting those objectives (Bouckaert 1992; Pidd 2012). As an example, when the FRS are meant to increase the safety of the public, the question of how to measure safety arises.

Although the terms efficiency and effectiveness are fairly specific, in 'real-life' they are frequently used in loose, vague and/or inconsistent ways (Pollitt and Bouckaert 2011: 15). Bouckaert (1992) emphasised that the efficiency and effectiveness might not be positively related – increased efficiency can lead to decreasing effectiveness. He states that this is related to the quality dimension, as producing outputs more efficiently may cause a loss in the quality of the final output and thus result in lesser effectiveness.

In addition to these three most common evaluation criteria – which are very popular in the NPM context – are many others. Depending on the public agency in focus, they have a higher or lower importance. A more versatile overview is given by Boyne (2002) and Pidd (2012: 24–25):

- Equality – evaluates whether the recipients of services are treated equally or the benefits of the service are divided unequally. In the case of FRS brigades, this might be the reaction time in case of an emergency (how far away the brigade is from the emergency location). In the equality dimension, horizontal and vertical equality are distinguished. In the case of horizontal equality, the treating all the consumers in the same way and, in case of vertical equality, the compliance of services to the needs of consumer is meant (horizontal equality means that for example everyone has the possibility to visit a doctor and vertical equality means that everyone receives a proper treatment for their case).
- Utility/efficacy – similar to the effectiveness, but more general and answers the question whether the service functions at all. In the case of FRS, this is whether the activities they do increases the safety of population.
- Ethicality – evaluates, whether the provided service follows the accepted ethical norms in the society.
- Productivity – total factor productivity is defined as a ratio of total outputs into total inputs.
- Process evaluation – usually related to workload or time, how quickly the inputs transform to outputs.
- Quality/satisfaction criteria of the service – consumers' (population) satisfaction with the service.

Conjointly, one could relate the evaluation criteria with inputs, activities, outputs and outcomes, which are provided in Table 1.1.

The above list of evaluation criteria is not all-embracing, and the definitions and division varies depending on the approach of different authors and fields of research. It is also important to emphasise that different criteria are relevant to different principals. Generally, the economy, efficiency and effectiveness ('the three E's') are associated with the management, and to the political traditions, mostly efficacy, ethicality and representativeness, quality and accountability are considered. From the legal point of view, the equality aspect is considered (Andrews and Van de Walle 2013: 5).

It is important to state that the performance of a public agency is a comprehensive phenomenon, which requires a PMS to adequately and

Table 1.1: The relationship between evaluation criteria and dimensions of publicly provided good/service supply

Evaluation criteria	Inputs	Activities	Outputs	Service quality	Outcomes
Economy	x				
Efficiency	x	x	x		
Effectiveness	x		x	x	x
Equality				x	x
Efficacy				x	x
Ethicality		x		x	x
Productivity	x		x		
Quality criteria of the service		x		x	

Source: Pidd (2012: 25).

systematically characterise its aspects and relations. In addition, every service provided by a public agency is a specific and unique case, so a standardised approach is complicated. The possible solutions are discussed in the following subsection.

Performance measurement systems

It is seldom possible or reasonable to use only one performance indicator. Poister (2008) proposes to systematise every performance measurement, thus creating a PMS, which includes data collection, analysis and implementation of indicators. The creator as well as user of the system are interested in the results – this however leaves open the priorities (weights), which are attributed to the results. The first option to solve this is to give the creator of the system the power to attribute the priorities based on his/her judgement. Alternatively, the creator might choose not to define the weights and leave it to the user of the PMS (see also Bogetoft and Otto 2010). Although these results might be more complicated to interpret, they are also more difficult to affect dishonestly (such as gaming, defined above) (Propper and Wilson 2003: 254).

One might also consider a 'mixed' option, when the creator of the PMS collects information from the users of the system to define the priorities. The users might be different managers, stakeholders, public agencies from different ranks, jurisdictions, etc. and their input is used as a source for social weights. This can be a result of negotiation. The priorities can be obtained by different methods, which follow, for example, multi-criteria decision-making possibilities (weighted sum model, weighted product model, Saaty analytic hierachy process (AHP), revised AHP method, ELECTRE, TOPSIS, among others) (Triantaphyllou 2000).

In the literature, it is often stressed that 'what you measure is what you get' (Poister 2008: 4). In case the goals of a public agency can be linked directly to a certain performance indicator, it is possible to focus on its priorities (by informing the employees and consumers of what is important from the strategic point of view). Therefore, it would be preferable to compile a system of performance indicators which targets concretely and comprehensively the nature of the organisation and possible developments, and allows to evaluate the performance. According to Poister (2008: 184), the suitable performance indicator system should satisfy the following conditions:

- Be a mix of process and output indicators which are fundamental for the organisation. Measuring the output indicators is mainly associated with the tactical level of planning and process indicators with operative planning.
- Should stress the general complex indicator, which is important for the agency as a whole, although its elements can be compiled by the results of different subunits' performance.
- Should contain nominal and qualitative as well as quantitative indicators.
- Should be linked to the KPI targets, so one can evaluate the progress.
- In some cases, the system should be managed by the top-down approach, so that all subunits can coordinate.

A successful benchmarking of performance measurement in a public sector follows three characteristics (Graham *et al.* 2005: 490):

- Firstly, it should focus on processes and strategical actions, and not outcomes, as they are difficult to measure and not directly influenced by the public agency. Fountain (2001: 58) stresses that the performance indicators should contain quality and trust indicators, because the public agency does not only have to be economically efficient, but also just, equal and trustworthy. Therefore, the benchmarking known from the private sector is not good enough, because it is too one-sided.
- Secondly, benchmarking should focus on the internal changes of a public agency, because there are seldom competitors suitable for comparison in a meaningful way. This point tends to diminish in time, because evermore the quasi-markets are introduced in the provision of public goods and services, where there exists some sort of competition between public agencies or third parties. As competition, comparison of subunits' performance measurement can also be considered.

- Thirdly, the results of benchmarking should be publicly available, to inform all stakeholders of the results and evaluation methodology.

One of the weaknesses of benchmarking becomes evident from the list above – it is difficult to evaluate the outcomes of a public agency as well as prioritising different indicators. This, however, is crucially important in the development of a performance indicator system when considering the strategical planning. Some management concepts reject the benchmarking due to that and deal with legality measures and delegation of competences instead (such as the Harzburg Model and bureaucratic concepts like OPA).

This section defined the dimensions of performance with respect to public agencies. As the public sector performance is a subject to many fields of research, some discords in definitions prevailed and were clarified. The performance evaluation criteria were analysed and integrated into the performance dimensions. For a more concrete approach, production theory is introduced in the following section.

1.3. Production theory as a basis of performance measurement

In this section, the definitions of efficiency and different theoretical possibilities to measure them are analysed from the production theory point of view. The discussed equations and explanations are mainly based on Bogetoft and Otto (2010) and Simar and Wilson (2013). Similarly, the theoretical economic considerations are clarified in more detail for example by Daraio and Simar (2007), Fried *et al.* (2008), Cooper *et al.* (2011), Simar and Wilson (2013), and many more, who in turn follow the seminal works of Koopmans (1951), Debreu (1951) and Farrell (1957).

Following the similar approach discussed in the last section from the public administration perspective, the basic production theory considers a producer who transforms inputs into outputs. There may be one or many inputs and outputs involved. Production is constrained by what is possible or feasible, so $x \in \mathbb{R}^p_+$ and $q \in \mathbb{R}^p_+$ denote the vectors of input and output quantities and let $T = \{(x, q) | x$ can produce $q\}$ denote the set of feasible combinations of inputs and outputs, also known as the production set. Any output quantities of q can be produced using input quantities x if and only if $(x, q) \in T$. Not all the points in T are equally desirable.

Standard microeconomic theory of the firm imposes the following three assumptions:

A1. T is closed. That ensures that the boundary of T is included in T.

A2. All production requires the use of some inputs: $(x, q) \notin T$ if $x = 0$ and $q \geq 0$. There are no 'free lunches' – one cannot produce something with nothing.

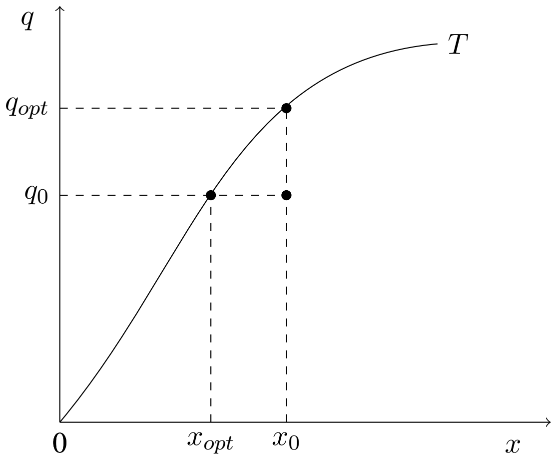


Figure 1.4: Relationship between outputs, inputs and production frontier

A3. Inputs and outputs are freely disposable: if $(x, q) \in T$, then for any (x', q') such that $x' \geq x$ and $y' \leq y$, $(x', q') \in T$. The free disposability assumption is sometimes also called strong disposability and is equivalent to an assumption of monotonicity of the technology. It also characterises the possibility of wasting resources (producing less with more resources).

Consequently, in Figure 1.4, a production set is the line T and everything that is interior.

In general, the total factor productivity (TFP) is defined as the production units' outputs quantities and inputs quantities ratio (Fried *et al.* 2008: 7). It is easy to evaluate productivity when there are only one type of input and output (see Figure 1.4, marked with a black dot in x_0, q_0). If multiple inputs are used to produce multiple outputs, aggregation is needed so the productivity can be described as the scalar ratio in analogy to the production units' outputs quantities and inputs quantities. Thereby it is possible to distinguish *partial* productivity, when only a single factor of production is considered and *total* productivity, when all factors of production are taken into account (Daraio and Simar 2007: 13). As one can see, $q_0/x_0 < q_0/x_{opt}$.

For efficiency measurement, the upper boundary of the set T is relevant as it describes the frontier of production. In other words, the efficient subset of points in T lies on the upper boundary – the locus of optimal production plans (e.g. minimal achievable input level for a given output, or alternatively maximal achievable output given the level of inputs). The upper boundary of T can be noted as:

$$T^{\sigma} = \{(x, q) \in T | (\gamma^{-1}x, \gamma q) \notin T \forall \gamma \in (1, \infty)\} \tag{1.1}$$

This is referred to as the *technology* or *production frontier*, and is given by the intersection of T and the closure of its complement. Technically inefficient firms (or in this case, public agencies) operate at points interior of T and technically efficient ones along the technology defined by T^{σ}.

Next, one should think of additional factors which influence the production frontier. Following O'Donnell (2016c), the *technology* is a technique, method or system for transforming inputs into outputs and it is common to make assumptions about the used technology. To make it more specific, the introduction of additional environmental factors z (variables that affect the production decisions and production, to a large extent known at the beginning of period when input decisions are made) and states of nature s (possible environmental factors, unknown at the beginning of period when input decisions are made) is necessary. The period-t metatechnology (a set of existing technologies) in a production environment characterised by z_t and state of nature s_t is defined as:

$T^t(z_t, s_t) = \{(x, q) : x$ can produce q in period t in environment z_t and state of nature $s_t\}$.

This could be referred to as a *period-environment-and-state-specific production possibilities set*. The boundary can be referred to as *period-environment-and-state-specific frontier*. It is assumed that all technologies available in period t are regular, thus:

T1: $(x, 0) \in T^t(z, s)$ for all $(x, z, s) \in \mathbb{R}_{+}^{\mathrm{M+J+S}}$ (inactivity)

Using a given set of inputs, producing nothing is possible.

T2: $\{q : (x, q) \in T^t(z, s)\}$ is bounded for all $(x, z, s) \in \mathbb{R}_{+}^{\mathrm{M+J+S}}$

Boundedness implies that one cannot produce unlimited levels of outputs with a given set of inputs.

T3: $q \geq 0 \Rightarrow (0, q) \notin T^t(z, s)$ (inputs weakly essential, no free lunch)

Non-zero output levels cannot be produced from zero levels of inputs.

T4: $(x, q) \in T^t(z, s)$ and $0 < \lambda \leq 1 \Rightarrow (x, \lambda q) \in T^t(z, s)$ (outputs weakly disposable)

If a vector of outputs, q, can be produced from a given input vector, x, then any contraction of q, λq, with $0 < \lambda < 1$, can also be produced with x inputs.

T5: $(x, q) \in T^t(z, s)$ and $\lambda \geq 1 \Rightarrow (\lambda x, q) \in T^t(z, s)$ (inputs weakly disposable)

If a vector of inputs, x, can produce a given output vector, q, then any increase of x, λx, with $\lambda > 1$, can also produce q outputs.

T6: the set $\{q : (x, q) \in T^t(z, s)\}$ is closed for all $(x, z, s) \in \mathbb{R}_{+}^{\mathrm{M+J+S}}$ and the set $\{x : (x, q) \in T^t(z, s)\}$ is closed for all $(q, z, s) \in \mathbb{R}_{+}^{\mathrm{N+J+S}}$

Closedness is essentially a mathematical requirement (Coelli *et al.* 2005).

Similarly, the **output set**, $P^t(z_t, s_t) = \{q : x$ can produce q in period t in environment z_{it} and state of nature $s_{it}\}$. Properties comply with the same regularity criteria as T1–T6.

And the **input set**, $L^t(z_t, s_t) = \{x : x$ can produce q in period t in environment z_{it} and state of nature $s_{it}\}$. Properties comply with the same regularity criteria as T1–T6.

Technical efficiency is a measure following on productivity – and can be defined as the distance (comparison) from the observable units' points of input-output to the production frontier (Daraio and Simar 2007; Fried *et al.* 2008), making it a relative measure. The comparison can be presented in two forms: the ratio of observable and optimal outputs in case of fixed inputs and the ratio of observable and optimal inputs in case of fixed outputs. Koopmans (1951: 60) defined the inputs-outputs vector technically efficient then, and only then, when increasing whichever output or decreasing whichever input is possible, only in the case when it follows a decrease in another output or increase in another input, respectively.

Debreu (1951) and Farrell (1957) decomposed efficiency assuming the lack of scale effect: technical efficiency, price efficiency (or allocative efficiency) and overall efficiency (cost-efficiency). The value of efficiency lies between (0;1], being 1, when the observed unit is fully efficient and no waste is produced. The price efficiency is calculated by comparing the observed units' average costs (AC) with average costs of those units, whose AC is the smallest (least-cost producers). The value also stays between (0;1], being 1 in case the observed unit uses cost-minimising input set. The overall efficiency is calculated by multiplying the two aforementioned parts of efficiency.

Coming back to the illustration (see Figure 1.4) of some problems with respect to efficiency evaluation, the following situation is created: a public agency has produced a certain amount of outputs (q_0) using a certain amount of inputs (x_0). Is it technically efficient (and how efficient)? To evaluate the technical efficiency, the production frontier might be used as a benchmark. When the frontier is as shown in the Figure, one can state that the public agency has been inefficient – one could have produced the same amount of outputs with less inputs, namely x_{opt} (or a bigger amount of outputs with the same inputs, or some combination of both possibilities). The excessive amount of inputs (vertical distance between the actual and minimal inputs) describes the absolute value of inefficiency. Technical inefficiency is in that case:

$$technical\ inefficiency = \frac{actual\ inputs - minimal\ inputs}{actual\ inputs} = \frac{x_o - x_{opt}}{x_0} \tag{1.2}$$

In a typical case, the smaller the technical inefficiency, the better the performance. Similarly, relative efficiency can be measured:

$$technical\ efficiency = \frac{minimal\ inputs}{actual\ inputs} = \frac{x_{opt}}{x_0} = 1 - inefficiency \quad (1.3)$$

In a typical case, the higher the technical efficiency, the better the performance.[1]

The input and output distance functions represent the reciprocals of technical efficiency measures in a generalised case. The input distance function (IDF) gives the reciprocal of the smallest fraction of inputs necessary for a public agency to produce its outputs and the output distance function (ODF) the largest factor by which a public agency can produce more outputs with the given inputs.

To understand the difference between input and output orientation, firstly the input and then the output-oriented efficiency is explained. Efficiency towards input minimisation can be defined as the smallest multiplier E of the input-output set (x, q) that is needed in case of x inputs to produce q outputs, or if the multiplier of x is smaller than E, it is not possible to produce $Ex = q$ amount of outputs. Thus,

$$E(x, q) = \min\{e|ex\ can\ produce\ q\} = \frac{|x_{opt}|}{|x_0|}. \quad (1.4)$$

On the other hand, the multiplier E can be viewed as a rate through which it is possible to save from inputs $(1 - E)x$, by still producing the amount of q outputs.

Similarly, the output-oriented efficiency can be found – it is defined as a biggest multiplier F of an input-output set (x, q), which is necessary to produce q in case of x amount of inputs, i.e. if the outputs q multiplier is bigger than F,

[1]Since the metatechnology is assumed to be regular, the equivalent representations of $T^t(z, s)$ include period-environment-and-state-specific distance functions, which are defined as:

Input distance function

$D_I^t(x, q, z, s) = \max\{\theta > 0 : x/\theta \in L^t(q, z, s)\}$, which has following properties:
$D_I^t(x, q, z, s)$ is non-decreasing in x and non-increasing in q;
$D_I^t(x, q, z, s)$ is linearly homogeneous in x;
$D_I^t(x, q, z, s)$ is concave in x and quasi-concave in q;
if x belongs to the production possibility set of q (i.e. $x \in L(q)$), then $D_I^t(x, q, z, s) \geq 1$; and distance is equal to unity (i.e. $D_I^t(x, q, z, s) = 1$) if x belongs to the 'frontier' of the input set.

Output distance function

$D_O^t(x, q, z, s) = \min\{\rho > 0 : q/\rho \in P^t(x, z, s)\}$, which has following properties:
$D_O^t(x, 0, z, s) = 0$ for all non-negative x;
$D_O^t(x, q, z, s)$ is non-decreasing in q and non-increasing in x;
$D_O^t(x, q, z, s)$ is linearly homogeneous in q;
$D_O^t(x, q, z, s)$ is quasi-convex in x and convex in q;
if q belongs to the production possibility set of x (i.e. $q \in P(x)$), then $D_O^t(x, q, z, s) \leq 1$; and distance is equal to unity (i.e. $D_O^t(x, q, z, s) = 1$) if q belongs to the 'frontier' of the production possibility set.

one cannot produce from x amount of inputs the $x = Fq$ amount of outputs. The output-oriented efficiency of unit (x_0, y_0) is then $F_{x_0} = \frac{q_{opt}}{q_0}$, following $q_{opt} = F_{x_0} q_0$. For the unit (x_0, y_0), it is technically possible to produce q_{opt} amount of output, thus the potential growth of output is $(F_{x_0} - 1)q_0$. The output-oriented efficiency is then:

$$F(x, q) = \max\{f|x\ can\ produce\ fq\} = \frac{|q_{opt}|}{|q_{x_0}|}. \quad (1.5)$$

To conclude, the smaller the E and the bigger the F, the more technically inefficient is the unit.

In the previous illustrated problem, the technical efficiency was evaluated in case of one input and one output. Often the aggregation of inputs or outputs is not possible nor wished. In addition, it is complicated to aggregate the utilities of the whole or a system. Thus, in the benchmarking the single input and single output cases are excluded and approached more systematically: the structural units use multiple inputs to produce multiple outputs, which might be connected and replaceable with each other. The structural unit is considered as a resource transformer to products and services. The transformation is affected by uncontrollable factors as well as (non)observable skills and contributions of a structural unit. Therefore, the main idea is to measure inputs, outputs and uncontrollable factors, and on the basis of that evaluate the (non)observable skills and contributions of a structural unit (the performance).

The simultaneous evaluation of multiple inputs and outputs makes benchmarking more difficult, because some structural units might receive a good result in one dimension and a bad one in another. In case of two inputs (see Figure 1.5, on the left) and two outputs (see Figure 1.5, on the right), an input isoquant for fixed outputs and output isoquant for fixed inputs are formed accordingly. Conjointly, the input set x and output set q of which the efficiency is evaluated, are marked. In both cases the sets are inefficient, because in the first case it is possible to spare inputs to produce the same amount of outputs and in the second case it is possible to produce more outputs with the same inputs. Since there are many possibilities to increase the efficiency, they should be summed.

In those cases, the Debreu-Farrell's approach is used, which is a generalised way of evaluating efficiency for multiple inputs or outputs (previously described with one input and output). Typically, proportional changes are in focus – i.e. different inputs and outputs would change by the same percentage point (dashed line). Thus, the input-oriented technical efficiency evaluates how much would be possible to decrease proportionately the inputs, to produce the same amount of output. Otherwise, multi-directional and hyperbolic efficiency analysis can also be conducted.

As long as the behaviour of a producer (such as public agency or its subunit) is known, which is stated through outputs and inputs, and an appropriate model can

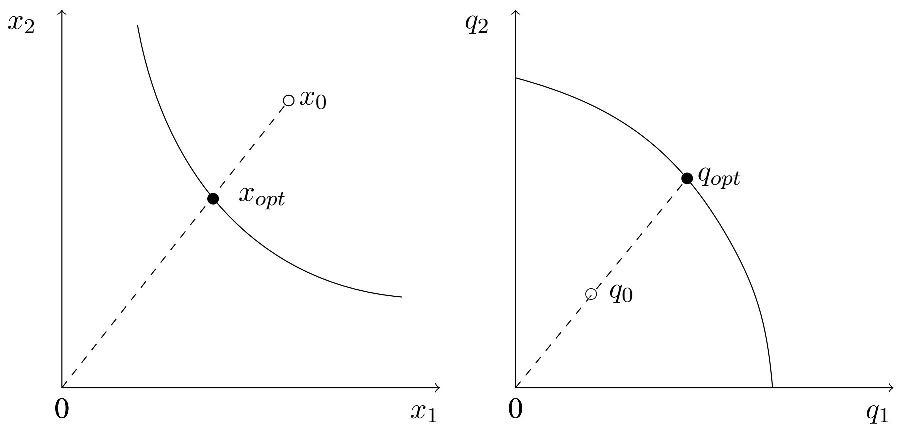


Figure 1.5: Input and output-oriented technical efficiency for two inputs and two outputs

be given, which declares the frontier, it is quite easy to evaluate the performance. Such approach is called rational ideal evaluation (Bogetoft and Otto 2010: 7) and it follows the theory of the firm. The evaluation is rational, because the behaviour (such as minimising the volume of inputs) and possibilities (such as production/cost frontier) are specified, trying to find the best way to fulfil the preferences. It is ideal, because it is assumed that all relevant information is available. Thus, it is just an optimisation problem.

Microeconomically, the effectiveness is described as the ability to choose the best way to achieve ones' targets (see Figure 1.6 for the two-output case). The effectiveness can be distinguished from efficiency by stating clear preferences in the form of utility function. In that sense, effectiveness can be treated as an evaluation on achieving the goals. In case of efficiency evaluation, only the (proportional) distance from full efficiency (T^{σ}) is considered, without making any assumption about the utility function. Possibilities are given by the production function T (in that case, a convex output isoquant). By definition, T is the highest possible volume of outputs with given inputs. The preferences are declared by the utility function $U(\cdot)$ (in that case the linear indifference curves), which express the combinations of outputs as equally beneficial. The rational ideal evaluation would then compare the actually achieved utility $U(A)$ and the maximal possible utility $U(ideal)$. In that case,

$$effectiveness = \frac{actual\ performance}{ideal\ performance} = \frac{U(A)}{maxU(y)} = \frac{U(A)}{U(ideal)}. \quad (1.6)$$

In the actual evaluation, this sort of approach is practically impossible to apply

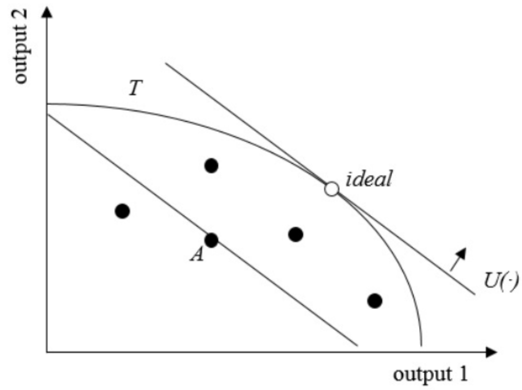


Figure 1.6: Rational ideal evaluation to maximise two outputs

– it is difficult to evaluate an adequate utility function U, also the production function T (or cost function) is unknown. Thus, none of the components are initially known, so one has to stick to a conceptual approach. In principle, benchmarking is an attempt to approximate this approach using real data. To an analyst, only the sample with inputs and outputs is available. To achieve this, the utility function is waived and relative comparison of different observations is applied (Bogetoft and Otto 2010).

The lack of knowledge of clear preferences and priorities is solved by 1) considering the efficiency as effectiveness, and 2) compensating *a priori* the technical information (production or cost function). To do that, weak and flexible assumptions are made by evaluating the function with respect to the best practices of the real data. In the following figure (Figure 1.7), ideally the effectiveness $\frac{U(A)}{U(ideal)}$ is evaluated, but it is impossible due to unknown U and T. The unknown utility function is compensated by focussing on the efficiency, i.e. how much the unit A could proportionately increase its outputs in the conditions of T. To do that, technical efficiency to the T is found, which is illustrated by $\frac{A}{\tilde{F}A}$ in the figure. The unknown production function is compensated by evaluating other units and finding the best of them. This in turn, is the basis for approximation of production function T^{*}, which becomes the benchmark. Following that, the technical efficiency to the T^{*} can be found, illustrated by $\frac{A}{FA}$. The value of the efficiency to T^{*} is higher than the efficiency to T, because the standards are lower – in reality no unit achieves the theoretically ideal result.

Evaluating the production frontier T is one of the most difficult processes. In case the data does not contain noise (deterministic), the production function consists of sets of achievable inputs and outputs – the volume of outputs they have produced from the inputs, is achievable in reality. This can be evaluated by using data envelopment analysis (discussed in the next chapter). On the other hand, this sort of estimation might not be enough for analysis, because

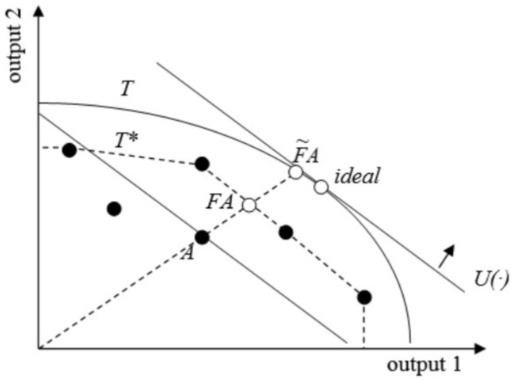


Figure 1.7: Differences between effectiveness and efficiency

every new observation is able to change the frontier. This is why the production function should be derived in a way that is more consistent – and that leads to making assumptions. Such possibility comes using the parametric method, such as stochastic frontier analysis (SFA, also discussed in the next chapter). The objective of analysis should be to make as few and flexible assumptions as possible to get estimates that are useful and valid. This is possible only in the case when the underlying assumptions are true.

Parametric models (including models based on DFA/SFA) assume the specification of functional shape. Typically, these are based on either a Cobb-Douglas or translog type of function (Jacobs 2001: 2). For an extensive comparison on the choice of functional form, see Giannakas *et al.* (2003) or Coelli *et al.* (2005: 211). In this thesis, due to multiple outputs and inputs, the constant elasticity of transformation (or CET) type of functions are used (for reasoning, see e.g. Powell and Gruen 1968; O'Donnell 2016c).

This section introduced the basic concept of technical efficiency and its derivation from effectiveness in the production theory framework and on which assumptions it is based. That concept is used to develop a more comprehensive efficiency measurement approach for the case of demand uncertainty in the following section and the possibilities to estimate the efficiency using real data in the next chapter.

1.4. The theoretical case of performance measurement in demand uncertainty

Many management control systems and operational research make the assumption that the future is known with a high degree of certainty and is well defined. However, most of the production (in the public as well as private sector) contains uncertainties of some sort. Uncertainty can be defined as a lack of ability to predict what the future will hold (Otley 2014) – the inability to forecast the events which may occur and the consequences of these events. Uncertainties originate from three sources: (a) externally from generalised uncertainty (shocks, unpredictability), or lacking the understanding of cause/effect in the culture at large, (b) contingency, in which the outcomes of organisational activities are in part determined by the actions of elements of the environment, and (c) internal source, as interdependencies of production components (Thompson 1967; Johnsen 2000).

In other words, the uncertainty occurs in many aspects and forms. The behaviour of nature is often uncertain (e.g. the weather) and one cannot predict it with certainty. Another source of uncertainty comes from the unpredictable actions of other people and firms, which is a study area of game theory and others. The collection of individual behaviours add another layer of uncertainty (government policies and market fluctuations). Lastly, one creates its own uncertainty by making different decisions (risk-seeking or risk-averse behaviour). Much uncertainty can be sourced in human ignorance, as in some cases of uncertainty some participants can have uncertainty about a given issue, while the others are fully informed (e.g. information asymmetry). Thus, uncertainty is more of a subjective state of mind than an objective property of the world (Quiggin 2012). To overcome the lack of information, more and more information is collected and processed.

Some of the uncertainties are addressed to random variations, which are tackled by stochastic processes and statistical analysis. As uncertainty is subjective, it has an objective counterpart in statistics, namely the variability in the values of the particular variable over time and space. However, variability is not a sufficient condition for the existence of uncertainty. The term risk can be used to refer to subjective beliefs with respect to the particular event, including individual realisations of a given random variable. Thus, the decisions that are made without knowing what realisation a given random variable will take and those that are made with this knowledge, are distinguished (Quiggin 2012). Without knowing the realisations, one faces the risk and variability and when knowing the realisations, one is subject only to variability. Thus, risk is the major economic problem associated with the uncertainty.

Alternatively, the term risk is distinguished in the statistical literature – risk occurs when the range of possible outcomes can be predicted and linked to a likelihood of probabilities. Such artificially closed systems occur rarely in the real world (lottery or casino) (Otley 2014). Thus, this terminology will not be used in the current thesis. Following Quiggin (2012: 4), the *'decisions under risk may be made with or without objective information on probabilities. The term ambiguity will be used for situations where probabilities are unknown and uncertainty will be used as a general term to cover risk, instability and ambiguity'*.

Ouchi (1979) combined the contingency theory, agency theory and new institutionalism and concluded that under conditions of ambiguity, of loose coupling, and of uncertainty, measurement with reliability and with precision was not possible. Most of the strategic planning is focussed on uncertainties, which cannot be solved by collecting more information or analysing it statistically. Such situation has been named 'deep uncertainty', *'that is, where analysts do not know, or the parties to a decision cannot agree on, (1) the appropriate conceptual models that describe the relationships among the key driving forces that will shape the long-term future, (2) the probability distributions used to represent uncertainty about key variables and parameters in the mathematical representations of these conceptual models, and/or (3) how to value the desirability of alternative outcomes'* (Lempert *et al.* 2003: xii). Thus, instead of determining the best predictive model and optimising the results, in case of (deep) uncertainty it might be better to choose an alternative that is most robust. It would achieve a reasonable level of goodness across situations and models. A robust plan yields satisfactory outcomes in a wide range of plausible states of nature (Walker *et al.* 2013). A robust plan from the principals' point of view is to reduce the possible (deep) uncertainty. This is achievable by an introduction of a minimum service level (MSL) as an insurance and to reduce the possible risks for themselves. Such a level introduces a minimal threshold, over which a service can be considered acceptable.

The uncertain environment and informational asymmetry can lead rational decision makers to make different production choices (O'Donnell *et al.* 2010). The performance measurement following the productivity and efficiency analysis can thus be systematically biased, when this is not considered (and typically, it is not). Conventional production frontiers implicitly impose the restriction that information differences have no effect on the framing and conditioning on producers' decision-making (such as input allocation) and so have no effect on outcomes. Thus, non-stochastic behaviour and non-stochastic technologies are typically estimated. Stochastic elements are taken into account only when it is econometrically beneficial. This, however, might lead rational and efficient *ex ante* decisions to be considered as inefficiency.

One should incorporate the demand uncertainty and risk-averse behaviour into the analysis to avoid unjustly labelling the input usage, which provides insurance against the demand uncertainty, as inefficiency. The uncertain demand will constrain the behaviour of decision makers. In the case of hospitals (as well as other public entities), the decision makers face social pressure to satisfy a large percentage of demand so they must ensure the ability to provide the service to consumers with the maximum feasible probability (Lovell *et al.* 2009). The informational differences due to stochastic elements in production are misclassified for differences in technical or allocative efficiency (O'Donnell *et al.* 2010). Therefore, it is necessary to consider whether the results of efficiency analysis have taken the uncertainties into account appropriately.

The cost and efficiency studies predominantly treat the demand of publicly provided services as known to the decision maker. Such an assumption is unrealistic and does not allow the researcher to uncover the essence of public agencies costs (Friedman and Pauly 1981; Gaynor and Anderson 1995). The demand uncertainty (alternatively stochastic demand) can be considered as an environmental uncertainty, which is beyond the direct control of the decision maker. The demand uncertainty in service provision is to an extent covered in the hospital services literature addressing the costs of hospitals (Lovell *et al.* 2009; Boutsioli 2010; Almeida and Cima 2015), but not widely in other sectors. One can claim that such concept could be used more widely, as the demand uncertainty prevails in many industries and affects the planning and allocation of resources.

Many service firms and public agencies (such as hospitals, fire and rescue boards, water and electricity suppliers) face a situation where it is important to have a sufficient standby capacity to keep the probability of excess demand below some desired level (Gaynor and Anderson 1995). In addition to direct services, such firms and public agencies also provide a standby service, which is an insurance or option demand, to ensure service availability if someone unexpectedly needs it. Providing such standby capacity raises the costs.

In the face of demand uncertainty, the decision makers must decide on a 'turn-away probability' (α) or in other words, on the probability that the future demand exceeds the capacity. Such probability measures the risk that a decision maker is willing to take, when predicting the future demand and whether it will exceed the supply. In estimating the distribution of the demand, Poisson (Joskow 1980) and normal (Baker *et al.* 2004) distribution have been assumed. Gaynor and Anderson (1995) used the mean and variance of the actual demand approximating the unpredicted demand. Hughes and McGuire (2003) and Lovell *et al.* (2009) used a simple autoregressive process – the difference between estimated and observed demand was then used for measuring the demand uncertainty. Following the decision on 'turn-away probability', the decision makers then minimise the costs subject to the constraint that the demand can exceed the capacity α percent of the time. Doing so, they will be providing a target standby capacity.

Some authors (e.g. Boutsioli 2010) distinguish the demand variability and demand uncertainty. The variability of demand can be predicted (i.e. there will be more emergencies during public holidays) and approximated by a distribution density function. The uncertainty, however, is considered unpredictable, as the variation cannot be described by any known probability distribution. As such, the unpredicted demand can be estimated only *ex post*. The demand variability can be taken into account more easily in the planning process, but the demand uncertainty complicates planning.

As noted before, there is not much empirical research undertaken that would consider the impact of demand uncertainty on costs of service provision. Even fewer studies have analysed the effect of uncertain demand on cost-efficiency. Boutsioli (2010) reviewed the relevant literature, twelve articles in total, on demand uncertainty in hospital costs. She concluded that there are many possibilities for further studies, namely the theory of hospital costs needs to take into consideration the demand variability and uncertainty. The literature focuses mainly on USA and UK cases, and might result in different conclusions in other areas and contexts. Furthermore, hospitals face more uncertainties than just demand fluctuations, i.e. uncertainty in diagnosis, treatment of illness or input supplies. According to Boutsioli, the excess capacity should be treated as insurance, while considering theoretically the economics of insurance. Another extension would be to develop the empirical methodology and use more disaggregated data (e.g. daily or monthly data). All these conclusions can be extended further than just the healthcare industry.

In addition to hospital cost analyses, the demand uncertainty has been discussed in the supply chain planning and inventory management context (see Mula *et al.* (2006) for a classification and literature review). Other more similar studies to the efficiency literature cover the case of port terminals (Rodríguez-Álvarez *et al.* 2011). For staffing allocation, an application for homecare services (Rodriguez *et al.* 2015) has been introduced. Such modest treatment on the potentially important aspect in many sectors indicates a severe research gap in efficiency analysis.

As a next step, such a concept is put into more specific terms, which would be the theoretical basis for the following estimation on FRS performance. This follows closely the development of the future working paper by O'Donnell and Puolokainen (2018), presented at e.g. O'Donnell and Puolokainen (2017a) and O'Donnell and Puolokainen (2017b). Considering a multilevel and multi-unit context, the production possibilities, behaviour of the decision makers and estimation is defined and discussed.

Production possibilities Irrespective of their objectives, central agencies and subunits can only choose inputs and/or outputs from a set of technically-feasible input-output combinations. To measure efficiency and under-resourcing, one needs to estimate the boundary of this so-called production possibilities set. A period-and-environment-specific production possibilities set is a set containing all input-output combinations that are possible in a given period in a given production environment:

$$T^t(z) = \{(x, q) : x \text{ can produce } q \text{ in period } t \text{ in environment } z\}.$$

It is assumed that all technologies available in period t are regular, thus:
T1: $(x, 0) \in T^t(z)$ for all $(x, z, s) \in \mathbb{R}_+^{\mathrm{M+J}}$ (inactivity)
T2: $\{q : (x, q) \in T^t(z)\}$ is bounded for all $(x, z) \in \mathbb{R}_+^{\mathrm{M+J}}$

T3: $q \geq 0 \Rightarrow (0, q) \notin T^t(z)$ (inputs weakly essential, no free lunch)
T4: $(x, q) \in T^t(z)$ and $0 < \lambda \leq 1 \Rightarrow (x, \lambda q) \in T^t(z)$ (outputs WD)
T5: $(x, q) \in T^t(z)$ and $\lambda \geq 1 \Rightarrow (\lambda x, q) \in T^t(z)$ (inputs WD)
T6: the set $\{q : (x, q) \in T^t(z)\}$ is closed for all $(x, z) \in \mathbb{R}_+^{\mathrm{M+J}}$ and the set $\{x : (x, q) \in T^t(z)\}$ is closed for all $(q, z) \in \mathbb{R}_+^{\mathrm{N+J}}$

For example, the production possibilities set can take the following form:

$$T^t(z) = \left\{ (x, q) : \left(\sum_{n=1}^{N} \gamma_n q_n^{\tau} \right)^{1/\tau} \leq A(t) \prod_{j=1}^{J} z_j^{\delta_j} \prod_{m=1}^{M} x_m^{\beta_m} \right\} \tag{1.7}$$

where $A(t) > 0$, $\beta = (\beta_1, ...\beta_M)' \geq 0$, $\gamma = (\gamma_1, ..., \gamma_N)' \geq 0$, $\tau \geq 1$ and $\gamma' \iota = 1$. Such a set assumes that the outputs and inputs are strongly disposable and it is homogeneous to degree r.

Under weak regularity conditions, this set can be represented using distance functions.

A period-and-environment-specific output distance function (ODF)[2] gives the reciprocal of the largest factor by which it is possible to scale up a given output vector when using a given input vector in a given period in a given production environment:

$$D_O^t(x, q, z) = \inf\{\rho > 0 : (x, q/\rho) \in T^t(z)\} \tag{1.8}$$

Assuming that the functional forms of relevant distance functions are known, then if the number of outputs is greater than one and outputs are strongly disposable, then the ODF cannot be translog function. If the number of outputs is greater than one and output sets are closed, then the ODF cannot be Cobb-Douglas type function (O'Donnell 2016a;c). Following on from the previous example, the ODF can take the following CET-type functional form:

$$D_O^t(x, q, z) = \left(A(t) \prod_{j=1}^{J} z_j^{\delta_j} \prod_{m=1}^{M} x_m^{\beta_m} \right)^{-1} \left(\sum_{n=1}^{N} \gamma_n q_n^{\tau} \right)^{1/\tau} \tag{1.9}$$

[2]ODF has following properties:
$D_O^t(x, 0, z) = 0$ for all non-negative x;
$D_O^t(x, q, z)$ is non-decreasing in q and non-increasing in x;
$D_O^t(x, q, z)$ is linearly homogeneous in q;
$D_O^t(x, q, z)$ is quasi-convex in x and convex in q;
if q belongs to the production possibility set of x (i.e. $q \in P(x)$), then $D_O^t(x, q, z) \leq 1$; and distance is equal to unity (i.e. $D_O^t(x, q, z) = 1$) if q belongs to the 'frontier' of the production possibility set.

A period-and-environment-specific input distance function (IDF)[3] gives the reciprocal of the smallest fraction of a given input vector that can produce a given output vector in a given period in a given production environment:

$$D_I^t(x, q, z) = \sup\{\theta > 0 : (x/\theta, q) \in T^t(z)\} \tag{1.10}$$

The example for IDF would be:

$$D_I^t(x, q, z) = \left(B(t) \prod_{j=1}^{J} z_j^{\kappa_j} \prod_{m=1}^{M} x_m^{\lambda_m} \right)^{-1} \left(\sum_{n=1}^{N} \gamma_n q_n^{\tau} \right)^{-1/\tau\eta} \tag{1.11}$$

Notation x_{it} = inputs allocated to subunit i in period t
w_{it} = prices of the inputs allocated to subunit i in period t
z_{it} = physical characteristics of jurisdiction i in period t
d_{it} = services demanded in jurisdiction i in period t
m_{it} = MSL in jurisdiction i in period t
q_{it} = services provided in jurisdiction i in period t
MSL represents the value such that $Pr(d_{it} \geq m_{it}) = \alpha$

Behaviour One is foremost interested in the following:
(a) the cost-efficiency of the central agency;
(b) any under-resourcing of subunits; and
(c) the output-oriented technical and mix efficiency of subunits.

In the face of uncertain demand, the central agency chooses inputs to minimise the cost of meeting the MSL in each jurisdiction. The period-t optimisation problem of the central agency is:

$$\min_{x \geq 0} \left\{ w_t' x : D_I^t(x_i, m_{it}, z_{it}) \geq 1 \text{ for } i = 1, \ldots, I \right\} \tag{1.12}$$

where $w_t' = (w_{1t}', \ldots, w_{It}')$ and $x' = (x_1', \ldots, x_I')$. The input vector that solves this problem is $x_t^{*\prime} = (x_{1t}^{*\prime}, \ldots, x_{It}^{*\prime})$. The i-th associated minimum cost is $C^t(w_{it}, m_{it}, z_{it}) = w_{it}' x_{it}^*$.

The cost-efficiency (CE) of the central agency in period t is

$$CE^t(x_t, w_t, m_t, z_t) = w_t' x_t^* / w_t' x_t \tag{1.13}$$

[3]IDF has following properties:
$D_I^t(x, q, z)$ is non-decreasing in x and non-increasing in q;
$D_I^t(x, q, z)$ is linearly homogeneous in x;
$D_I^t(x, q, z)$ is concave in x and quasi-concave in q;
if x belongs to the production possibility set of q (i.e. $x \in L(q)$), then $D_I^t(x, q, z) \geq 1$; and distance is equal to unity (i.e. $D_I^t(x, q, z) = 1$) if x belongs to the 'frontier' of the input set.

where $x_t' = (x_{1t}', \ldots, x_{It}')$. If $D_I^t(x_{it}, m_{it}, z_{it}) \geq 1$ for all i, then $CE^t(x_t, w_t, m_t, z_t) \leq 1$. On the other hand, if $D_I^t(x_{it}, m_{it}, z_{it}) < 1$, then the subunit i is under-resourced, meaning it would have not had sufficient amount of inputs that the target standby capacity assumed. For an illustration, Figure 1.8 considers a two input case, where point A indicates an over-resourcing (not cost-efficient), point S indicates a cost-optimal point for MSL and point X indicates a case when $q_{it} > m_{it}$ and has thus been under-resourced. The cost of meeting the MSL is minimised at S. Any subunit operating below the curve passing through S is under-resourced.

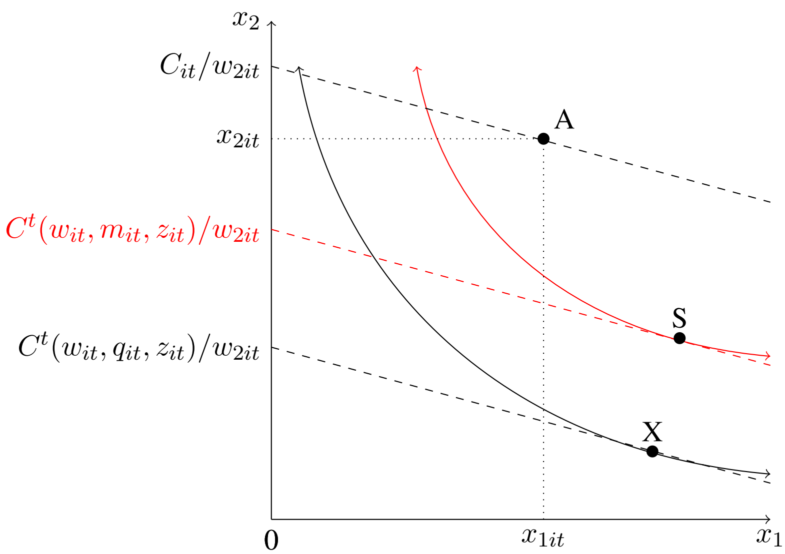


Figure 1.8: The cost of meeting the MSL

After the central agency has allocated the inputs, the subunit i seeks to use their allocated inputs to provide the services demanded in a jurisdiction. Their period-t optimisation problem is:

$$\max_q \{Q(q) : q \leq d_{it}, D_O^t(x_{it}, q, z_{it}) \leq 1\} \tag{1.14}$$

where $Q(.)$ is a non-negative, non-decreasing, linearly homogeneous, scalar-valued aggregator function with weights that represent the values the subunit places on outputs. The output vector that solves this problem is $\hat{q}_{it} \equiv \hat{q}^t(x_{it}, d_{it}, z_{it})$. The associated aggregate output is $Q(\hat{q}_{it})$.

The output-oriented technical and mix efficiency (OTME) of subunit i in

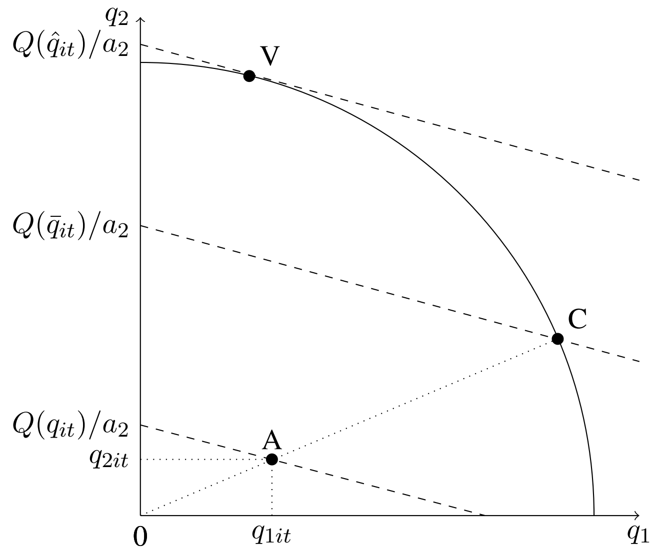


Figure 1.9: Output technical and mix efficiency

period t is:

$$OTME^t(x_{it}, q_{it}, d_{it}, z_{it}) = Q(q_{it})/Q(\hat{q}_{it}). \tag{1.15}$$

In case the number of services provided meets exactly the demand for services ($q_{it} = d_{it}$), then the subunit can be considered efficient ($OTME^t(x_{it}, q_{it}, d_{it}, z_{it}) = 1$). For an illustration, consider Figure 1.9. If $q_{it} = d_{it}$, then the feasible output set is given by the rectangle with vertices at the origin and point A, as a subunit is unable to store the services (cannot provide more services than there is demand for). Thus, a subunit is considered efficient, even though the target standby capacity would have allowed to provide more services, when needed.

The current thesis considers FRS as an example of public agency and public service provision – the top manager (in this case, the chief fire officer/minister, etc.) has to make decisions beforehand to allocate resources (such as labour, equipment and vehicles, covering the maintenance of facilities and administration of brigades) between different units in different regions they are responsible for without knowing how big is the demand for services (how many emergencies occur) in that region. Secondly, the units (FRS providers) have to function and respond to emergencies with fixed input bundles. As the occurrence of emergencies (such as fires, traffic accidents, but also false alarms) is unknown and stochastic in nature, the best a decision maker can do is to predict the occurrences in a region based on historical data and taking the possible environmental factors

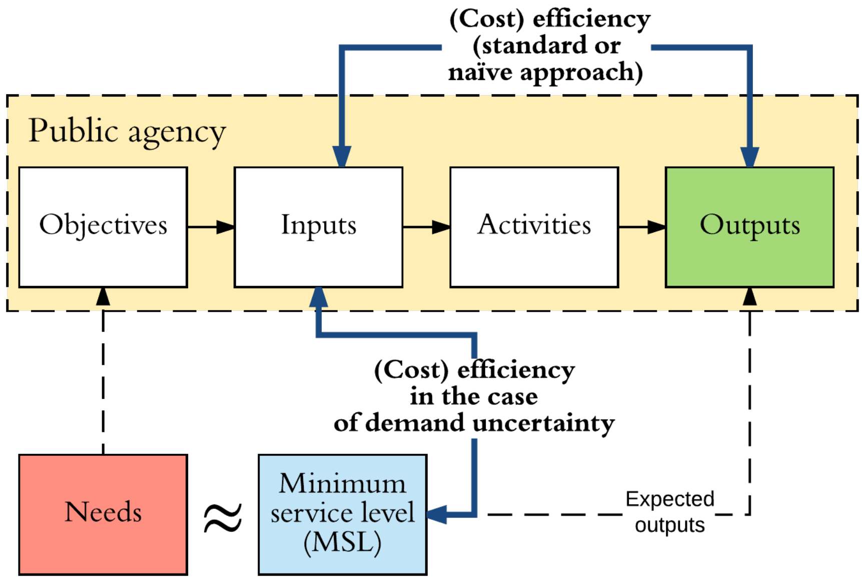


Figure 1.10: The conceptional model of demand uncertainty (Source: Author's compilation)

into account as best as possible.

Incorporating the demand uncertainty into the public sector management framework

This subsection will conclude the chapter by demonstrating how the proposed demand uncertainty in the production theory would fit into the context of public sector management field. In other words, the subsection merges the two fields – public sector management approach (see Chapters 1.1–1.2) and production theory (see Chapters 1.3–1.4).

For illustration, a stripped-down version of the proposed input-output model (originally Figure 1.1) is of use (see Figure 1.10). In a standard (or from now on a naïve) case, the (cost) efficiency would have been evaluated using the observed outputs and inputs. As noted previously, such simplification would not take into account the fact that the demand for services (needs) is uncertain at the time of resource allocation.

Therefore, one might be interested in approximating the (expected) outputs for the next period for which the resource allocation decisions must be made. For that, there are many possibilities. The current case argues for a minimum service level (MSL), which is based on a P-th percentile of the occurred outputs in a respective jurisdiction. Thus, with a probability of P (e.g. 95%), the occurring outputs of the period under consideration would stay lower than the proposed MSL. Such threshold would indicate the potential need (demand) for the services. However, providing a context for the MSL is more open – it can be a result of negotiation, voting, etc. In other words, the decision maker can freely choose the indicator value for MSL.

While accounting for such demand uncertainty, the cost-efficiency analysis would reckon that some of the excess capacity of resources in the jurisdictions were there 'just in case', in order to meet the potential upsurges in demand. In a naïve case, such prudence of the decision maker is labelled as inefficiency, which might be unjust.

For the study, a performance measurement system (PMS) with three components was developed. First, using the MSL as an indicator for potential needs (demand for services), the cost-efficiency of the central agency is of interest. Second, after the allocation of resources by the central agency, one is interested in any under (or over)-resourcing of the subunits, meaning that based on the resources the subunit received, would it be able to provide the amount of services dictated by the MSL. Third, one is interested in the potential of providing more services in case there would be a need for them, i.e. would a subunit have the resources to provide more services (and how much more) in comparison to the best performing subunits? The methods to estimate such indicators will be introduced in the next chapter.

For a public sector manager, such concept can be of use in many ways (see Chapter 1.1.3). For example:

- Planning purposes – introducing the MSL would provide support for decision-making in order to allocate resources in a cost-efficient way. In further cases, such MSL can be used for contracting purposes, e.g. the subunit should be prepared to provide services at a fixed MSL and the central agency allocates resources accordingly.

- Improvement purposes – the PMS allows the central agency to make improvements in the resource allocation (e.g. reducing the number of over-resourced subunits, increasing the cost-efficiency of central agency, evaluating the potential of subunits to provide more services).

- Evaluation and benchmarking – for external purposes, the PMS can be used to rank different subunits based on their efficiency scores. This opens opportunities to scrutinise the least-performing subunits (or central agency, if it has allocated too many resources).

- Monitoring and controlling – the central agency can use MSL as an indication on how well a subunit is doing, thus monitoring the progress of a subunit with respect to a target indicator.

A limitation of introducing such MSL instead of using the occurred amount of provided services is the lack of theoretical ground in choosing an optimal threshold. In other words, the concept takes into account the demand uncertainty, but actually choosing an optimal level of services is up for discussion and out of the scope of this thesis.

The first chapter of the thesis focussed on introducing the concept of performance measurement in the public sector by defining the performance and its subcomponents. This was followed by an analysis of the possible goals why one should conduct a performance measurement in public agencies and what could be the possible limitations of doing it. The public sector management approach was clearly defined with production theory from microeconomics by introducing the efficiency analysis as a possibility to measure the performance. Uncertainty, as an important factor in the public sector management, was integrated into the production theory framework. The next chapter will investigate the possible methods how the performance should be quantitatively measured.

2. METHODS FOR MEASURING THE PERFORMANCE IN THE PUBLIC MANAGEMENT

I had a lot of research done for me at Central House because I was unable to get clear statistics out of my own Department. Shocking!

They continually change the basis of comparative figures from year to year, thus making it impossible to check what kind of bureaucratic growth is going on.

'Humphrey', I began, fully armed with chapter and verse, 'the whole National Health Service is an advanced case of galloping bureaucracy.'

Humphrey seemed unconcerned. 'Certainly not,' he replied. 'Not galloping. A gentle canter at the most.'

— J. Lynn & A. Jay, Yes Minister, The Compassionate Society

2.1. Possible methods for performance measurement

2.1.1. Introduction

The previous chapter analysed what is performance measurement and why one would be interested in measuring the public agencies performance. The current chapter provides a systematic overview and analysis of the possible methods and how they can be used to measure performance quantitatively.

The opportunities to assess public agencies' performance quantitatively depend on the nature of information and which stakeholder is conducting the analysis. This reflects the possibilities of methods used and aspects of performance assessed. Performance measurement is highly dependent on the available and collectable data. The measurement will provide more useful insights when much effort has been put into conceptualisation of the essence of the provided service. Murphy and Greenhalgh (2013) have identified four stages of the development of data and information availability and its use in maturing performance management regimes:

1. Poor data – lack of comparability and national perspective; poor quality assurance.

2. Rich data – available quantitative and qualitative data; subjective and objective measurements; absolute and relative indicators and measures; differentiation of inputs, outputs, and outcomes; operationally focussed.

3. Intelligent data – real time and remote data access; specified standards for absolute and objective measures; robust and comprehensive measurement tools, detailed trend and comparison data readily available.

4. Self-regulation – facilitates independent academic and operational research; identifies and disseminates best practices and innovations; facilitates robust international comparisons; independent quality assured.

Each of the following phases gives new and more meaningful opportunities to measure the performance. A wide range of performance evaluation methods have been developed (see Figure 2.1, which doubles as a guide to the structure of the chapter). Thus, this chapter does not provide only necessary information about the methods used to solve the problems introduced in the last chapter, but also gives a systematic overview of advantages and disadvantages of most of the common methods available for performance measurement. The discussed methods differ from required information, procedures of data analysis and aspects from which they evaluate the performance. Therefore, it is relevant to adapt adequate techniques and modelling processes to reflect the nature of performance from the aspects one is interested in, while having the constraint of limited information. In the following, these methods, which can and have been used to measure the performance of public agencies, are discussed. Smith (2006) gives an extensive overview of some used methods to analyse adjusted performance indicators: cluster analysis, risk adjustment, regression analysis (including multilevel analysis and seemingly unrelated regression (SUR) equations). In addition to these methods, this thesis also compares the ratio analysis (including cost-unit analysis), composite index analysis, Four Quadrant model and BSC. Frontier analysis methods, which are the main focus of this thesis, are considered in more depth in the next section.

Figure 2.1 subjectively characterises different methods which have been used to measure the performance of public agencies. They are distinguished by their main domain – whether a method is more inclined to performance management or productivity analysis, and their subjective and relative technical complexity on applying. All illustrated methods are discussed as follows: their setting, possible limits and former use in the public sector performance measurement literature.

2.1.2. Unadjusted absolute single performance indicator

Several methods have been implemented to assess the performance in public agencies quantitatively. The simplest method to rank different subunits on

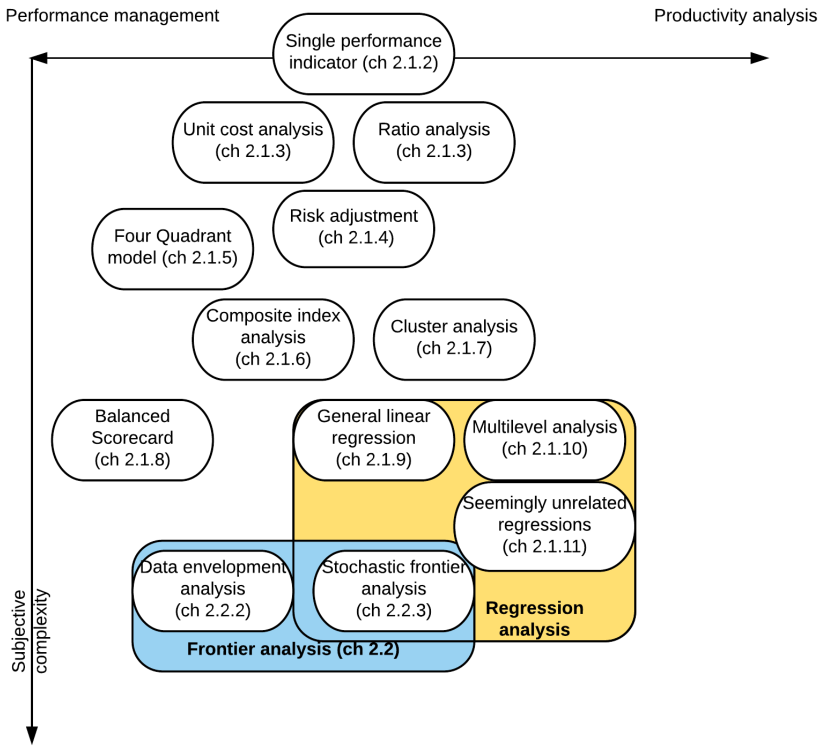


Figure 2.1: Different measurement methods used to measure performance in a public agency (Source: Author's compilation)

the basis of performance characteristics would be to use an unadjusted single performance indicator (the ranking of units is often called a 'league table', mainly a sports term). As a single performance indicator, one could classify any quantitative (but also qualitative) piece of information that reflects the performance of the public agency in one way or another. Using single indicators, the evaluator can describe the volumes of inputs, outputs and also environmental factors independently from each other. Simple descriptive statistics on performance indicator can be of value for performance measurement. Single performance indicators form the basis for using other more complicated methods for performance measurement in a systematic way.

Using only a single indicator might lead to biases in interpreting the results, as subunits operate in different environments, which they usually cannot choose freely. The local variations caused by uncontrollable environmental factors should be considered for fairer and more valid evaluation. This comes into practice, as the differences in environment should be compensated by the top-level manager (in the current case, the central rescue board). In order to provide the service to a certain standardised minimum level, a proper funding system that would take the environmental factors into account as well as possible, should be developed. If such funding scheme is already present, it complicates the proper performance measurement even further – for evaluation, the differences in environment as well as funding, should be considered. This is why a single indicator is usually not good enough to characterise the performance of a (multi-unit multilevel multi-goal) public agency as it does not reflect the performance comprehensively.

2.1.3. Ratio analysis

Many decision makers use some fixed ratios, such as staffing ratio, for performance measurement and management. Thus, a ratio analysis can be introduced. Ratio analysis is based on the idea that the absolute indicator is not of importance, but some ratio.

Ammons (2008) starts the ratio analysis with an example of staffing ratio, which can be described as a number of staff per x population. Such proxy would be appropriate to estimate the demand for services if every one group of x population would be similar to each other. Consequently, a production ratio would be a ratio of staffing and amount of services provided, which would be a better proxy for describing the needs for service provision.

A similar approach is unit cost analysis, which alternatively considers costs instead of staffing – such as costs per capita or costs per provided service. For the publicly provided goods, finding proper costs might not be that straightforward, as the outputs can be affected by many different entities, e.g. the same resources can be providing multiple services simultaneously.

From unit cost analysis, one could introduce engineered standards, which would declare the amount of effort (resources or time) needed to complete a certain task. The standards can be based on historical data as well as expert opinion. Such 'flat rates' can then be used as a proxy for assessing the appropriate size of needed resources to meet the demand of services in a given region. If fewer resources are needed than the provided standard states, then such 'efficiency rating' would be greater than 100%. Such standards can describe and provide a comparison of public agencies performance, but the comparison is fragmented between different ratios and do not take into account the (uncontrollable) environmental factors.

2.1.4. Risk adjustment

When measuring performance of a public agency, the customer (user of the service), is of utmost importance. The characteristics of service users, or all of the population in the service area, have an impact on provided service. Thus, the desired outcomes of public services depend on the users. For example, in schools, the examination results depend on schooling as well as individual capabilities (intellect). To take such characteristics into account and adjust the observed outcomes for the nature of the population 'at risk', various techniques have been developed (Iezzoni 1997).

The goal of such techniques is to construct a ratio of observed and expected outcomes amongst consumers. Given the served population, an estimation of expected outcomes for an organisation is calculated. Observed outcomes will be then ranked to the expected outcomes.

Smith (2006) uses the policing example, when calculating the national average prosecution rate p_i for each crime type i. If the number of each crime type i in an area in a certain period is n_i, the expected number of prosecutions in that period is $n_1p_1 + n_2p_2 + ...$. This estimate can be used as the denominator in the risk-adjusted measure of prosecution success. The numerator is the actual number of prosecutions, and the ratio gives the organisation's performance after adjusting for the types of crime being investigated.

Such risk adjustment methods have been mostly used in health care, such as Iezzoni *et al.* (1996). As a critique, the choice of most appropriate risk adjustment methodology is a subject for considerable debate and controversy.

2.1.5. Four Quadrant model

As an intermediary technique, a simple Four Quadrant model (Ballanti *et al.* 2014) can be considered. This is a joint graphical analysis of two measures, using on the one hand the differences between actual and standard inputs, and on the other hand the actual and standard level of outputs. As a result, four classes (quadrants) of subunits can be distinguished: over-standard (input and output gaps are positive); under-standard (input and output gaps are negative); efficient (output gap is positive and input gap is negative) and non-efficient (input gap is positive and output gap is negative). Such classification would be the basis for performance improvement.

Such a method would be relatively easy to implement on various datasets and gives an opportunity to interpret the results graphically (as well as for combining the results with other methods in graphically accessible setting). This is a highly attractive property for empirical work and policymaking. As a critique, excessive simplification should be stated – the four classes are divided only on whether they are above stated standard or below. For complex assessments, such analysis might be insufficient as there might be different input-output sets that are of interest as well as taking into account the influence of environmental factors.

The Four Quadrant model has been proposed by Ballanti *et al.* (2014) and applied to Italian local governments.

2.1.6. Composite index analysis

One category of performance measurement techniques attempts to aggregate different performance indicators into a single index. Arguments in favour of using a single composite index are (Smith and Street 2005):

1. A global index of performance can offer a rounded assessment of system performance. This is important when inputs cannot be attributable to specific activities.

2. A global index leaves the decision maker the possibility to set their own priorities and seek improvements in various dimensions, which is not possible for piecemeal examination (one would want to improve across all indicators).

3. A global index can be used to support other objectives, such as allocation of resources or identifying the priority subunits for further inspection (best practices versus lagging subunits).

4. A global index can be a basis for a 'league table'. It is argued that publishing such rankings have a positive effect on the provided service as well as increasing the accountability of organisations (Hibbard *et al.* 2003).

Such index creation is also complicated and controversial. Smith and Street (2005) and Hauck and Street (2006) state altogether four main challenges. Let one assume that an organisation is pursuing a set of objectives, Y, which has i separate objectives, $Y = \{y_1, y_2, ..., y_i\}$. To create a single index characterising the overall performance, the separate objectives have to be combined, so that in additive case $Y = w_1 y_1 + ... + w_i y_i, \sum iw = 1$.

First, there is no certain way how to choose the included indicators and how one should prioritise (weight) the different performance indicators. Thus, the issue of giving values to weights, w, arises. How should they be determined? Determination of the weight size can be left to analytical techniques, but they are ultimately a political judgement. One option would be to impose a single set of weights, which origins are dependent on the context of analysis (such as analysis of historical data, expert opinions, political ambitions, etc.). The creation of composite performance measures relies crucially on securing estimates of the relative value placed by society on each performance indicator (Smith and Goddard 2002). This possibility will be contradicted in case organisations' subunits are more independent (decentralised) in decision-making and might value priorities differently from each other. Another extreme would be to allow weights vary freely, so that they are specific to each subunit. Such setting is applied through data envelopment analysis (DEA), discussed below. DEA method chooses weights that would maximise the result for each subunit. This has been regarded as a positive feature (Cooper *et al.* 2000) but it also limits the analytical discriminatory power, such that subunits are more difficult to distinguish in terms of their relative overall performance (Hauck and Street 2006).

Second, the link between weights and the scale on which the performance is measured – each weight must apply to all levels of attainment. This implies that a unit increase in one of the performance indicators value is deemed equally valuable from whatever starting point and in whatever form the performance indicator is represented.

The third issue related to the single index approach is the loss of information, as the possibilities for improvement are not apparent from a single bit of information. For example, in applying DEA with unconstrained weights, it is possible to place an organisation in its best possible light simply by assigning a zero weight to those objectives on which it performs poorly (Hauck and Street 2006). The simplicity of ranking may result in a superficial approach for judging different subunits' performance. Thus, much effort has to be put on designing an index to be understandable and useful.

The fourth issue is related to exogenous constraints. One of the main purposes of performance measurement is to distinguish the organisational effort from observed performance. The observed performance may be partially explained by factors over which the organisation has little or no control, defined previously as environmental factors. When the performance is assessed using a single index, only the average effect over different performance indicators can be taken into account. It is probable that the effect of a specific constraint varies over performance indicators. A middle way between the analysis of performance indicators in isolation and the creation of a single index is recommended by Hauck and Street (2006) by implementing the seemingly unrelated regression (SUR) techniques (Zellner 1962), discussed below.

2.1.7. Cluster analysis

Cluster analysis has been defined by Aldenderfer and Blashfield (1984: 6) as *'the generic name for a wide variety of procedures that can be used to create a classification. These procedures empirically form clusters' or groups of highly similar entities. More specifically, a clustering method is a multivariate statistical procedure that starts with a data set containing information about a sample of entities and attempts to reorganise these entities into relatively homogeneous groups'*.

Various forms of cluster analysis are one of the most widely deployed approaches in performance comparison (Smith 2006). Clustering enables the local environmental factors relevant to the service in question to be taken into account. Based on such factors, a measure of similarity between units can be calculated. The units can be then classified/clustered, based on the measured similarity into more homogenous groups. Various algorithms can be used for clustering, such as the ‘nearest neighbour’, and base the comparison of performance in a cluster.

One of the difficulties of cluster analysis is to decide on the cut-off criterion – which units are similar enough to be essentially comparable. An inaccurate decision can lead to unfair comparisons, as some units are more similar to each other than the other. Moreover, some units may be on the edge of a cluster, having more similarities to some units outside of the cluster. For that reason, in policymaking the nearest-neighbour approach is preferred. In addition, some units might be so different from every group that comparison is impossible. By clustering, some information is lost (comparison to units in other clusters).

The proponents of cluster analysis emphasise its transparency – when the analysis is complete, the results are straightforward. One usually compares the unit with respect to the cluster average, which, however might be a crude’ method for adjusting for variations in environment. Potentially useful factors might not be included as comparators.

An example of clustering/classification in the FRS field is undertaken by Estonian Rescue Board (ERB). The ERB divides FRS brigades into three groups, based mainly on the population of their service area. This is basis for providing different services. As extraordinary events occur more often in more populated places, they require also more diverse service lists.

2.1.8. Balanced Scorecard

Another widely used management tool, which was initially developed for performance measurement, is the Balanced Scorecard (BSC) (Kaplan and Norton 1992; 2001). The BSC has thriven because of the systematic and logical structure. At first targeted to the for-profit businesses, BSC was a supplement to conventional financial reporting. In addition to the financial performance, the framework included three areas: a company’s relationship with customers, key internal processes, and learning and growth (Kaplan and Norton 1992). From the commencing framework, the BSC has developed in many variations.

The BSC presents a multi-dimensional view of performance across different objectives and stakeholders, which is appropriate for many public sector organisations. Furthermore, the BSC makes a case for key performance indicators (KPIs), and as such, directs managerial attention to important drivers of organisational results. This confirms the management by objectives (MBO) approach, introduced by Peter Drucker in 1954 (Drucker 1995) and links the KPIs in causal relationships with desired outcomes.

The BSC approach has been used more often by policymakers than academics. For example, the Comprehensive Performance Assessment (CPA) of English local governments conducted by Audit Commission used a BSC approach (see e.g. Woods and Grubnic 2008). The CPA ranked local governments based on a number of stars achieved by performance measurement. CPA was regarded foremost as an incentive scheme, on which depended the allocated resources by central government. Lockwood and Porcelli (2013) concluded that it had failed in stimulating efficiency improvements.

2.1.9. Regression analysis

Organisations performing benchmarking sometimes use regression analysis to control for selected variables when identifying top performers (for a technical explanation of regression analysis, see, for instance Draper *et al.* 1966). Regression analysis covers many techniques for explaining the variability in a dependent variable based on the values of one or more independent variables. In addition, it allows to control for the environmental factors. The result can be obtained by comparing the differences between the observed and estimated results, addressing the differences as differences in performance. As regression analysis is a major topic in econometrics, including productivity analysis, a plethora of these methods have been applied also to public sector performance measurement (see e.g. Smith 2006).

A range of technical difficulties can be attributed to the regression analysis, such as the specification of the model and whether it complies with underlying statistical assumptions. In addition, for performance measurement, the standard approaches of using a t-test or an analogous criterion for variable inclusion, is not sufficient. A variable might pass such a test, because it is exogenous influence on performance; or it is correlated with inefficiency, either accidentally or because of deliberate managerial choices. In the second case, its inclusion to the model is inappropriate as it would effectively 'excuse' any element of inefficiency correlated with the variable (Smith 2006). Such technical choices might have a profound impact on the estimates of performance, so it is critical to make clear which choices have been made so that they can be scrutinised.

2.1.10. Multilevel analysis

Public agencies are often structured hierarchically, having more than one level of management. In such situations, it might be plausible to analyse the units at multiple levels and determine the effect of each level. Boyne *et al.* (2005) state in their overview article that levels within organisations have different impacts. Multilevel analysis is a statistical approach which allows the simultaneous modelling of processes across such different units of analysis.

Following Luke (2004), the goal of a multilevel model is to predict values of some dependent variable based on a function of predictor variables at more than one level. Most typically, such a method is applied to evaluating the performance of schools (e.g. pupil-class-school-jurisdiction-state as levels). Smith (2006) states the key policy question: to what level of the hierarchy are variations in individual outcomes attributable?

The basic idea of multilevel modelling as a performance measurement tool is to analyse to which extent the outcomes can be attributed to various levels of hierarchy. From a statistical point of view, these models are variations and extensions of regression analysis. Having two levels of hierarchy, such a model would add an 'organisational' effect to an individual attainment.

In the schooling context, multilevel models have been developed to a high level of refinement, but their use on other public services are rather scarce (some examples include drug abuse treatment, employment and training, see Heinrich and Lynn 2001). For the example of schools, one might be interested in the value-added models (Goldstein and Thomas 1996; Goddard and Goddard 2001; Bock 2014). To clarify a schools' role in a student's education attainment (as the dependent variable), one might be willing to take into account additional variables, such as environment, which is out of direct control of a school as well as organisational (school) effect. The model gives an estimate of that effect after adjusting for prior attainment, and could be interpreted as a value added by the attainment of the school.

2.1.11. Seemingly unrelated regression methods

When considering the performance of a public agency, important relationships may exist between individual performance measures that are lost if they are analysed only independently. Martin and Smith (2005) state that also for the decision makers the dominant interest is in indicators of performance in specific service areas, rather than aggregate measures of organisational performance. Smith (2006) highlights five reasons for variations in the performance: environmental factors, resource levels, efficiency, substitution, and data quality (discussed above). To take these reasons into account simultaneously, seemingly unrelated regression methods might be appropriate.

Seemingly unrelated regression or seemingly unrelated regression equations (SUR) is defined by Tien (2003) as a method used when two or more separate regression equations that have different sets of independent variables are related to each other through correlated disturbance terms. Estimation with SUR produces more efficient coefficients than estimation with separate regressions, especially when the disturbances are highly correlated and the sets of independent variables are not highly correlated.

The SUR approach models such covariances by incorporating a latent variable, which can be regarded as an implicit unmeasured 'organisational' effect on performance across all indicators. This can be defined as any influence on overall organisational performance, whether or not it is within the direct control of the organisation. SUR is best applied when each equation in the set of equations has a different set of independent variables (Tien 2003). In budgeting, this allows to recognise that not all subunits are alike and to group the subunits by type in different models. Putting subunits into separate equations allows for agencies to be influenced differently by the same independent variables and to be influenced by different independent variables.

According to Martin and Smith (2005) such simultaneous modelling of performance measures might be useful because it economises on the need for detailed modelling of individual performance measures. Instead of analysing multiple indicators with respect to the possible ‘organisational’ effect, it is considered as a latent variable. Consequently, it also makes it simpler for the need to measure factors that affect performance across all performance measures, such as environmental variables, as they are implicitly included. As the important information is exploited in the covariances between performance measures, it reduces the large confidence intervals frequently observed in single equation models, which are in part caused by omitted or poorly measured explanatory variables. More sensitive modelling of interactions may result in different inferences about the level of an organisation’s performance on specific indicators. Compared to the piecemeal modelling, the SUR models reduce the standard errors, and thus provide more secure/robust performance rankings, without recourse to additional data or the highly questionable aggregation of performance indicators which is the case in traditional productivity models.

SUR is not widely used in evaluating the performance of public agencies (Martin and Smith 2005), although examples include health care (Martin and Smith 2005; Hauck and Street 2006; Lovell *et al.* 2009).

2.1.12. Conclusion

Several techniques have been developed to measure performance as adequately as possible, most of which were briefly discussed above. As one can see, there is no perfect method which would cover all the necessary aspects in an unbiased and comprehensive way. The unadjusted absolute single performance indicator is a starting point for measuring performance, but as it is absolute, it only describes the volumes of inputs, outputs or other factors in an isolated way. A little more comprehensive is ratio analysis, which also allows to take into account the relative measures such as ‘per capita’ or ‘per provided service’. However, the ratio analysis also lacks comprehensiveness due to fragmented ratios and lack of accounting for the environmental factors. The risk adjustment makes assumptions on the expected values based on historical information and compares them to the observed values. Such methods have received much critique due to the difficulties on choosing an appropriate methodology.

The Four Quadrant model is appealing due to its clarity and graphical representation but this comes from the (over-)simplification of the underlying problem of performance measurement. The composite index analysis aggregates different performance information and thus introduces a comprehensive approach, but the negative side is finding the correct weights for different indicators. By aggregating the information, some is also lost. Cluster analysis groups similar entities and thus provides opportunities to take the environmental factors into account, but as a trade-off the selection of cut-off criterion complicates the interpretation of the results. Popular as a public sector management tool, the BSC allows a comprehensive and balanced approach to measure performance but some issues arise when one would like to evaluate efficiency and compare different subunits (benchmark).

Regression analysis can be used to comprehensively take into account various indicators as well as control for environmental factors. However, the general linear regression estimates the average of the observations and not the efficient frontier. Multilevel analysis adds a possibility to analyse the performance hierarchically in different levels of service providers. SUR simultaneously takes into account environmental factors, resource levels, efficiency, substitution, and data quality and thus provides a comprehensive method to assess the performance in public agencies. However, SUR has not yet seen many applications. In addition, frontier analysis methods have gained popularity in recent decades, as they provide the possibility to measure the performance of public agencies comprehensively by following the microeconomic production theory. The frontier analysis methods and their pros and cons are analysed more thoroughly in the following section.

2.2. Frontier analysis methods

2.2.1. The possibilities and limits of frontier analysis methods

The frontier analysis methods, which are mainly used in the framework of productivity analysis, have a common purpose of modelling the frontier of feasible performance. The frontier can be estimated under various underlying assumptions and estimation methods. As the next step, the observed organisation's performance indicator is then compared to such frontier and relative efficiency is found. Frontier analysis methods compute the observations' distance from the frontier – this means measuring the efficiency or performance as maximal-minimal proportionate feasible changes in an activity with given technology (Simar and Wilson 1998).

Frontier analysis methods have evolved rapidly over the past few decades and reached a high level of technical refinement (Smith and Street 2005), while there is still no consensus on which method would be more appropriate to implement in practice. Two schools of thought are mainly distinguished: some prefer econometric methods which use stochastic and parametric models (e.g. Battese and Coelli 1993; Kumbhakar and Lovell 2003; Greene 2005), and others who prefer linear programming methods which use mainly deterministic and non-parametric models (e.g. Simar and Wilson 1998; Thanassoulis 2001; Cooper *et al.* 2011). A few articles investigate the problems of sensitivity in case different methods of frontier analysis are used on the same data (see Simar and Wilson 1998; Jacobs 2001; O'Neill *et al.* 2008; Aubyn *et al.* 2009). More recently, the two schools of thought have introduced synthesised methods taking advantage of both sides, such as StoNED (Kuosmanen 2006) and non-parametric stochastic analysis (Henderson and Parmeter 2015).

Frontier analysis methods also have a great relevance in modelling public sector performance (Martin and Smith 2005), particularly DEA, as it accommodates multiple inputs and multiple outputs in a single model, which deals with the issue of goal ambiguity in the public sector. For some examples, where frontier analysis and other methods are used, see Table 2.1. For an extensive list of empirical applications and a literature review, see Fried *et al.* (2008).

Frontier analysis methods have also received vast critique (Smith and Street 2005). Although popular amongst scholars, the frontier analysis methods are rarely used as direct policy tools. In addition to general reluctance of policymakers towards statistical methods, such scarce use can be attributed to the limits of these techniques (Daraio and Simar 2007). First is the impossibility to extend the efficiency analysis beyond the current regression sample – making it useless for subunits not included in the sample (Daraio and Simar 2007). Similarly, comparison of two different efficiency studies is of little use.

Second, the several assumptions made about the production function and inefficiencies cannot be successfully verified, and ensuring the robustness of the results is complicated (Daraio and Simar 2007). Similarly, the lack of knowledge about the 'true' production process restricts the development of a convincing theoretical model (Martin and Smith 2005).

Third, the initial result of efficiency analysis is a single composite measure (see limits from Subsection 2.1.6), which might not be helpful from a managerial point of view (Martin and Smith 2005).

Fourth, considering the environmental influences and dynamic effects is of importance (Smith and Street 2005: 411–414). As they argue, in whatever environmental conditions the public agencies operate, there are some agencies which have preferable conditions. This leads to the case where the production possibilities frontier would be different for agencies in different environments. Thus, one has to account for these differences in the efficiency analysis. However, this is complicated in two ways: first, it is impossible to simplify environmental factors to the level suitable for modelling, and second, many public agencies are already (imperfectly) trying to take the environmental factors into account in the resource allocation process and so complicating the *ex post* analysis. In the case of DEA, there are two main possibilities to take the environmental factors into account. They can be considered as inputs, and so the observable subunit would be compared to other subunits that are in the same or more

Table 2.1: Frontier analysis and other methods used to benchmark the performance of the public sector

	Frontier analysis methods		Other methods
	SFA	DEA	
Rescue services	Jaldell (2002)	Jaldell (2002); Choi (2005); Sánchez (2006); Lan *et al.* (2009); Horton (2011); Peng *et al.* (2014)	Bouckaert (1992)
Schools/ universities	Aubyn *et al.* (2009)	Abbott and Doucouliagos (2003); Afonso and Aubyn (2006); Aubyn *et al.* (2009)	Thanassoulis (1993); Johnes and Johnes (2009); Salgado *et al.* (2014)
Hospitals and health care	Jacobs (2001); Greene (2004)	Jacobs (2001)	Iezzoni (1997); Majeed *et al.* (2001); Leyland and Groenewegen (2003); Hauck and Street (2006); Riewpaiboon *et al.* (2007); Gyrd-Hansen *et al.* (2012)
Other	Container ports Cullinane *et al.* (2006)	Forest management - Kao (2000), Quality of life in counties – Põldaru and Roots (2014)	

Source: Author's compilation.

adverse environmental conditions. Another possibility is to first estimate the model without environmental factors and then incorporate them in a second-stage analysis. There is no consensus among scholars as to which way would provide more acceptable results. In the case of SFA, typical regression analysis concerns arise (selection of model, omitted variables). In addition, the environmental factors can be considered as part of the random error or part of the inefficiency error. In terms of dynamic effects, the public agencies are path-dependent, meaning that the performance for the current period is also to an extent dependent on the performance from the last period(s). To quantify such issue is a very complex task. One can consider such effect also as an environmental factor, but so far, such challenging task has not been given much attention in the modelling process.

The most common methods of frontier analysis are discussed in the next subsections: data envelopment analysis (DEA), which is a deterministic and non-parametric method; stochastic frontier analysis (SFA), which is a stochastic and parametric method; and other methods, which attempt to fit the advantages of both approaches.

2.2.2. Data envelopment analysis

DEA is a popular data-oriented approach applied to the evaluation of relative efficiency in the public agencies. The linear programming approach enables the use of multiple inputs to produce multiple outputs (Charnes *et al.* 1978; Cooper *et al.* 2011). DEA was created specifically with the work of non-profit organisations and public agencies in mind (Charnes *et al.* 1978: 429). With available quantities technical efficiency can be evaluated, when quantities and prices are available, the economic efficiency can be evaluated and decomposed into technical and allocative efficiencies.

No assumptions are made about the shape of the (production) function when using DEA. The method can handle multiple numbers of inputs and outputs simultaneously. This relies on general regularity assumptions, such as free disposability, convexity and assumptions related to returns to scale. Another similar estimator named free disposal hull (FDH) (Deprins *et al.* 1984) assumes only free disposability (so it is non-convex).

The FDH can be solved using a series of minimax problems. For input efficiency in a FDH technology set, the first step would be to identify decision-making units (DMUs) that weakly dominate (they produce at least as much of each output with no more of each input). This formulates the frontier of FDH. From there, the radial efficiency can be measured to the most dominating observation (peer) (De Borger *et al.* 1994). An appealing characteristic of using FDH in public management is that is uses the 'real' observations as peers, so every analysed DMU will get one clear-cut comparison (unlike DEA, which assumes convexity and thus the comparison can be to a 'virtual' peer, which comprises of multiple 'real' peers).

However, the lack of assumptions causes many possible problems for the DEA and FDH estimators. The deterministic performance frontier is extremely sensitive to the number and distribution of observations as well as errors in the data: in a case where the best efficiency rating results from data errors, the other agencies' efficiency is evaluated at levels too low compared to the actual efficiency of their work. However, there are no statistical tests to verify the suitability of the model used in the prospective dataset. Therefore, according to its opponents, the method does not have sufficient justification for benchmarking work efficiency for regulatory purposes or for setting efficiency improvement targets (Smith 2006: 86–87).

All sorts of problems related to performance (including efficiency) measurement arise from the multiplicity of inputs and outputs: how to provide various inputs and outputs with target weights characterising their importance; which result (output) is more important to achieve; and how and in what proportions the inputs (outputs) are interchangeable. In the DEA method, the algorithm normally generates the goal's formal weights. These weights, however, do not necessarily coincide with society's perceived priorities in evaluating the usefulness of the service (Jacobs 2001: 20). Since there is no single solution to this problem (there are developed methods, which constrain the size of these limits), this is another reason why the DEA method is not applied daily on shaping the policy decisions (Pidd 2012: 291-292).

2.2.3. Stochastic frontier analysis

SFA was proposed by Aigner *et al.* (1977) and Meeusen and van Den Broeck (1977). The aim of parametric efficiency evaluation methods is to establish a cost frontier model (aggregated cost indicator is modelled as a function of output or outcome indicators) or a production frontier model (aggregated output or outcome indicator is modelled as a function of various input indicators). Public agencies costs are generally observable, measurable, and aggregatable, but it is difficult to create an aggregated output indicator because there are usually no market prices for outputs. Therefore, cost frontier modelling is usually preferred in benchmarking the efficiency of public services (Smith 2006: 85).

In contrast to the traditional statistical methods, a common feature of SFA models is to take special interest in the residual. In the early applications, the frontier was assumed to be deterministic (DFA – all relevant quantities, prices, etc. are observed and measured without error and the functional forms are known), which means that it could be estimated using the least squares estimation method and all of the error term was then considered as inefficiency. Such modelling is named the corrected ordinary least squares (COLS) approach. The frontier is estimated by shifting the regression line so it passes the fully efficient observation. Next, the inefficiency of other subunits is estimated as the distance from the corrected regression line (difference of the largest residual). This, however, is a stiff solution, leaving no alternative interpretation possibilities, such as measurement error or omitted variables. To allow such possibility, SFA was developed.

SFA decomposes the residual into two with zero covariance: the traditional two-sided symmetrical element that characterises the random variation (noise), and the one-sided asymmetric element that characterises the concerned public agency's relative inefficiency (Bogetoft and Otto 2010: 204). Function parameters can be assessed using the Maximum Likelihood (ML) method. The *ex ante* determination of the u_i distribution is a prerequisite in cross-sectional data – exponential, half-normal, or gamma distributions are typically used, on which the model parameter vector β is assessed. There is no theoretical basis for favouring one asymmetric distribution over another (Schmidt 1985). The efficiency estimates are then obtained on the basis of the residuals.

SFA has the advantage of assuming the existence of random shocks and measurement errors in the data set. In the deterministic model, the distance from the cost frontier is interpreted as relative inefficiency, but in the stochastic model, the residual is decomposed as random variation (noise) and inefficiency, which should give a more precise evaluation of the efficiency of an agency (Porcelli 2009: 17). Another strength of the method is the opportunity to examine more closely the structure of the phenomenon and the impact of various indicators on the structure of efficiency (the parameters are interpretable and decomposable). Consequently, the SFA method has a stronger theoretical economic background than the DEA method. As a weakness, the strong assumption set is highlighted: knowledge of the functional relationship (the shape of the function) and the distribution of relative inefficiency is assumed. In reality, however, the *a priori* set characteristics are unknown. In addition, it is highly likely that assumptions result in an error in evaluations, as inefficiencies may not actually be following a known distribution.

In conclusion, Table 2.2 compares the advantages and disadvantages of DEA and SFA. The disadvantage of using the DEA method is that potential random shocks and measurement errors are disregarded – it is assumed that the distance from the frontier is fully characterised as inefficiency. However, there are no specification tests in DEA because the functional shape of the relationship between the inputs and outputs is not assumed, and furthermore, the distribution of (in)efficiency is not assumed. Thus, the DEA is based only on the used data set ('the data speaks for itself') (Cullinane *et al.* 2006: 355–356).

In several previous studies, the benchmarking results obtained using DEA and SFA are compared (see Jacobs 2001; O'Neill *et al.* 2008; Aubyn *et al.* 2009). In a case where the functional shape of the frontier should be possible to be clearly defined, the SFA is expected to give more accurate results. In practice, however, it is very difficult to define and determine the shape of the function, and often there is no panel data available for comparable agencies to confirm the results. Another fact hindering the evaluation of the efficiency of public agencies using the SFA method is the multitude of objectives. Since both linear programming and econometric methods have their own advantages and disadvantages, various

Table 2.2: The comparison of data envelopment analysis and stochastic frontier analysis

Data Envelopment Analysis	**Stochastic Frontier Analysis**
Deterministic (distance from the frontier is considered inefficiency)	Stochastic (decomposition of inefficiency and the error term)
Non-parametric	Parametric
Multiple inputs and outputs concurrently	Multiple inputs and one output or vice versa
No assumptions about the functional shape	Specification of the cost (production) function and the distribution of inefficiency is assumed
Sensitive about outliers and data errors	Assumes the existence of random shocks and measurement errors in the data set
No specification tests	The parameters are interpretable and can be tested, therefore it has a stronger theoretical economic background
Problems about weighting the importance of inputs/outputs	Too strong assumptions that cannot be justified in the real world

Source: Author's compilation (based on the literature discussed above).

methods are used in the empirical study of evaluating the performance of FRS in the next chapter.

2.2.4. Further developments of frontier analysis methods

Building a bridge between the SFA and DEA approaches has been an important task in the productivity analysis, and has seen much progress. For a profound overview of the developments, see Kuosmanen and Kortelainen (2012). The semi/non-parametric SFA is based on the contributions from the SFA method, when replacing the parametric frontier function with a non-parametric specification, which can be estimated by kernel regression (Fan *et al.* 1996; Kneip and Simar 1996) or local (polynomial) ML (Kumbhakar *et al.* 2007; Simar and Zelenyuk 2011). Although such a model is parametrised similarly to the standard SFA model, all of the model parameters are approximated by local polynomials.

Based on the DEA methods, Banker and Maindiratta (1992) considered ML estimation of the stochastic frontier model subject to the global free disposability and convexity axioms. Although the theoretical model combines the features of DEA and SFA models, solving this problem has proved extremely difficult and has not seen any empirical studies (Kuosmanen and Kortelainen 2012). A non-parametric version of the COLS method (discussed above) was proposed by Kuosmanen and Johnson (2010), which as a positive feature has a generally higher discriminatory power than basic DEA, but as a negative feature the deterministic frontier shifting method is more sensitive to stochastic noise.

Stochastic non-smooth envelopment of data

One of the methods that has received more feedback from policymakers, mainly in the field of regulating electricity distribution networks, is Stochastic Non-smooth Envelopment of Data (StoNED) (Kuosmanen and Kortelainen 2012). The semiparametric frontier model combines the DEA-type non-parametric frontier, which satisfies monotonicity and concavity, with the SFA-style stochastic homoscedastic composite error term.

This two-stage method estimates in the first stage the shape of the frontier without any assumptions about its functional form or smoothness (similar to DEA) using the convex non-parametric least squares (CNLS). CNLS identifies the function that best fits the data from the family of continuous, monotonic increasing, concave functions that can be non-differentiable. Unlike the DEA frontier, for which the small number of influential observations make it sensitive to outliers and data errors, the StoNED frontier uses information from full sample of observations and infers the expected value of inefficiency in a probabilistic fashion.

As next step, based on the skewness of the CNLS residuals, the variance parameters of the stochastic inefficiency and noise terms are estimated. The noise term is assumed symmetric, so similarly to the SFA, the skewness of the regression residuals is part of the inefficiency estimate. The inefficiency and noise terms are assumed to follow a parametric distribution, and thus the variance parameters are estimated using the method of moments or pseudolikelihood techniques. The conditional expected value of the inefficiency can be obtained using the JLMS estimator (Jondrow *et al.* 1982).

The StoNED method also has some setbacks (Kuosmanen and Kortelainen 2012). Similar to the DEA, the StoNED is affected by the curse of dimensionality, which means that the sample size needs to be very large when the number of input variables is high. In addition, the assumptions similar to the SFA method imposed on the error term and inefficiency are rather restrictive, and might often be inadequate.

In general, such merging of different approaches is becoming more and more popular and technically evermore advanced. As these methods are rather new and evolve in various directions, they have not seen many empirical applications nor been applied by policymakers.

2.3. The case of uncertainty in productivity analysis methodology

Most production technologies are stochastic in their nature. However, as can be concluded by the following analysis, the efficiency literature has not typically taken this into account, as efficient frontier is generated in a representation of a non-stochastic technology. Such practice places strong and untested *a priori* restrictions on stochastic technologies. The real-world data, however, will reflect the multiple sources of behavioural differences across firms. Consequently, when one attempts to estimate a supposed common frontier technology, problems arise due to ignoring the other sources of stochasticity in production technologies (O'Donnell and Shankar 2010).

The issue that environmental conditions can have an effect of underestimating the efficiency of a producer, is recognised, but most of the literature lacks an explicit recognition that production invariably takes place under conditions of uncertainty (O'Donnell and Shankar 2010). Even though the SFA models are regarded as stochastic, the stochastic elements arise mostly from econometric concerns such as measurement error and missing variables. Only a few models take into account that stochasticity can be a response to the stochastic decision environment in which firms or public agencies actually operate. Such models usually use latent variables representing uncertainty and it is subsumed in the noise and/or inefficiency error terms (Battese *et al.* 1997; Kumbhakar 2002). This, however, does not take into account the role that a stochastic decision environment has on decision-making and thus observed outcomes.

Therefore, analysing the production (in the private as well as public sectors) under uncertainty requires another set of concepts and techniques. Uncertainty can be considered as two different concepts:

1. The validation of the model and uncertainties due to the measurement and model specification errors (uncertainty about how good a model is), and

2. Due to lack of knowledge about the future, and its effect on decision-making.

With respect to performance measurement in the public sector, the Royal Statistical Society has made a strong note that all performance data should come with appropriate measures of uncertainty (Bird *et al.* 2005) in the meaning of declaration, how and why the model works out on a certain case. Thus, focussing on the first concept of uncertainty. Second, the paper states that it is statistically not wise to increase the targets of performance indicators progressively requiring the next year's performance to be better than this year's. In such a case, failure is inevitable, as uncertainty has been ignored also in this year's result for performance indicator.

The uncertainties of the first kind are taken into account usually when it is convenient for statistical analysis (easier to apply on parametric models) but the second kind of uncertainties, which can be considered as lack of *ex ante* knowledge about the demand for services (outputs) and environment, have received much less attention in the field of efficiency studies. This thesis focuses on the second concept, as there is a need to develop estimation techniques that will divide the differences in (technical) efficiency from difference caused by the stochastic nature of production (O'Donnell and Shankar 2010). In some previous cases, the policy recommendations derived from inefficiencies might have been caused by production uncertainties and not technical inefficiencies.

Expected utility model

The traditional approach to take uncertainty/risk into account is the EU model. For a comprehensive treatment of the approach, see for example Quiggin (2012). In applied studies, the approach has simplified into mean-variance models, where EU is maximised as a function of the expected mean and variance of profits, based on stochastic production function. Such model is only valid when either the utility function is quadratic or profits are distributed normally, and such restrictive assumptions have been criticised profoundly (O'Donnell and Woodland 1995). Another problem is that the traditional approach typically does not consider the interaction between the uncontrolled (uncertain) variables and the decision variables controlled by the decision maker. This has been criticised by Chambers and Quiggin (2000), who introduce the state-contingent approach as an alternative.

In a further development, O'Donnell and Woodland (1995) make an assumption that EU maximising producers face price and yield risk and have access to a stochastic production technology which is multiplicative in its deterministic and stochastic components. Based on these assumptions, the producers will choose such inputs that would minimise the cost of producing the so-called 'planned output'. In the application, the input cost share and cost functions can be expressed in terms of actual outputs. The method incorporates the stochastic component of the technology in a disturbance term in the cost function. To estimate the variance of the stochastic term in the production function and to obtain a consistent estimate of this term for each time-series observation in the sample, they exploit the error component structure of the disturbances.

State-contingent approach

The foundations for developing the state-contingent production under uncertainty approach was introduced by Chambers and Quiggin (2000), which is an alternative way of describing and analysing production decisions under uncertainty. The state-contingent production approach applies analytical tools from microeconomics in a stochastic production setting, given that *ex ante* preferences and production technologies are defined properly. Chambers and Quiggin (2000) applied the theory to analyse problems of choice under uncertainty, including the problems of moral hazard, incentive regulation and portfolio choice. They describe different types of state-contingent production technologies, including state-allocable. Such technology describes the possibility for a producer to manage uncertainty by allocating inputs to different states of nature. In conventional stochastic production theory, the role of inputs remains the same, regardless of which state of nature applies and it does not allow to substitute the state-contingent outputs. Such a technology, which does not permit any substitutability between state-contingent outputs is referred as output-cubical (e.g. Leontief).

As an advantage, the state-contingent approach considers the interaction between controllable and uncontrollable inputs (the uncertain states of nature). The state-contingent production approach states that '*insight into problems involving production under uncertainty is best gained by consistent use of the concept of states of nature and of actions having different consequences in different states of nature. Production decisions are thus best viewed as choices between bundles of state-contingent goods.*' (Chambers and Quiggin 2000: 17). The state of nature is theoretically defined as a comprehensive set of mutually exclusive and exhaustive set of possible descriptions of the state of the world (Chambers and Quiggin 2000: 17–18). This contrasts with approaches where decisions are modelled as choices between random variables indexed by input (effort) levels or between probability distributions over a finite set of possible outcomes. These alternatives are referred to as parametrised distribution formulation and the outcome-state formulation of production under uncertainty.

The state-contingent approach has not seen many empirical implementation, mostly because the *ex ante* production choices of firms (allocations of inputs to different states of nature) are only partially observed (Rasmussen 2006; O'Donnell and Shankar 2010). Recent literature has put much effort on estimating state-contingent technologies that have involved predicting unobserved states of nature and/or *ex ante* production choices (Shankar and Quiggin 2013). The combination of these predictions with observed input and output data, allows the possible estimation of the technology using conventional econometric techniques.

Rasmussen derives the criteria for risk-averse producers for production under uncertainty for single input (Rasmussen 2003) and multi-variable input case (Rasmussen 2006). Based on a formal definition of 'good' and 'bad' states of nature, he found that the use of inputs and levels of production of strictly risk-averse and those of risk-neutral producers depend on the type of input. Criteria were derived for different types of input, including state-specific and state-allocable input. He concludes that specific criteria for these two types of inputs are redundant; the general criteria derived will cover any type of input.

He shows that the approach has an advantage of being based on marginality principles and optimisation tools, but suffers the lack of empirical applicability as the state-contingent production functions are not typically available. The article also demonstrates that optimal production decisions under uncertainty may be identified without knowledge of the state-contingent utility function, when there are markets for state-contingent insurance contracts. In addition, for a comparison of the state-contingent approach and EU model, with respect to choice of utility function and the description of production technology (production function), see Rasmussen (2006).

O'Donnell and Griffiths (2006) show how to empirically estimate unobserved states of nature and the parameters of output-cubical state-contingent technologies in a Bayesian finite mixtures framework. As a result, output-cubical technologies are found inconsistent with important stylised facts concerning behaviour in the presence of risk. In addition, they note that in situations, where state-contingent uncertainty plays a role, the SFA approach may lead to overestimation of inefficiency. Chavas (2008) estimates the cost function defined over predicted state-contingent outputs and thus the parameters of a more flexible state-contingent technology. O'Donnell and Shankar (2010) proposed a two-state case to overcome the problem of lack of data in the two-state case. Such model can be theoretically estimated by sampling theory or Bayesian methodology. In their application using Philippine rice data, the sampling theory failed due to an inability to dynamically control a non-linear least squares optimisation algorithm, but the Bayesian framework resulted in more plausible results.

O'Donnell *et al.* (2010) estimated a stochastic frontier which would allow state-allocable inputs. As a result, they show that the failure to account properly for the stochastic elements in planning, the standard efficiency analysis methods such as SFA and DEA may produce biased findings of inefficiency. Using the state-contingent framework, the producers are estimated to be fully efficient, while having just encountered an unfavourable state of nature.

This is followed by an empirical study by Nauges *et al.* (2011), where they specify a constant elasticity of substitution (CES)-type production function. The proposed model is a generalisation of the single-input model of O'Donnell *et al.* (2010) – output in a particular state of nature can still be non-zero even when none of the input has been allocated to that state (such an input is said to be state-general). They apply the multiple-input state-allocable model in a frontier framework and estimate levels of input-allocability and technical efficiency for farm data from Finland. They reject the assumption that production technology is output-cubical. Such findings are important because the value of timely information about the state of nature can be maximised under such technology. In an output-cubical model, the producers could change the scale of production as a response to new information, in the state-allocable model, they can reallocate inputs towards more likely states of nature. Thus, the state-allocable production technology has the capacity to actively manage production risks and integrates technological and financial approaches to risk management.

Another generalisation for O'Donnell *et al.* (2010) is proposed by Shankar and Quiggin (2013), in which they model production technology in a state-contingent framework that obviates the need to predict unobserved state-contingent outputs. The model is not explicit about the nature of producer risk preferences, as it is captured by the risk-neutral probabilities they assign to the different states of nature. In addition, they develop an econometric methodology to estimate the risk-neutral probabilities and the parameters of stochastic technology when there are two states of nature and only one of which is observed. They confirm the previously presented insights that rational producers, who use the same stochastic technology can make significantly different production choices determined by their risk attitudes and beliefs involving the relative probabilities of different states of nature. The obtained estimates on a simulated data using conventional OLS are found to be biased, which can be attributed to the misspecification of the underlying stochastic technology. The simulation shows that the estimation bias for the productivity parameter in each of the two states of nature and the elasticity of scale parameter is minimum when technology exhibits a low degree of substitutability between state-contingent outputs.

Shankar (2015) develops the approach further, and shows that conventional efficiency estimators such as DEA, DEAS (state-dependent DEA), SFA and output-cubical are systematically biased. As almost all conventional frontier models are output-cubical, such restrictive representation of technology can have a serious impact on policy implications. Mistakenly, the fully efficient units are considered inefficient, which indicates the possibility for efficiency gains that in reality are implausible.

Uncertainty is not widely considered in the efficiency studies. As the short literature overview indicated, there are many substantial issues involved, as the typical efficiency evaluation might cause biased results, followed by flawed policy implementations. The next section reviews the literature that has analysed the efficiency in the field of FRS.

2.4. Literature review of frontier analysis methods used in the field of fire and rescue services

Based on the analysis of Sánchez (2006), the first studies addressed to the efficiency of the FRS were based on non-frontier methodologies, estimating uni-equational cost functions (Hirsch 1959; Hitzhusen 1972; Kristensen 1983; Southwick Jr and Butler 1985; Duncombe 1991; 1992; Duncombe and Yinger 1993), production functions (Simon *et al.* 1943; Getz 1979), discriminate analysis (Coulter 1979) and productivity ratios (Schaenman and Swartz 1974).

Although frontier analysis methods are used widely to evaluate public agencies' performance, they are not commonly used for FRS. The main problem might be the lack of comparable data – typically the FRS brigades are managed and maintained by local municipalities, and the data collected thus varies depending on the municipality. DEA has been used on several occasions to analyse the efficiency of FRS brigades (Jaldell 2002; Choi 2005; Sánchez 2006; Lan *et al.* 2009; Horton 2011; Peng *et al.* 2014), and SFA has been applied only in a few cases (Jaldell 2002; Holmgren and Weinholt 2016). For the list of articles, see Appendix 4.

When using the frontier analysis methods, the different costs (Jaldell 2002; Choi 2005; Lan *et al.* 2009; Horton 2011; Peng *et al.* 2014; Holmgren and Weinholt 2016), staffing levels (Jaldell 2002; Sánchez 2006; Lan *et al.* 2009), and number of fire engines (Sánchez 2006; Lan *et al.* 2009) have been used as inputs.

As outputs, number of fires and emergencies (in some cases the total number and in others per person) (Jaldell 2002; Choi 2005; Sánchez 2006; Lan *et al.* 2009; Horton 2011; Holmgren and Weinholt 2016), number of turnouts (Holmgren and Weinholt 2016), fire deaths and losses (Choi 2005; Horton 2011; Peng *et al.* 2014), response times (Jaldell 2002; Holmgren and Weinholt 2016), number of prevention-related campaigns (Sánchez 2006), prevention workers (Holmgren and Weinholt 2016), and size of fire (Sánchez 2006) (characterised by the factors that affect the firefighting service: a) the conditions the buildings were in, represented by their age; b) the presence of high risks in the property which were combined in the percentage of buildings over two floors high and the economic activity index; and c) the probability of the fire spreading, which was represented using the variable density of population) have been used.

The current thesis is mainly based on the thesis conducted by Jaldell (2002) by widening the conception, introducing the uncertainty and deepening the analysis and discussion about the indicators, methods and models. The study analysed the Swedish rescue services and used both the DEA and SFA methods (although the methods were not evaluated in a comparative way). By using the SFA method, the staffing level was modelled as a function of risk (number of fires, emergencies, risky industries) and environment (population, size of the area). Only the size of the population had an impact on the staffing level. In addition, the results showed that the mean input saving potential is 30% (the mean efficiency score was 0.7) and there has been no improvement in efficiency over time. Using the DEA method, the aggregated costs were used as inputs and the turnout time (number of people reached within x minutes) and suppressing power (total number of firemen reaching the fire within x minutes) as outputs. In addition, the environmental factors (population, size of the area, population density, number of population centres with at least 1,000 inhabitants) were taken into account. The results showed that the larger municipalities had higher efficiency scores, and no optimal size of production (comparing the returns to scale for the frontier units) could be found.

Another study, closely related to the current thesis, is from Holmgren and Weinholt (2016). The aim of the article was to assess whether the organisational changes in the Swedish FRS have had an effect on the targeted cost-efficiency gains. As the Swedish FRS are provided at municipal level, the cooperation between neighbouring municipalities has intensified, leading to a merger (formalised cooperation) of the service provision. In addition, other actors such as home care personnel or security officers are collaborating more in order to reduce the response time. Another initiative changing the functioning of FRS is the introduction of the first response person, meaning that the unit leader heads straight to the scene without first going to the rescue station. Lastly, the FRS have gained more assignments, such as helping the ambulance (providing first aid until the arrival of ambulance). Holmgren and Weinholt (2016) used SFA on Swedish municipalities for the years 2009–2012. As inputs, the cost of capital and wage of firefighters are used; as outputs the number of turnouts, number of fires per person, number of people employed for preventive work and response time. In addition, the population, income, assessed values for real estates and area are used for environmental factors. As a result, they found that none of the policy changes have had an effect on cost-efficiency. The industry-level average efficiency was similar to the findings by Jaldell (2002), being around 70%.

Choi (2005) studied the rescue departments' efficiency in 60 counties in Florida using the DEA method. The article described four models and the correlations between efficiency scores and exogenous factors (population, density, manufactures, property, tax millage rate). Sánchez (2006) focussed on estimating the effect of environmental conditions on the technical efficiency of FRS in 29 Spanish municipalities with over 50,000 inhabitants by constructing a 'basic' and 'complete' DEA model, which incorporated environmental variables that by Tobit regression could be considered as having effect on the performance. As an interesting follow-up for an efficiency evaluation, Lan *et al.* (2009) provided decision-making strategies for reallocating the resources based on DEA results. Horton (2011) evaluated whether using the Citistat PMS in the management of a municipality improves the effectiveness and efficiency of FRS brigades. It was found that there were differences in the organisational behaviour but that these did not result in differences in performance according to the efficiency scores evaluated by the DEA method. Peng *et al.* (2014) evaluated differences and changes over time of aggregated rescue services between countries and not FRS brigades; thus, the results are not directly comparable nor relevant for the present study. Comparing the results of different studies does not give the full picture, because the results matter only on the chosen sample. Thus, when one study finds that the mean efficiency would be 60% and another finds it would be 80%, no conclusions can be made about comparable efficiency level in these samples because one does not know the true frontier. On the other hand, one can analyse the variation between the efficiency estimates.

2.5. Estimation of frontiers in case of demand uncertainty

Following on the problem statements introduced in the Chapter 1.4, one must make some assumptions in order to estimate these. In the efficiency literature, it is common to assume that:

1. all relevant quantities, prices and environmental variables are observed and measured without error;
2. production frontiers are piecewise linear;
3. outputs, inputs and environmental variables are strongly disposable;
4. production possibilities sets are convex.

Under these assumptions, the cost minimisation problem of the central agency introduced in the first chapter can be solved using DEA (Charnes *et al.* 1978). If $Q(.)$ is linear, then output maximisation problem of each subunit can also be solved using DEA. Relaxing the convexity assumption, these problems can be solved using free disposal hull (FDH) (Deprins *et al.* 1984). Using the same notation, as in Chapter 1.4, in order to estimate the cost-efficiency at the industry level, any under-resourcing of subunits and maximum aggregate output is solved as follows.

Cost-efficiency

In order to estimate the minimum cost of meeting the MSL (Equation 1.12) of subunit i in period t using DEA, the following problem is to be solved:

$$\min_{x,\lambda_{11},\ldots,\lambda_{It}} \Big\{ w'_{it}x : m_{it} \leq \sum_{h=1}^{I}\sum_{r=1}^{t} \lambda_{hr}q_{hr}, \ \sum_{h=1}^{I}\sum_{r=1}^{t} \lambda_{hr}x_{hr} \leq x, \\ \sum_{h=1}^{I}\sum_{r=1}^{t} \lambda_{hr}z_{hr} \leq z_{it}, \ \sum_{h=1}^{I}\sum_{r=1}^{t} \lambda_{hr} = 1, \ \lambda_{hr} \geq 0 \Big\}.$$

If one replaced m_{it} with q_{it}, then this would be a standard (or naïve) cost-minimisation problem.

In case one assumes the constant elasticity of transformation for DFA method, the cost function can be written as:

$$-\ln(C_{it}/w_{1it}) = \sum_{m=1}^{M} \lambda_m \ln\left(\frac{w^*_{mit}}{\lambda_m}\right) + \frac{1}{\tau\eta}\ln\left(\sum_{n=1}^{N}\gamma_n q^{\tau}_{nit}\right) - \sum_{j=1}^{J}\kappa_j \ln z_{jit} - \xi(t) - u_{it} \quad (2.1)$$

where $w^*_{mit} \equiv w_{mit}/w_{1it}, \xi(t) = \ln B(t)$, $\eta > 0$, $\lambda = (\lambda_1, ..., \lambda_M)' \geq 0$, $\gamma = (\gamma_1, ..., \gamma_N)' \geq 0$, $\tau \geq 1$. $\gamma'\iota = 1$ and $u_{it} \equiv -\ln CE_{it} \geq 0$. From there, the $CE_{it} = \exp(-u_{it})$.

Under-resourcing

To determine whether subunit i was under-resourced in period t, using DEA, one should solve the following problem:

$$\min_{x,\lambda_{11},...,\lambda_{It}} \Big\{\mu : m_{it} \leq \sum_{h=1}^{I}\sum_{r=1}^{t}\lambda_{hr}q_{hr},\ \sum_{h=1}^{I}\sum_{r=1}^{t}\lambda_{hr}x_{hr} \leq \mu x_{it},\ \sum_{h=1}^{I}\sum_{r=1}^{t}\lambda_{hr}z_{hr} \leq z_{it},\ \sum_{h=1}^{I}\sum_{r=1}^{t}\lambda_{hr} = 1,\ \lambda_{hr} \geq 0\Big\}.$$

The value of μ at the optimum is an estimate of $1/D^t_I(x_{it}, m_{it}, z_{it})$. If one replaced m_{it} with q_{it}, then this would be a standard (or naïve) ITE problem.

In case one assumes the constant elasticity of transformation for DFA method, the input-oriented model can be written as:

$$-\ln x_{1it} = \xi(t) + \sum_{j=1}^{J}\kappa_j \ln z_{jit} + \sum_{m=1}^{M}\lambda_m \ln x^*_{mit} - \frac{1}{\tau\eta}\ln\left(\sum_{n=1}^{N}\gamma_n q^{\tau}_{nit}\right) - u_{it} \quad (2.2)$$

where $x^*_{mit} \equiv x_{mit}/x_{1it}, \xi(t) = \ln B(t)$, $\eta > 0$, $\lambda = (\lambda_1, ..., \lambda_M)' \geq 0$, $\gamma = (\gamma_1, ..., \gamma_N)' \geq 0$, $\tau \geq 1$. $\gamma'\iota = 1$ and $u_{it} \equiv -\ln ITE_{it} \geq 0$. From there, the $ITE_{it} = \exp(-u_{it})$.

Maximum aggregate output

To estimate the maximum aggregate output (Equation 1.14) that subunit i can produce in period t, using DEA, one should solve the following problem:

$$\max_{q,\lambda_{11},\dots,\lambda_{It}} \Big\{ Q(q) : \; q \leq \sum_{h=1}^{I}\sum_{r=1}^{t} \lambda_{hr} q_{hr}, \; \sum_{h=1}^{I}\sum_{r=1}^{t} \lambda_{hr} x_{hr} \leq x_{it}, \; q \leq d_{it}, \\ \sum_{h=1}^{I}\sum_{r=1}^{t} \lambda_{hr} z_{hr} \leq z_{it}, \; \sum_{h=1}^{I}\sum_{r=1}^{t} \lambda_{hr} = 1, \; \lambda_{hr} \geq 0 \Big\}.$$

If one deleted the constraint $q \leq d_{it}$, then this would be a standard (or naïve) aggregate-output-maximisation problem.

In case one assumes the constant elasticity of transformation for DFA, the output-oriented model can be written as:

$$\ln q_{1it} = \alpha(t) + \sum_{j=1}^{J} \delta_j \ln z_{jit} + \sum_{m=1}^{M} \beta_m \ln x_{mit} - \\ \frac{1}{\tau} \ln \left(\sum_{n=1}^{N} \gamma_n q_{nit}^{*\tau} \right) - u_{it} \quad (2.3)$$

where $q^*_{nit} \equiv q_{nit}/q_{1it}$, $\alpha(t) = \ln A(t)$, $\beta = (\beta_1, ..., \beta_M)' \geq 0$, $\gamma = (\gamma_1, ..., \gamma_N)' \geq 0$, $\tau \geq 1$. $\gamma'\iota = 1$ and $u_{it} \equiv -\ln OTE_{it} \geq 0$. From there, the $OTE_{it} = \exp(-u_{it})$.

This chapter firstly mapped and analysed the most common methods to measure the performance of a public agency. As the analysis of such methods was directing towards frontier analysis methods to be most suitable in order to solve the problem in hand (see Chapter 1.4), more emphasis was put on them, as they are in the focus of the current thesis. In addition, the chapter analysed two components relevant to the thesis: firstly, the literature on the previous use of the uncertainty in production and secondly, the literature on efficiency studies of FRS. Both of these components are rather understudies, so the current thesis is novel in terms of empirical concept and application. The last section of the chapter provided the estimation techniques in order to solve the problems introduced in the Chapter 1.4. The next chapter benchmarks the performance of FRS in the case of Estonia, Finland and Sweden.

3. ESTIMATING THE PERFORMANCE OF ESTONIAN, FINNISH AND SWEDISH FIRE AND RESCUE SERVICES

'Wellington House has no fire escape or fire doors and the fabric of the building would not stand the alteration, so it can't be sold as offices.'
'Then how can **we** *use it?' enquired Frank aggressively.*
'Government buildings do not need fire safety clearance.'
'Why?' demanded Frank.
'Perhaps,' Humphrey offered, 'because Her Majesty's Civil Servants are not easily inflamed.'
— J. Lynn & A. Jay, Yes Minister, The Economy Drive

3.1. Methodology for estimating the performance of the fire and rescue services

The performance measurement in a public agency has many possible usages, as was analysed in the first chapter of the thesis. In addition, multiple methods can be used to estimate the performance from different perspectives and priorities. These methods were compared in the second chapter in the thesis. To assess empirically the performance in a public agency, the fire and rescue services (FRS) field was chosen. This is the main topic of the current, third chapter. The FRS have seen some performance measurement and management attempts, mostly in the Anglo-Saxon countries – the UK (England, Wales and Scotland), the United States, Australia and New Zealand (see Chapters 1.2.2, 2.4 or Appendix 2). Many countries also provide some statistics about FRS, such as number of emergencies and casualties. But not many analyses have been undertaken to evaluate the FRS at a subnational level and then comparing the results in multiple countries.

Following O'Donnell (2016b), the policy-oriented performance analysis follows the steps:

1. Identifying the subjects of interest – the firms, public agencies, etc;
2. Listing all inputs, outputs and environmental variables;
3. Determining the economic quantities of interest;

4. Making regularity assumptions about technology sets;
5. Collecting/assembling price, quantity and/or value data;
6. Selecting functional representations of technology set;
7. Choosing an estimation approach;
8. Estimating the model and testing the model assumptions;
9. Seeing whether the results are robust to the assumptions and choices made in steps 4, 6 and 7.

The subjects of interest (1) in this case are the FRS brigades/regions and the rescue board, supervising and controlling their subunits (a two-level hierarchy in management). To understand in-depth the concept of FRS, one should analyse the organisations:

1. with respect to law, statutes, orders, acts, etc. in order to determine which services the FRS brigades have to deliver;
2. the form of law under which they operate;
3. the decision-making structure, coordination, finance, procurement, production, delivery, features of services;
4. the restrictions under which they act.

This assessment and analysis of FRS in the Estonian, Finnish and Swedish case is undertaken below (see 3.2).

To determine the relevant inputs, outputs and environmental variables (2), one should explore and compare the assignments and work of Estonian, Finnish and Swedish FRS agencies. For that, one should analyse different approaches of public sector management applied in the FRS. From the literature, one can define indicators and verify approaches that are applicable for performance benchmarking. As a last step, one should identify relevant performance indicators which have a theoretical ground. This has been done as well below (see Subsection 3.2).

To develop the appropriate benchmarking system, one should consider the goals of the subjects of interest and thus the economic quantities of interest (3). The current thesis proposes three main objectives for evaluating the performance of FRS. First, the cost-efficiency of the FRS at the industry level, which evaluates how much it would be possible to save resources from a central agencies' perspective, while taking the uncertain demand into account. Second, how many subunits have received too few resources (labour, vehicles, etc.) to provide the proposed MSL. Third, have the subunits provided the services efficiently by measuring the output technical (and mix) efficiency. For the theoretical justification see the first chapter (Subsection 1.4).

As the model is only a mere reflection of reality, one should make simplifying assumptions (4) about its nature. Thus, the technology sets are assumed to follow certain regularity assumptions (see Subsection 1.4).

The following step would be to collect the necessary data for performance assessment by measuring the indicators and creating the database (5). The descriptive statistics and analysis of FRS are provided in the following sections. The database consists of Estonian, Finnish and Swedish datasets, which have been collected with the help of officials in 2015–2017.

In the case of DEA/FDH, the functional form (6) is not assumed. In the case of DFA, the constant elasticity of transformation production function has been chosen due to multiple outputs. The estimation (7) is achieved using first the linear planning (for DEA/FDH) and second non-linear least squares (for DFA). The results of estimation (8) are then considered for each country. Thereafter, the robustness (9) of the models is discussed based on the results of different models.

For estimation, R (up to v3.4.0) language in the RStudio (up to v1.0.143) was used. The packages 'lpSolve', 'Benchmarking', 'frontier', 'plm', 'minpack.lm' and 'plyr' were also used for the analysis, while the packages 'ggplot2' and 'corrgram' were used to create graphs and 'stargazer' to receive tables.

As the last step, one is able to give suggestions for improving the performance by introducing performance indicators and benchmarking in the FRS field on the following:

1. which indicators to use;
2. how to conduct the benchmarking;
3. the possibilities how to use the results in planning, realisation and monitoring;
4. eventually the suggestion of reform.

In this chapter the performance of FRS in three countries, Estonia, Finland and Sweden is evaluated. The following section analyses the FRS systems in different countries and then the estimation results, based on the proposed models, are analysed. First, the number of emergencies in each country are analysed. This would be the basis for assessing the outputs of FRS. Second, the cost-efficiency at industry level is estimated and analysed for each country. Third, the under-resourcing of FRS brigades is analysed and fourth, the output-oriented technical and mix efficiency of FRS brigades is estimated and analysed. Finally, the comparison of different countries and possibilities for further research are given. In addition, the limits and setbacks of the study are analysed.

3.2. Fire and rescue services in Estonia, Finland and Sweden

3.2.1. National strategies – performance indicators used by Estonia

The following section analyses the available performance indicators used by Estonian policymakers, namely the Estonian Ministry of the Interior, which represents the governments' view and the Estonian Rescue Board (ERB – Päästeamet in Estonian) as the public agency. The collected indicators are named in the appendices (see Appendix 3). The aim of this subsection is to analyse which documents are relevant for the performance measurement of FRS and which of them are eligible directly to the national FRS brigades as subunits of ERB. This analysis is an input to generate the models to evaluate the efficiency of FRS brigades by frontier analysis methods.

In Estonia many indicators that describe a public agency's performance are lacking. Estonian public agencies so far have been mainly in the stage of building up organisational structures and therefore a constant object of reform, which has an effect on consistency of data collection. This situation has essentially not given many opportunities to implement performance indicators and measurement possibilities in public management for a longer time period (have been only part of a once-only project). Increasingly, however, the structure of public agencies in Estonia is stabilising and therefore, the opportunities to implement performance measurement arise. The structural reform of the ERB was completed in 2012, and thus the comparable year-to-year information about their activities has been collected. An original database has been created containing the annual data from the ERB. Therefore, it is possible to prepare a meaningful performance measurement framework.

The ERB is a centralised organisation, which has four different regional centres (North – PEPK, East – IEPK, South – LõPK, West – LäPK). Under them functions a network of 72 national FRS brigades, of which six are considered distant brigades (on small islands, rural areas, etc.) and therefore operate in a somewhat different way. The number of FRS brigades has been stable since the last reform in 2012, when nine FRS brigades were closed, although there have been some rearrangements and moving between locations[1]. The ERB is the third largest public agency in Estonia in 2017. In addition, the network of national FRS brigades is supplemented by 115 voluntary FRS brigades and four reserve rescue squads.

The ERB has a leading role in planning and operating the preparedness

[1] Due to the relocation, the Tartu and Tõrvandi and Tartu and Annelinna FRS brigades have been merged in the dataset, which results in 65 FRS brigades in total (72-6 distant brigades-1 due to merging=65). Additionally, the Lasnamäe and Mustamäe brigades have been considered as one, although it is a result of relocation.

for emergencies and responding to emergencies. It is also responsible for the development and implementation of national rescue policies. The ERB is a governmental organisation under the jurisdiction of the Ministry of Interior.

All of the main publicly available strategic plans relevant to the ERB are analysed with respect to possible (quantifiable) indicators. The most important document is the Development Plan for Internal Security 2015–2020, approved by the Estonian government on 5 February 2015 (Siseministeerium 2015b). Another valuable document is the Strategy of the Estonian Rescue Board 2015–2025 (approved at the end of 2014) (Päästeamet 2014). Since the role of volunteers is increasing in the future (for discussion, see Puolokainen *et al.* 2018), the publication National Orientations in the Development of the Voluntary Rescue 2013–2016 should be considered in determining the indicators (Siseministeerium 2013).

The different (quantifiable) performance indicators used in Estonian strategic documents, which would be the basis for the measurement of performance, are pointed out. The Development Plan for Internal Security 2015–2020 (Siseministeerium 2015b) is the most general document to cover all aspects and trends of internal security development. Since it is all-embracing, it rather lacks concrete quantifiable indicators but describes the progress in general (or qualitative) form.

The most important and comprehensive strategic guide for the ERB is their strategy for 2015–2025 (Päästeamet 2014). The strategy points out a few important trends in the environment that ERB has to function in. As part of the socio-economic environment one has to deal with the issue of ageing population, which leads to tougher competition in the work-force (as the rescuers should be very fit and rather young). In addition, elderly people are causing more fires and might be unable to deal with the fire extinguishing themselves. Another issue is the urbanisation process – mostly, younger people are moving into cities, leaving no proper workforce in the rural areas. In addition, the development of rescue services is highly connected to the attitudes of the population – the alcohol, drugs, tobacco and healthcare policies are relevant in the prevention process[2]. This leads to the public opinion of rescue services and awareness (there are a few public polls to determine the value) and prevention through volunteers. The document indicates a need for using the right channels for communications (social media for younger people, etc.).

The document highlights four main fields in which performance should be measured and explains and justifies the use of these measures:

- impact to society;
- processes supporting the achievement of the impact;

[2]Eg in 2013, 68% of people died in a fire were using alcohol and 47% of fires with fatalities were started by careless smoking.

- employee development supporting the achievement of the impact;
- finances assisting in achievement of the impact.

The Development Plan for the Area of Government of the Ministry of Interior Affairs 2015–2018 (Siseministeerium 2015a) contains all of the performance indicators measured by the ministry in the following period. Although there are many indicators that are multidisciplinary, a few indicators are mostly covered by the services of ERB.

Although the National Orientations in the Development of the Voluntary Rescue 2013–2016 document focuses only on the volunteer rescuers, it is still important to understand the trends in rescue services as a whole.

The methodology for risk assessment of FRS brigades' service areas (Päästeamet 2012b) was the basis for the last reform which resulted in the closure of nine FRS brigades. The document contained a formula, which standardised and summarised the risks and workload for every FRS brigade. The following criteria were included (the maximum sum of all indicators were 100 points – the work load was 55% and the risk assessment 45%):

- total departures (30%);
- departures for the building fires (25%);
- population in the service area (20%);
- density of traffic in the service area (10%);
- number of hazardous objects in the service area (10%);
- density of railway in the service area (5%).

The value of every indicator was calculated as the percentage of the brigade with the highest value and then summarised (e.g. one brigade has the highest number of total departures (x_1), so it will get 30 points, the following will get $\frac{x_2}{x_1} * 30$).

Although not specifically containing performance indicators, another important framework in which ERB operates in, is the Rescue Act (Riigikogu 2010). Section two of the Act states the fundamentals of activity of rescue service agency: *The aim of the activity of a rescue service agency is to establish and maintain a safe living environment, prevent threats and render prompt and professional assistance.* This is in line with the theoretical approach argued in Subsection 1.2.2. The priority is to prevent emergencies from happening and secondly to alleviate the effects of the rescue events. The planning should take place based on regional hazards. The Act also states that the rescue service can use volunteers and may enter into civil law contracts with legal persons to organise the voluntary work. Furthermore, §5 (1) states the functions of the Rescue Board, of which providing the rescue work on land and inland water bodies; supervising the planning and building process of constructions; and preventing rescue events are of highest importance.

The final document which discusses potentially important performance indicators for the ERB is the FRS brigades' restructuring plan (which was applied in 2012) (Päästeamet 2012a). The main reasons for restructuring were that the FRS brigades' location and staffing was not in accordance with the relocated risks in the regions; the low staffing was unable to provide necessary assistance in case of emergency; the budget is decreasing, yet the potential costs for rescue services are increasing; the safety of the rescuers was not guaranteed. The aim of the restructuring was to ensure rescue services within 15 minutes to as many people as possible, by responding with a team with at least four rescuers.

The justification and suitability of the performance indicators as a basis for performance measurement framework is discussed in Subsection 3.2.4.

3.2.2. National strategies – performance indicators used by Finland

The following subsection analyses the available performance indicators used by Finnish policymakers. The aim of this subsection is to analyse which documents are relevant for the performance measurement of FRS and which of them are eligible directly to the FRS brigades. This analysis would be an input to generate the models to evaluate the efficiency of FRS by frontier analysis methods.

The Finnish FRS system is more decentralised than the Estonian one. The Finnish Ministry of Interior's Department for Rescue Services is responsible for directing, steering and supervising rescue services and maintaining the oversight of their coverage and quality. For regional coverage and quality monitoring, the regional state administrative agencies are used (jointly cooperating with municipalities). The rescue service duties are carried out by 22 rescue departments, under which operate 370 FRS brigades, 523 contracted fire brigades and 105 industrial or institutional fire brigades (Emergency Services College 2015). It has been noted that the FRS system will go through a reform in 2018–2019.

The Finnish Ministry of the Interior published the new strategy in 2016 (Sisäministeriö 2016), emphasising cooperation as the main actor in achieving its goals. Similar to Estonia, the changes in (international) operation environment and society, such as changes in climate, demographic trends, urbanisation, cross-border migration, food security, access to energy, resource scarcity, and changes in the nature of conflicts as well as tougher economic conditions are highlighted. The increasing number of elderly living alone received special attention. The strategy follows objectives set by the Government Programme and Report on Internal Security. The targets set by the Government Programme are:

national rescue authorities will further enhance the national direction, planning, guidance, supervision and coordination of rescue services. The command and coordination of actions by rescue and civilian authorities to combat wide-ranging threats will be clarified. The cost-effectiveness of rescue services and career paths in the sector will be improved by reforming the training system and by further developing contract FRS brigades. The Government Report on Internal Security stated the following objectives: the structures of the rescue services will be reformed and national steering of their operations enhanced. Command capability of civilian authorities and cooperation between authorities in rescue operations and under emergency conditions will be strengthened.

In a sparsely populated country like Finland, the use of contract fire brigades that rely heavily on volunteer activity and part-time contract personnel is necessary for arranging efficient rescue operations. For financial reasons alone, it is impossible to build rescue operations solely upon the use of full-time personnel. Operating under fiscal constraints, the cost-efficiency of operations is stressed, as the saving cannot come from compromises of service quality (standard of service). As an opportunity to achieve this, the centralisation of operations, and harmonisation of services as extensively as possible are mentioned. Prioritisation is another measure to reach efficiency, as a very extensive provision of services is not always possible.

Seven targets are set in the strategy: *'(1) The rescue services have an overall picture of the risks in society, based on continuous analysis. (2) The rescue services are prepared to meet the risks within their own area of operation. (3) The rescue services are a strong coordinator of civil emergency preparedness and a trusted partner. (4) Services have been arranged in a high standard and in a cost-effective and consistent manner. (5) We are all aware of, and take responsibility for our own and our community's safety and the safety of surrounding areas. (6) The rescue services actively develop their practices and procedures. (7) Personnel's well-being is a priority'* (Sisäministeriö 2016: 12).

In conclusion, the performance indicators for Finnish rescue services are similar to Estonian ones, but more diffuse. No concrete targets are set in strategy, which might be due to decentralised management – every rescue department sets its own targets with respect to general strategy.

3.2.3. National strategies – performance indicators used by Sweden

This subsection considers the system of Swedish FRS. In comparison to Estonia and Finland, the Swedish system is even more decentralised, which also means that the possible performance indicators are more general in the national context and vary across different service providers. The FRS in Sweden is a responsibility of, and therefore organised by, the municipalities. There are 290 municipalities in Sweden, but many municipalities have formed (and it is an on-going process) a formal cooperation with their neighbours to provide FRS. Thus, in 2015, there were about 165 different FRS authorities. The most important act is the Civil Protection Act (LSO).

The municipalities are free to plan and organise the FRS, as long as they meet the obligations stated by the law (LSO, Chapter 3). Each municipality must decide on the 'action plan for civil protection', which describes how the FRS are carried out. Typically, the service provision is divided into two: the operative response and the prevention. The plan should specify the target for municipal operations and the risks of accidents in the municipality that could lead to the need of emergency response. The programme also describes how the municipality's preventive activity is planned and organised.

In general, the larger municipalities have full-time firefighters or a mix of full-time and part-time firefighters. The proportion of part-time firefighters is higher in smaller municipalities (Holmgren and Weinholt 2016). Full-time firefighters have gone through longer training and are more experienced than part-time firefighters. The biggest difference between these two approaches is the response time – the full-time firefighters are on average 5 minutes faster, as the part-time firefighters are either at work or home at the time when the alarm is received (Mattsson and Juås 1997, cited in Holmgren and Weinholt 2016). The turnouts across municipalities have typically followed a similar standard: four firefighters and a unit leader are sent to the emergency scene, regardless of the type of accident. Recently, this procedure has become more flexible.

According to Holmgren and Weinholt (2016), the fire managers in Sweden are struggling with obtaining the necessary resources to operate, so the efficiency gains are highly valued. Swedish FRS provision in the last decade has seen some shifts in organisational structure, as concluded by Holmgren and Weinholt (2016):

- Introduction of formalised cooperation between neighbouring municipalities. Due to the rising service costs, some municipalities have merged the provision of FRS. Most such co-operations have been created in the larger cities and areas with higher population density. The main aim for that is the gains in cost-efficiency through economies of scale. As a challenge, extra transaction costs should be considered, to see whether they are lower than the efficiency gains.

- Cooperation with third parties. In order to reduce the response time, a small fraction of municipalities is collaborating with other actors who would help to respond to the emergencies. For instance, the collaboration has started with home care personnel and security officers.

- Changes in the structure of rescue teams. To reduce the response time, using smaller units for first response has been introduced. Some municipalities with part-time firefighters use the concept of the First Response Person –

the unit leader, instead of first going to the station, heads straight to the emergency scene to break or mitigate the development of the situation at an early stage. The effects of such concept have been evaluated and it was concluded that it saved valuable response time and was profitable from the society's perspective (Lång 2012, cited in Holmgren and Weinholt 2016).

- Increase in the number of tasks performed by FRS. The collaboration with ambulance services has increased, as the FRS units also respond to some emergencies intended for ambulance services. The FRS provides basic rescue skills while waiting for the ambulance to arrive. In addition, studies have shown that such collaboration has saved several lives (Sund *et al.* 2012, cited in Holmgren and Weinholt 2016).

The FRS authorities are overlooked by the Swedish Civil Contingencies Agency (Myndigheten för samhällsskydd och beredskap – MSB), which is a national government agency responsible for the legislation in the field. According to the MSB website, the organisation *'covers issues concerning civil protection, public safety, emergency management and civil defence as long as no other authority has this responsibility. Responsibility refers to measures taken before, during and after an emergency or crisis. The concept of MSB is to work with knowledge-building, support, training, exercises, regulations, supervision and operational work in close collaboration with municipalities, county councils, other agencies, and the private sector and organisations.* [Their] *goal is to achieve a safe and functioning society across all societal levels, from the local to the global'* (MSB 2017). The MSB trains and educates the FRS personnel and in cooperation with the County Administrative Boards and other stakeholders, collects information and carries out supervision over the municipalities' specific responsibilities stated in the LSO. The MSB itself is steered by the Swedish Government by annual appropriation and via a body of instructions, which specify the objectives, reporting requirements and allocated resources.

The MSB provides many guidelines how to organise the FRS in municipalities and does research to ground the insights. Although the Swedish system builds their preparedness from below, they also have some elements of planning some activities 'from above'. One example is the national risk assessment (MSB 2016), which covers the risks in society in a holistic manner and considers threats and risks that would have a nation-wide impact. This links the prevention, preparation, handling and follow-up elements and combines the bottom-up approach with a top-down approach, thus filling potential gaps in the system.

In addition to the FRS provided by municipalities and MSB, the Swedish Fire Protection Association is of importance. They mainly undertake research, provide information and educate/train the public. This is a voluntary organisation which is not funded by the government nor municipalities, so they have no judicial mandate or obligations.

From performance measurement perspective, the LSO (Justitiedepartementet L4 2003) states the national objectives for FRS. Life, health, property and the environment – with consideration given to local conditions are of importance and the provision of equivalent, satisfactory and comprehensive civil protection. It also noted that the rescue services should be provided within an acceptable time and using effective methods, without being more specific on these. The rescue services should prevent or limit the injuries to individuals, damage to property and the environment in case of emergencies or imminent danger of emergencies. Rescue services should provide a rapid response, while considering the importance of that which is threatened, the cost of operations and other circumstances. The LSO also marks (Chapter 3) that it is the municipality's responsibility to prevent fires and damages due to fires and, without infringing on other agencies' responsibilities, work to provide protection against accidents other than fires, and provide the rescue and subsequent services. No quantifiable performance measures are given.

To conclude, the performance indicators in the Swedish case are not defined at the national level, although the MSB provides many guidelines and step-by-step guides on how to plan FRS at a municipality level. The municipalities are autonomous to organise and cooperate with neighbouring municipalities in the service provision. The only obligation is to provide rescue services equivalently, satisfactorily and comprehensively within an acceptable time limit using effective methods.

3.2.4. Comparison of national approaches

This subsection will analyse the similarities and differences of FRS systems in Estonia, Finland and Sweden, keeping in mind what would affect the following performance assessment. Estonia has centralised the FRS provision; thus, it values a standardised approach, achieving a certain service level (for some reasoning, see Puolokainen 2017). Finland and Sweden provide the services on rescue departments' and municipalities' level, respectively, so it is a decentralised approach, giving more freedom and flexibility to adjust to regional peculiarities. As Finland and Sweden also have a bigger population and service area than Estonia, the decentralised approach might be influenced by that.

The goal of FRS provision is similar in all discussed countries and is specified in rescue acts (or similar laws) – to reduce the threat on human lives, assets, environment, etc., through preventing and responding to emergencies as quickly as possible.

Estonia and Finland have general strategies they want to achieve in a medium-term perspective. Estonia has set quite many quantifiable targets, Finland has clarified only priorities and directions of changes. The MSB in Sweden provides guidelines how to plan and organise the FRS but no clear targets are set at a national level, leaving more autonomy to municipalities to decide and plan out the service levels.

All discussed countries are cooperating with different entities. There is a tendency to use more help from volunteers, which is more popular in Sweden (part-time firefighters) and Finland (contracted fire brigades), but also increasing in Estonia. In Estonia, the voluntary rescuers have been acknowledged and their involvement prioritised since the reform in 2012. Many countries have merged the FRS with the ambulance service, but in the current cases, the ambulance service is provided by a different entity, although there are some common features, as the FRS provide the first aid in case the ambulance lags to the emergency.

The FRS provide many different services, of which fire extinguishing is only one. The FRS in Estonia, Finland and Sweden also provide help during road accidents, drownings and chemical accidents/spills, to name but a few. This complicates the performance analysis, as various types of emergencies require different skills and equipment and thus affects the number of inputs and resources as well as potential outputs (e.g. quality of service).

Estonia, Finland and Sweden are providing insights in the changing environment where they are planning the services. Estonia focuses on it in the ERB's strategy, Finland in the Ministry of Interior's strategy and Sweden through the materials provided by the MSB and a risk assessment at municipality level.

To conclude, Estonia has most performance indicators with set targets, due to the centralised system. Finland would be somewhere in the middle, as it gives clear directions on what to achieve at a national level through strategic documents, but the 22 fire departments have the autonomy to plan the provision of their services on their own. Sweden does not have any quantifiable indicators at national level, as it has the most decentralised system – the FRS are mostly provided at municipality level with full autonomy in planning and organising the services, including cooperating with their neighbours to develop a unified service provision.

3.2.5. Justification for the choice of indicators

Based on the literature and policy materials covered in the last subsection, the following subsection discusses and justifies the selection of performance indicators for FRS and its subunits. Structuring follows the main stages of the service provision, introduced in the first chapter.

Costs and other inputs: As was pointed out in many documents, the budget for FRS is decreasing (the financial constraint was covered by all countries under analysis), yet the costs for rescue services (per emergency, per capita) are increasing. Although it was said that the current network of FRS brigades is rather optimal in the Estonian case, the ERB strategy states other priorities. In the ERB strategy it is stated that: financing of volunteers will increase, the proportion of investments in budget will increase and the salary will increase. This has to lead to a decrease in the number of FRS brigades or a decrease in the workforce of the headquarters. As around 80% of the running costs of a FRS already goes on salaries, the potential decrease of maintenance costs as a result of investments is marginal. The need for fiscal austerity in the light of the last economic crisis in Estonia (and other Baltic states) has also been affirmed by Masso and Espenberg (2013). The costs of providing FRS in Estonia has been analysed previously by Grünvald (2016), who concluded that continuing in the same way would not be sustainable in the long run, meaning that the service provision should become more cost-efficient, alternated thoroughly or receive an increase in financing. Similar tendencies are widespread elsewhere as well (for the Swedish case, see Holmgren and Weinholt 2016). Therefore, it is essential to include running costs to the models, as cost-efficiency is becoming one of the most important subjects.

As other inputs, the size and age of the fleet and other gear should be discussed. The main problem is to develop a standardised index to aggregate and take the inputs into account fairly. There is no such index readily available.

Processes: One of the indicators that describes the processes/activities quality is the average time of the first responder to arrive to the emergency site (all emergencies) or responder with lifesaving capability (dwelling fires, traffic accidents, water accidents). There are different opportunities to include the average time in the models (aggregated average time or detailed time for the first responder/responder with lifesaving capability and with different emergency event types). Since the average time might not be the best indicator, different quantiles (e.g. median time or 90% line) should be discussed. Schaenman and Swartz (1974) emphasised that as a productivity measure, one should use the faster response time in less loss, *ceteris paribus*, and that the response of FRS brigades ultimately reflects the responsiveness of the government in the eyes of citizens, while indicating their own security. The response time is affected by the time of the day when the fire starts, when it is detected, how efficiently the rescuers prepare for it and head to the fire, traffic situation and weather.

As it is important to have at least four rescuers responding to the emergency (although the standard varies from country to country, four rescuers is a bare minimum to effectively put out a fire), the number of on-call staff should be taken into account. Yet another goal in the future is to involve the on-call staff in the prevention activities – to raise the efficiency, the number of hours spent on prevention per FRS brigade, should be discussed.

For quality as an employer, the voluntary turnover (%) and the commitment of staff (index) can be taken into account, as pointed out in the Strategy of the ERB (Päästeamet 2014).

To include the volunteers in the Estonian case, there are two major changes in near future – volunteers should take part of the trainings by national FRS brigades and also contribute as on-call staff. These hours can also be tracked and analysed. In the case of Finland and Sweden, such data on volunteers is already collected and can be taken into account.

Environment: The most important factor for a FRS brigade is to be near people – therefore as an environmental factor, the population of service area is crucial. The documents highlight that the criterion to arrive to the scene in Estonia is within 15 minutes. Closely related to the population of the area is the size of the service area as well as the density.

Other important environmental factors which describe the service area are the number of hazardous objects (schools, hospitals, industry, etc.), the density of traffic and density of railways. All those factors should be correlated with the size of the population in the service area.

Outputs: Typically different sort of attendances to emergencies are considered (emergencies themselves can be considered as 'bad' outputs) as outputs. The statistics here are very detailed, so in the models all sorts of emergency types (all rescue events [including/excluding misreports], fires, building fires, residential fires, different causes of fires, traffic accidents, drownings, environmental accidents, etc.) can be considered in detail as well as in an aggregated way. Choosing the optimal level of aggregation is the most difficult task to justify. Schaenman (1977) has suggested that larger, more serious fires should be counted separately from smaller, no-casualty fires and that the actual fire rates should be compared to the expected fire rates, while taking into account the local community's characteristics. Schaenman and Swartz (1974) caution on using fire rates by population size, which might be misleading in cases where the day and night populations differ to a great extent.

Closely related to the number of emergencies is the number of fatalities, injuries and saved people. One of the goals is to reduce the property damage caused by building fires.

Outcomes: Outcomes of the FRS are the most difficult task to measure as a good accident is an accident that never happened. As a proxy of outcomes, one can consider the feeling of safety among population and managers of firms (which is a non-transactional good, thus complicating the measurement). In the Estonian case, there are different possibilities to approximate it – trustworthiness of the ERB (as a survey), also the safety-related awareness of the population (fire and water safety awareness index), which evaluates the impact of prevention actions. Unfortunately, these indices are collected only at the county level and cannot be strictly related to the service area of a FRS brigade. Nonetheless, one can use it in an aggregated level (all of the FRS brigades in one county would get the same rating).

One could also use the willingness to pay approach in determining the value for FRS. Such approach has not been used in the FRS case in the considered countries.

Another proxy to discuss is the number of volunteers in the service area – it can be argued that the number of volunteers is high because of the high status of rescue services, or quite the opposite – people are frustrated by the lack of safety and form their own units to increase the safety in their neighbourhood.

Table 3.1 concludes the discussed performance indicators and provides an indication of whether such data would be available for analysis at the level of detail proposed by the thesis.

Table 3.1: Performance indicators for FRS discussed in the literature and strategic documents

Type of measurement	Possible performance indicator	Data availability		
		Estonia	Finland	Sweden
Costs and inputs	Total costs	x	x	x
	Salaries	x	x	x
	Maintenance costs	x	x	x
	Number of employees	x	x	x
	Size and age of the fleet, other equipment	x	(x)	
Processes	Average time of the first responder to arrive to the emergency	x	x	x
	Average time of the first responder with lifesaving capabilities to arrive to the emergency	x		
	Number of on-call staff	x		
	Number of hours spent on prevention activities	(x)		
	Voluntary turnover and commitment of staff			
	Number of hours of volunteers taking part in the on-call			
Environment	Population of the service area	x	x	x
	Percentage of people reached within x (e.g. 15) minutes	x	x	
	Size of the service area	x	x	x
	Population density of the service area	x	x	x
Outputs	Number of attended emergencies	x	x	x
	... fires	x	x	x
	... fires in buildings	x	x	x
	... other fires	x	x	x
	... traffic accidents	x	x	x
	... drownings	x	x	x
	... other emergencies	x	x	x

Table 3.1 Continued:

Type of measurement	Possible performance indicator	Data availability		
		Estonia	Finland	Sweden
	Number of fatalities	x	x	x
	... fire fatalities	x	x	x
	... injuries	x	x	x
	... saved	x		
	Property damage caused by emergencies	(x)		
Outcomes	Feeling of safety			
	... trustworthiness of the public agency	(x)		
	... populations' awareness of (fire) safety	(x)		
	Number of volunteers	x	x	x

Note: Brackets denotes data with limited availability.
Source: Author's compilation.

As one can see, much (easily) quantifiable data is collected. This applies mainly to different costs, numbers of departures to emergency events, some environmental background statistics, and numbers of volunteers. However, this does not mean that the information would be directly comparable between countries (one can assume that the data is comparable within one country in different years). There are some issues to consider. First, the data is collected at different levels of management, e.g. at the FRS brigade level in Estonia, fire department level (that has multiple FRS brigades under it) in Finland and municipal level in Sweden (which cooperate with each other and also have multiple FRS brigades under it). This affects, for example, the number of departures. In the Estonian case, when multiple FRS brigades respond to the same event, it is counted by as many FRS brigades responded. In the Finnish and Swedish cases, however, the events would be counted as one, as though FRS brigades from one entity responded. Second, one should consider the administration costs for services. For the Estonian case, the costs cover expenses made by the FRS brigades (and thus top-level management costs, including expenses of many prevention activities, are left out). As fire departments in Finland and municipalities in Sweden have higher autonomy in decision-making, these management costs would also be reflected there. Third, there are different standards for operating the FRS and different standards for collecting and classifying the data. For example, in the Finnish case, traffic accidents as a type of emergency are often attended by the FRS, and thus, the figure seems high in comparison to other countries. This, however, does not mean that the traffic in Finland is worse, but it means that the likelihood of FRS unit to respond to every traffic accident is higher.

The current subsection combined and synthesised: 1) theoretical frameworks in the field of FRS; 2) performance indicators of FRS used in the literature; 3) performance indicators used by policymakers, including suggestions from researchers to improve them. Finally, the proposed performance indicators were discussed.

3.3. Comparative descriptive statistics of Estonian, Finnish and Swedish fire and rescue services

3.3.1. Introduction

The current section analyses the data on which the following performance evaluation is undertaken. In addition to a general introduction, an analysis of inputs, outputs and environmental factors is given to compare the Estonian, Finnish and Swedish FRS case.

Data required to empirically assess the ERB was collected and merged in cooperation with many ERB officials. The created database consists of different expenditures, brigades' network statistics (how many people are reached within every subsequent minute), staffing levels, emergency cases, rescued people, fatalities, etc., at a FRS brigade level. The data is available for 2011-2015. In total, 65 FRS brigades have been included, totalling 325 observations (5x65). For location information about the FRS brigades in Estonia, see Estonian Rescue Board (2017).

As the focus of the thesis is to evaluate the subunits which are directly under the control of ERB, only the national FRS brigades are analysed. For a more comprehensive approach, the voluntary FRS brigades should be included, as the importance of volunteers is rising and more functions and capabilities are given to them. The promotion of the voluntary rescue service started from 2013. Before then, not much effort was given to quantifying their contribution in activities (in 2015, the voluntary FRS participated in around 18% of all emergencies, which is a large increase compared to previous periods, e.g. more than triple in comparison to 2011, the first year of the analysis in this thesis). For now, one of the setbacks of including the volunteers in the evaluation is the lack of (statistics about) their activities in a longer time period. The systematic and good-quality data gathering started from 2013–2015 depending on the variable, and is improving since.

A database required to analyse the Finnish FRS system was acquired through PRONTO, which is the statistics system of Finnish rescue services and managed by the Finnish Emergency Service College in Kuopio. The PRONTO database covers the accident reports, information about costs in fire departments and numbers of prevention activities. Most of the data is available at the fire department level, which is the basis for the current analysis. The data is available for 2004–2015. Twelve periods times 22 fire departments equals a total of 264 observations. For location information about the fire departments in Finland, see Pelastustoimi (2017).

The data for Swedish FRS was collected from IDA, which is a national database managed by MSB. Additionally, extra information was collected with the help of officials from MSB and Statistics Sweden. Most of the data is available at the municipality level, which is chosen as the basis for further analysis. The data is available from 2005–2015 but it is unbalanced. Thus, for the 11 periods, and a maximum of 282 municipalities, the database includes 2,962 observations for all municipalities and 1,629 observations for the case when only municipalities with professional rescuers are considered. For location information about the FRS brigades in Sweden, see MSB (2017).

3.3.2. Notation

For the analysis, the number of departures to fires in buildings, number of other fires, traffic accidents and other emergencies are considered. In the following, the notation of the used variables is used:
q_{1it} = fires in buildings
q_{2it} = other fires
q_{3it} = traffic accidents
q_{4it} = other emergencies

x_{1it} = labour (number of employees)
x_{2it} = other inputs (assumed proportional to number of vehicles (Estonia and Finland), FRS brigades in service area (Sweden))

z_{it} = 1/area
$d_{nit} = q_{nit}$ (i.e. all demands for service were met)
m_{nit} = the value such that $Pr(d_{nit} \geq m_{nit}) = 0.05$

c_{1it} = population reached within 15 minutes (target level) (Estonia), population reached within 20 minutes (target level) (Finland), population in service area (Sweden)
c_{2it} = average time of arriving to the emergency scene after receiving the message (faster is better)

To illustrate such notation on the formerly proposed input-output model, Figure 3.1 builds on the model from Figure 1.10.

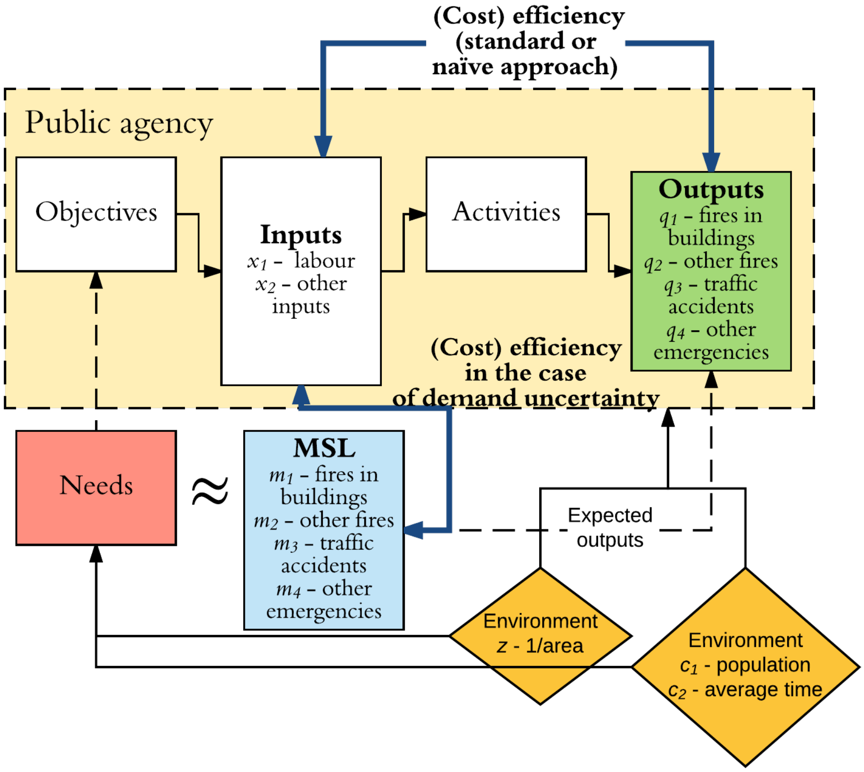


Figure 3.1: The proposed model to measure the performance of FRS using MSL as an indicator for demand uncertainty (Source: Author's contribution)

3.3.3. Outputs

To understand the framework in which the FRS (brigades) operate, one should first analyse the number of emergency departures as a whole and by major response types in FRS brigades, to identify potential differences in the response to the emergencies as pointed out by Flynn (2009).

To compare different countries that have been analysed, the number of emergency departures per 100,000 population are illustrated in Figure 3.2. Estonia has the most emergency departures per capita but the trend from 2010 has shown a decrease. Yet, the number of emergency departures per capita is still around 2.5 times higher than in Sweden and around 30% higher than in Finland. Sweden has shown a very stable number of emergency departures with respect to its population, staying around 1,000 emergency departures a year per 100,000 people. It must be noted, however, that these numbers are not directly comparable due to the differences in the structure of the fire authorities. In the Estonian case, the departures are based on FRS brigades, which are smaller units than municipalities (in Sweden) and fire departments (in Finland) – so, one would count it as a different departure, when multiple brigades respond to an emergency, which is more likely when a subunit is smaller (one FRS brigade versus one department with many FRS brigades for example).

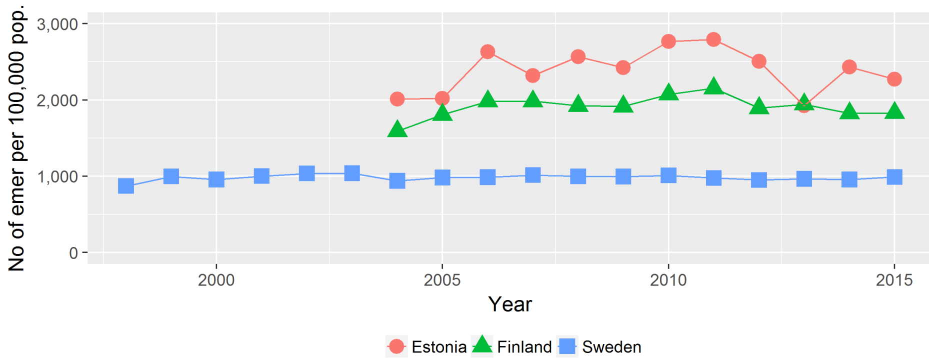


Figure 3.2: Number of emergency departures per 100,000 population in Estonia, Finland and Sweden (Source: Estonian Rescue Board 2016; PRONTO 2016; IDA 2016; Author's calculations)

Since 2003, the total number of attended emergencies in Estonia increased steadily until 2011, from where on the trend has been downwards. The total number of departures to emergencies in a year has fluctuated between 25,000 and 37,000 and it varies to an extent in random (which can be attributed for example to the weather as well as uncertainties in the environment). Some of the decrease in the number of emergencies can be explained by the greater focus on emergency safety (media campaigns, home visits to map the fire risks of the households, compulsory chimney sweeping, installation of smoke detectors for early detection, etc.). This includes only the emergency departures that have been attended by a national FRS brigade. Since the ERB functions as four rescue centres, the division of departures across centres were analysed. The trends are similar to Estonia as a whole, so one cannot differentiate the progress by differences in regions, but one could argue that the peak of departures has been reached in the early 2010s and further on the number has declined, most rapidly in the North rescue centre. For a more specific distribution of departures to emergency situations between different FRS brigades, see Appendix 5. Appendix 6 highlights the variation in the number of emergencies attended by a FRS brigade. The variation of departures is higher in the large-sized FRS brigades.

The FRS brigades attend many different rescue situations. The ERB publishes annual statistics and analyses on the emergencies, fires, building fires, drowning victims, and fire casualties. According to the ERB yearbook (ERB 2016), the most common rescue situations in Estonia in 2015 were fires, mistaken calls, automated false alarms, provision of help, natural phenomena events and traffic accidents. The mistaken calls and automated false alarms accounted for 27% of all events. The most typical causes of building fires were negligence with the use of open fire and carelessness while smoking. In addition, malfunctions in electrical installations and heating devices caused many building fires.

In addition to official emergency statistics, there have been a few studies covering the analysis of fire deaths, such as Randoja and Käerdi (2010; 2011), and special topics such as distant rural areas (Espenberg *et al.* 2013). In addition, a few masters' theses have been conducted on the topic as well as coverage by trade press.

The number of emergencies in Finland varies between 82,000 and 115,000 annually. The trend is similar to Estonia, as the number of emergencies increased until the beginning of 2010s and decreased from there on. That includes emergencies attended by any FRS brigade, including voluntary, part-time, industrial, etc. brigades. Since 2004, the total number of attended emergencies in Sweden increased steadily until 2010, from where on the number of emergencies has stabilised. The total number of departures to emergencies in a year has been between 84,000 and 97,000. This includes all emergencies.

The next tables compare the absolute (and per 100,000 population) numbers of emergency departures[3] to different types of emergencies in each country as well as the share of different types of emergencies attended. The share of fires in buildings, other fires, traffic accidents and other emergencies are considered. From here on, there will be no distinction between the number of emergencies and emergency departures.

Estonia has the highest emergency departures per 100,000 population for fires in buildings (see Table 3.2). The trend is decreasing in all considered countries but it varies to a greater extent in Estonia. In addition, in Estonia, the building fires cover the highest share of all emergency departures in comparison, being around one-third of all departures in contrast to less than one-tenth in Sweden and one-fiftieth in Finland. The attendance to fires in buildings might be high due to the following reasons: (1) as building fires are a priority, the FRS brigades respond with bigger forces (multiple FRS brigades to a single fire); (2) the general development level of the society – people are unaware of the potential risks and have a risky behaviour (e.g. smoking indoors); (3) the general economic development – the quality of building materials is lower as well as the

[3]In Estonian case, there is the number of emergency departures available for each FRS brigade and this has been used for the analysis; for Finland and Sweden, there is the number of emergencies in the respective service area. One can argue that FRS brigades as smallest subunits cooperate more (can attend emergencies in neighbouring service areas) and municipalities in Sweden and fire departments in Finland are bigger subunits, for which the cross-service area cooperation happens more seldomly.

requirements for fire safety in buildings are softer. The first reason corresponds to the differences in the management structure and the last two to the fact that Estonia encounters a higher number of building fires.

The number of departures to other fires is more comparable in the countries than the building fires (see Table 3.3). In the last few analysed years Estonia has encountered around three times more emergency departures to other fires than Finland or Sweden. The reasons are probably similar to the previous case. One should note that the share of departures to other fires is decreasing in the Estonian case and more stable in the other two countries.

In contrast, Estonian FRS brigades depart least (as per 100,000 population) to the traffic accidents compared to Finland and Sweden (see Table 3.4). The lower number in comparison might be due to the fact that in Sweden, and especially in Finland the standards on how to respond to traffic accidents differ, meaning that the FRS would depart to every traffic accident (so the threshold for response is lower). In addition, as a share of all emergency departures, Estonia has about one-third of a share of the other countries. This might be due to different standards – the FRS brigades depart only when they are considered needed and not for every event. The share of departures to traffic accidents has steadily increased in Estonia. An upwards trend is also apparent for Finland and Sweden.

As for other emergency departures, the share varies to a greater extent (see Table 3.5). This is due to the reason of different responsibilities and standards of each country. It is highest in the Finnish case.

Table 3.2: The number of departures to fires in buildings, number of departures per 100,000 population and share of all emergency departures in Estonia, Finland and Sweden, 2004–2015

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Estonia												
Number	8,987	9,560						6,601	12,568	8,707	9,292	7,183
Number per capita	657.8	703.5						496.4	948.4	659.5	706.2	547.0
% of emergencies	32.7	34.9						17.8	37.9	34.2	29.0	24.1
Finland												
Number	3,313	3,671	3,998	4,025	4,485	2,736	2,786	2,543	2,413	2,341	2,291	2,008
Number per capita	63.7	70.5	76.1	76.6	85.0	51.6	52.3	47.6	44.9	43.4	42.2	36.9
% of emergencies	4.0	3.9	3.8	3.9	4.4	2.7	2.5	2.2	2.4	2.2	2.3	2.0
Sweden												
Number	9,684	10,321	10,500	11,110	10,886	11,078	11,246	10,265	10,419	10,201	9,665	9,497
Number per capita	107.5	114.1	115.2	121.0	117.6	118.6	119.4	108.2	109.0	105.8	99.2	96.4
% of emergencies	11.5	11.6	11.7	11.9	11.8	11.9	11.8	11.1	11.5	11.0	10.4	9.7

Source: Estonian Rescue Board 2016; PRONTO 2016; IDA 2016; Author's calculations.

Table 3.3: The number of departures to other fires, number of departures per 100,000 population and share of all emergency departures in Estonia, Finland and Sweden, 2004–2015

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Estonia												
Number	9,853	8,109						13,343	5,192	5,068	7,598	5,437
Number per capita	721.2	596.8						1003.5	391.8	383.9	577.4	414.0
% of emergencies	35.8	29.6						35.9	15.6	19.9	23.7	18.2
Finland												
Number	8,400	9,306	13,582	10,131	10,159	12,258	12,206	12,194	9,390	11,080	11,734	9,212
Number per capita	161.6	178.6	258.7	192.9	192.5	231.3	229.3	228.0	174.8	205.3	216.3	169.1
% of emergencies	10.2	9.9	13.1	9.7	10.0	12.1	11.1	10.6	9.2	10.6	11.8	9.2
Sweden												
Number	14,936	15,727	16,604	16,092	17,804	16,379	13,311	14,544	12,246	15,209	14,598	13,288
Number per capita	165.7	173.8	182.2	175.2	192.3	175.4	141.4	153.4	128.2	157.7	149.8	134.9
% of emergencies	17.7	17.7	18.5	17.3	19.3	17.6	14.0	15.7	13.5	16.4	15.7	13.6

Source: Estonian Rescue Board 2016; PRONTO 2016; IDA 2016; Author's calculations.

Table 3.4: The number of departures to traffic accidents, number of departures per 100,000 population and share of all emergency departures in Estonia, Finland and Sweden, 2004–2015

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Estonia												
Number	598	672	963	1,058	932	1,084	1,467	1,675	1,599	1,161	1,576	1,753
Number per capita	43.8	49.5	71.3	78.8	69.6	81.2	110.0	126.0	120.7	87.9	119.8	133.5
% of emergencies	2.2	2.5	2.7	3.4	2.7	3.4	4.0	4.5	4.8	4.6	4.9	5.9
Finland												
Number	8,391	10,205	10,912	12,874	13,401	13,349	13,472	13,630	14,163	13,140	13,724	14,107
Number per capita	161.4	195.9	207.8	245.1	253.9	251.9	253.1	254.9	263.6	243.4	253.0	259.0
% of emergencies	10.1	10.8	10.5	12.4	13.2	13.2	12.2	11.8	13.9	12.5	13.9	14.2
Sweden												
Number	12,370	12,723	12,778	13,905	14,394	15,856	16,211	14,380	16,904	15,745	16,305	18,327
Number per capita	137.3	140.6	140.2	151.4	155.5	169.8	172.2	151.6	176.9	163.2	167.3	186.0
% of emergencies	14.6	14.3	14.2	14.9	15.6	17.1	17.1	15.5	18.6	16.9	17.5	18.8

Source: Estonian Rescue Board 2016; PRONTO 2016; IDA 2016; Author's calculations.

Table 3.5: The number of departures to other emergencies, number of departures per 100,000 population and share of all emergency departures in Estonia, Finland and Sweden, 2004–2015

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Estonia												
Number	8,056	9,083	19,872	9,701	13,723	14,492	15,600	15,532	13,821	10,489	13,555	15,463
Number per capita	589.6	668.4	1471.2	722.4	1025.3	1084.9	1170.0	1168.1	1042.9	794.5	1030.2	1177.4
% of emergencies	29.3	33.1	55.9	31.2	40.0	44.8	42.3	41.8	41.7	41.3	42.3	51.8
Finland												
Number	62,590	70,917	75,571	77,115	73,440	73,096	81,927	86,798	75,696	78,282	71,319	74,334
Number per capita	1204.3	1361.1	1439.2	1468.3	1391.4	1379.5	1538.9	1623.2	1408.8	1450.2	1314.7	1364.8
% of emergencies	75.7	75.4	72.6	74.0	72.4	72.1	74.2	75.4	74.5	74.7	72.0	74.6
Sweden												
Number	47,558	50,118	49,927	52,041	49,284	49,673	54,240	53,428	51,235	51,838	52,611	56,367
Number per capita	527.8	553.9	547.9	566.7	532.4	531.8	576.1	563.4	536.2	537.5	539.7	572.2
% of emergencies	56.2	56.4	55.6	55.9	53.4	53.4	57.1	57.7	56.4	55.7	56.5	57.8

Source: Estonian Rescue Board 2016; PRONTO 2016; IDA 2016; Author's calculations.

In Estonia, the proportions of attendance to different main emergency situations do not vary much across the years or FRS brigades (see Appendix 7). In 2015, the share (and number) of fires has declined, while share (and number) of other emergencies attended consequently increased. For other emergencies, the number of emergencies on a water body had a large increase in 2010, when the count almost tripled (from 180 emergency events in 2009 to 489 in 2010) and after that has remained steady at a higher level. The number of chemical accidents (average around 75 events per year) and oil spills (average around 550 per year) have been fairly stable and no trends prevail. The number of gas leakages has increased remarkably in the last few years (98 in 2009 and peaked with 1,040 in 2014). Emergencies related to animals is showing a decreasing trend (peaked in 2010 with 1,635 emergencies, 1,081 in 2015).

For Finland, Appendix 8 highlights the emergency attendances across fire departments and Appendix 9 the variation in the number of emergencies attended by fire departments. As the fire departments are larger subunits than the FRS brigades, the variation of emergencies is lower. In addition, the differences between largest and smallest fire departments are not as large as between the FRS brigades in the Estonian case (less than ten-fold). The proportions of attendance for main different emergency situations do not vary much across the years nor fire departments (see Appendix 10). In comparison to Estonia, Finnish fire departments deal more with other emergencies, as the share of fires in all emergency calls is low.

The proportions of attendance for main different emergency situations do not vary much across the years in Sweden. The shares of different emergencies are more similar to the Estonian case, e.g. the other emergencies account for about half of departures. In different municipalities, the shares vary to a greater extent than in other countries cases (see Appendix 11).

For the multiple outputs case, the number of different attended emergencies are considered (see Tables 3.6, 3.7, 3.8 and 3.9 for descriptive statistics). The differences between the attended emergencies vary to a large extent as the smallest FRS subunits have to attend only a few emergencies in a year while the largest ones' deal with thousands. The differences between the attended emergencies vary less in the Finnish fire departments as in Estonian FRS brigades. It should be noted that fire departments respond to fewer fires (in buildings or other) than some biggest Estonian FRS brigades.

For the current case, the FRS subunits are expected to be able to respond to each emergency (such that $d_{nit} = q_{nit}$) and thus meeting the demand. The minimum service levels (MSL – see Chapter 1.4 for explanation) are predicted as a 95% quantile of the respective emergency type based on each FRS subunits' data. These predicted numbers of emergencies by type are also highlighted in the tables. In most cases, the actually occurred emergency attendances are lower than the predicted MSL. Situations where MSL has been estimated lower than the actually occurring number of emergencies, are also noted. Such number is higher in cases where the number of emergencies vary to a greater extent across years. This happens more often in smaller-sized subunits – FRS brigades (Estonia). Statistically, the actually occurred number of emergencies is on average around 70% of the expected MSL.

Table 3.6: The descriptive statistics of outputs (departures of the FRS subunit to fires in buildings) in analysed countries

Statistic	Mean	St. Dev.	Min	Max
Estonia (n=325)				
Population	19,816.0	27,491.4	2,562	167,246
Number	134.1	175.4	4	1,282
MSL	184.0	226.0	27.6	1,222.4
Number/per capita	908.7	653.7	81.0	5,691.6
Number/MSL	0.7	0.2	0.1	1.1
MSL< number (n=63)				
Finland (n=264)				
Population	241,668.6	138,081.6	79,280	622,240
Number	138.7	73.8	33	477
MSL	200.8	94.0	75. 0	469.3
Number/per capita	62.9	22.2	18.8	127.0
Number/MSL	0.7	0.2	0.3	1.2
MSL< number (n=22)				
Sweden (n=2,962)				
Population	32,549.1	62,205.6	2,516	923,516
Number	36.9	57.7	1	688
MSL	46.8	63. 4	7.0	678.2
Number/per capita	137.7	63.9	9.0	573.7
Number/MSL	0.7	0.2	0.1	1.2
MSL< number (n=254)				

Source: Estonian Rescue Board 2016; PRONTO 2016; IDA 2016; Author's calculations.

3.3.4. Inputs

As inputs, the number of employees and the number of vehicles or FRS brigades in a municipality (as approximation of other inputs) are used (see Table 3.10).

Table 3.7: The descriptive statistics of outputs (departures of the FRS subunit to other fires) in analysed countries

Statistic	Mean	St. Dev.	Min	Max
Estonia (n=325)				
Number	109.7	154.1	9	1,152
MSL	184.0	227.2	27.8	1,023
Number/per capita	689.4	538.1	15.3	5,001.5
Number/MSL	0.6	0.3	0.04	1.2
MSL$<$ number (n=64)				
Finland (n=264)				
Number	491.1	236.1	140	1,193
MSL	611.1	258.6	235.2	1,109.4
Number/per capita	215.4	48.6	114.7	373.1
Number/MSL	0.8	0.1	0.5	1.2
MSL$<$ number (n=22)				
Sweden (n=2,962)				
Other fires	52.9	110.7	0	1,624
MSL of other fires	70.4	132.9	7.2	1,506.5
Number/per capita	166.1	72.0	0.0	628.3
Number/MSL	0.7	0.2	0.0	1.4
MSL$<$ number (n=264)				

Source: Estonian Rescue Board 2016; PRONTO 2016; IDA 2016; Author's calculations.

Table 3.8: The descriptive statistics of outputs (departures of the FRS subunit to traffic accidents) in analysed countries

Statistic	Mean	St. Dev.	Min	Max
Estonia (n=325)				
Number	23.2	23.6	1	176
MSL	30.5	27.2	5.0	172
Number/per capita	172.7	101.2	10.2	644.5
Number/MSL	0.7	0.2	0.1	1.2
MSL$<$ number (n=63)				
Finland (n=264)				
Number	573.4	276.1	157	1,574
MSL	697.8	322.0	291.9	1,440.9
Number/per capita	256.8	64.0	51.5	446.6
Number/MSL	0.8	0.1	0.3	1.1
MSL$<$ number (n=22)				
Sweden (n=2,962)				
Traffic accidents	54.2	69.4	1	874
MSL of traffic accidents	69.6	82.5	7.5	844.5
Number/per capita	206.2	90.3	6.7	775.9
Number/MSL	0.8	0.2	0.04	1.3
MSL$<$ number (n=261)				

Source: Estonian Rescue Board 2016; PRONTO 2016; IDA 2016; Author's calculations.

Table 3.9: The descriptive statistics of outputs (departures of the FRS subunit to other emergencies) in analysed countries

Statistic	Mean	St. Dev.	Min	Max
Estonia (n=325)				
Number	208.5	282.3	13	1,587
MSL	252.9	325.1	32.8	1,577
Number/per capita	1,104.9	511.3	236.6	3,522.6
Number/MSL	0.8	0.2	0.3	1.1
MSL< number (n=63)				
Finland (n=264)				
Number	3,413.2	1,761.9	759	9,036
MSL	3,954.3	1,940.0	1,478.1	9,027.2
Number/per capita	1,470.7	263.7	838.4	2,309.0
Number/MSL	0.9	0.1	0.5	1.1
MSL< number (n=22)				
Sweden (n=2,962)				
Other emergencies	182.4	324.1	0	4,221
MSL of other emergencies	215.7	347.3	15.3	4,215.4
Number/per capita	587.3	235.8	0.0	2,087.6
Number/MSL	0.8	0.2	0.0	1.3
MSL< number (n=276)				

Source: Estonian Rescue Board 2016; PRONTO 2016; IDA 2016; Author’s calculations.

In Estonia, the number of employees varies from nine to 87 and the number of vehicles from one to 19. Over the brigades, it averages to 25 full-time employees and almost five vehicles per brigade. The ratio of labour costs to number of employees is on average 11,400 euros (2011 prices, which applies to every cost) and for other inputs around 31,400 euros per vehicle.

In Finnish fire departments, the number of employees varies from 17 to 479 and the number of vehicles from 26 to 227. It must be noted that the number of vehicles has limited data availability, so it presents the state in the beginning of 2017. Over the departments, it averages to 172 full-time employees and 87 vehicles per department. For comparison, one fire department has 17–146 FRS brigades in its service area, averaging up to 60 FRS brigades (which is similar to the total of national FRS brigades in Estonia). The ratio of labour costs to number of employees in Finland is on average 69,100 euros (2011 prices, which applies to every cost) and for other inputs around 91,100 euros per vehicle. Thus, the labour is six times and the other inputs around three times more expensive than in Estonia.

In Swedish municipalities, the number of employees varies from one to 524, being 58 on average. This includes all employees, including full-time, part-time and voluntary rescuers. Unfortunately, the number of vehicles is not readily available, so that variable has been replaced by the number of FRS brigades in a municipality. In the average municipality, there are three FRS brigades, which varies from a minimum of one to 15 in biggest municipality. The ratio of labour costs to number of employees is on average 12,300 euros (2011 prices, which applies to every cost). Such ratio covers all rescuers (full-time, part-time, voluntary), so for the analysis, one must make an assumption that the ratio of different types of rescuers should stay constant for the analysis to be adequate. Therefore, the ratio is similar to the Estonian case (which included only professional rescuers). For other inputs, the price for capital is expressed by other costs for a FRS brigade, which is around 740,000 euros per brigade.

For the total costs (see Table 3.11), keeping a professional FRS brigade in Estonia operating amounts to 420,000 euros annually on average. The labour costs amount to 292,000 euros and other costs to 128,800 euros. Thus, the labour costs are around 70% of total costs. Since Estonian FRS brigades are entitled to different service provision, the special functions vary to an extent across brigades. The ERB follows the service-based management, which means that every type of service offered has been described and follows a certain standard.

The FRS brigades then vary on the operational capabilities, meaning the larger FRS brigades have the necessary equipment and training to provide extra services. This also has an effect on the costs of FRS brigades. Based on the offered services and service areas (especially population size in the service area, see Päästeamet (2012a) and Päästeamet (2012b) for further elaboration), the FRS brigades are divided into three groups, the smallest ones are in group 1 and the largest in group

Table 3.10: The descriptive statistics of inputs of FRS subunits in analysed countries

Statistic	Mean	St. Dev.	Min	Max
Estonia (n=325)				
Employees (number)	25.3	121	9	87
Employees per capita (number/100,000 pop)	256.5	147.3	26.7	699.3
Total vehicles (#)	4.6	3.2	1	19
Total vehicles per capita (number/100,000 pop)	44.5	35.5	3.0	307.2
Labour costs/employees ('000 of 2011 €)	11.4	1.7	2.7	18.8
Capital costs/total vehicles ('000 of 2011 €)	31.4	40.1	3.2	555
Finland (n=264)				
Employees (number)	173.7	106.1	37	479
Employees per capita (number/100,000 pop)	72.1	13.6	11.5	112.0
FRS brigades (number)	59.5	31.8	17	146
Total vehicles (number, 2017)	86.9	57.6	26	227
Total vehicles per capita (number/100,000 pop)	46.2	34.7	8.2	117.3
Labour costs/employees ('000 of 2011 €)	65.7	22.1	30.7	258.8
Capital costs/total vehicles ('000 of 2011 €)	91.1	80.3	8.7	357.4
Sweden (n=2,962)				
Employees (number)	57.9	52.4	1	524
Employees per capita (number/100,000 pop)	292.1	235.0	7.1	2,259.6
FRS brigades (number)	3.3	2.2	1	15
FRS brigades per capita (number/100,000 pop)	20.1	18.3	0.0	119.2
Labour costs/employees ('000 of 2011 €)	12.3	20.2	0	495.8
Capital costs/FRS brigade ('000 of 2011 €)	741.5	804.6	7.9	8,913.0

Source: Estonian Rescue Board 2016; PRONTO 2016; IDA 2016; Author's calculations.

3 (from here on, the group 1 FRS brigades are referred to as 'small-sized FRS brigades', group 2 as 'middle-sized FRS brigades' and group 3 as 'large-sized FRS brigades'). The differences in the performance in-between groups will be analysed further. The division between groups also indicates the population density in the service areas of the FRS brigades, meaning that the large-sized FRS brigades are operating in bigger cities, middle-sized FRS brigades in smaller towns and bigger boroughs, and the small-sized FRS brigades in rural or distant areas.

For Finland, the annual cost of keeping a fire department operating amounts to 15.9 million euros on average. The labour costs amount to 10.6 million euros and other costs to 5.3 million euros. Thus, the share of labour costs is a little lower than in Estonia, around 66.7% of total costs.

Swedish municipalities spend on average about 3 million euros a year to keep the FRS operating in a municipality. The labour costs amount to 760,000 euros and other costs to 2.24 million euros. Thus, the share of labour costs is lower than in other cases, but it must be noted that there are many volunteers involved in the work-force, which cuts down the labour costs.

For a comparison of the countries, the costs per 100,000 people are pointed out (thousands of 2011 euros) (see Figure 3.3). The Swedish FRS is the costliest, closely followed by Finland in the last few years. The costs for Estonian FRS are around 40% of the Finnish and Swedish costs. A similar result is in the figure that contrasts the costs per 1,000 emergencies (see 3.4). As Sweden showed the least emergencies per population, its costs are also the highest. This might indicate that although there are less emergencies occurring, the costs do not reduce (which indicates a need to keep the FRS on standby, so it needs resources but does not produce outputs). Lastly, the costs per FRS brigade are compared (see Figure 3.5). As some Estonian FRS brigades react to more emergencies than do the Finnish fire departments (with multiple FRS brigades), the costs are higher. In addition, one should note that the Finnish data includes all FRS brigades that are under the fire department (voluntary, industrial, etc.) but Estonian data includes only the national FRS brigades.

3.3.5. Environmental factors

The environmental factors in the models are represented by the reciprocal of the service area (see Table 3.12), assuming that the larger areas are more difficult to service in case the subunit is in a single location of its service area. In the case of fire departments, however (in contrast to the FRS brigades, when it is centrally located with respect to its service area), the relationship is not so clear, as a fire department has multiple FRS brigades allocated in various locations.

The service areas' sizes vary to a great extent. The fire departments in Finland cover the largest service areas, 17,000 km^2 on average. The smallest average

Table 3.11: The descriptive statistics of costs of FRS subunits in analysed countries

Statistic	Mean	St. Dev.	Min	Max
Estonia (n=325)				
Total costs ('000 of 2011 €)	421.1	288.4	149	2,546
Total costs per capita ('000 of 2011 €/100,000 pop.)	3,877.9	2,224.0	409.9	14,224.3
Labour costs ('000 of 2011 €)	292.3	160.1	57	1,185
Labour costs per capita ('000 of 2011 €/100,000 pop.)	2,818.3	1,558.3	301.7	8,743.2
Other costs ('000 of 2011 €)	128.8	167.6	10	1,543
Other costs per capita ('000 of 2011 €/100,000 pop.)	1,059.6	967.6	21.4	7,127.5
Finland (n=264)				
Total costs ('000 of 2011 €)	15,858.9	9,548.6	4,096	52,893
Total costs per capita ('000 of 2011 €/100,000 pop.)	6,709.9	1,439.3	3,936.4	10,015.8
Labour costs ('000 of 2011 €)	10,597.5	6,217.4	2,789	36,944
Labour costs per capita ('000 of 2011 €/100,000 pop.)	4,528.2	960.1	2,752.5	6,998.1
Other costs ('000 of 2011 €)	5,261.4	3,471.6	995	18,228
Other costs per capita ('000 of 2011 €/100,000 pop.)	2,181.6	671.8	578.3	4,625.8
Sweden (n=2,962)				
Total costs ('000 of 2011 €)	3,003.1	4,381.7	54	63,235
Total costs per capita ('000 of 2011 €/100,000 pop.)	11,111.3	5,800.7	356.0	76,412.4
Labour costs ('000 of 2011 €)	761.7	1,412.6	0	28,462
Labour costs per capita ('000 of 2011 €/100,000 pop.)	3,299.6	3,958.5	0	29,808.7
Other costs ('000 of 2011 €)	2,241.4	3,702.0	54	47,444
Other costs per capita ('000 of 2011 €/100,000 pop.)	7,811.7	3,290.2	356.0	68,271.7

Source: Estonian Rescue Board 2016; PRONTO 2016; IDA 2016; Author's calculations.

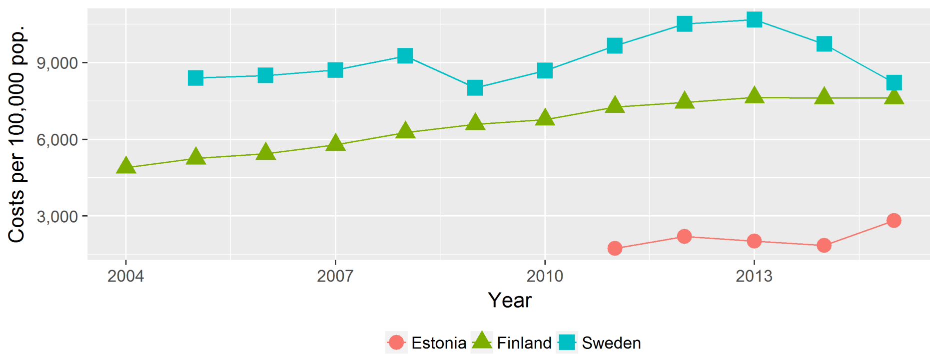


Figure 3.3: Costs ('000 of 2011 €) of FRS per 100,000 population in Estonia, Finland and Sweden (Source: Estonian Rescue Board 2016; PRONTO 2016; IDA 2016; Author's calculations)

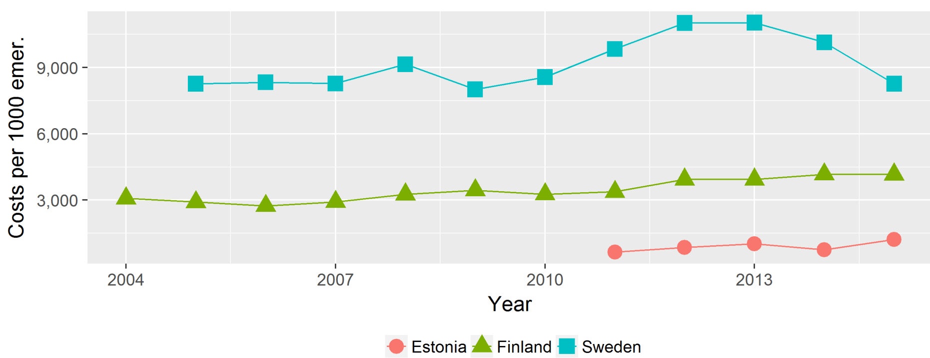


Figure 3.4: Costs ('000 of 2011 €) of FRS per 1,000 emergencies in Estonia, Finland and Sweden (Source: Estonian Rescue Board 2016; PRONTO 2016; IDA 2016; Author's calculations)

service area is for Estonian FRS brigades, averaging around 624 km^2. A service area for an average Swedish municipality is about 2.5 times bigger than for Estonian FRS brigade. The smallest service areas are for the largest cities (or districts of a city) in every country and naturally the largest service areas are in the most distant rural areas.

In addition, to control whether the estimated results are correlated to some other variables, the average time to reach the emergency scene after receiving the message is considered. For Estonia, it averages over the brigades to less than 14 minutes, in Finland 11 minutes and in Sweden less than 7 and half minutes.

Furthermore, the population reached within 15 or 20 minutes (or the whole service area population), which in most cases would be the target level, has

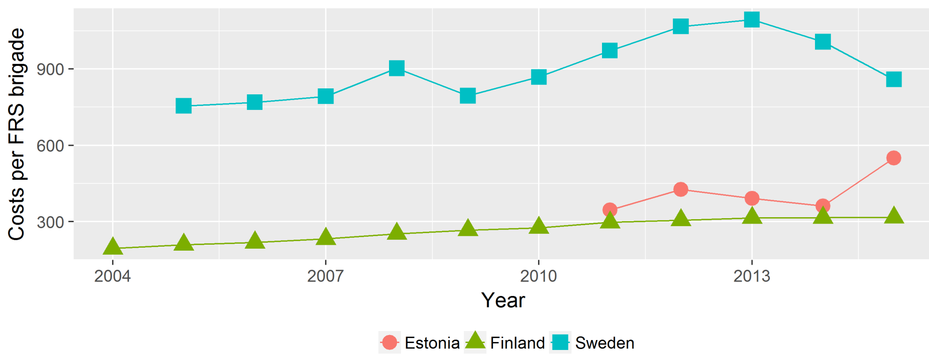


Figure 3.5: Costs ('000 of 2011 €) of FRS per brigade in Estonia, Finland and Sweden (Source: Estonian Rescue Board 2016; PRONTO 2016; IDA 2016; Author's calculations)

been included. The more people in the close vicinity of the FRS brigade, the more emergencies are likely to happen, which also should have an effect on the efficiency of FRS brigade.

This section described and analysed the indicators used to empirically evaluate the performance of FRS in three countries. The outputs, inputs and environmental factors were compared and multiple ratios indicated, to introduce the context of FRS. The following section estimates the performance based on the theoretical framework introduced in the theory part (see Subsection 1.4).

3.4. Estimation and results

3.4.1. Introduction

In the current case, the indicators discussed in the previous subsection were evaluated for FRS systems in three countries. However, one must note that the service provision differs at the organisational level and thus constrains even further the possibilities for comparison. In the Estonian case, the system is centralised and only the professional national FRS brigades were included in the analysis. In the Finnish case, the FRS are provided by fire departments, which have the autonomy to plan their service provision and have multiple FRS brigades under their management. The data on Finnish FRS brigades is limited, which is why the analysis was undertaken at fire department level. Under fire departments, there are professional, part-time and voluntary rescuers, which indicates the complexity of the system. For the Swedish case, the system is even more decentralised, as the FRS are provided by the municipalities. To complicate the analysis further, the municipalities cooperate with each other. In

Table 3.12: The descriptive statistics of environmental factors of FRS subunits in analysed countries

Statistic	Mean	St. Dev.	Min	Max
Estonia (n=325)				
Service area (km^2)	623.6	323.3	15	1,883
Population reached within 15 minutes	17,310	26,951	1,511	167,246
Average time (s)	821.2	253.3	391	2,067
Finland (n=264)				
Service area (km^2)	17,218.9	19,563.3	716	100,368
Population reached within 20 minutes	196,423.9	141,447.8	26,034	564,190
Average time (s)	660.0	93.0	495	985
Sweden (n=2,962)				
Service area (km^2)	1,412.5	2,474.9	19	19,140
Population of the service area	32,758.8	62,764.9	2,516	923,516
Average time (s)	441.9	120.9	60	1,860

Source: Estonian Rescue Board 2016; PRONTO 2016; IDA 2016; Author's calculations.

Sweden, much FRS is also provided by voluntary and part-time rescuers, which are included in the database.

The data limitation hinders the complexity of the performance measurement, as for instance, not many indicators that would allow the effectiveness of FRS to be measured are available. Thus, one has to constrain the analysis mostly to the efficiency measurement. As such, one can assess the FRS systems cost-efficiency (can one do the same amount with less finances) and output efficiency (would there be a possibility to provide more services in case of increased demand with the same resources). This lessens the importance of the most crucial question – whether the FRS provision actually does what it should be doing. In addition, the current models suffer from the lack of 'other inputs'. In the Estonian case, this is assumed to be proportional to the total number of vehicles that a FRS brigade has. In the Finnish case, only the total number of vehicles are available on a rolling basis (and so, the status in the beginning of 2017 was used) and in the Swedish case, such data is not readily available and has been substituted with the number of all FRS brigades in the municipality.

Different used methods make different assumptions about the production possibilities set. DEA makes an assumption about convexity (any weighted average of feasible production plans is feasible as well.) so the production possibilities set is larger and thus the analysed units will probably receive lower efficiency estimates. In the FDH case, the convexity assumption is relaxed, so the frontier is formed only by a 'better self'. In case there is no such comparable unit, the method would use the unit itself as a benchmark (resulting in an efficiency estimate of 1). The production possibilities set is smaller and thus the units will receive a higher efficiency estimate in comparison to DEA. In the DFA case, only one unit (with the biggest residual) will receive the efficiency estimate of 1. Thus, that method is highly responsive to any (measurement) errors that might occur during the analysis. As a result, the mean and median efficiency estimates are probably the lowest among the three discussed methods. To overcome such problem, the noise term should be introduced (SFA).

3.4.2. Industry-level cost-efficiency of FRS

Introduction

The necessity to provide cost-efficient FRS was stressed in every analysed country's strategic documents. For industry-level cost-efficiency one is interested in using as few finances as possible to provide an optimal (fixed) level of services at a country-wide level. The measure of cost efficiency also indicates the potential savings when the subunit would be functioning fully efficiently.

Following the theoretical case (see the first chapter), the central agency (or government) allocates resources (different inputs) to different subunits, while taking into account that the demand for their services is uncertain. Thus, the allocation involves previously made costs (amount of inputs times price for inputs). The industry-level cost-efficiency in a standard case would measure how cost-efficiently the resources in financial terms were allocated when the amount of real outputs are clarified (*ex post*), which in the following models have been named as the naïve models. As in reality, the amounts of outputs are unknown due to uncertain demand, the central agency has to decide in the first place on a hypothetical expected demand, which would be called minimum service level (MSL). Such expected demand will probably be higher than the real demand, as the decision maker is assumed to be risk averse (to avoid public scrutiny due to the possibility of unmet higher levels of demand). Therefore, this sort of cost-efficiency would take the uncertainty into account and does not label such input usage as inefficiency (when the real amount of outputs were less than the expected MSL). The models would use the real outputs to define the technology (plausible assumption) and then the expected MSL amount of outputs would be used as the observations for which the optimal, cost minimising input vector is to be estimated.

The cost-efficiency has been estimated using the DEA, FDH and DFA (see Chapter 2.5). For each estimation method a country comparison is consequently undertaken. To point out the differences between the models that take into account the uncertain demand (with the need to provide a certain standby capacity), naïve models are also estimated, where m_{it} has been replaced by q_{it}, so the estimation would indicate the results of the standard cases.

Efficiency estimation using DEA

The cost efficiency is first estimated using the DEA. A model that uses the previously described inputs and outputs are formalised and evaluated. The model does not allow technical regress, meaning that for each year a new estimation is undertaken, allowing the frontier to be formalised also by the observations with the characteristics of previous years (if a subunit would have been able to respond with the same amount of inputs to the determined outputs, it is possible to do the same in following years).

In the case of Estonia, the results from DEA (see Table 3.13) are quite consistent between the naïve model (from here on the naïve CE) and the model that accounts for uncertain demand (CE). The cost-efficiency is estimated at 0.772-0.843 for the naïve CE across the years and 0.906-1.058 for the CE. The CE is higher than naïve CE for every year for about 7–26%, as the MSL has exceeded the observed outputs. In 2011, this would have meant that the FRS brigades were estimated as unable to meet the expected MSL (the CE was estimated greater than 1). In 2013, there were unexpectedly few emergency departures (see Figure 3.2), which means that accounting for the uncertain demand (the risk averse behaviour to assume higher demand) had a bigger effect on the CE estimates.

However, as there were no proper peers for some FRS brigades, the CE estimate was unavailable for some large-sized FRS brigades[4] (which in naïve case resulted mostly a fully efficient estimate).

Table 3.13: The cost-efficiency and potential aggregate savings (thousands of 2011 euros) of ERB using the DEA

Year	Real costs	Naïve CE	CE	Naïve savings	...share of labour	Savings	...share of labour
	('000 €)	(0,1]	(0,...)	('000 €)	(%)	('000 €)	(%)
2011	22,776	0.838	1.058	3,698	84.2	−1,011	98.4
2012	28,111	0.843	0.902	4,411	72.8	2,324	73.1
2013	25,857	0.753	0.930	6,389	60.8	1,518	52.8
2014	23,820	0.814	0.919	4,438	66.1	1,650	61.0
2015	36,304	0.772	0.906	8,291	27.6	2,723	8.6

Source: Estonian Rescue Board 2016; Author's calculations.

Comparing the cost-efficiency in different groups of FRS brigades, both models indicate that the most cost-efficient are large-sized FRS brigades (more than 10 percentage points on average), second the small-sized FRS brigades, and the least cost-efficient are the middle-sized brigades. This means that the FRS brigades in large cities and rural areas are estimated to be more cost-efficient. This holds true for most of the years except 2013–14, when the average CE for middle-sized FRS brigades is estimated at higher than for the small-sized FRS brigades. On average, there are no large differences between small- and middle-sized FRS brigades. The high CE for large-sized FRS brigades can be explained by the high variability in the number of emergencies in that group, which also increases the levels of the MSL. The average cost-efficiency is the most stable across the years in the middle-sized FRS brigades group in both models. The small-sized FRS brigades group has a V-shaped trend (lowest CE in 2013). With respect to the rescue centres, the naïve CE is similar across the centres (slightly lower in the Western (0.78) and higher in the Eastern centre (0.80)), a similar order of results is obtained for the model that includes demand uncertainty (lowest in the Western centre (0.88) and highest in the Eastern (0.98)).

Analysing the models on the single FRS brigade-level, the estimates correlate between different DEA models and years highly (0.66 for 2013 up to 0.90 for 2012). In addition, the CE estimates are quite consistent for FRS brigades across years (the correlations for naïve CE are 0.52 to 0.75 and 0.56 to 0.86 for CE). A similar result applies for ranking: the FRS brigades which received a higher

[4] Jõhvi, Kesklinna, Lilleküla, Tartu

CE estimate in one year would also do well in another year, so they are operating rather stable.

Hypothesising that every FRS subunit would achieve the full cost-efficiency, would indicate that these CE estimates respond to potentially 3.7 to 8.3 million euros on aggregate saved annually (2011 euros) according to the naïve CE. When taking the demand uncertainty into account, only a maximum of 2.7 million euros in 2015 would be potentially saved, and the FRS brigades were estimated not to able to meet the MSL in 2011 (they needed an extra million euros to meet the MSL). The potential save has been estimated highest in 2015, when the occurring real costs were also higher. The potential savings rate (cost inefficiency) would be 16.2% to 24.7% per annum according to the naïve models, and −4.4% to −8.3% for the CE models.

When deconstructing the potential savings between the two types of costs (labour and other inputs), one can see that the share of labour costs in potential savings is decreasing in time in both models. On average, the share of labour costs to total costs were 70% (discussed in Chapter 3.3.4), so from 2013 onwards, the share of labour costs to potential savings was under-represented. Thus, in time, the potential savings have turned from being more labour-related to being more other inputs related. This is most true for 2015, when the costs of other inputs increased to a great extent and the models then estimated this to be an inefficient move.

For Finland, in the case of DEA (see Table 3.14), the naïve CE is estimated at over 0.86 in 2004 and the trend from there on has been slightly negative. The CE model estimated a higher level of cost-efficiency than the naïve counterpart. The downwards trend in the CE can be attributed to the higher prices for inputs and change in the structure of outputs – the share of fires in buildings has decreased to less than 60% of the former level. The potential savings, when bringing all of the fire departments to the cost-efficient frontier, would have thus a positive trend in both models, averaging to 78 million euros a year for the naïve model and 28 million euros for the model that accounts for demand uncertainty.

Table 3.14: The cost-efficiency and potential aggregate savings (thousands of 2011 euros) of fire departments in Finland using the DEA

Year	Real costs	Naïve CE	CE	Naïve savings	...share of labour	Savings	...share of labour
	('000 €)	(0,1]	(0,...)	('000 €)	(%)	('000 €)	(%)
2004	254,668	0.867	1.279	33,984	71.4	-20,823	83.9
2005	273,739	0.794	1.073	56,455	77.4	-9,249	221.1
2006	285,402	0.860	0.936	40,043	44.4	9,597	-78.8
2007	303,891	0.799	0.870	60,958	58.8	26,196	15.3
2008	330,752	0.827	0.866	57,238	51.6	29,150	12.8
2009	349,350	0.789	0.856	73,577	53.9	33,492	16.0
2010	360,720	0.810	0.838	68,396	51.0	38,685	15.7
2011	388,767	0.791	0.829	81,067	50.6	44,472	12.0
2012	399,842	0.725	0.835	109,894	58.2	44,024	21.6
2013	412,404	0.739	0.849	107,569	61.3	41,539	17.7
2014	412,853	0.698	0.818	124,876	60.2	51,526	28.6
2015	414,368	0.709	0.816	120,772	55.8	52,265	26.5

Source: PRONTO 2016; Author's calculations.

In the naïve case, the correlations between the estimates across years are rather strong among subsequent years (0.74–0.98, with the exception of 2004, which has a lower correlation with the next year (0.37)[5]). Results are similar for the case with demand uncertainty (0.75–0.96; 0.39 for 2004). Across models, the correlations for the same year estimates are also strong (0.64–0.94; 0.04 for 2004).

For some fire departments, the MSL has been estimated as too high for the analysis, which results being out of the frontier and thus unable to calculate the CE estimates. This is true for Helsinki, Lappi, Pirkanmaa and Varsinais-Suomi. In the naïve case, these fire departments continuously outperformed other fire departments, as these four averaged the CE around 0.93, while the rest averaged just 0.72. This indicates that the proposed models will have difficulties with dealing with the cases of the biggest fire departments, as there will be no good comparison opportunities for them. Typically, an increasing return to scale will prevail in these models and the MSL would be further away from the frontier.

Similar to Estonia, the share of labour costs in potential aggregate savings is slightly decreasing in time for the naïve case. For Finland, the share of labour costs in total costs were on average 67%, being thus slightly under-represented in

[5]This is due to the model specification – in the first year, the frontier is estimated from 22 observations, which is clearly too few for adequate results. The next year is estimated using the data for both years (e.g. 44 observations) and thus not allowing the technical regress.

potential savings. For the model with demand uncertainty, the potential savings in 2004–2005 are estimated to be negative, meaning the fire departments would have not been able to meet MSL. This was caused by the under-financing of the labour. Thus, the share of labour costs as potential savings fluctuates. From 2007 onwards, the share of labour costs in potential savings is very low, although the trend is upwards. This indicates that the staffing levels of Finnish fire departments are estimated to be well allocated and more potential to cut costs lies in other inputs.

Estimating the cost-efficiency for Swedish municipalities using DEA, the naïve CE has resulted in around 0.6 and it has been stable across years, fluctuating only a few percentage points. The model that accounts for demand uncertainty varies to a greater extent, from 0.56 to 0.72. In most cases, these estimates are little higher than in the naïve CE model. This results in potential savings (all municipalities provide the fully cost-efficient service) of around 237–426 million euros annually for the naïve model and 154–351 million euros for the model that accounts for demand uncertainty.

Table 3.15: The cost-efficiency and potential aggregate savings (thousands of 2011 euros) in Swedish municipalities providing FRS using the DEA

Year	Real costs	Naïve CE	CE	Naïve savings	...share of labour	Savings	...share of labour
	('000 €)	(0,1]	(0,...)	('000 €)	(%)	('000 €)	(%)
2005	606,727	0.608	0.676	237,596	29.3	161,594	26.9
2006	616,076	0.608	0.695	241,633	29.4	159,483	24.9
2007	637,221	0.605	0.726	251,483	28.1	154,602	29.8
2008	820,405	0.604	0.663	324,741	33.7	213,556	35.9
2009	736,833	0.606	0.612	290,427	29.3	232.211	27.4
2010	804,747	0.584	0.563	334,684	29.0	284,112	28.5
2011	891,434	0.569	0.563	384,578	28.4	320,669	28.5
2012	988,109	0.582	0.571	413,316	27.8	351,139	27.5
2013	1,014,872	0.580	0.583	426,557	27.2	350,139	26.4
2014	934,328	0.572	0.578	400,218	25.4	336,592	26.1
2015	797,123	0.581	0.561	334,341	30.8	298,234	31.2

Source: IDA 2016; Author's calculations.

The obtained results are remarkably consistent across years. The naïve case correlates among subsequent years with 0.91–0.98 and the uncertain demand case 0.67–1.00. This also holds for the comparisons across models, correlating from 0.76 to 0.98.

The share of labour costs in potential savings are very stable across years and models, being between 25.4% and 33.7% for the naïve and 24.9% and 35.9% for the one that accounts for demand uncertainty. As the share of labour costs to total costs in Sweden was around 25% on average, this is in line or slightly over-represented in potential savings.

Efficiency estimation using FDH

The results from FDH differ more (see Table 3.16 for Estonia), as due to the relaxed convexity assumption the production possibilities set is smaller. Thus, for each FRS subunit only one peer could be possible (and no mix of two or more peers). For management, however, this simplifies the decisions, as for each subunit there is one certain benchmark to seek. On the other hand, if the set of subunits is small, mostly a proper peer is missing, so one would compare the subunit with oneself. Therefore, the naïve CE in Estonia is near one and the model that accounts for uncertain demand has a higher CE than one, meaning that the costs should have been higher than they really occurred in order to meet the hypothesised MSL. Typically, if oneself is the peer for comparison, but the higher than real outputs MSL is the benchmark, the FRS brigade would be considered to have insufficient resources (funds). The FRS brigades operated with insufficient funds, thus unable to meet the MSL. In different years, the extra required funds were between 7.4 and 13.2 million euros. According to the FDH model, the most insufficiently funded (meaning they were estimated the highest scores) were the large-sized FRS brigades, and the Eastern centre. The naïve case resulted in 330,000–1,600,000 euros for potential savings.

Table 3.16: The cost-efficiency and potential aggregate savings (thousands of 2011 euros) of ERB using the FDH

Year	Real costs	Naïve CE	CE	Naïve savings	...share of labour	Savings	...share of labour
	('000 €)	(0,1]	(0,...)	('000 €)	(%)	('000 €)	(%)
2011	22,776	0.985	1.781	334	74.5	−13,200	73.2
2012	28,111	0.976	1.626	670	44.2	−13,163	69.1
2013	25,857	0.962	1.634	980	16.4	−12,037	61.8
2014	23,820	0.975	1.411	587	34.0	−7,397	67.7
2015	36,304	0.956	1.470	1,606	11.5	−11,529	49.3

Source: Estonian Rescue Board 2016; Author's calculations.

The small production possibilities set hinders the possible analysis in the naïve case, as the majority of FRS brigades would not have a peer to compare to. Thus, in majority of the cases (57–75% across years) the comparable peer is itself, resulting in full efficiency, while the correlations between different models across years are also lower than in case of DEA, 0.34–0.59. Similar results come from correlations between different years for the naïve case (0.21–0.65), but the results from the proposed model are highly correlated (0.45–0.91, the correlation is higher between subsequent years). The share of labour costs in potential savings fluctuates to a greater extent than it did for using the DEA method (especially for the naïve case).

For Finland, the FDH case (see Table 3.17) is very similar to the DEA, resulting in higher cost-efficiency estimates, which are due to the underlying assumptions of the models. The naïve CE estimates are near efficient but show a downwards trend. This corresponds to potential savings of 6–47 million euros a year in the naïve case. For the uncertain demand case, the fire departments would have not been able to meet the MSL according to estimates (on average, an extra 53 million euros a year would have been needed).

Table 3.17: The cost-efficiency and potential aggregate savings (thousands of 2011 euros) of fire departments in Finland using the FDH

Year	Real costs	Naïve CE	CE	Naïve savings	...share of labour	Savings	...share of labour
	('000 €)	(0,1]	(0,...)	('000 €)	(%)	('000 €)	(%)
2004	254,668	0.976	2.211	6,196	76.4	-71,057	80.3
2005	273,739	0.934	1.651	17,976	64.1	-68,928	107.7
2006	285,402	0.968	1.615	9,238	20.0	-70,894	95.0
2007	303,891	0.951	1.423	14,791	27.2	-59,894	109.6
2008	330,752	0.949	1.324	16,922	27.4	-49,779	120.1
2009	349,350	0.944	1.245	19,654	-6.3	-40,280	137.1
2010	360,720	0.971	1.244	10,547	-39.5	-41,305	134.1
2011	388,767	0.953	1.254	18,282	12.8	-46,864	134.5
2012	399,842	0.905	1.287	38,020	50.8	-54,905	125.1
2013	412,404	0.903	1.205	39,942	62.8	-40,544	123.0
2014	412,853	0.905	1.227	39,397	41.7	-46,292	119.5
2015	414,368	0.885	1.216	47,556	35.0	-44,215	132.0

Source: PRONTO 2016; Author's calculations.

Across naïve and demand uncertainty models, the CE estimates correlate from a minimum of 0.07 in 2004 to 0.59 in 2009. The naïve CE estimates correlate 0.11–0.85 in subsequent years (2004–2005 is the lowest correlation) and the CE estimates 0.37–1.00 (again, 2004–2005 was the lowest, when excluded, the average correlation was 0.94 across years), thus the fire departments are estimated to be fairly stable in rankings, and not fluctuating from the least to most efficient in the next year. However, one must note that 8–13 fire departments out of 22 were unable to find a peer in the uncertain demand case (13 for 2004, see the footnote for DEA case), resulting in no estimate for CE. Out of fire departments, Pirkanmaa and Varsinais-Suomi received the highest estimates in most of the years (thus they would have not been able to meet the MSL), and a bit lower estimates were for Kanta-Häme. According to the FDH method, there was some under-financing for labour in 2009–2010 (thus the minus sign in the share of labour costs in potential aggregate savings). The share also fluctuates to a greater extent than it did for the DEA method. The model that takes demand uncertainty into account estimates the share of labour costs in potential aggregate savings to be over 100%, meaning that the under-financing has been estimated to originate from the labour. Thus, there was the potential to save from other inputs (or rather, to reallocate the finances from other inputs into labour).

Similarly, the FDH results in higher CE estimates than in DEA for the Swedish municipalities as well (see Table 3.18), as one can assume from the models' specification. The model that accounts for demand uncertainty estimates higher CE, being in each case higher than 1. Again, the naïve CE estimates are stable across years as for DEA. The naïve case estimates potential savings into the interval of 75–172 million euros and the one that accounts for demand uncertainty, results an under-finance in all of the years, which indicates a need for extra funding from 48–160 million euros.

The correlations between different models are quite strong, 0.48–0.75 across years. The naïve CE estimates correlate 0.51–0.92 in subsequent years and the CE estimates 0.62–0.99. As there are many municipalities in Sweden, most of them found a peer to compare to, leaving only 6–22% of municipalities without an estimate.

The share of labour costs in aggregate potential savings is stable for the naïve case, being in line with the share of labour costs to total costs. The model that accounts for demand uncertainty estimates 2011 and 2015 to have been slightly over-financed the labour (thus the minus sign).

Efficiency estimation using DFA

Unlike the DEA and FDH, the deterministic frontier analysis assumes a functional form. As the current case involves multiple inputs and outputs, the constant elasticity of transformation is assumed and an appropriate CET-type function has been used. In the DFA case, initially an ITE model was estimated using non-linear least squares (Levenberg-Marquardt algorithm) and then the coefficients were used to estimate the cost-efficiency using the values of q_{nit} for naïve case and m_{nit} for the model that accounts for demand uncertainty.

Table 3.18: The cost-efficiency and potential aggregate savings (thousands of 2011 euros) of Swedish municipalities providing FRS using the FDH

Year	Real costs	Naïve CE	CE	Naïve savings	...share of labour	Savings	...share of labour
	('000 €)	(0,1]	(0,...)	('000 €)	(%)	('000 €)	(%)
2005	$606,727$	0.875	1.328	$75,963$	23.2	$-150,011$	33.6
2006	$616,076$	0.866	1.295	$82,631$	30.3	$-140,064$	29.4
2007	$637,221$	0.871	1.316	$82,235$	27.7	$-160,062$	21.9
2008	$820,405$	0.852	1.197	$121,161$	38.2	$-113,049$	12.7
2009	$736,833$	0.852	1.115	$109,039$	29.2	$-62,211$	20.4
2010	$804,747$	0.834	1.093	$133,853$	26.3	$-56,059$	10.1
2011	$891,434$	0.815	1.073	$164,735$	28.8	$-48,347$	-3.4
2012	$988,109$	0.826	1.093	$171,598$	22.7	$-69,405$	28.4
2013	$1,014,872$	0.830	1.133	$172,528$	29.5	$-102,082$	21.9
2014	$934,328$	0.829	1.141	$159,616$	22.2	$-104,881$	28.8
2015	$797,123$	0.826	1.117	$139,072$	36.0	$-73,283$	-4.6

Source: IDA 2016; Author's calculations.

For a comparison, two models were estimated for each country using formula 2.2. The results of the initially estimated models are given in Table 3.19. From there, the coefficients of the q_{nit} model has been used for the basis. The functional form is theoretically plausible, and the results indicate that the coefficients have signs that are consistent with prior expectations. An exception is the naïve case for Estonia, where the technological trend ξ is evaluated as negative, due to the decrease in number of outputs as less emergencies occur. Some of the γ-s are statistically not significant (as the 'other emergencies' account to the highest share in the emergencies, it also resulted in the highest γ value). Somewhat surprisingly, for the Finnish case, the κ is also estimated as negative, but as was pointed out, in fire departments the relation between the FRS and service area is not so clear-cut due to multiple FRS brigades. Most of the coefficients are also statistically significant. In both of the models for Estonia and Sweden, the elasticity of scale (η) is somewhat surprisingly high. All of the models estimate an increasing return to scale as $\eta > 1$.

The cost-efficiency can be found from these estimates using the following

Table 3.19: Non-linear least squares estimates for CET-type production function

	Dependent variable:					
	-ln(x_1)					
	Estonia		Finland		Sweden	
	(q)	(m)	(q)	(m)	(q)	(m)
ξ	-0.012^{***}	0.016^{***}	0.047	0.022^{***}	0.023^{***}	0.039^{***}
	(0.003)	(0.004)	(0.002)	(0.003)	(0.002)	(0.002)
κ	0.098^{***}	0.090^{***}	-0.066^{**}	-0.076^{**}	0.189^{***}	0.177^{***}
	(0.012)	(0.012)	(0.017)	(0.023)	(0.008)	(0.008)
λ	0.094^{***}	0.100^{***}	0.191^{***}	0.188^{***}	0.411^{***}	0.412^{***}
	(0.021)	(0.020)	(0.019)	(0.022)	(0.016)	(0.017)
η	2.590^{***}	2.514^{***}	1.120^{***}	1.133^{***}	2.010^{***}	1.847^{***}
	(0.080)	(0.082)	(0.043)	(0.050)	(0.047)	(0.047)
γ_1	0.258^{***}	0.254^{***}	0.139^{**}	0.106	0.162^{***}	0.088
	(0.040)	(0.056)	(0.053)	(0.092)	(0.042)	(0.060)
γ_2	0.235^{***}	0.054	0.037	0.008	0.228^{***}	0.343^{***}
	(0.045)	(0.054)	(0.073)	(0.127)	(0.042)	(0.055)
γ_3	0.044	0.028	0.235^{**}	0.054	0.204^{***}	0.164^{**}
	(0.040)	(0.051)	(0.073)	(0.148)	(0.046)	(0.056)
γ_4	0.463^{***}	0.664^{***}	0.589^{***}	0.832^{***}	0.406^{***}	0.398^{***}
	(0.061)	(0.069)	(0.101)	(0.159)	(0.049)	(0.060)
Obs	325	325	264	264	1,629	1,629
Res. SE	0.158	0.151	0.212	0.232	0.359	0.366

Note: *p<0.1; **p<0.05; ***p<0.01
ξ – time, κ – 1/area, λ – other inputs/labour,
η – elasticity, γ_1 – fires in buildings,
γ_2 – other fires, γ_3 – traffic accidents, γ_4 – other emergencies

Source: Estonian Rescue Board 2016; PRONTO 2016; IDA 2016; Author's calculations.

formula:

$$-\ln(C_{it}/w_{1it}) = \lambda \ln\left(\frac{w_{it}^*}{\lambda}\right) + \frac{1}{\eta}\ln\left(\sum_{n=1}^{N}\gamma_n q_{nit}\right) - \kappa \ln z_{it} - \xi(t) - u_i t \quad (3.1)$$

where $w_{it}^* \equiv w_{2it}/w_{1it}$,$\xi(t) = \ln B(t)$, $\eta > 0$, $\lambda \geq 0$, $\gamma = (\gamma_1, ..., \gamma_N)' \geq 0$, $\gamma'\iota = 1$ and $u_{it} \equiv -\ln CE_{it} \geq 0$.

For Estonian FRS (see Table 3.20), the industry-level naïve CE is estimated around 0.36–0.49 and the CE is a little higher in every respective year, being 0.41–0.54. The CE estimates are consistently higher than in the case of naïve CE, suggesting that the potential savings would be lower than would be expected by a standard cost-efficiency analysis. This corresponds to a potential saving of 11–23 million euros for the naïve model and 11–21 million euros for the model that accounts demand uncertainty.

Across centres, there are no remarkable differences, only in 2015 the Southern centre outperforms other centres in both models. In the groups, the large-sized FRS brigades were estimated a little lower average cost-efficiency in both models. Between models, the correlations are very strong, across years 0.89–0.97. The results are also fairly stable across years, as the correlations between subsequent years are from 0.58–0.81 for the demand uncertainty case and 0.66–0.87 for the naïve case.

Table 3.20: The cost-efficiency and potential aggregate savings (thousands of 2011 euros) of ERB using the DFA

Year	Real costs	Naïve CE	CE	Naïve savings	Savings
	('000 €)	(0,...)	(0,...)	('000 €)	('000 €)
2011	22,776	0.491	0.529	11,601	10,737
2012	28,111	0.476	0.530	14,743	13,203
2013	25,857	0.418	0.511	15,045	12,638
2014	23,820	0.489	0.549	12,168	10,737
2015	36,304	0.360	0.411	23,243	21,372

Source: Estonian Rescue Board 2016; Author's calculations.

As in the Estonian case, the DFA models for Finland were estimated using non-linear least squares (see the estimates in Table 3.19). The estimated coefficients were used in formula 3.1. Similar to the Estonian case, in the DFA framework the naïve estimates are lower than the ones accounting for demand uncertainty (see Table 3.21). However, the naïve CE is unexpectedly low in the Finnish fire departments, just above 0.2. The model accounting for demand uncertainty results in a little, but not remarkably, higher cost-efficiency. The correlations between the models across years are very strong (0.86–0.99). However, the results are not very stable, as the correlations between subsequent years for the demand uncertainty case vary greatly – in some years, the correlations are low (0.18 for 2004–2005) but in most years strong (excluding 2004–2005, from 0.63–0.97; 0.81 on average). A similar, although slightly lower, result is true for the naïve case.

Table 3.21: The cost-efficiency and potential aggregate savings (thousands of 2011 euros) of fire departments in Finland using the DFA

Year	Real costs	Naïve CE	CE	Naïve savings	Savings
	('000 €)	(0,...)	(0,...)	('000 €)	('000 €)
2004	254,668	0.241	0.341	193,371	167,819
2005	273,739	0.219	0.275	213,801	198,489
2006	285,402	0.231	0.270	219,472	208,222
2007	303,891	0.231	0.260	233,634	224,900
2008	330,752	0.226	0.253	256,021	247,003
2009	349,350	0.210	0.253	275,983	261,010
2010	360,720	0.222	0.250	280,555	270,627
2011	388,767	0.231	0.254	299,001	289,984
2012	399,842	0.215	0.256	314,024	297,443
2013	412,404	0.215	0.256	323,647	306,676
2014	412,853	0.203	0.251	329,008	309,311
2015	414,368	0.207	0.253	328,790	309,378

Source: PRONTO 2016; Author's calculations.

For the Swedish municipalities' cost-efficiency analysis, using the DFA method, only municipalities that had professional rescuers were included. The estimated coefficients were used in formula 3.1 to get the cost-efficiency estimates (see Table 3.22). As expected, the naïve estimates are lower than the ones that take the demand uncertainty into account. The cost-efficiencies fluctuate between years to a greater extent than for other methods, being 0.38–1.17 for the naïve case and 0.41–1.22 in the case when demand uncertainty is accounted for. However, the correlations between different models are strong across years (around 0.99) and in the same models between the years – for the naïve case, from 0.81–0.99 between subsequent years and 0.84–0.99 for the demand uncertainty case.

Table 3.22: The cost-efficiency and potential aggregate savings (thousands of 2011 euros) of Swedish municipalities that have professional FRS brigades (wages > 0, n=1,629) using the DFA

Year	Real costs	Naïve CE	CE	Naïve savings	Savings
	('000 €)	(0,...)	(0,...)	('000 €)	('000 €)
2005	419,223	0.450	0.510	230,506	205,325
2006	424,628	0.872	0.942	54,230	24,651
2007	425,782	0.596	0.666	171,604	142,069
2008	595,635	0.450	0.499	324,739	296,033
2009	517,872	1.173	1.219	−89,472	−113,587
2010	497,894	0.634	0.704	181,882	147,178
2011	540,274	0.595	0.690	219,076	167,390
2012	608,409	0.586	0.686	252,026	190,962
2013	626,980	0.683	0.788	198,594	132,803
2014	529,573	0.450	0.491	291,308	269,581
2015	365,770	0.380	0.408	213,584	203,943

Source: IDA 2016; Author's calculations.

Comparison of methods used

The cost-efficiency was estimated using different methods, namely DEA, FDH and DFA. The aim is to get similar estimates, which would indicate that the results are robust and reflect the data in similar ways. Comparing the different models, the correlations and scatterplots are highlighted (see Figure 3.6 for Estonia, Figure 3.7 for Finland and Figure 3.8 for Sweden). For the Estonian case, all of the correlations are above 0.35, which is in line with different similar studies on efficiency estimates with different methods. For Finnish fire departments, the estimates from different models are also strongly correlated between different methods, being 0.24 at the lowest. For Swedish municipalities, the CE estimates in different models are strongly correlated with their naïve counterparts, as well as with different methods. The weakest correlation is 0.36.

Insights

In all of the models for Estonian FRS brigades, the cost-efficiency for the industry level is decreasing, which might be due to the fact that the number of total emergencies have decreased as well (especially in 2013, which has also quite low CE estimates) and the costs were increasing in 2015, which adds up as the

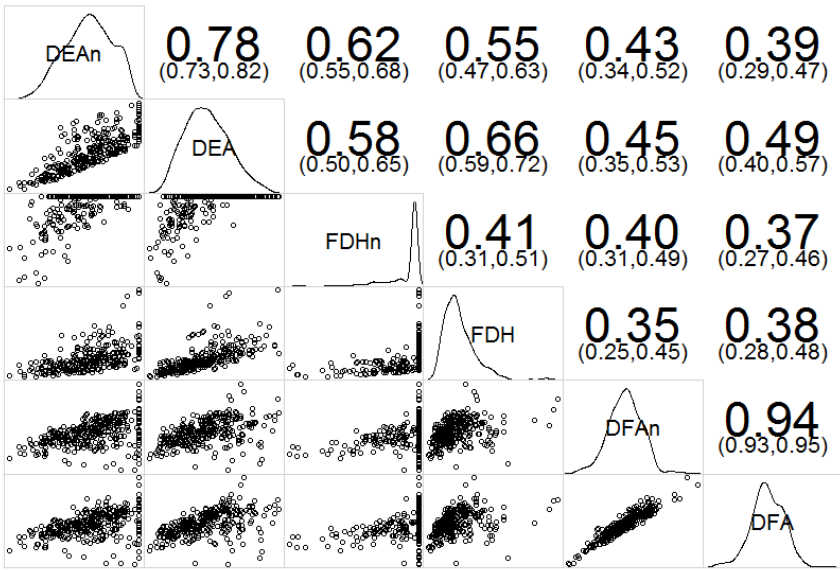


Figure 3.6: The Pearson correlations, densities and scatterplots of cost-efficiency estimates for the Estonian FRS brigades using different methods (Source: Estonian Rescue Board 2016; Author's calculations)

lowest CE estimates across models. The correlations between CE estimates and population reached within 15 minutes are weakly negative (-0.20 to -0.35), which indicates that the higher CE estimates would be gained by brigades with a lesser population in their close vicinity. The correlation between CE estimates and with the average time to the scene is in most cases weakly positive (-0.09 to 0.16), indicating that the brigades which average a higher time are more cost efficient (one can argue that being faster is costlier, which is not indicated in the estimated models).

For Finnish fire departments, the cost-efficiency is also steadily decreasing for a similar reason to Estonia. The CE estimates in different models are very weakly positively correlated with the population reached within 20 minutes in the DEA/FDH models (-0.04 to 0.17) and weakly negatively in the DFA models (-0.18 to -0.19). This indicates that as there are more people reached faster, the CE is estimated higher. For the average time to the scene, the estimates for DEA/FDH are weakly positively (0.05–0.16, except FDH with demand uncertainty, for which it is -0.20) and not at all for DFA estimates correlated. As the correlations are only weak, one cannot make solid conclusions.

For Swedish municipalities, the cost-efficiency is very stable across years

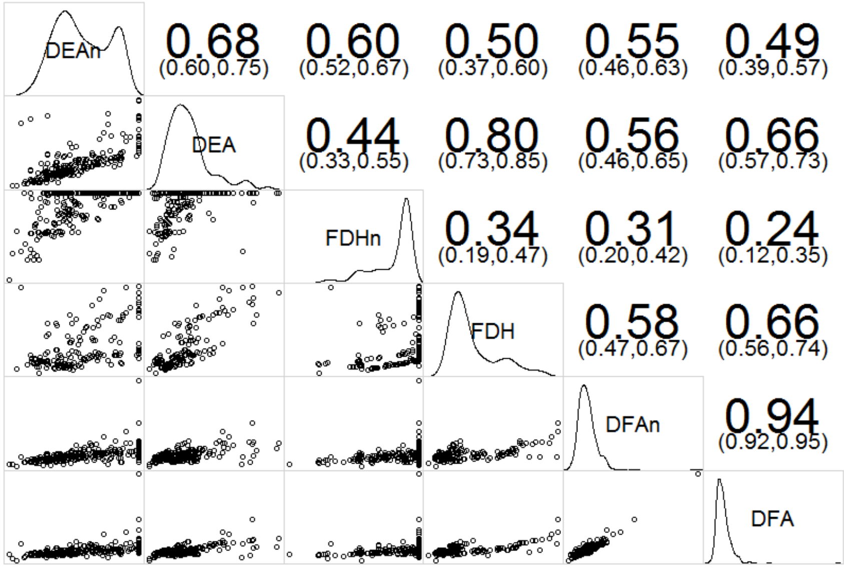


Figure 3.7: The Pearson correlations, densities and scatterplots of cost-efficiency estimates for the Finnish fire departments using different methods (Source: PRONTO 2016; Author's calculations)

estimated by DEA and FDH. For DFA (a subset of municipalities), the CE varies to a greater extent. Some of it can be attributed to the case that some municipalities are cooperating and thus the costs vary from year to year (according to agreements). The estimated cost-efficiencies are weakly positively correlated with the population in the service area (0.14–0.32) and not correlated with the average time to the scene (-0.03 to 0.02). There are no statistically significant differences between voluntary-only municipalities (where the cost of labour is zero) and municipalities that have paid rescuers (for DEA and FDH cases).

3.4.3. Under-resourcing of subunits

Introduction

After the central agency (or government/municipality) has allocated the resources between different subunits, the subunits have to respond to emergencies with the given input bundles (a fixed number of rescuers and vehicles). For the standard case, the input-oriented technical efficiency (ITE) would be calculated in order to analyse whether the subunits would have been able to respond to the observed number of emergencies with fewer amounts of inputs. However, this hindsight

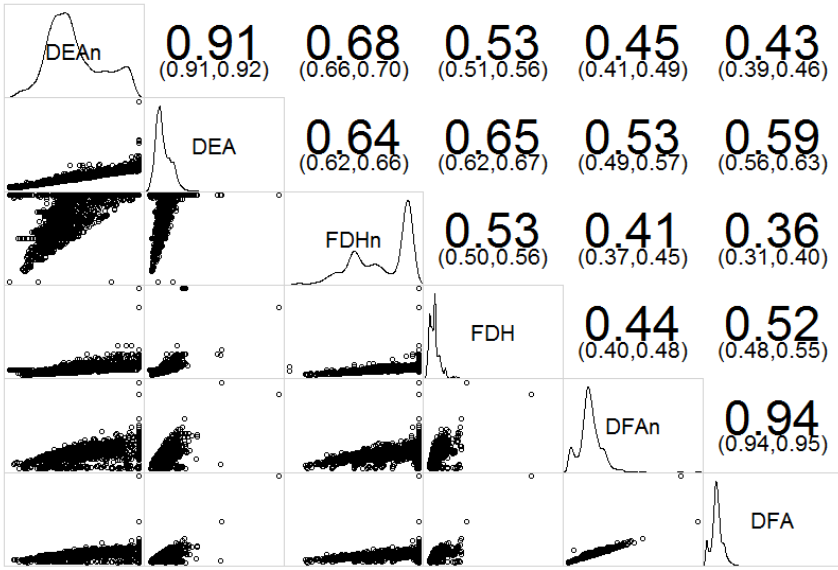


Figure 3.8: The Pearson correlations, densities and scatterplots of cost-efficiency estimates for the Swedish municipalities using different methods (Source: IDA 2016; Author's calculations)

will not capture the whole effect of the service provision, as there is demand uncertainty. Introducing the MSL to the ITE framework, one is able to distinguish the FRS subunits that would have not been able to meet the expected MSL, in case the demand would have been higher as it were observed.

Therefore, as the next step, the under-resourced FRS brigades are identified in the models, which take the demand uncertainty into account (as the concept of being under-resourced in the naïve case is not possible). A FRS brigade is considered under-resourced when it has received less inputs than the MSL would suggest. By estimating the ITE using the MSL, the FRS brigades would result in IDF less than 1 (or ITE more than 1), in case it is under-resourced (see Figure 1.8). Table 3.23 illustrates the share of under-resourced FRS subunits in Estonia, Finland and Sweden, estimated by DEA, FDH and DFA (see Chapter 2.5).

After identifying the under-resourced FRS subunits, the decision makers can use such indication to reassess and reallocate the resources in the following periods at the expense of reducing the resources of most over-resourced FRS subunits (assuming they would have a fixed total budget as a constraint). This would result in a more balanced resource allocation, while considering the MSL as a target.

Table 3.23: The percentage of under-resourced FRS subunits in Estonia, Finland and Sweden, estimated by DEA, FDH and DFA

	DEA			FDH			DFA		
	EST	FIN	SWE	EST	FIN	SWE	EST	FIN	SWE
2004		40.9			40.9			4.5	
2005		27.3	14.2		59.1	63.0		0	3.4
2006		22.7	12.7		59.1	61.1		0	4.1
2007		27.3	13.0		59.1	62.2		0	4.3
2008		22.7	12.4		54.5	56.9		0	2.4
2009		27.3	10.2		59.1	55.8		4.5	3.7
2010		18.2	8.4		59.1	54.4		4.5	5.9
2011	61.5	18.2	7.6	87.7	63.6	52.7	16.9	4.5	4.7
2012	36.9	18.2	8.9	87.7	59.1	55.7	16.9	4.5	7.3
2013	33.8	18.2	8.8	87.7	54.5	54.9	16.9	4.5	6.0
2014	32.3	18.2	8.8	80	59.1	54.6	13.8	4.5	6.3
2015	29.2	18.2	7.7	78.4	54.5	55.3	13.8	4.5	10.1

Source: Estonian Rescue Board 2016; PRONTO 2016; IDA 2016; Author's calculations.

For Estonia, the share of under-resourced FRS brigades is decreasing (similarly to the CE, as these two are interlinked). The results are more similar between the DEA and DFA, in the case of the FDH, most of the brigades are estimated under-resourced. The DEA estimates that around 29% to 62% FRS brigades across years are under-resourced, while only about 14% to 17% for the DFA. For DFA, the results are fairly stable across years. In the case of DEA, in terms of groups, the large-sized FRS brigades are more likely to be under-resourced and it is decreasing in time (around 60% of large-sized FRS brigades in contrast to around 20% of the aggregate of small and middle-sized FRS brigades in 2015). *Vice versa* is true for the FDH, as mostly large-sized FRS brigades are the only ones not under-resourced (although most of the FRS brigades are under-resourced in total). In the case of DFA, mostly the smaller and middle-sized FRS brigades were found under-resourced, the same applies for the Western and Southern rescue centres. In terms of centres, for DEA, more under-resourced FRS brigades are in Eastern and Northern rescue centres (around half of all FRS brigades in these centres were under-resourced in 2015), as there are more large-sized FRS brigades in larger cities.

According to the DEA estimates, in Finnish fire departments around every fourth was under-resourced until 2009 from when the under-resourcing has decreased to a lower level. In the FDH case, more than half of the fire departments were estimated to be under-resourced. The DFA estimates indicate that no (in 2005–2008) fire departments were under-resourced and in other years one fire department (in 2004 Etelä-Karjala, other years Kainuu) was under-resourced. As the estimated CE is quite low, there are also not many under-resourced fire departments (as ITE and CE are interlinked).

For Sweden, in the DEA case, the percentage of under-resourced municipalities has decreased in time, the same is true for the FDH case, although the share of under-resourced municipalities is much higher. The opposite is true for the DFA case, where the percentage of under-resourced municipalities has rather increased (it should be noted that this includes only municipalities with professional rescuers).

Comparison of methods used

The following tables (see 3.24 for Estonia, 3.25 for Finland and 3.26 for Sweden) show the division of under-resourced FRS brigades across different models. These tables indicate that the results vary between models to a large extent – a FRS subunit that is estimated to be under-resourced according to one estimation method receives an alternative estimation in another.

Table 3.24: Division of under-resourced FRS brigades in Estonia across different models

	DEA vs FDH		DEA vs DFA		FDH vs DFA	
	Not UR	UR	Not UR	UR	Not UR	UR
Not UR	45	154	156	43	49	9
UR	13	113	118	8	225	42
χ^2 test	7.139***		12.449***		0	

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Source: Estonian Rescue Board 2016; Author's calculations.

Table 3.25: Division of under-resourced fire departments in Finland across different models

	DEA vs FDH		DEA vs DFA		FDH vs DFA	
	Not UR	UR	Not UR	UR	Not UR	UR
Not UR	99	104	203	0	114	0
UR	15	46	53	8	142	8
χ^2 test	10.212***		23.173***		4.586**	

Note: *p<0.1; **p<0.05; ***p<0.01
Source: PRONTO 2016; Author's calculations.

Table 3.26: Division of under-resourced municipalities in Sweden across different models (for DFA only municipalities that have professional FRS brigades – wages > 0, n=1,629)

	DEA vs FDH		DEA vs DFA		FDH vs DFA	
	Not UR	UR	Not UR	UR	Not UR	UR
Not UR	1343	1341	1336	72	695	44
UR	75	228	175	11	816	39
χ^2 test	68.794***		0.082		1.288	

Note: *p<0.1; **p<0.05; ***p<0.01
Source: IDA 2016; Author's calculations.

Insights

Although the indications of under-resourced subunits of FRS are not very robust and alternate between different models, in most cases the trend has been to better resource allocation (with the exception of Swedish municipalities estimated by DFA). This means that with time, the share of under-resourced subunits has decreased. That might be due to fewer emergencies, as the number of inputs has remained quite steady across years (outputs decrease as inputs stay constant). This complies with the assessment to the cost-efficiency of FRS.

3.4.4. Output efficiency of subunits

Introduction

In addition to cost-efficiencies and under-resourcing, the central agency is interested in how well the resources are utilised by the local subunits in different jurisdictions in comparison to their most efficient counterparts. For that, the output-oriented technical (and mix) efficiencies (OTME) should be estimated. This indicates how many more emergencies the subunits could have responded to, in case there would have been demand for them. The OTMEs are estimated for each country in different years and methods (DEA, FDH, DFA, see Chapter 2.5). When taking the demand uncertainty into account one can argue that the OTME should be one (the FRS subunits are efficient) if they are able to respond to every emergency in their service area – which is the current case (demand does not exceed the supply). While taking this into account, one cannot expect that a FRS subunit would increase its outputs (as services cannot be stored). In other words, even if a FRS subunit would have been able to respond to more emergencies, there was no demand for that (and one should not label this as inefficiency). Therefore, only the naïve OTMEs will be estimated.

Efficiency estimation using DEA

With DEA, the naïve OTME was estimated (see Figures 3.9 for Estonia, 3.10 for Finland and 3.11 for Sweden). In Estonia, the median OTME is estimated at around 0.5–0.6 across years, the lowest median OTME was estimated for 2013, with a few fully efficient exceptions (Kohila, Tartu). The lowest OTMEs are estimated at 0.2, which means that some FRS brigades would have been able to respond to five times as many emergencies when compared to their more efficient peers. The highest estimates are constantly received by the large-sized FRS brigades (they average to around 0.77, while the small and middle-sized FRS brigades average to about 0.5), meaning that there is not much more potential left to increase the workload of large-sized FRS brigades, but there is a possibility to increase the workload for other FRS brigades. In accordance, the Eastern and Northern centres that have more large-sized FRS brigades also have higher efficiency scores on average.

In Finland, the median OTME has a slightly decreasing trend across years. As fire departments are larger subunits (and thus the number of inputs and outputs are more comparable in-between units, as was discussed above) than in the Estonian FRS brigades, the differences between the more and less efficient fire departments are also smaller. This means that only in a few occasions the lowest OTME estimated by DEA is near 0.4.

For Sweden, the OTME is estimated to be the lowest. This is due to the fact that the municipalities differ in size and number of emergencies are manifold, so a few fully efficient municipalities move the frontier far away. The median is around 0.25 and slightly increasing over time. In addition, the median is very stable and do not fluctuate greatly in different years (as there are quite many observations included).

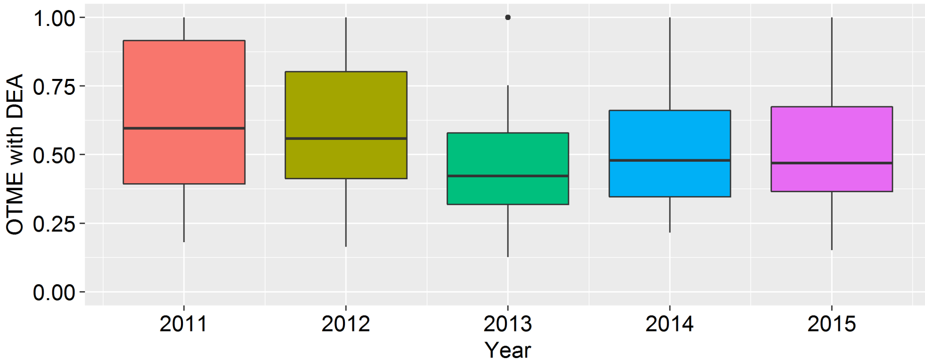


Figure 3.9: The boxplots of estimated OTMEs in Estonian FRS brigades using DEA (Source: Estonian Rescue Board 2016; Author's calculations)

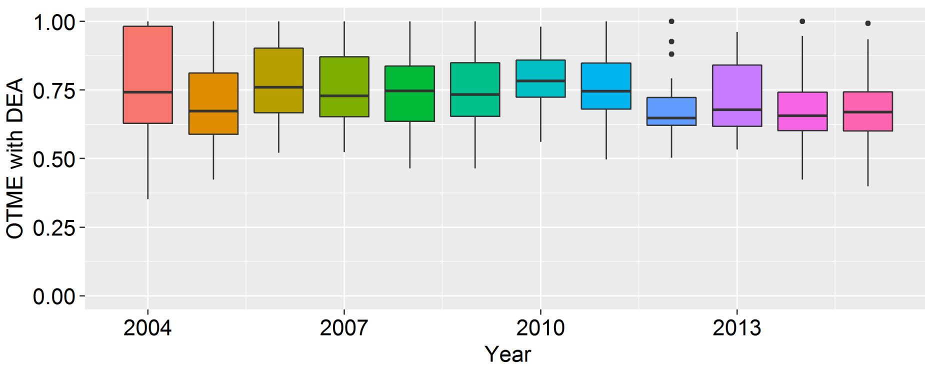


Figure 3.10: The boxplots of estimated OTMEs in Finnish fire departments using DEA (Source: PRONTO 2016; Author's calculations)

Efficiency estimation using FDH

With FDH, the naïve OTME was estimated (see Figure 3.12 for Estonia, 3.13 for Finland and 3.14 for Sweden). As for CE and ITE, the estimates produced by FDH are the highest, which is an expected result. In Estonia, more than every fourth FRS brigade is estimated to the frontier (for 2011–2014), and again 2013 has the lowest median. Similar to the DEA method, the higher scores were received by the large-sized FRS brigades, and Northern and Eastern rescue centres.

For Finland, the median of OTME is decreasing, being near efficient for the first four analysed years, and from there on having a slight decrease.

Figure 3.11: The boxplots of estimated OTMEs in Swedish municipalities using DEA (Source: IDA 2016; Author's calculations)

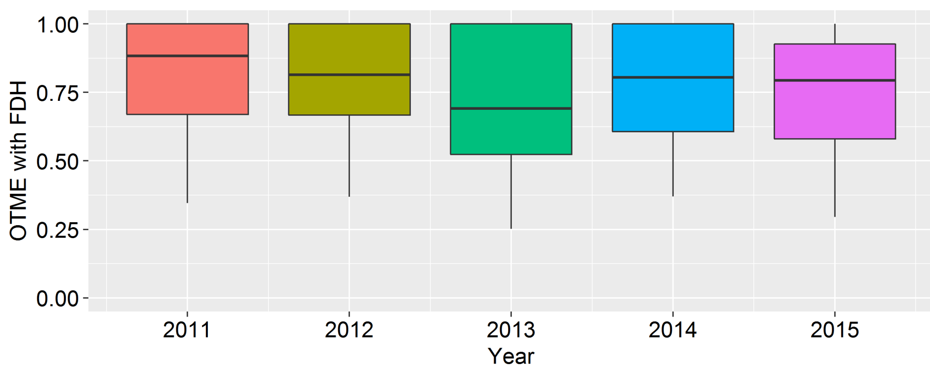


Figure 3.12: The boxplots of estimated OTMEs in Estonian FRS brigades using FDH (Source: Estonian Rescue Board 2016; Author's calculations)

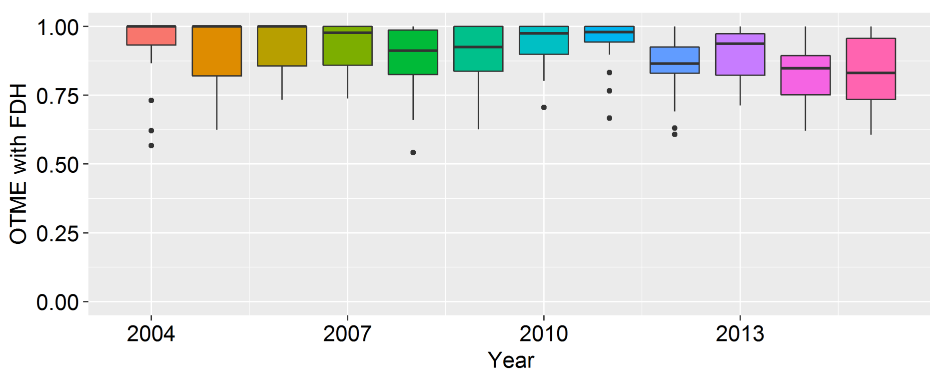


Figure 3.13: The boxplots of estimated OTMEs in Finnish fire departments using FDH (Source: PRONTO 2016; Author's calculations)

For Sweden, the median OTME estimated by FDH has been fairly stable across years, seeing a very slight decrease in trend. In addition, over time the results are converging (as the box plot is getting comparatively shorter, although the whiskers are getting longer).

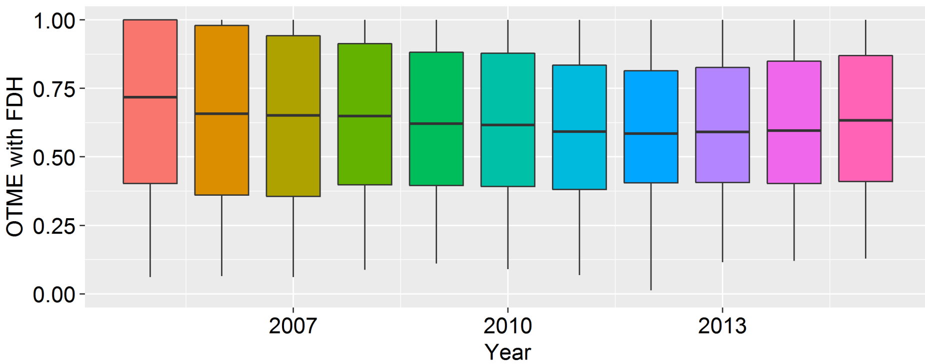


Figure 3.14: The boxplots of estimated OTMEs in Swedish municipalities using FDH (Source: IDA 2016; Author's calculations)

Efficiency estimation using DFA

For the DFA, an output-oriented model following equation 2.3 was used to be estimated by the non-linear least squares. The estimates for each country are given in the following table (see Table 3.27). The coefficients are in accordance with the theory as well as with the previous input-oriented CET-type model. With DFA, the naïve OTE was estimated (see 3.15 for Estonia, 3.16 for Finland and 3.17 for Sweden).

For Estonia, the lowest estimates resulted in the DFA having a few more efficient outliers across the years influencing the frontier. The median OTE is around 0.3–0.4, being the lowest in 2013, just like the two previous estimation methods. In terms of groups, the large-sized FRS brigades receive a somewhat higher score, but not by a wide margin. In terms of rescue centres, there are no significant differences in scores.

In Finnish fire departments, the median is even lower than in the Estonian case. This is due to one efficient outlier in 2004 (Keski-Uusimaa), which influences the frontier to a great extent.

For Sweden, the median estimated by DFA is lower still than in two other countries, being around 0.1. The median OTE is stable across years, with few more efficient outliers.

Table 3.27: Non-linear least squares estimates for output-oriented CET-type production function

	Dependent variable:		
	$\ln(q_1)$		
	Estonia	Finland	Sweden
α	-0.039^{***}	-0.003	0.001
	(0.008)	(0.003)	(0.003)
δ	0.266^{***}	-0.073^{***}	0.197^{***}
	(0.027)	(0.017)	(0.013)
β_1	1.800^{***}	0.720^{***}	0.773^{***}
	(0.076)	(0.024)	(0.027)
β_2	0.218^{***}	0.085^{***}	0.340^{***}
	(0.048)	(0.021)	(0.028)
τ	1.432^{***}	1.320^{***}	1.610^{***}
	(0.075)	(0.088)	(0.086)
γ_1	0.432^{***}	0.195^{*}	0.266^{***}
	(0.063)	(0.080)	(0.048)
γ_2	0.163^{**}	0.371^{***}	0.504^{***}
	(0.053)	(0.087)	(0.054)
γ_3	0.405^{***}	0.434^{***}	0.230^{***}
	(0.073)	(0.100)	(0.054)
Observations	325	264	1,629
Residual Std. Error	0.366	0.203	0.538
Degrees of Freedom	318	257	1,622
Achieved convergence tolerance	1.49e-08	1.49e-08	1.49e-08

Note: *p<0.1; **p<0.05; ***p<0.01

α – time, δ – 1/area, β_1 – labour, β_2 – other inputs,
τ – elasticity, γ_1 – other fires/fires in buildings,
γ_2 – traffic accidents/fires in buildings, γ_3 – other emergencies/fires in buildings
Source: Estonian Rescue Board 2016; PRONTO 2016; IDA 2016;
Author's calculations.

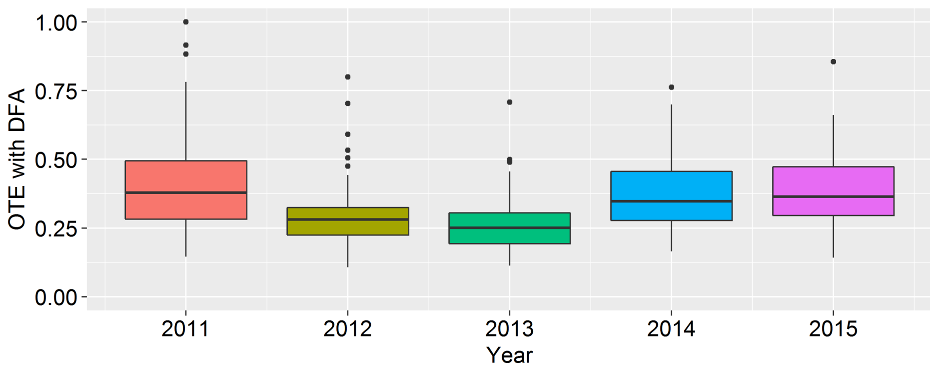


Figure 3.15: The boxplots of estimated OTEs in Estonian FRS brigades using DFA (Source: Estonian Rescue Board 2016; Author's calculations)

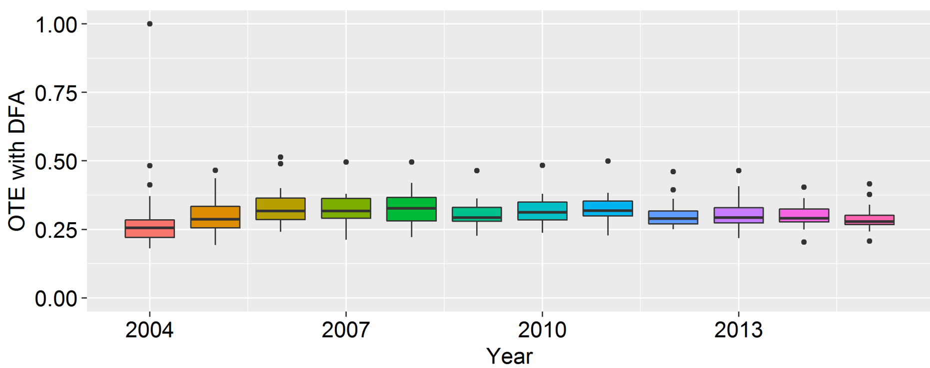


Figure 3.16: The boxplots of estimated OTEs in Finnish fire departments using DFA (Source: PRONTO 2016; Author's calculations)

Comparison of methods used

As the different methods should be measuring the same phenomenon, it is important to evaluate the robustness between the estimates. For Estonia, the output-oriented efficiencies are consistent in terms of fluctuations between years in different models, the lowest estimates are in 2013 (when the least emergencies also occurred). The correlations between the estimates of different methods are positive and strong (0.48–0.73).

In Finnish fire departments, the output-oriented efficiencies are largely consistent in terms of fluctuations between years in different models. The median OTMEs are slightly decreasing from 2008 onwards. The OTE are estimated to be very low with DFA, due to a few very efficient fire departments. To overcome such issue, a stochastic framework should be introduced. The correlations between the estimates of different methods are fairly strong (0.39–0.69).

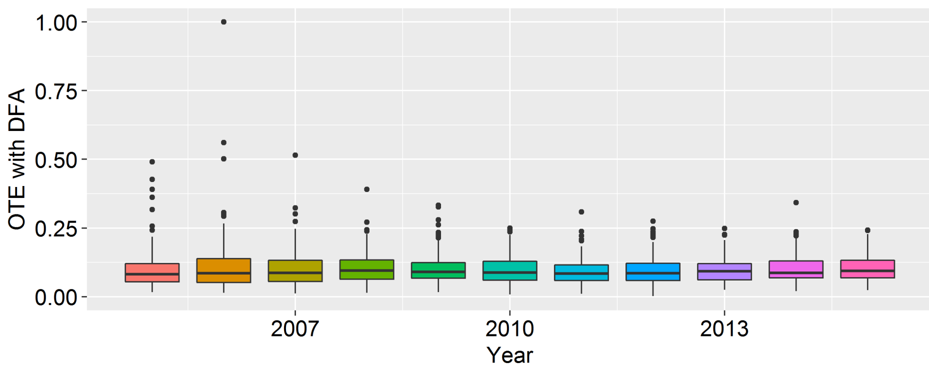


Figure 3.17: The boxplots of estimated OTEs in Swedish municipalities using DFA (Source: IDA 2016; Author's calculations)

For the Swedish municipalities, the medians of output-oriented efficiencies are very stable across years, only slightly increasing in the DEA and DFA cases and decreasing in the FDH case. All of the median estimates are very low in comparison to the results of other countries. The OTE are estimated to be extremely low with DFA, which is due to a few very efficient municipalities. The estimates by different methods are positively and strongly correlated (0.56–0.74).

Insights

As the output-oriented efficiencies are calculated for the naïve case, the years that observed fewer emergencies will be evaluated lower. This is true for each analysed country. Irrespective of the method, there are some years that had lower efficiency scores (e.g. in Estonia for 2013). Thus, to an extent this can be caused by the need to provide a higher standby capacity due to uncertain demand for the services. The results for Estonia are positively correlated with the population within the 15-minute arrival zone (0.46 for DEA, 0.32 for FDH, but only 0.04 for DFA) and negatively correlated to the average time to the scene (-0.26 to -0.30 with different methods). This means that the more there are people in the nearby of the FRS brigade, the higher OTME estimate a FRS brigade would get. The negative correlation with the average time to the emergency scene indicates that the faster the FRS brigade is, the higher OTME estimate it would get.

In the Finnish case, the results are very weakly positively correlated with the population in the 20-minute service area (0.28 for DEA, 0.07 for FDH, 0.07 for DFA) and weakly positively correlated to the average time to the scene (0.11 for DEA, 0.17 for FDH, but -0.16 for DFA). This indicates a similar result to the Estonian output-oriented efficiency estimates for the DFA case.

For Swedish municipalities, the results are positively correlated with the population in the service area (0.26–0.43) but not correlated to the average time to the scene (-0.01 to -0.03 with different methods). Thus, these correlations are rather similar to the estimates of the previous two countries. In most cases, the possible environmental factors indicated a uniform (but weak) relationship to the models.

3.5. Discussion

This section discusses the previous results with respect to different countries and methods used. The introduced PMS which takes the demand uncertainty into account was a mix of public sector management field with the efficiency and productivity field from microeconomics.

The provided PMS started from a centralised public agency, whose first task was to allocate the resources (input bundles) to its subunits in different jurisdictions. The central agency is interested then, how cost-efficient such allocation turned out, given that the demand for services was unknown during the time of allocation. To predict the potential demand, a MSL was introduced, which would be an appropriate tool to help to take such possibilities into account in resource allocation. In the current case it was estimated to be the 95% quantile of the observed number for the corresponding time period of respective emergencies. However, such threshold can be set to other levels (depending on how risk averse the decision maker is), based on contracts or other criteria (negotiation). Introducing this concept for planning purposes (outsourcing services) would increase the appeal of efficiency studies, as it can indicate to the decision maker beforehand the needed resources for a certain threshold level of service provision.

In the empirical study, such MSL approach captured a large (random?) drop in the number of emergencies (e.g. Estonia in 2013), which in a naïve case would have been labelled as a low (cost) efficiency score. Thus, as the decision maker would be unaware of the ‘good’ year beforehand, they should not get the blame for extensive use of resources afterwards. Introducing the demand uncertainty to the efficiency analysis framework brings the finding that the expected potential savings suggested by standard cost-efficiency analysis might be overestimated. Due to the demand uncertainty, some resources were allocated to ensure the necessary means to respond to potential upsurges in demand, indicated by MSL. In all of the cost-efficiency models, the cost-efficiency was estimated to be higher when taking such demand uncertainty into account. Thus, one can argue that not all of the inefficiency was a result of excessive mismanagement but rather a risk averse behaviour to ensure enough resources for potential leaps in demand.

The estimated models were quite robust, as the correlations between different methods and efficiencies are positive and well in line with other efficiency studies comparing the methods (see e.g. O’Neill *et al.* 2008). In addition, the estimated coefficients of CET-type production functions by non-linear least squares are theoretically plausible across countries and mostly statistically significant. The concept introduced works plausibly well in each country, thus the results obtained can be analysed further to be of use in the policymaking.

In addition, after the expenses have been made and resources allocated, the central agency is interested in whether the subunits would have been able to respond to the emergencies with less inputs. For that, ITE is calculated. In terms of demand uncertainty, one is interested in whether there were subunits, which would not have had enough resources to respond to the number of emergencies dictated by MSL. In all of the countries, such subunits existed, but varied to a great extent. In every analysed country, the resource allocation has improved, as the share of under-resourced FRS subunits has decreased in years (the only exception was Sweden in the case of DFA estimates, which indicated an increase in the share of under-resourced subunits). A policy implication is that, these under-resourced subunits should be evaluated further to identify any necessary adjustment to their input bundles.

Finally, one can be interested in the OTE, which indicates whether with given input bundles the FRS subunits would have been able to respond to more emergencies. While taking into account that in the current case, the outputs are services (which cannot be stored) and all the demand was met (all emergencies were attended), such efficiency measure should be equal to one. However, in a standard case an OTE can be calculated while using the occurred emergencies as outputs. This gives an indication of the potential of FRS subunits. The estimated models were quite robust between the methods and years, which would indicate that there are some subunits that constantly outperform other subunits, giving a possibility to learn from them. These efficiency estimates correlated also weakly to the control variables, indicating that the higher OT(M)E estimates also have a higher population in the close vicinity and respond on average quicker to the emergencies.

As a manager of a public agency, one might be interested in:

- Using the concept of MSL in outsourcing or contracting purposes, as dictating the MSL opens the possibility to predict to a certain degree the costs of providing such level of services.

- Reallocating the resources in order to improve the cost-efficiency at the industry level.

- Identifying the under-resourced subunits and reallocating more resources for them at the expense of the over-resourced subunits.

- Identifying the possibility to increase the amount of provided services based on the OT(M)E.

As such, due to simplifications and data issues, this PMS would be a casualty of a few undesirable effects highlighted in the first chapter of the thesis (see

Chapter 1.1.4). First and foremost, these models do not cover the whole spectrum of activities conducted by FRS and so promote tunnel-vision. As prevention and supervision is becoming increasingly important, the subunits would focus on providing these services (in a hope of reducing the outputs of the current models). In addition, analysing the efficiency is only a part of the performance for FRS. There are other possible criteria to consider, e.g. productivity growth, which would enable better comparison between different sets of data. However, then again, one would need to deal with the possible convergence issues of socio-economic levels between different countries (and decreasing number of emergencies as outputs). Since the FRS has a rather standard-based service provision, one might be interested in the equality of the service provision (especially from the service receivers' point of view), e.g. whether every taxpayer is receiving a similar service independent of the location. This is topical in the analysed countries as there are many rural areas, which depopulate rapidly. Overall, as the goal of FRS is the feeling of safety of the population, many (objectively) non-measurable indicators should be taken into account. Thus, the current analysis over-prioritises measurable indicators.

As a limitation of efficiency studies, the direct comparison between different studies is restricted, since the results are dependent on the analysed sample from which the frontier is estimated. This hinders the possibilities for comparative studies – which is also the case for this study. The different analysed countries face various situations that are incomparable within the same framework.

First, the structure of the management and differences in the levels of management. In the Estonian case one can treat the issue as a centralised public agency with subunits. But for Finnish and Swedish case, the responsibilities and rights would go to a lower management level, indicating a higher independence from a benevolent principal. Thus, one might be interested to address this rather as industry-level efficiency analysis and not a single public agency analysis. Yet again, this might ease or complicate the development of suggestions for managers – for a single centralised public agency, one could provide a detailed development plan that can be applied as the agency has the needed power to implement those. For a decentralised system, one would indicate rather a plan for each subunit as there is no single decision maker.

Second, growing out from the first point, are the differences in data sets – one would not be able to put FRS brigades (Estonia) and fire departments (Finland) into the same models as they are not directly comparable (size, responsibilities, assignments). Thus, a comparison would be based on two different models and thus limits making conclusions. One can compare, however, the variations between different countries, which would indicate how standardised the service provision would be (or how it is affected by external environmental factors).

Analysing the used methods, one can indicate some setbacks. Although mostly robust, the methods (DEA, FDH) would get into trouble estimating the efficiencies for bigger subunits, whose MSL would be predicted too high and incomparable to other subunits' levels. Thus, such observations would get no score for efficiency taking the demand uncertainty into account. In the naïve case, such observations would mostly be evaluated to be fully efficient. Another issue is the small production possibilities set provided by FDH – one could only argue, whether the assumption of convexity is plausible or not (and then, which method would be preferable – DEA or FDH?). Finally, the efficiency scores are dependent on the number of observations – it should be large enough to have peers for subunits. In the Finnish case, where the FRS is organised by 22 fire departments, the number of observations might be too low to conduct a meaningful statistical analysis and thus the analysis suffers. In the Swedish case, where the FRS are provided at a municipal level, it is the other way around – as there are 290 municipalities it would be overwhelming for the management to comprehensively deal with such analysis. This would leave Estonia to be of optimal size of subunits (65 for the analysis – acceptable in terms of statistical analysis and not too much to capture them from a managerial point of view) to conduct such an analysis.

This chapter applied the proposed concept of performance benchmarking in demand uncertainty on Estonian, Finnish and Swedish FRS. For each country, the cost-efficiency of the industry was estimated, followed by identifying the under-resourced subunits and aggregate possible OTE calculations. For the cost-efficiency estimations, naïve models were also estimated, which do not account for the demand uncertainty. The estimation was undertaken using three methods: DEA, FDH and DFA. The estimated efficiencies can be considered quite robust, as the correlations of estimates between different methods were (strongly) positive.

CONCLUSIONS

This study has explored the possibilities and limitations of benchmarking the performance of public agencies in the presence of demand uncertainty. The proposed concept rooted on the public administration and microeconomic production theory. For estimation, various frontier analysis methods were used and the concept was applied using Estonian, Finnish and Swedish fire and rescue services as an example.

Performance of a public agency has been an interest in many fields of research. When one is able to measure the performance, the gained information can be used to improve the management decisions in increasing the public agencies' performance. Thus, the performance measurement can be defined as a system or protocol for giving an assessment of how well a public agency is working and fulfilling its goal. The performance of public agencies might differ because they have different objectives, needs, ways of service provision, interactions with other organisations, efficiency, accounting, reporting and measuring methodology as well as random fluctuations. Thus, it might be useful to conduct a performance measurement in order to plan and improve the work of public agencies or to evaluate and benchmark against other public agencies. Planning and improving as an aim for performance measurement is of most value for internal purposes. Evaluation and benchmarking, on the other hand, is mostly targeted at external audiences, who, in turn can influence the operations of subunits. The purpose of learning is focussed on the future, the steering/controlling on the present and evaluating/benchmarking on the past.

The performance measurement does not come without limitations. Namely, these can be classified into three bigger groups: 1) technical issues (measuring in a wrong way), which are related to the choice, collection, analysis, and interpretation of performance indicators; 2) systemic or conceptual issues (measuring the wrong thing), which are related to the integration of the measurement to the management, as well as the lack of strategic focus, goal ambiguity, etc.; and 3) involvement issues (applying the measurement in an unwanted way), which are related to the human factor – how willing is the management to adapt the results.

To start measuring the performance, one must understand the work processes of a public agency. For conceptualising the work process, an input-output model is of use. An input-output model illustrates the mechanism of a public agency from used inputs into provided services. A public agency uses inputs such as labour, capital, time, etc. to create outputs, which can be the provided services. In-between are necessary activities for producing these outputs. From this concept, the performance indicators can be derived, which can be used to characterise the public agencies performance. In the case of FRS, the inputs can be the number of rescuers, vehicles and other equipment. As outputs, one might consider the number of attended emergencies (fires, traffic accidents, other emergencies). In addition, one should note that the FRS operate in a changing environment, which suggests that the performance of a public agency might depend also on the characteristics of the service area, such as the structure of dwellings, socio-economic status, weather, and behaviour of residents.

In order to relate the different performance indicators to each other, many possible evaluation criteria have been proposed. Depending on the aim, one might be interested in the efficiency, effectiveness, economy, equality, utility, ethicality, etc. For this thesis, the interest was targeted mostly on efficiency, as it has a specific economic definition, although in practice is ambiguously used. Using different performance indicators and evaluation criteria, a PMS was developed in order to assess the performance of a public agency comprehensively.

The efficiency is defined in the production theory. Technical efficiency is a measure, which can be defined as the distance (comparison) from the observable units' points of input-output to the production frontier (Daraio and Simar 2007; Fried *et al.* 2008), making it a relative measure. The comparison can be presented in two forms: the ratio of observable and optimal outputs in case of fixed inputs and the ratio of observable and optimal inputs in case of fixed outputs. Koopmans (1951: 60) defined the inputs-outputs vector technically efficient then and only then, when increasing whichever output or decreasing whichever input is possible only in the case, when it follows a decrease in another output or increase in another input, respectively.

When planning the allocation of resources in public agencies, the demand for services is often unknown and prone to uncertainty. Without having beforehand the full information of possible demand, the decision maker will insure oneself with additional standby capacity against upsurges in demand. Cost and efficiency studies predominantly assume known demand, which is unrealistic and hinders understanding the essence of service provision in public agencies. In many cases, it has probably resulted in underestimation of efficiency. The observed excess capacity can partly be explained by risk averse behaviour and should be incorporated to the analysis to avoid unjustly labelling such input usage as inefficiency. *Ex ante*, given expected demand, the resource allocation is optimal, but not *ex post*, given realised levels of demand. The standard case does not consider such standby capacity, so agency appears to operate inefficiently. The challenge is to distinguish the necessary standby capacity from excessive mismanagement. To evaluate the efficiency of a multi-unit and multilevel service providing public agency in the presence of demand uncertainty, one is interested in: (a) the cost-efficiency of the central agency, (b) any under-resourcing of subunits, and (c) the OTE of each subunit in jurisdictions.

There are many possible methods to estimate the performance of public agencies performance. The second chapter systematised and analysed these methods. Starting from the most straightforward analysis of a single unadjusted absolute performance indicator, the level of complexity can increase and the methods used also depend on the user of these methods, having a clearer focus on policy implications, when used by the field of performance management and a more theoretical conceptualisation, when used in the field of productivity analysis. The list of potential methods is extensive: unit cost analysis, ratio analysis, risk adjustment, Four Quadrant model, composite index analysis, cluster analysis, BSC, general linear regression, multilevel analysis, SUR, DEA and SFA. Such a list is not exhaustive, but includes all main methods used so far for the performance measurement.

The frontier analysis methods, which are mainly used in the framework of productivity analysis, have a common purpose of modelling the frontier of feasible performance. The frontier can be estimated under various underlying assumptions and estimation methods. As the next step, observed organisation's performance indicator is then compared to such frontier and relative efficiency is found. Frontier analysis methods compute the observations distance from the frontier – this means measuring the efficiency or performance as maximal-minimal proportionate feasible changes in an activity with given technology (Simar and Wilson 1998). Although popular amongst scholars, frontier analysis methods are rarely used as direct policy tools. The most popular estimation methods are: DEA, which is a deterministic and non-parametric method; SFA, which is a stochastic and parametric method; and other methods, which attempt to fit the advantages of both approaches. The demand uncertainty has not been addressed often in productivity analysis, and when it has, it has been modelled as EU model or using the state-contingent approach. Only a few studies have analysed the performance of FRS and even fewer have used the frontier analysis methods, so the use of these methods in the current application was novel.

For an application for the proposed concept that would evaluate the performance of a public agency in the presence of demand uncertainty, the FRS in three countries, namely Estonia, Finland and Sweden were used. FRS are usually provided by multiple subunits in different jurisdictions. Typically, the rescue authority decides beforehand to allocate resources (rescuers, fire trucks, etc.) between different subunits in jurisdictions without knowing how many emergencies will occur. The subunits must maintain a certain level of standby capacity to be able to react to emergencies. Sustaining the readiness, however, is the most expensive component in the budget, so it would be important to allocate the resources without much waste, e.g. minimising the level of readiness to provide an optimal service.

The structure of FRS systems differs across countries: in Estonia, the system is centralised and provided by one authority, in Finland, the service is provided by 22 fire departments, and in Sweden, the provision of FRS is a task for local municipalities. Estonia has a strategical target to reach the level of the Nordic FRS by 2025, which would mean a decrease in number of emergencies and fatalities. The costs for keeping the FRS operating is most expensive in Sweden and least expensive in Estonia. In addition, the cost per emergency in Sweden is higher. Although the aim and general provision is similar amongst rescue authorities, they are not directly comparable units, when considering the cost-efficiency, under-resourcing of subunits and output efficiency. Therefore, to ensure that the concept would be meaningful, an independent analysis was conducted for each country and afterwards the results and possible improvements were compared.

The empirical analysis was undertaken using three frontier analysis methods, namely DEA, FDH and DFA. As different methods make different assumptions, the results varied to an extent from one another. First, the cost-efficiency at the industry level (principal of the fire authority) was estimated. The cost-efficiency was estimated in two cases – one that would take the uncertain demand into account by using the concept of MSL (expected number of outputs) and a naïve case which would be a standard approach using the real occurred number of outputs. The results showed convincingly that the models which included the demand uncertainty, estimated the cost-efficiency higher independent of the method and country under consideration. This indicates that the decision makers might be providing a certain level of standby capacity to be able to respond to upsurges in demand. So far, this observation has not seen much attention and has been labelled as inefficiency. The cost-efficiency has been estimated quite stable across years and models, meaning that the subunits that did well in one year were ranked highly the next year as well. This opens opportunities to learn from the best, indicating a possibility of positive learning introduced in theory as a goal of performance benchmarking.

Second, the concept of under-resourcing was estimated. In case a subunit received an input bundle that would not have been sufficient to provide an amount of output covering the MSL, the subunit would be labelled as under-resourced. As such concept would be impossible in the standard naïve case, only the models that took the MSL into account were estimated. For every country and method used, the share of under-resourced subunits was decreasing in time. Unfortunately, the results were not very robust, meaning that the subunits that were considered under-resourced by one method, received an alternative estimation by another.

Lastly, the output-oriented technical (and mix) efficiency for the subunits was estimated. This would indicate whether the subunits would still have some potential left to provide more services (in comparison to their more efficient counterparts) in case there would be demand for them. In case of demand uncertainty such concept would not make sense if the demand was met (all the emergencies were attended). Therefore, for that, the subunits should have been considered fully efficient. In the standard case, however, estimating such OTME would be possible. As the subunits are of different nature, there is potential for improvement, resulting in quite low efficiency estimates. The medians of the scores fluctuated in the same directions across methods, indicating that the results are rather robust.

The current study has also a practical output – for example, the concept of demand uncertainty can be applied in the planning process, while allocating the resources. The estimation of MSL supports the improvement of resource allocation (e.g. for contracting purposes), meaning that the need for resources is based on an expected demand for the services. In addition, the proposed PMS can be applied for performance improvement (resource allocation, improving cost-efficiency). Finally, the PMS provides the opportunity to benchmark the subunits based on the efficiency estimates, which opens new possibilities to improve the management decisions (e.g. appreciation of the best performers, helping out the low performers).

In conclusion, introducing the concept of demand uncertainty in the form of MSL into efficiency analysis has provided some useful insights. Foremost, the decision makers are allocating the resources without the full knowledge of outputs. Thus, as they are risk averse, they would be providing a sufficient amount of resources to cover the unexpected upsurges in demand. The empirical study illustrated that applying such estimation to different types of management (centralised, semi-centralised, decentralised) provides possibilities for better understanding of the phenomenon. In addition, the models that would take the demand uncertainty into account were developed using the DEA, FDH and DFA. As a next logical step, one would be interested in introducing the noise term to the analysis, among others.

FUTURE RESEARCH

The thesis is the first stepping stone into a variety of further possible developments. In terms of theory, the uncertain demand as a concept can be exploited further by analysing, how exactly it would alter the behaviour of decision makers and managers and how this in turn can affect the performance of a public agency. In addition, multiple levels of decision makers (and the effect of cooperation between different actors) can be analysed to address the issue comprehensively, as the possible risk aversive behaviour might accumulate. The idea of using minimum service level (MSL) as a prediction for future demand can be of interest in contracting and outsourcing, as it also enables the potential costs for such service provision to be predicted.

Yet again, such MSL approach introduces the negotiation issues between various decision makers and thus can the different goals become of interest. Different levels of decision makers value goals that can be considered contradictory (e.g. to gain votes) as well as the development and change of the demographic and environmental background. If such contradictory goals can be quantified and described, these results can be extended and applied to optimise some decision makers utility. Due to the changes in the society, the service also alternates qualitatively (the structure of various emergencies) and quantitatively (the decreasing trend in total number of emergencies).

Another possible stream on which to focus, would be the environmental factors. As the public agencies operate in vast differences of environment, it surely influences the performance (as well as the rankings based on performance). The environment causes also uncertainty for the decision maker. For the current empirical case, many socio-economic and climate-related environmental factors are of importance (see e.g. Jennings 2013) but difficult to take into account due to data issues (the service areas of FRS do not follow administrative borders). To overcome such an issue, a comprehensive use of GIS advantages could be of help.

Considering the methods, the uncertain demand was only considered in the frontier analysis framework with no noise. It would be of interest to develop models that take the potential measurement errors and misspecification of models into account (namely, SFA). For management, other methods may be also of interest (risk adjustment, among others). Although popular amongst scholars, the frontier analysis methods are not widely used by policymakers. Another possibility would be to design these methods to be more accessible to wider audiences (by designing a user-friendly application, etc.).

In addition to analysing the FRS, other public agencies should be considered, such as the ambulance, police and coastguard, which have the most similarities in their service provision. The uncertain demand is a common feature for public agencies that provide services and even further for the service industry as a whole, as their outputs cannot be stored for later usage.

The current models applied to FRS can be improved by introducing other activities closely related to the service provision as other outputs, these would be activities related to the prevention – education/schooling and supervision. In terms of assumptions made, one might also consider the perfectly-adjustable inputs (e.g. Lovell *et al.* 2009), government budget constraints, factors affecting MSLs (e.g. prevention activities, choice of α), spillovers (i.e. providing services in another jurisdiction), regularity assumptions (strong disposability), environmental uncertainty (e.g. weather and population), and other economic quantities of interest (e.g. values for 'turn-away' rates). One opportunity to increase the comparability of the analysed countries would be to focus on productivity growth instead of efficiency.

Alternatively, one might turn to individual cases (subunits) of the analysis. In order to gather a deep understanding of why some subunit received a score like it did, would be to benchmark it separately against others and complement the results with interviews and consultations of the managers of these subunits (and/or managers of the central agency). This would provide enough information to come up with a development plan to improve the performance of the analysed subunit. Such task would be greatly beneficial from the managerial point of view.

Bibliography

Abbott, M. and Doucouliagos, C. (2003). The efficiency of Australian universities: a Data Envelopment Analysis. *Economics of Education Review*, 22(1):89–97.

Adcroft, A. and Willis, R. (2005). The (un) intended outcome of public sector performance measurement. *International Journal of Public Sector Management*, 18(5):386–400.

Afonso, A. and Aubyn, M. S. (2006). Cross-country efficiency of secondary education provision: A semi-parametric analysis with non-discretionary inputs. *Economic Modelling*, 23(3):476–491.

Ahlbrandt, R. (1973). Efficiency in the provision of fire services. *Public Choice*, 16(1):1–15.

Aigner, D., Lovell, C. K., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6(1):21–37.

Aldenderfer, M. S. and Blashfield, R. K. (1984). *Cluster analysis. Sage University paper series on quantitative applications in the social sciences*. Number 044 in 07. Beverley Hills, CA: SAGE Publications.

Almeida, A. S. and Cima, J. F. (2015). Demand uncertainty and hospital costs: an application to Portuguese public hospitals. *The European Journal of Health Economics*, 16(1):35–45.

Ammons, D. N. (2008). Analyzing performance data. In de Lancer Julnes, P., Berry, F., Aristigueta, M., and Yang, K., editors, *International handbook of practice-based performance management*, pages 145–170. Los Angeles, CA: SAGE Publications.

Andrews, R. (2010). The impact of modernisation on fire authority performance: an empirical evaluation. *Policy & Politics*, 38(4):599–617.

Andrews, R., Ashworth, R., and Meier, K. J. (2014). Representative bureaucracy and fire service performance. *International Public Management Journal*, 17(1):1–24.

Andrews, R. and Van de Walle, S. (2013). New public management and citizens' perceptions of local service efficiency, responsiveness, equity and effectiveness. *Public Management Review*, 15(5):762–783.

Arrow, K. J. (1951). *Social Choice and Individual Values*. John Wiley & Sons.

Aubyn, M. S., Pina, A., Garcia, F., and Pais, J. (2009). Study on the efficiency and effectiveness of public spending on tertiary education. Technical report, Brussels: Directorate General Economic and Financial Affairs (DG ECFIN), European Commission.

Audit Scotland (2012). Learning the lessons of public body mergers. review of recent mergers. Technical report, Audit Scotland.

Audit Scotland (2015). The Scottish fire and rescue service. Technical report, Audit Scotland.

Baker, L. C., Phibbs, C. S., Guarino, C., Supina, D., and Reynolds, J. L. (2004). Within-year variation in hospital utilization and its implications for hospital costs. *Journal of Health Economics*, 23(1):191–211.

Ballanti, D., Dispotico, R., Porcelli, F., and Vidoli, F. (2014). A simple four quadrants model to monitor the performance of local governments. *CESifo Working Paper Series*, 5062.

Banker, R. D. and Maindiratta, A. (1992). Maximum likelihood estimation of monotone and concave production frontiers. *Journal of Productivity Analysis*, 3(4):401–415.

Battese, G. E. and Coelli, T. J. (1993). A stochastic frontier production function incorporating a model for technical inefficiency effects. *Working Papers in Econometrics and Applied Statistics*, 69. Department of Econometrics, University of New England Armidale.

Battese, G. E., Rambaldi, A., and Wan, G. (1997). A stochastic frontier production function with flexible risk properties. *Journal of Productivity Analysis*, 8(3):269–280.

Behn, R. D. (2003). Why measure performance? Different purposes require different measures. *Public Administration Review*, 63(5):586–606.

Bird, S. M., David, C., Farewell, V. T., Harvey, G., Tim, H., and Peter, C. (2005). Performance indicators: good, bad, and ugly. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(1):1–27.

Bock, R. D. E. (2014). *Multilevel analysis of educational data*. Amsterdam: Elsevier.

Bogetoft, P. and Otto, L. (2010). *Benchmarking with DEA, SFA, and R*, volume 157. New York, NY: Springer Science & Business Media.

Bolton, M. (2003). Public sector performance measurement: delivering greater accountability. *Work Study*, 52(1):20–24.

Bouckaert, G. (1992). Productivity analysis in the public sector: The case of the fire service. *International Review of Administrative Sciences*, 58(2):175–200.

Bouckaert, G. (2009). Public sector reform in Central and Eastern Europe. *Halduskultuur-Administrative Culture*, 10:94–104.

Boutsioli, Z. (2010). Demand variability, demand uncertainty and hospital costs: a selective survey of the empirical literature. *Global Journal of Health Science*, 2(1):138.

Boyne, G. A. (2002). Theme: Local government: concepts and indicators of local authority performance: an evaluation of the statutory frameworks in England and Wales. *Public Money and Management*, 22(2):17–24.

Boyne, G. A., Meier, K. J., O'Toole, L. J., and Walker, R. M. (2005). Where next? Research directions on performance in public organizations. *Journal of Public Administration Research and Theory*, 15(4):633–639.

Boyne, G. A., Meier, K. J., O'Toole Jr, L. J., and Walker, R. M. E. (2006). *Public service performance. Perspectives on measurement and management*. Cambridge: Cambridge University Press.

Brignall, S. and Modell, S. (2000). An institutional perspective on performance measurement and management in the 'new public sector'. *Management Accounting Research*, 11(3):281–306.

Brueckner, J. K. (1981). Congested public goods: The case of fire protection. *Journal of Public Economics*, 15(1):45–58.

Burkhead, J. and Hennigan, P. J. (1978). Productivity analysis: A search for definition and order. *Public Administration Review*, 38(1):34–40.

Carlsson, F., Daruvala, D., and Jaldell, H. (2012). Do administrators have the same priorities for risk reductions as the general public? *Journal of Risk and Uncertainty*, 45(1):79–95.

Carvalho, J., Fernandes, M., Lambert, V., and Lapsley, I. (2006). Measuring fire service performance: a comparative study. *International Journal of Public Sector Management*, 19(2):165–179.

Chambers, R. G. and Quiggin, J. (2000). *Uncertainty, production, choice, and agency: the state-contingent approach*. Cambridge: Cambridge University Press.

Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6):429–444.

Chavas, J.-P. (2008). A cost approach to economic analysis under state-contingent production uncertainty. *American Journal of Agricultural Economics*, 90(2):435–466.

Choi, S. O. (2005). Relative efficiency of fire and emergency services in Florida: an application and test of data envelopment analysis. *International Journal of Emergency Management*, 2(3):218–230.

Christensen, T. (2012). Global ideas and modern public sector reforms: A theoretical elaboration and empirical discussion of a neoinstitutional theory. *The American Review of Public Administration*, 42(6):635–653.

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., and Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*. New York, NY: Springer Science & Business Media.

Cooper, W., LM, S., and Tone, K. (2000). *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. Boston, MA: Kluwer Academic Publishers.

Cooper, W. W., Seiford, L. M., and Zhu, J. (2011). Data envelopment analysis: history, models, and interpretations. In Cooper, W. W., Seiford, L. M., and Zhu, J., editors, *Handbook on data*

envelopment analysis, pages 1–39. New York, NY: Springer.
Coulter, P. B. (1979). Organizational effectiveness in the public sector: The example of municipal fire protection. *Administrative Science Quarterly*, 24:65–81.
Cullinane, K., Wang, T.-F., Song, D.-W., and Ji, P. (2006). The technical efficiency of container ports: comparing data envelopment analysis and stochastic frontier analysis. *Transportation Research Part A: Policy and Practice*, 40(4):354–374.
Daraio, C. and Simar, L. (2007). *Advanced Robust and nonparametric methods in efficiency analysis: methodology and applications*. New York, NY: Springer Science & Business Media.
De Borger, B., Kerstens, K., Moesen, W., and Vanneste, J. (1994). Explaining differences in productive efficiency: An application to Belgian municipalities. *Public Choice*, 80(3):339–358.
Debreu, G. (1951). The coefficient of resource utilization. *Econometrica: Journal of the Econometric Society*, 19(3):273–292.
Deming, W. E. (1952). *Elementary principles of the statistical control of quality: a series of lectures*. Japan: Nippon Kagaku Gijutsu Remmei.
Demsetz, H. (1967). Toward a theory of property rights. *The American Economic Review*, 57(2):347–359.
Deng, T.-J., Hsieh, C.-H., Yang, C., and Sheu, H.-J. (2001). A conceptual framework for improving fire-fighting service quality of a public fire department. *International Journal of Public Administration*, 24(4):405–422.
Denhardt, R. B. and Denhardt, J. V. (2000). The new public service: Serving rather than steering. *Public Administration Review*, 60(6):549–559.
Deprins, D., Simar, L., and Tulkens, H. (1984). Measuring labor-efficiency in post offices. In Marchand, M., Pestieau, P., and Tulkens, H., editors, *The Performance of Public Enterprises: Concepts and Measurement*, pages 243–267. Amsterdam: Elsevier.
Dixit, A. (2002). Incentives and organizations in the public sector: An interpretative review. *Journal of Human Resources*, 37(4):696–727.
Draper, N. R., Smith, H., and Pownell, E. (1966). *Applied regression analysis*, volume 3. New York, NY: Wiley.
Drucker, P. F. (1995). *People and performance: The best of Peter Drucker on management*. New York, NY: Routledge.
Duncombe, W. and Yinger, J. (1993). An analysis of returns to scale in public production, with an application to fire protection. *Journal of Public Economics*, 52(1):49–72.
Duncombe, W. D. (1991). Demand for local public services revisited: the case of fire protection. *Public Finance Review*, 19(4):412–436.
Duncombe, W. D. (1992). Costs and factor substitution in the provision of local fire services. *The Review of Economics and Statistics*, 74(1):180–184.
Dunleavy, P. and Hood, C. (1994). From old public administration to new public management. *Public Money and Management*, 14(3):9–16.
Dupuit, J. (1844). On the measurement of the utility of public works (De la mesure de l'utilité des travaux publics, 1844, translated by R. H. Barback). *International Economic Papers*, 2(1952):83–110.
Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., and Shale, E. A. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research*, 132(2):245–259.
Eerma, D. and Friedrich, P. (2016). Welfare-oriented social accounting. In Andessner, R., Greiling, D., and Vogel, R., editors, *Public Sector Management in a Globalized World*, pages 59–93. Wiesbaden: Springer.
Eichhorn, P. and Friedrich, P. (1976). *Verwaltungsökonomie. 1. Methodologie und Management der öffentlichen Verwaltung*. Nomos-Verlag-Ges.
Emergency Services College (2015). *Finnish rescue services' pocket statistics 2010-2014*. Kuopio: Emergency Services College Publication.
ERB (2016). Estonian Rescue Board Yearbook 2015. Technical report, Estonian Rescue Board.
Espenberg, K., Puolokainen, T., and Varblane, U. (2013). Abikaugetes piirkondades päästealase

ennetustöö, ohutusjärelevalve ning päästetöö teenuste optimaalsete osakaalude määratlemine ja sellealase planeerimismudeli väljatöötamine. Technical report, University of Tartu.

Estonian Rescue Board (2016). A combined database from internal sources. A statistics database gathered especially for the thesis in cooperation with the Estonian Rescue Board.

Estonian Rescue Board (2017). Päästekomandode ajatsoonid (Time zones for Estonian FRS brigades). http://www.arcgis.com/home/webmap/viewer.html?webmap=45f7c0b650e54bbda980c181b5a8e8dd&extent=21.3488,57.5497,28.7535,59.719. [Accessed: 27.11.2017].

Fan, Y., Li, Q., and Weersink, A. (1996). Semiparametric estimation of stochastic production frontier models. *Journal of Business & Economic Statistics*, 14(4):460–468.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120(3):253–290.

Fimreite, A. L. and Laegreid, P. (2009). Reorganizing the welfare state administration: Partnership, networks and accountability. *Public Management Review*, 11(3):281–297.

Fiszbein, A., Ringold, D., and Rogers, F. H. (2011). Making services work: indicators, assessments, and benchmarking of the quality and governance of public service delivery in the human development sectors. *World Bank Policy Research Working Paper Series*.

Flynn, J. D. (2009). *Fire service performance measures*. Quincy, MA: National Fire Protection Association, Fire Analysis and Research Division.

Fountain, J. E. (2001). Paradoxes of public sector customer service. *Governance*, 14(1):55–73.

Fried, H. O., Lovell, C. K., and Schmidt, S. S. (2008). *Efficiency and productivity*, volume 3, pages 3–91. New York, NY: Oxford University Press.

Friedman, B. and Pauly, M. (1981). Cost functions for a service firm with variable quality and stochastic demand: the case of hospitals. *The Review of Economics and Statistics*, pages 620–624.

Fryer, K., Antony, J., and Ogden, S. (2009). Performance management in the public sector. *International Journal of Public Sector Management*, 22(6):478–498.

Gaynor, M. and Anderson, G. F. (1995). Uncertain demand, the structure of hospital costs, and the cost of empty hospital beds. *Journal of Health Economics*, 14(3):291–317.

Getz, M. (1979). *The economics of the urban fire department*. Baltimore, MD: Johns Hopkins University Press.

Giannakas, K., Tran, K. C., and Tzouvelekas, V. (2003). On the choice of functional form in stochastic frontier modeling. *Empirical Economics*, 28(1):75–100.

Goddard, R. D. and Goddard, Y. L. (2001). A multilevel analysis of the relationship between teacher and collective efficacy in urban schools. *Teaching and Teacher Education*, 17(7):807–818.

Goldstein, H. and Thomas, S. (1996). Using examination results as indicators of school and college performance. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 159(1):149–163.

Graham, F., Lawrence, S., and Wynn-Williams, K. (2005). Performance assessment and benchmarking in the public sector: An example from New Zealand. *Benchmarking: An International Journal*, 12(5):482–492.

Greene, W. (2004). Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the World Health Organization's panel data on national health care systems. *Health Economics*, 13(10):959–980.

Greene, W. (2005). Fixed and random effects in stochastic frontier models. *Journal of productivity analysis*, 23(1):7–32.

Gregory, R. (2006). Theoretical faith and practical works: de-autonomizing and joining-up in the New Zealand state sector. In Christensen, T. and Laegreid, P., editors, *Autonomy and regulation: Coping with agencies in the modern state*, pages 137–161. Cheltenham: Edward Elgar.

Grünvald, O. (2016). Päästeameti finantseerimise mõju tõhusa päästevõimekuse tagamisele, sealhulgas organisatsiooni jätkusuutlikkusele (in English: The impact of financing of ERB to sustain the efficient rescue capabilities, including the sustainability of the organisation).

Technical report, Finantsakadeemia OÜ.

Gulick, L. (1937). Notes on the theory of organization. *Classics of Organization Theory*, 3:87–95.

Gyrd-Hansen, D., Olsen, K. R., and Sørensen, T. H. (2012). Socio-demographic patient profiles and hospital efficiency: Does patient mix affect a hospital's ability to perform? *Health Policy*, 104(2):136–145.

Hall, J. R., Flynn, J., and Grant, C. (2008). Measuring code compliance effectiveness for fire-related portions of codes. Technical report, Quincy, MA: National Fire Protection Association and Fire Protection Research Foundation.

Hatry, H. P., Schaenman, P. S., Fisk, D. M., Hall Jr., J. R., and Snyder, L. (2007). *How Effective Are Your Community Services?: Procedures for Performance Measurement*. Washington, D.C.: The Urban Institute.

Hauck, K. and Street, A. (2006). Performance assessment in the context of multiple objectives: A multivariate multilevel analysis. *Journal of Health Economics*, 25(6):1029–1048.

Heinrich, C. J. (2012). Measuring public sector performance and effectiveness. In Guy Peters, B. and Pierre, J., editors, *The SAGE Handbook of Public Administration*, pages 32–49. Thousand Oaks, CA: Sage Publishing.

Heinrich, C. J. and Lynn, L. E. (2001). Means and ends: A comparative study of empirical methods for investigating governance and performance. *Journal of Public Administration Research and Theory*, 11(1):109–138.

Henderson, D. J. and Parmeter, C. F. (2015). *Applied nonparametric econometrics*. Cambridge: Cambridge University Press.

Hibbard, J. H., Stockard, J., and Tusler, M. (2003). Does publicizing hospital performance stimulate quality improvement efforts? *Health Affairs*, 22(2):84–94.

Hirsch, W. Z. (1959). Expenditure implications of metropolitan growth and consolidation. *The Review of Economics and Statistics*, 41:232–241.

Hitzhusen, F. J. (1972). *Some policy implications for improved measurement of local government service output and costs: The case of fire protection*. PhD thesis, Cornell University. Unpublished PhD thesis.

Ho, A. T.-K. (2008). Reporting public performance information: The promise and challenges of citizen involvement. In Van Dooren, W. and Van de Walle, S., editors, *Performance Information in the Public Sector*, pages 192–210. London: Palgrave Macmillan.

Höhn, R. (1967). *Das Harzburger Modell in der Praxis: Rundgespräch über die Erfahrungen mit dem neuen Führungsstil in der Wirtschaft*, volume 8. Verlag für Wissenschaft, Wirtschaft und Technik.

Holmgren, J. and Weinholt, Å. (2016). The influence of organisational changes on cost efficiency in fire and rescue services. *International Journal of Emergency Management*, 12(4):343–365.

Holmström, B. (1979). Moral hazard and observability. *The Bell Journal of Economics*, 10(1):74–91.

Hood, C. (1991). A public management for all seasons? *Public Administration*, 69(1):3–19.

Hood, C. (1995). Emerging issues in public administration. *Public Administration*, 73(1):165–183.

Horton, J. (2011). *An examination of the applicability of the Citistat performance management system to municipal fire departments*. PhD thesis, University of Texas at Arlington.

Hughes, D. and McGuire, A. (2003). Stochastic demand, production responses and hospital costs. *Journal of Health Economics*, 22(6):999–1010.

IDA (2016). Statistics system of Swedish rescue services (IDA). A statistics system developed and maintained by the Swedish Civil Contingencies Agency [https://ida.msb.se/].

Iezzoni, L. I. (1997). *Risk adjustment for measuring healthcare outcomes*. Chicago: Health Administration Press.

Iezzoni, L. I., Shwartz, M., Ash, A. S., Hughes, J. S., Daley, J., and MacKiernan, Y. D. (1996). Severity measurement methods and judging hospital death rates for pneumonia. *Medical Care*, 34(1):11–28.

Ingraham, P. W. (2005). Performance: Promises to keep and miles to go. *Public Administration*

Review, 65(4):390–395.

Jacobs, R. (2001). Alternative methods to examine hospital efficiency: Data envelopment analysis and stochastic frontier analysis. *Health Care Management Science*, 4(2):103–115.

Jaldell, H. (2002). *Essays on the performance of fire and rescue services*. PhD thesis, Göteborg University,.

Jaldell, H. (2017). How important is the time factor? Saving lives using fire and rescue services. *Fire Technology*, 53(2):695–708.

Jennings, C. R. (1999). Socioeconomic characteristics and their relationship to fire incidence: A review of the literature. *Fire Technology*, 35(1):7–34.

Jennings, C. R. (2013). Social and economic characteristics as determinants of residential fire risk in urban neighborhoods: A review of the literature. *Fire Safety Journal*, 62:13–19.

Jensen, M. C. and Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4):305–360.

Johnes, G. and Johnes, J. (2009). Higher education institutions' costs and efficiency: Taking the decomposition a further step. *Economics of Education Review*, 28(1):107–113.

Johnsen, A. (2000). Public sector performance measurement uncertainty, ambiguity and epistemic communities. Department of Accounting and Business Method, The University of Edinburgh, 9th version.

Jondrow, J., Lovell, C. K., Materov, I. S., and Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, 19(2-3):233–238.

Joskow, P. L. (1980). The effects of competition and regulation on hospital bed supply and the reservation quality of the hospital. *The Bell Journal of Economics*, 11(2):421–447.

Jung, C. S. (2011). Organizational goal ambiguity and performance: Conceptualization, measurement, and relationships. *International Public Management Journal*, 14(2):193–217.

Justitiedepartementet L4 (2003). Lag (2003:778) om skydd mot olyckor (the Swedish Civil Protection Act (2003:778)). *Regeringskansliet*, 2003(778).

Kao, C. (2000). Measuring the performance improvement of Taiwan forests after reorganization. *Forest Science*, 46(4):577–584.

Kaplan, R. S. and Norton, D. P. (1992). The balanced scorecard-measures that drive performance. *Harvard Business Review*, 70(1):71–79.

Kaplan, R. S. and Norton, D. P. (2001). Transforming the balanced scorecard from performance measurement to strategic management: Part i. *Accounting Horizons*, 15(1):87–104.

Kloot, L. (2006). Budget reporting in local government: Problem or solution? *Australian Accounting Review*, 16(40):72–81.

Kloot, L. (2009). Performance measurement and accountability in an Australian fire service. *International Journal of Public Sector Management*, 22(2):128–145.

Kneip, A. and Simar, L. (1996). A general framework for frontier estimation with panel data. *Journal of Productivity Analysis*, 7(2):187–212.

Koopmans, T. C. (1951). *Activity analysis of production and allocation*. 13. New York, NY: Wiley.

Krasuski, A., Kreński, K., and Łazowy, S. (2012). A method for estimating the efficiency of commanding in the State Fire Service of Poland. *Fire Technology*, 48(4):795–805.

Kristensen, O. P. (1983). Public versus private provision of governmental services: The case of Danish fire protection services. *Urban Studies*, 20(1):1–9.

Kumbhakar, S. C. (2002). Specification and estimation of production risk, risk preferences and technical efficiency. *American Journal of Agricultural Economics*, 84(1):8–22.

Kumbhakar, S. C. and Lovell, C. K. (2003). *Stochastic frontier analysis*. Cambridge: Cambridge University Press.

Kumbhakar, S. C., Park, B. U., Simar, L., and Tsionas, E. G. (2007). Nonparametric stochastic frontiers: A local maximum likelihood approach. *Journal of Econometrics*, 137(1):1–27.

Kuosmanen, T. (2006). Stochastic nonparametric envelopment of data: combining virtues of SFA and DEA in a unified framework. *MTT Discussion Paper*, 3(2006).

Kuosmanen, T. and Johnson, A. L. (2010). Data envelopment analysis as nonparametric least-squares regression. *Operations Research*, 58(1):149–160.

Kuosmanen, T. and Kortelainen, M. (2012). Stochastic non-smooth envelopment of data: Semi-parametric frontier estimation subject to shape constraints. *Journal of Productivity Analysis*, 38(1):11–28.

Lan, C.-H., Chuang, L.-L., and Chen, Y.-F. (2009). Performance efficiency and resource allocation strategy for fire department with the stochastic consideration. *International Journal of Technology, Policy and Management*, 9(3):296–315.

Lång, E. (2012). Är införande av förstainsatsperson samhälsekonomiskt lönsamt?: En kostnadsnyttoanalys av fip-verksamheten i Jönköpings kommun (in English: Is introduction of first responders economically beneficial to society?: A cost-benefit analysis of the project of first responders in the municipality of Jönköping). Karlstad: Karlstad University.

Leckie, G. and Goldstein, H. (2009). The limitations of using school league tables to inform school choice. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(4):835–851.

Leibenstein, H. (1966). Allocative efficiency vs. "x-efficiency". *The American Economic Review*, 56(3):392–415.

Lempert, R. J., Popper, S. W., and Bankes, S. C. (2003). *Shaping the next one hundred years: new methods for quantitative, long-term policy analysis*, volume MR-1626. Santa Monica, CA: Rand Corporation.

Leyland, A. H. and Groenewegen, P. P. (2003). Multilevel modelling and public health policy. *Scandinavian Journal of Public Health*, 31(4):267–274.

Linna, P., Pekkola, S., Ukko, J., and Melkas, H. (2010). Defining and measuring productivity in the public sector: managerial perceptions. *International Journal of Public Sector Management*, 23(5):479–499.

Lockwood, B. and Porcelli, F. (2013). Incentive schemes for local government: Theory and evidence from Comprehensive Performance Assessment in England. *American Economic Journal: Economic Policy*, 5(3):254–286.

Lovell, C., Rodríguez-Álvarez, A., and Wall, A. (2009). The effects of stochastic demand and expense preference behaviour on public hospital costs and excess capacity. *Health Economics*, 18(2):227–235.

Luke, D. A. (2004). *Multilevel modeling*, volume 143. Thousand Oaks, CA: Sage.

Lynn, L. E., Heinrich, C. J., and Hill, C. J. (2000). Studying governance and public management: Challenges and prospects. *Journal of Public Administration Research and Theory*, 10(2):233–262.

MacLeod, W. B. and Urquiola, M. (2009). Anti-lemons: School reputation and educational quality. Technical report, Cambridge, MA: National Bureau of Economic Research.

Majeed, A., Bindman, A. B., and Weiner, J. P. (2001). Use of risk adjustment in setting budgets and measuring performance in primary care i: How it works. *British Medical Journal*, 323(7313):604–607.

Mannion, R. and Goddard, M. (2001). Impact of published clinical outcomes data: case study in NHS hospital trusts. *British Medical Journal*, 323(7307):260–263.

Martin, S. and Smith, P. C. (2005). Multiple public service performance indicators: Toward an integrated statistical approach. *Journal of Public Administration Research and Theory*, 15(4):599–613.

Masso, J. and Espenberg, K. (2013). Early application of fiscal austerity measures in the Baltic States. In Vaughan-Whitehead, D., editor, *Public Sector Shock. The Impact of Policy Retrenchment in Europe*, pages 84–133. Edward Elgar Publishing.

Mattsson, B. and Juås, B. (1997). The importance of the time factor in fire and rescue service operations in Sweden. *Accident Analysis & Prevention*, 29(6):849–857.

Mayne, J. and Zapico-Goñi, E. (2007). *Monitoring performance in the public sector: Future directions from international experience*. New Brunswick, NJ: Transaction Publishers.

McDavid, J. C. and Hawthorn, L. R. (2006). *Program evaluation & performance measurement: An*

introduction to practice. Thousand Oaks, CA: Sage.
Meeusen, W. and van Den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, 18(2):435–444.
Meyer, M. W. and Zucker, L. G. (1989). *Permanently failing organizations*. Thousand Oaks, CA: Sage Publications.
Mintzberg, H. (1994). The fall and rise of strategic planning. *Harvard Business Review*, 72(1):107–114.
Modell, S. (2003). Goals versus institutions: The development of performance measurement in the Swedish university sector. *Management Accounting Research*, 14(4):333–359.
Moore, M. H. (1995). *Creating public value: Strategic management in government*. Harvard University Press.
MSB (2016). *A summary of risk areas and scenario analyses 2012-2015*. Karlstad: The Swedish Civil Contingencies Agency (MSB).
MSB (2017). MSB Swedish Civil Contingencies Agency. https://www.msb.se/en/About-MSB/ [Accessed: 08.03.2017].
MSB (2017). MSB Brandstationsdata. https://gisapp.msb.se/apps/kartportal/enkel-karta_brandstationer.html. [Accessed: 27.11.2017].
Mula, J., Poler, R., Garcia-Sabater, J., and Lario, F. C. (2006). Models for production planning under uncertainty: A review. *International Journal of Production Economics*, 103(1):271–285.
Murphy, P. and Greenhalgh, K. (2013). Performance management in fire and rescue services. *Public Money and Management*, 33(3):225–232.
Musgrave, R. A. (1959). *Theory of public finance; a study in public economy*. New York, NY: McGraw-Hill.
Nauges, C., O'Donnell, C. J., and Quiggin, J. (2011). Uncertainty and technical efficiency in Finnish agriculture: A state-contingent approach. *European Review of Agricultural Economics*, 38(4):449–467.
Neely, A., Gregory, M., and Platts, K. (1995). Performance measurement system design: a literature review and research agenda. *International Journal of Operations & Production Management*, 15(4):80–116.
NFPA (1977). *Municipal fire service workbook*. Research Triangle Institute, International City Management Association, Quincy, MA: National Fire Protection Association.
NFPA (2016). NFPA Overview. http://www.nfpa.org/overview [Accessed: 20.08.2016].
O'Donnell, C. J. (2016a). Deterministic frontier analysis a. A lecture on the Deterministic Frontier Analysis, ECON3340/7341 Productivity and Efficiency Analysis, University of Queensland, Brisbane, 16 September, 2016.
O'Donnell, C. J. (2016b). Practical issues. A lecture on the practical issues, ECON3340/7341 Productivity and Efficiency Analysis, University of Queensland, Brisbane, 21 October, 2016.
O'Donnell, C. J. (2016c). Using information about technologies, markets and firm behaviour to decompose a proper productivity index. *Journal of Econometrics*, 190(2):328–340.
O'Donnell, C. J., Chambers, R. G., and Quiggin, J. (2010). Efficiency analysis in the presence of uncertainty. *Journal of Productivity Analysis*, 33(1):1–17.
O'Donnell, C. J. and Griffiths, W. E. (2006). Estimating state-contingent production frontiers. *American Journal of Agricultural Economics*, 88(1):249–266.
O'Donnell, C. J. and Puolokainen, T. (2017a). Efficiency analysis in the presence of demand uncertainty. Presented at the School of Business and Economics, Loughborough University, 21 June, 2017.
O'Donnell, C. J. and Puolokainen, T. (2017b). Efficiency analysis in the presence of demand uncertainty. Presented at the 2017 CEPA International Workshop on Performance Analysis: Theory and Practice, University of Queensland, Brisbane, 2-3 November, 2017.
O'Donnell, C. J. and Puolokainen, T. (2018). Efficiency analysis in the presence of demand uncertainty.
O'Donnell, C. J. and Shankar, S. (2010). Estimating state-allocable production technologies when

there are two states of nature and state allocations of inputs are unobserved. *Centre for Efficiency and Productivity Analysis Working Papers WP04/2010. University of Queensland*.

O'Donnell, C. J. and Woodland, A. (1995). Estimation of Australian wool and lamb production technologies under uncertainty: An error-components approach. *American Journal of Agricultural Economics*, 77(3):552–565.

ODPM (2005). Measurement of output and productivity of the fire and rescue service. a conceptual framework. Technical report, London: Office of the Deputy Prime Minister.

O'Neill, L., Rauner, M., Heidenberger, K., and Kraus, M. (2008). A cross-national comparison and taxonomy of DEA-based hospital efficiency studies. *Socio-Economic Planning Sciences*, 42(3):158–189.

Osborne, D. and Gaebler, T. (1993). *Reinventing government: How the entrepreneurial spirit is transforming the public sector*. New York, NY: Addison-Wesley.

Otley, D. (2014). Management control under uncertainty: Thinking about uncertainty. In Otley, D. and Soin, K., editors, *Management control and uncertainty*, pages 83–96. London: Palgrave Macmillan.

Ouchi, W. G. (1979). A conceptual framework for the design of organizational control mechanisms. *Management Science*, 25(9):833–848.

Päästeamet (2012a). Päästekomandode ümberkorralduste kava (in English: FRS brigades' restructuring plan). Technical report, Tallinn: Päästeamet.

Päästeamet (2012b). Päästekomandode väljasõidupiirkondade ohutegurite ja riskide hindamise metoodika (in English: The methodology for risk assessment of FRS brigades' service areas). Technical report, Tallinn: Päästeamet.

Päästeamet (2014). Strategy of the Estonian Rescue Board 2015–2020. Technical report, Tallinn: Päästeamet.

Pelastustoimi (2017). Pelastuslaitokset. http://www.pelastustoimi.fi/pelastustoimi/pelastuslaitokset. [Accessed: 27.11.2017].

Peng, M., Song, L., Guohui, L., Sen, L., and Heping, Z. (2014). Evaluation of fire protection performance of eight countries based on fire statistics: An application of data envelopment analysis. *Fire Technology*, 50(2):349–361.

Pidd, M. (2012). *Measuring the performance of public services: Principles and practice*. Cambridge: Cambridge University Press.

Poister, T. H. (2008). *Measuring performance in public and nonprofit organizations*. San Francisco, CA: John Wiley & Sons.

Põldaru, R. and Roots, J. (2014). A PCA–DEA approach to measure the quality of life in Estonian counties. *Socio-Economic Planning Sciences*, 48(1):65–73.

Pollitt, C. (2003). Joined-up government: A survey. *Political Studies Review*, 1(1):34–49.

Pollitt, C. (2006). Performance management in practice: A comparative study of executive agencies. *Journal of Public Administration Research and Theory*, 16(1):25–44.

Pollitt, C. (2013). The logics of performance management. *Evaluation*, 19(4):346–363.

Pollitt, C. and Bouckaert, G. (2011). *Public Management Reform: A comparative analysis-new public management, governance, and the Neo-Weberian state*. Oxford: Oxford University Press.

Porcelli, F. (2009). Measurement of technical efficiency. a brief survey on parametric and non-parametric techniques. University of Warwick, 11.

Powell, A. A. and Gruen, F. (1968). The constant elasticity of transformation production frontier and linear supply system. *International Economic Review*, 9(3):315–328.

Power, M. (1997). *The audit society: Rituals of verification*. Oxford: Oxford University Press.

PRONTO (2016). Statistics system of Finnish rescue services (PRONTO). A statistics system developed and maintained by the Emergency Services College in Kuopio, Finland [https://prontonet.fi/].

Propper, C. and Wilson, D. (2003). The use and usefulness of performance measures in the public sector. *Oxford Review of Economic Policy*, 19(2):250–267.

Puolokainen, T. (2017). Reforming fire and rescue services: A comparative study of Estonia and

Georgia. *International Journal of Public Sector Management*, 30(3):227–240.

Puolokainen, T., Jaansoo, A., and Klaos, M. (Under review, 2018). Providing rescue services in remote areas of Estonia. *International Journal of Emergency Services*.

Putnam, R. D., Leonardi, R., and Nanetti, R. Y. (1994). *Making democracy work: Civic traditions in modern Italy*. Princeton, NJ: Princeton University Press.

Quiggin, J. (2012). *Generalized expected utility theory: The rank-dependent model*. Amsterdam: Springer Science & Business Media.

Randoja, P. and Käerdi, H. (2010). Tulesurmade aegridade analüüs (in English: The time-series analysis of fire deaths). *Sisekaitseakadeemia toimetised*, 9:85–107.

Randoja, P. and Käerdi, H. (2011). Tulekahjudes hukkunute ajaline, piirkondlik ja sündmuskohaga seonduv statistiline analüüs (in English: The statistical analysis of fire deaths with respect to time, area and scene). *Sisekaitseakadeemia toimetised*, 10:151–177.

Rasmussen, S. (2003). Criteria for optimal production under uncertainty: The state-contingent approach. *Australian Journal of Agricultural and Resource Economics*, 47(4):447–476.

Rasmussen, S. (2006). Optimizing production under uncertainty generalization of the state-contingent approach and com-parison with the EV model. IFRO Working paper 5, Department of Food and Resource Economics (IFRO), University of Copenhagen.

Rasmussen, S. (2013). *Production Economics: The Basic Theory of Production Optimisation*. Springer Texts in Business and Economics. Berlin, Heidelberg: Springer-Verlag, 2 edition.

Riewpaiboon, A., Malaroje, S., and Kongsawatt, S. (2007). Effect of costing methods on unit cost of hospital medical services. *Tropical Medicine and International Health*, 12(4):554–563.

Riigikogu (2010). Rescue act (päästeseadus). *Riigi Teataja*, 24(115).

Rodriguez, C., Garaix, T., Xie, X., and Augusto, V. (2015). Staff dimensioning in homecare services with uncertain demands. *International Journal of Production Research*, 53(24):7396–7410.

Rodríguez-Álvarez, A., Tovar, B., and Wall, A. (2011). The effect of demand uncertainty on port terminal costs. *Journal of Transport Economics and Policy (JTEP)*, 45(2):303–328.

Ross, J. P. and Burkhead, J. (1974). *Productivity in the local government sector*. Lexington, MA: Lexington Books.

Salgado, M., Marchione, E., and Gilbert, N. (2014). Analysing differential school effectiveness through multilevel and agent-based modelling. *Journal of Artificial Societies and Social Simulation*, 17(4):3.

Sánchez, I. M. G. (2006). Estimation of the effect of environmental conditions on technical efficiency: The case of fire services. *Revue d'Économie Régionale & Urbaine*, 4:597–614.

Schacter, M. (1999). *Means... ends... indicators: performance measurement in the public sector*. Ottawa: Institute on Governance.

Schaenman, P. S. (1977). *Procedures for improving the measurement of local fire protection effectiveness*. Washington, D.C.: Urban Institute Press.

Schaenman, P. S. and Swartz, J. (1974). *Measuring fire protection productivity in local government*. Quincy, MA: National Fire Protection Association.

Schmidt, P. (1985). Frontier production functions. *Econometric Reviews*, 4(2):289–328.

Schubert, T. (2009). Empirical observations on new public management to increase efficiency in public researchboon or bane? *Research Policy*, 38(8):1225–1234.

Shankar, S. (2015). Efficiency analysis under uncertainty: A simulation study. *Australian Journal of Agricultural and Resource Economics*, 59(2):171–188.

Shankar, S. and Quiggin, J. (2013). Production under uncertainty: A simulation study. *Journal of Productivity Analysis*, 39(3):207–215.

Simar, L. and Wilson, P. W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science*, 44(1):49–61.

Simar, L. and Wilson, P. W. (2013). Estimation and inference in nonparametric frontier models: Recent developments and perspectives. *Foundations and Trends® in Econometrics*, 5(3–4):183–337.

Simar, L. and Zelenyuk, V. (2011). Stochastic FDH/DEA estimators for frontier analysis. *Journal* *of Productivity Analysis*, 36(1):1–20.

Simon, H. A., Shepard, R. W., and Sharp, F. W. (1943). *Fire losses and fire risks*. Berkeley, CA: Bureau of Public Administration, University of California.

Sisäministeriö (2016). A safe and resilient Finland – rescue services strategy 2025. Technical report, Helsinki: Sisäministeriö.

Siseministeerium (2013). Riiklikud suunad vabatahtliku pääste arengus 2013–2016 (in English: The National Orientations in the Development of the Voluntary Rescue 2013–2016). Technical report, Tallinn: Siseministeerium.

Siseministeerium (2015a). Siseministeeriumi valitsemisala arengukava 2015-2018 (in English: The Development Plan for the Area of Government of the Ministry of Interior Affairs 2015–2018). Technical report, Tallinn: Siseministeerium.

Siseministeerium (2015b). Siseturvalisuse arengukava 2015–2020 (in English: The Development Plan for Internal Security 2015–2020). Technical report, Tallinn: Siseministeerium.

Smith, P. (1990). The use of performance indicators in the public sector. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 153:53–72.

Smith, P. C. (2006). Quantitative approaches towards assessing organizational performance. In Boyne, G. A., Meier, K. J., OToole Jr, L. J., and Walker, R. M., editors, *Public service performance: perspectives on measurement and management*, pages 75–91. Cambridge: Cambridge University Press.

Smith, P. C. and Goddard, M. (2002). Performance management and operational research: a marriage made in heaven? *Journal of the Operational Research Society*, 53(3):247–255.

Smith, P. C. and Street, A. (2005). Measuring the efficiency of public services: the limits of analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(2):401–417.

Southwick Jr, L. and Butler, R. J. (1985). Fire department demand and supply in large cities. *Applied Economics*, 17(6):1043–1064.

Steers, R. M. (1975). Problems in the measurement of organizational effectiveness. *Administrative Science Quarterly*, 20(4):546–558.

Sund, B., Svensson, L., Rosenqvist, M., and Hollenberg, J. (2012). Favourable cost-benefit in an early defibrillation programme using dual dispatch of ambulance and fire services in out-of-hospital cardiac arrest. *The European Journal of Health Economics*, 13(6):811–818.

Thanassoulis, E. (1993). A comparison of regression analysis and data envelopment analysis as alternative methods for performance assessments. *Journal of the Operational Research Society*, 44(11):1129–1144.

Thanassoulis, E. (2001). *Introduction to the theory and application of data envelopment analysis*. Boston, MA: Springer.

Thompson, J. D. (1967). *Organizations in action: Social science bases of administrative theory*. New Brunswick, NJ: Transaction publishers.

Thornhill, D. (2006). Productivity attainment in a diverse public sector. Paper presented at the Institute of Public Administration Seminar on Promoting Productivity in a Diverse Public Sector, Dublin, 21 April 2006f.

Tien, C. (2003). Seemingly unrelated regression. In Lewis-Beck, M., Bryman, A. E., and Liao, T. F., editors, *The Sage encyclopedia of social science research methods*, page 1010. Thousand Oaks, CA: Sage Publications.

Triantaphyllou, E. (2000). *Multi-criteria Decision Making Methods: A Comparative Study*. Boston, MA: Springer.

Turu-uuringute AS (2014). Eesti elanikud siseturvalisust toetavast vabatahtlikust tegevusest: teadlikkus ja usaldus, 2013/2014 (in English: Estonian population about the voluntary activities towards internal security: The knowledge and trust, 2013/2014). Technical report, Tallinn: Turu-uuringute AS.

Van de Walle, S. (2008). Comparing the performance of national public sectors: Conceptual problems. *International Journal of Productivity and Performance Management*, 57(4):329–338.

Van Dooren, W., Bouckaert, G., and Halligan, J. (2015). *Performance management in the public* *sector*. London: Routledge.

Van Thiel, S. and Leeuw, F. L. (2002). The performance paradox in the public sector. *Public Performance and Management Review*, 25(3):267–281.

Walker, W. E., Haasnoot, M., and Kwakkel, J. H. (2013). Adapt or perish: A review of planning approaches for adaptation under deep uncertainty. *Sustainability*, 5(3):955–979.

Weinholt, Å. (2015). *Exploring Collaboration Between the Fire and Rescue Service and New Actors: Cost-efficiency and Adaptation*. PhD thesis, Linköping University.

Woods, M. and Grubnic, S. (2008). Linking comprehensive performance assessment to the balanced scorecard: Evidence from Hertfordshire County Council. *Financial Accountability & Management*, 24(3):343–361.

Zeckhauser, R. J. and Pratt, J. W. (1985). *Principals and agents: The structure of business*. Boston, MA: Harvard Business School Press.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American statistical Association*, 57(298):348–368.

APPENDICES

Appendix 1: Reasons to measure performance according to different authors

Author	**Pidd (2012)**	**Bird *et al.* (2005)**	**Behn (2003)**	**Poister (2008)**	**Van Dooren *et al.* (2015)**
1	Planning and improvement	To see what works	To learn, improve	Quality and process management	To learn
2	Monitoring and control		To control, motivate	Monitoring and reporting, programme management	To steer and control
3	Evaluation and comparison	To identify competences	To evaluate	Programme evaluation, contract management, external benchmarking	
4	Accountability	Public accountability	To promote	Public communications	To give account
5	Financial budgeting and planning		To budget	Strategic planning, budgeting	
6	Individual performance management		To celebrate	Performance management	

Source: Pidd (2012: 31); Author's additions.

Appendix 2: A selection of performance indicators used in fire and rescue services' planning process

Type	Indicator	Remarks	Source
Environment	% of calls from individuals (stratify by call type – fire, medical, other, etc.)	The source of the incident call can affect the total response time from fire start to department response. This measure can be used in concert with a measure of average response time. This will help identify if the call source is or is not impacting the fire spread outcome. Alternatively, departments could analyse response time separately by source of call. This will relate combined response time to percentage of calls by source.	Flynn (2009)
Environment	% of calls from alarm monitoring companies (stratify by call type – fire, medical, other, etc.)		Flynn (2009)
Environment	% of calls from some other source (stratify by call type – fire, medical, other, etc.)		Flynn (2009)
Environment	Number of calls by local government area	Metropolitan Fire and Emergency Services Board	Kloot (2009)
Environment	Days of total fire bans (fires cannot be lit in open air)	Country Fire Authority	Kloot (2009)
Input/Environment	Calls by type (fires, hazardous conditions, rescues, service calls, false alarms)	Metropolitan Fire and Emergency Services Board	Kloot (2009)
Outcome/Environment	False alarms, breakdown by reason (good intent, due to apparatus, malicious)	ONS measures in the UK in 2006	ODPM (2005)
Input/Outcome	Number of volunteer and professional firefighters	Country Fire Authority	Kloot (2009)

Appendix 2 Continued:

Input	Number of qualified personnel in various categories	Country Fire Authority	Kloot (2009)
Input	Expenditure per 1,000 people	Commission	Kloot (2009)
Process	Turnout time	Shorter response times can usually be expected in urban areas compared to rural areas, due to the size of the community, number of facilities, such as fire stations, population and resources within the community. This needs to be taken into consideration when comparing response time performance across communities.	Flynn (2009)
Process	Travel time		Flynn (2009)
Process	Total response time		Flynn (2009)
Process	Time of control of fire* *This is time from arrival to fire control, not the point in time where control occurs		Flynn (2009)
Process	Time spent on scene		Flynn (2009)
Process	Average response time per fire incident call		Flynn (2009)
Process	% of times that are less than x minutes		Flynn (2009)
Process	% of times that are more than x minutes		Flynn (2009)

Appendix 2 Continued:

Process	Average time to control spread or confirm spread has stopped, by size on arrival and type of occupancy		Flynn (2009)
Process	% of firefighters with completed, uptodate training	This measure is a proxy for quality of service provided. It is assumed that a high percentage of responders with completed training and certification are providing high quality service when responding to calls. The measures should be stratified by response type and certification or training required by responder for that response.	Flynn (2009)
Process	% of firefighters that are certified		Flynn (2009)
Process	% of responses to fire calls which met the national target for speed of attendance and number of appliances, for areas in four risk categories – high, substantial, moderate and low	Audit Scotland (2005)	Carvalho *et al.* (2006)
Process	For fire calls to rural areas, the % of responses which met the local target, and the average time for attendance	Audit Scotland (2005)	Carvalho *et al.* (2006)
Process	% of rider shifts lost due to sickness absence and light duties	Audit Scotland (2005)	Carvalho *et al.* (2006)
Process	% of working time lost to sickness for all other staff directly employed by the brigade	Audit Scotland (2005)	Carvalho *et al.* (2006)

Appendix 2 Continued:

Process	Average time taken to handle calls to incidents	Audit Scotland (2005)	Carvalho *et al.* (2006)
Process	% handled within the following time periods: less than one minute; between one and two minutes; and two minutes or more	Audit Scotland (2005)	Carvalho *et al.* (2006)
Process	Response times in minutes (90th percentile)	Metropolitan Fire and Emergency Services Board	Kloot (2009)
Process	Response times to structure fires (50th and 90th percentiles)	Commission	Kloot (2009)
Process/Outcome	Service delivery standards (unspecified)	Country Fire Authority	Kloot (2009)
Output	Number of reported fires per 1,000 population	Fire prevention is an important function of fire departments. Ideally, the department would want to measure how many fires have been prevented, but that is not possible. Instead, fire departments can use fire rate measures of reported fires by population or by buildings.	Flynn (2009)
Output	Number of reported fires per 1,000 buildings (by occupancy)		Flynn (2009)
Output	Average number of accidental dwelling fire per 10,000 population over the five-year period finishing at the end of the reporting year	Audit Scotland (2005)	Carvalho *et al.* (2006)
Output	Total incidents by category (fires, rescues, hazardous conditions, service calls, false alarms)	Country Fire Authority	Kloot (2009)

Appendix 2 Continued:

Output	Number of fires	DSE	Kloot (2009)
Output	Number of turnouts to fire and other incidents	Metropolitan Fire and Emergency Services Board/Country Fire Authority	Kloot (2009)
Output	Fuel reduction	DSE	Kloot (2009)
Output	Total incidents by category	Commission	Kloot (2009)
Output	Total fire incidents	Commission	Kloot (2009)
Output	Total hours spent attending fires, by type (e.g. primary, secondary, chimney, false alarms).	ONS measures in UK 2006	ODPM (2005)
Output	Total hours spent on fire prevention inspections (full inspections, re-inspections, examination of plans)	ONS measures in UK 2006	ODPM (2005)
Output	Total hours spent on community safety work, consisting mainly of giving talks to voluntary groups and advice to individuals	ONS measures in UK 2006	ODPM (2005)
Output	Total hours spent attending road accidents (extricating people from wreckage, clearing chemical spills)	ONS measures in UK 2006	ODPM (2005)
Output	Total hours spent attending other special services (such as rescuing trapped people and clearing up floods)	ONS measures in UK 2006	ODPM (2005)

Appendix 2 Continued:

Output	Responding to fires: a weighted measure of response to a mix of fire incidents, quality adjusted using response times.	Proposed measures; By weighting each type of incident by the value of lives or property that could be affected by fire, multiplied by the likelihood of a fire occurring	ODPM (2005)
Output	Special Service Incident response: an activity index, representing the volume of non-fire incidents attended	Proposed measures	ODPM (2005)
Outcome	% of fires extinguished before department arrival	A high percentage of fires extinguished before departments arrive might show successful campaigns for fire detection and/or fire sprinkler installations. By removing fires that were extinguished before arrival, the fire department is able to actually measure the effect of their actions taken. Accelerants or the presence of hazardous materials can contribute to fire spread and these factors should be taken into consideration when possible.	Flynn (2009)
Output	% of fires not extinguished before department arrival		Flynn (2009)
Outcome	% of fires responded to that spread beyond room of origin before fire department arrival		Flynn (2009)
Output	% of total fires responded to that spread beyond room of origin after fire department arrival		Flynn (2009)

Appendix 2 Continued:

Output/Outcome	Number of firefighter fatalities (injuries) per 1,000 fire fighters	These measures identify the relative risk that departments face. On scene firefighter injuries and deaths are bad on their own, but there is also a potential that suppression is affected. On scene firefighter injuries and deaths impact the suppression of the fire, fire department expenditures, medical bills and workers' compensation, and staffing, all of which ultimately impact fire department performance. As with fire spread, it may make sense to distinguish between all injuries, and injuries requiring treatment or resulting in lost time. The presence of EMS personnel on-site can impact the severity of an injury.	Flynn (2009)
Output/Outcome	Number of firefighter fatalities (injuries) per 1,000 fires		Flynn (2009)
Output/Outcome	Average number of fire casualties per 10,000 population over the five-year period finishing at the end of the reporting year	Audit Scotland (2005)	Carvalho *et al.* (2006)
Output/Outcome	Human resources work injury claims	Metropolitan Fire and Emergency Services Board	Kloot (2009)
Output/Outcome	Number of firefighter injuries and deaths per 100 fires		NFPA (1977)
Outcome	Number of civilian deaths (or injuries) per 100,000 population	Fire rescue and public education programmes affect these measures and can be evaluated by them. Death rates can be problematic for analysis because most communities and even many states do not have enough fire deaths to compensate for generic fluctuations.	Flynn (2009)
Outcome	Number of civilian deaths (injuries) per 1,000 fires		Flynn (2009)

Appendix 2 Continued:

Outcome	% of fires in which a person or people were rescued from the building by firefighters	Measures of ”saves” or rescues show what portion of the community’s fire incident calls required rescue. There is obvious risk to firefighters entering buildings to remove victims; this measure tries to identify that risk. It also helps to identify the demand on the department for this service. This measure can be a very small number compared to the total number of fires responded to and would be less reliable and a low indicator of performance in communities where there are few rescues made.	Flynn (2009)
Outcome	Number of saves versus number of casualties		Flynn (2009)
Outcome	Rate of saves per incident involving at least one save		Flynn (2009)
Outcome	Total $ saved, in terms of structure and contents	In order to be reliable and valid, buildings and dollars saved must be looked at realistically. For example, if the fire department is alerted to a confined cooking fire in a very large building with expensive contents, or an alarm activation with nothing found, it is not realistic to say that the firefighters saved millions of dollars’ worth of property from being destroyed. These measures can be determined by using dollar estimates collected in NFIRS, but remember to be realistic in estimates and discuss this issue when using these measures. Averages can be significantly raised by the inclusion of a single major loss.	Flynn (2009)
Outcome	Average $ saved per fire		Flynn (2009)

Appendix 2 Continued:

Outcome	% of fires in which $ damage to the building was greater than x amount of dollars		Flynn (2009)
Outcome	% of fires in which $ saved was greater than x amount of dollars		Flynn (2009)
Outcome	Fire containment to room of origin	Metropolitan Fire and Emergency Services Board/Commission (State of Victoria, Australia)	Kloot (2009)
Outcome	Preventable fire fatalities	Metropolitan Fire and Emergency Services Board (State of Victoria, Australia)	Kloot (2009)
Outcome	Community fireguard meetings and fireguard groups	Country Fire Authority (State of Victoria, Australia)	Kloot (2009)
Outcome	Hectares burned due to fire	DSE (State of Victoria, Australia)	Kloot (2009)
Outcome	Fire death rate per million people	Commission (State of Victoria, Australia)	Kloot (2009)
Outcome	Fire injury rate per 100,000 people	Commission (State of Victoria, Australia)	Kloot (2009)
Outcome	% of households with an operational smoke alarm	Commission (State of Victoria, Australia)	Kloot (2009)
Outcome	Property loss from structure fire, $ per person	Commission (State of Victoria, Australia)	Kloot (2009)
Outcome	% of households with a fire safety measure installed	Commission (State of Victoria, Australia)	Kloot (2009)
Outcome	Inspection of property for hazards, and the level of compliance with standards and building codes	Commission (State of Victoria, Australia)	Kloot (2009)

Appendix 2 Continued:

Outcome	% of commercial structures with sprinklers	Commission (State of Victoria, Australia)	Kloot (2009)
Outcome	Total $ of property loss due to fire per year		NFPA (1977)
Outcome	$ of property loss per $1,000 of market value of the property		NFPA (1977)
Outcome	$ of property loss per fire		NFPA (1977)
Outcome	Number of civilian injuries and deaths per 100,000 population protected		NFPA (1977)
Outcome	Number of civilian injuries and deaths per 100 fires		NFPA (1977)
Outcome	Fire safety and prevention: a basket of outcome-based measures, using a set of weights to derive a composite measure	Proposed measures	ODPM (2005)

Source: Author's compilation.

Appendix 3: Performance indicators used in Estonian fire and rescue services' planning process

Type	Indicator	Past value	Target value	Source
Output	Number of unnatural deaths* *This includes traffic accidents, fire deaths, overdoses, homicides, drownings, suicides and work accidents. Most important for the ERB are fire deaths and drownings.	508 (2014)	385 (2020)	Development plan for internal security 2015–2020
Output	Number of fire deaths	54 (2014)	38 (2020)	Development plan for internal security 2015–2020
Output	Number of drownings	68 (2014)	48 (2020)	Development plan for internal security 2015–2020
Output	Number of building fires	1833 (2014)	1525 (2020)	Development plan for internal security 2015–2020
Outcome	Trustworthiness rating towards the volunteer rescuers (opinion poll)			Development plan for internal security 2015–2020
Outcome	Trustworthiness rating towards the ERB (opinion poll)* *The poll is ordered by the Ministry of Defence. ERB was the most reliable public agency in 2014.	95.5% (2014)	>95% (2020)	Development plan for internal security 2015–2020
Input	Costs* *The costs should be decreasing since the ERB is publicly funded, the budget strategy dictates the plan.	77.14 million euros (2015)	60.93 million euros (2020)	Development plan for internal security 2015–2020
Output	Number of fire deaths	47 (2013)	<12 (2025)	Strategy of the Estonian Rescue Board 2015–2025
Output	Number of water-related accidents	56 (2013)	<20 (2025)	Strategy of the Estonian Rescue Board 2015–2025

Appendix 3 Continued:

Output	Number of detonations of explosives	0 (2013)	0 (2025)	Strategy of the Estonian Rescue Board 2015–2025
Output	Number of chemical accidents	0 (2013)	0 (2025)	Strategy of the Estonian Rescue Board 2015–2025
Output	Number of rescue events	20,195 (2013)	<15,500 (2025)	Strategy of the Estonian Rescue Board 2015–2025
Output	Number of fires in buildings	1,627 (2013)	<1,300 (2025)	Strategy of the Estonian Rescue Board 2015–2025
Output	Number of fires in residential buildings	892 (2013)	<700 (2025)	Strategy of the Estonian Rescue Board 2015–2025
Output	Number of environmental accidents and the amount of environmental damage. Number of forest and wildfires	1,275 (2013)	<500 (2025)	Strategy of the Estonian Rescue Board 2015–2025
Output/ Outcome	Property damage caused by building fires	13,575,000 (2013)	<10,000,000 (2025)	Strategy of the Estonian Rescue Board 2015–2025
Outcome	Trustworthiness rating towards the ERB (opinion poll)	95.5% (2014)	>95% (2025)	Strategy of the Estonian Rescue Board 2015–2025
Process	Number of partners involved in prevention efforts	69 (2013)	>100 (2025)	Strategy of the Estonian Rescue Board 2015–2025
Outcome	Fire safety awareness index	60 (2013)	>70 (2025)	Strategy of the Estonian Rescue Board 2015–2025
Outcome	Water safety awareness index	66 (2013)	>75 (2025)	Strategy of the Estonian Rescue Board 2015–2025
Process	Number of working hours of on-call staff spent on prevention efforts	0.03% (2013)	>1% (2025)	Strategy of the Estonian Rescue Board 2015–2025
Process	Average response time to a rescue event	9:40 (2013)	<10:00 (2025)	Strategy of the Estonian Rescue Board 2015–2025

Appendix 3 Continued:

Process	Response time in reaching the victim in case of a rescue event		Reduced	Strategy of the Estonian Rescue Board 2015–2025
Process/ Outcome	Commitment level of the staff of the Rescue Board		Increased	Strategy of the Estonian Rescue Board 2015–2025
Process	Voluntary turnover of Rescue Board employees (On-call staff)	3% (2013)	<3% (2025)	Strategy of the Estonian Rescue Board 2015–2025
Process	Voluntary turnover of Rescue Board employees (Staff on 8-hour shifts)	7.75% (2013)	<7% (2025)	Strategy of the Estonian Rescue Board 2015–2025
Process/ Outcome	Number of members in volunteer rescue organisations	1,868 (2013)	>4,000 (2025)	Strategy of the Estonian Rescue Board 2015–2025
Process/ Outcome	Number of certified volunteer rescuers	1,281 (2013)	>3,000 (2025)	Strategy of the Estonian Rescue Board 2015–2025
Process/ Outcome	Experience and competence level of Rescue Board staff and volunteer rescuers		Increased	Strategy of the Estonian Rescue Board 2015–2025
Input	Financing the volunteers to ensure the sustainability		Increased	Strategy of the Estonian Rescue Board 2015–2025
Input	Proportion of investments in the budget		Increased	Strategy of the Estonian Rescue Board 2015–2025
Input	A fair and competitive salary			Strategy of the Estonian Rescue Board 2015–2025
Output	Number of unnatural deaths (fire deaths, drownings, traffic accidents: 3-year average)			The development plan for the Ministry of Internal Affairs 2015–2018
Process/ Outcome	Number of voluntary rescuers/certified voluntary rescuers/volunteers			The development plan for the Ministry of Internal Affairs 2015–2018

Appendix 3 Continued:

Output	Number of fires in buildings/residential houses/caused by smoking			The development plan for the Ministry of Internal Affairs 2015–2018
Output	Percentage of public school teachers, who have been trained in fire safety to provide prevention (10%)			The development plan for the Ministry of Internal Affairs 2015–2018
Process	Average time of the first responder to arrive to the emergency site (all emergencies)/responder with life saving capability (building fires, traffic accidents, water accidents)			The development plan for the Ministry of Internal Affairs 2015–2018
Process	Number of areas to localise and liquidate the big wild fires simultaneously (in hectares)			The development plan for the Ministry of Internal Affairs 2015–2018
Input	Number of new fleet (40 main trucks, 40 tank trucks, 6 ladder trucks)			The development plan for the Ministry of Internal Affairs 2015–2018
Input	Introducing a new national FRS brigade in Lasnamäe, Tallinn			The development plan for the Ministry of Internal Affairs 2015–2018
Process	Number of voluntary FRS brigades eligible to perform fire extinguishing	95 (2012)	110 (2016)	National orientations in the development of the voluntary rescue 2013–2016
Process	Number of voluntary FRS brigades in the rural areas (further than 14 minutes from national FRS brigades)	58 (2012)	70 (2016)	National orientations in the development of the voluntary rescue 2013–2016

Appendix 3 Continued:

Process	Number of voluntary FRS brigades in the rural areas with more than 12 members (further than 14 minutes from national FRS brigades)	8 (2012)	36 (2016)	National orientations in the development of the voluntary rescue 2013–2016
Process	Number of voluntary rescuers and number of voluntary rescuers eligible to perform fire extinguishing	640: 535 (2012)	1,600: 1,220 (2016)	National orientations in the development of the voluntary rescue 2013–2016
Process	Number of all volunteers	1,238 (2012)	2,000 (2016)	National orientations in the development of the voluntary rescue 2013–2016
Process	Number of all volunteers under age of 35	175 (2012)	400 (2016)	National orientations in the development of the voluntary rescue 2013–2016
Process	Number of youth rescue societies led by volunteers	37 (2012)	41 (2016)	National orientations in the development of the voluntary rescue 2013–2016
Process	Number of volunteers doing prevention tasks	195 (2012)	254 (2016)	National orientations in the development of the voluntary rescue 2013–2016
Process	Number of voluntary FRS brigades doing prevention tasks	41 (2012)	47 (2016)	National orientations in the development of the voluntary rescue 2013–2016
Process	Number of voluntary FRS brigades which have joined the representative organisation	59 (2012)	100 (2016)	National orientations in the development of the voluntary rescue 2013–2016

Appendix 3 Continued:

Input	Total budget for voluntary rescue	455,472 euros (2012)	1.5 million euros (2016)	National orientations in the development of the voluntary rescue 2013–2016
Input	Percentage of local governments supporting the voluntary rescue	36% (2012)	50% (2016)	National orientations in the development of the voluntary rescue 2013–2016
Input	Percentage of voluntary FRS brigades supported by private sector	? (2012)	100% (2016)	National orientations in the development of the voluntary rescue 2013–2016
Input	Percentage of self-financing in the budget	? (2012)	20% (2016)	National orientations in the development of the voluntary rescue 2013–2016
Input	Percentage of fixed costs covered by the ERB budget for the voluntary FRS brigades in the rural areas	78% (2012)	85% (2016)	National orientations in the development of the voluntary rescue 2013–2016
Input	Number of volunteers with full personal equipment	119 (2012)	800 (2016)	National orientations in the development of the voluntary rescue 2013–2016
Input	Average age of fleet	29 years (2012)	25 years (2016)	National orientations in the development of the voluntary rescue 2013–2016
Input	Percentage of voluntary FRS brigades' buildings fulfilling the necessary requirements (heating, everyday conditions, etc.)	no standard (2012)	30% (2016)	National orientations in the development of the voluntary rescue 2013–2016

Appendix 3 Continued:

Process	Volunteers provided with further training with respect to fire safety or water safety	252: 97 (2012)	328: 126 (2016)	National orientations in the development of the voluntary rescue 2013–2016
Output	Hours of prevention activities by volunteers	2,829 hours (2012)	3,678 hours (2016)	National orientations in the development of the voluntary rescue 2013–2016
Process	Percentage of volunteers who have participated in national FRS brigades trainings at least for 10 hours	0% (2012)	30% (2016)	National orientations in the development of the voluntary rescue 2013–2016
Process	Percentage of national FRS brigades which involve volunteers in the trainings	0% (2012)	50% (2016)	National orientations in the development of the voluntary rescue 2013–2016
Process	Percentage of national FRS brigades which involve volunteers as on-call staff	0% (2012)	30% (2016)	National orientations in the development of the voluntary rescue 2013–2016
Outcome	The knowledge of the public about the possibilities of voluntary rescue	? (2012)	60% of people (2016)	National orientations in the development of the voluntary rescue 2013–2016
Outcome	Trustworthiness of public towards the voluntary rescue	? (2012)	80% (2016)	National orientations in the development of the voluntary rescue 2013–2016

Source: Author's compilation.

Appendix 4: Articles using frontier analysis methods on FRS

Author (Year)	Method	Country (Subject)	Inputs	Outputs	Environmental variables	Main results
Jaldell (2002)	DEA/ SFA	Sweden (288 municipalities)	SFA: Manning level	SFA: Traffic accidents, Total fires, Employees in risky industries, Part-time (=1) / full-time (=0) rescuers	Population, Size of the area, Population density, Number of population centres with at least 1,000 inhabitants	Only the size of the population had an impact on the staffing level. In addition, the results showed that the mean input saving potential is 30% (the mean efficiency score was 0.7) and there has been no improvement in efficiency over time. The results showed that the larger municipalities had higher efficiency scores, and no optimal size of production (comparing the returns to scale for the frontier units) could be found.
			DEA: Total cost of firemen	DEA: Population reached within 5 minutes, within 10 minutes, Total number of firemen reaching the fire within x minutes		

Appendix 4 Continued:

Choi (2005)	DEA	United States (rescue departments in 60 counties in Florida)	Costs: Fire control personnel services, Fire control operating expenses, Fire control capital outlay, Fire control others, Emergency and disaster relief, Ambulance and rescue service	Total fires, Rescue/EMS, Other emergencies, Dollar loss, Injuries and deaths	Population, Density, Manufactures, Property, Tax millage rate	First, the majority of Florida counties (about 40%) appear to have low efficient scores, below 65% of the efficient score. Second, in order to achieve efficient fire and emergency services, overall both Rescue/Emergency Medical Services (76.91%) and Other emergency (21.52%) needs to be investigated for potential improvements. Finally, this DEA method will be helpful for fire and emergency service practitioners to check existing status of their own county's fire and emergency services in a sense that managerial strategies, focussing on both economic loss and injuries and deaths.

Appendix 4 Continued:

Sánchez (2006)	DEA	Spain (29 municipalities with population over 50,000)	Number of staff and vehicles	Fires, Other emergencies, Number of prevention campaigns	Unemployment rate, Level of income, Population with no studies, Area (not statistically significant: Rainfall, Amount of civil disturbance, Number of traffic accidents with victims, % of youth, Temperature, Population)	This study investigated the technical efficiency of the municipal firefighting service, contrasting the effect that the conditions of the environment have on the efficiency indices. To achieve the objectives, the 'basic' and 'complete' (which incorporated as non-controllable inputs those conditions of the environment which, using Tobit analysis, could be considered statistically as variables that explain efficiency) DEA models were designed. The comparison of the two models revealed that the specific environmental conditions (combined) of each town had a strong effect on the performance of the service analysed.

Appendix 4 Continued:

Lan *et al.* (2009)	Stochastic DEA	Taiwan (firefighting branches in Tainan County)	Number of on-duty personnel, On-duty cost, Number of fire engines, Vehicle maintenance fee	Number of fires and emergencies	-	This research proposes a constructive and quantitative fire resource allocation method and further establishes and executive prototype of the new era to pursue higher efficiency.

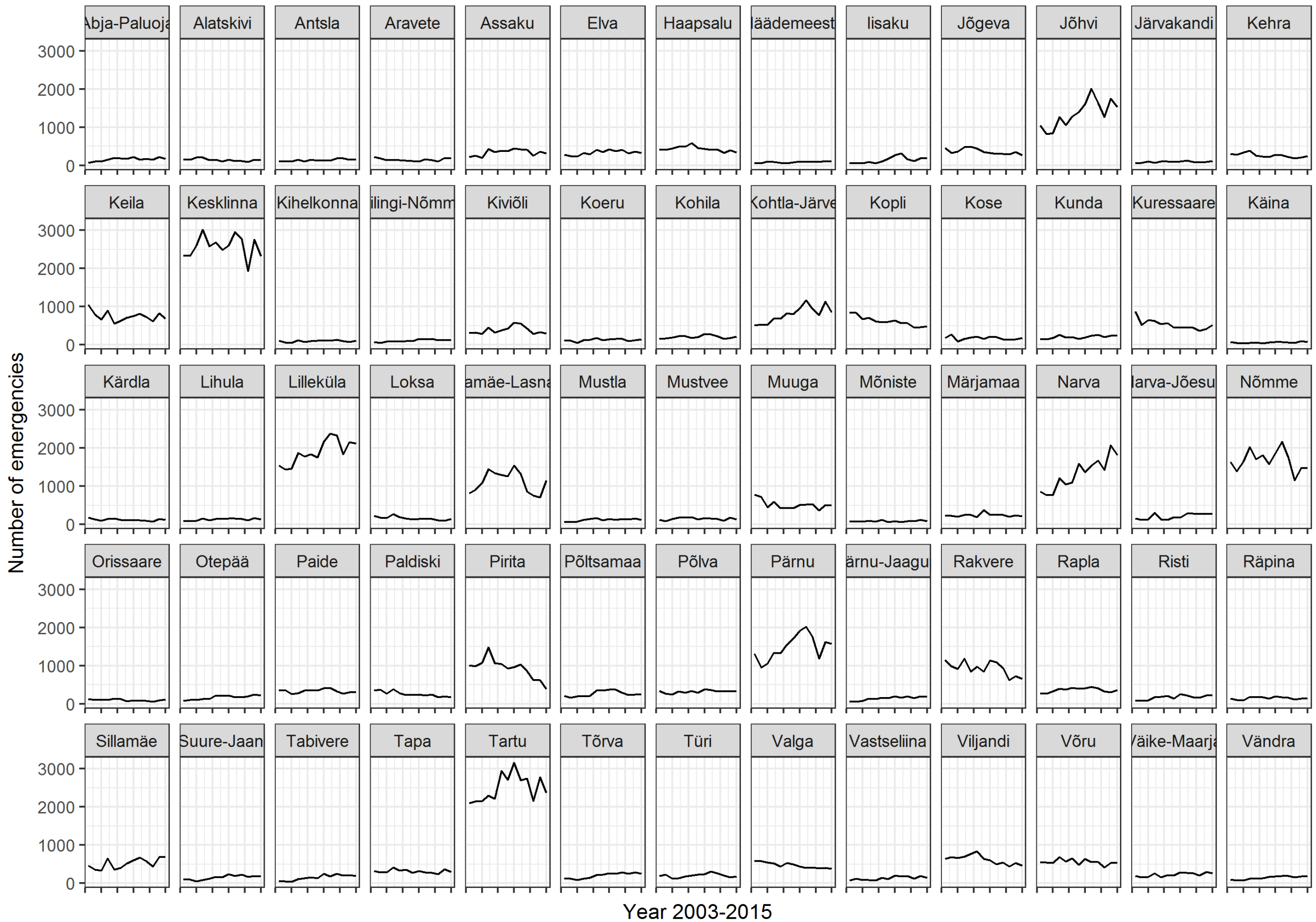
Appendix 4 Continued:

Horton (2011)	DEA	United States (254 municipal fire departments)	Costs	Number of fires and emergencies, fire deaths and losses	Socio-economic: percent of the population below the poverty level, median household income, percent of the population with children living with both parents, percent of female headed households. Demographic: age (% <5, % <17, and, % >65), population density, percent of the population who are African American. Building Stock: percent of property that is owner occupied, percent of housing built prior to 1940, percent of property that is vacant. Climate: the average annual rainfall. Management: Citistat type PMS, Insurance Service Organisation (ISO) rating.	This study evaluated whether using the Citistat PMS in the management of a municipality improves the effectiveness and efficiency of FRS brigades. It was found that there were differences in the organisational behaviour but that these did not result in differences in performance according to the efficiency scores evaluated by the DEA method.

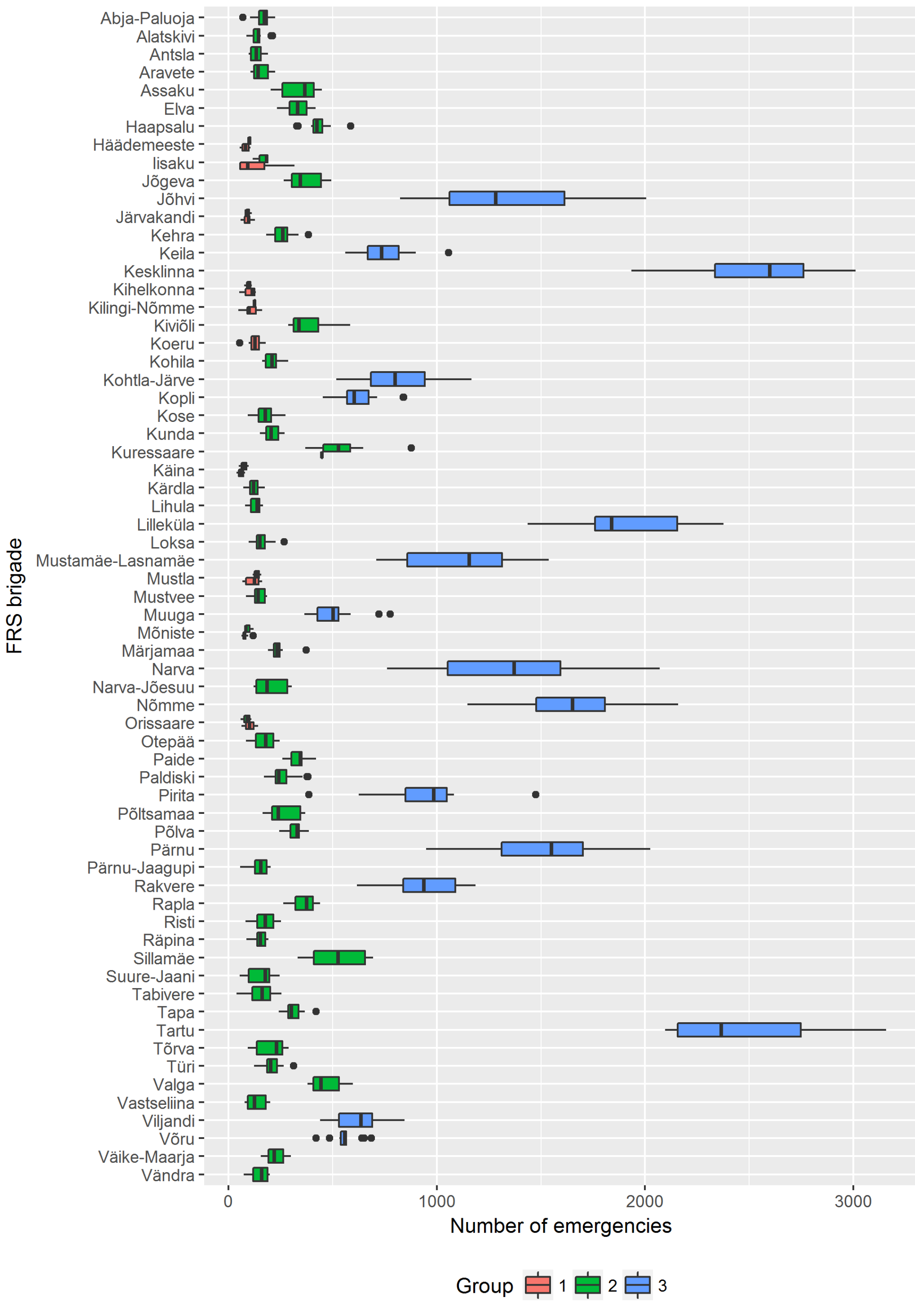
Appendix 4 Continued:

Peng *et al.* (2014)	DEA	International aggregated data (8 countries)	Cost of Firefighting, Organisations, Fire Insurance Administration and Fire Protection for Buildings	Cost of Direct Fire Losses and Fire Deaths	-	The need for international technical exchange and cooperation in fire protection has been identified. In addition, Singapore and Slovenia could be set as good examples for other countries. Denmark, Norway, and the UK need to reallocate resources and refine the operation mode. The United States should reduce its scale. Japan and Sweden need both scale increase and technique adjustment.
Holmgren and Weinholt (2016)	SFA	Sweden (205 out of 290 municipalities)	Wage of firefighters, Cost of capital	Number of turnouts, Number of fires per person, No. of people employed for preventive work, Response time	Population, Income, Assessed values for real estates, Area	The recent policy changes (formalised cooperation between municipalities, cooperation with other actors, first response person, waiting for ambulance, as well as the effect of using part-time firefighters) have not had any effect on the cost-efficiency.

Source: Author's compilation.

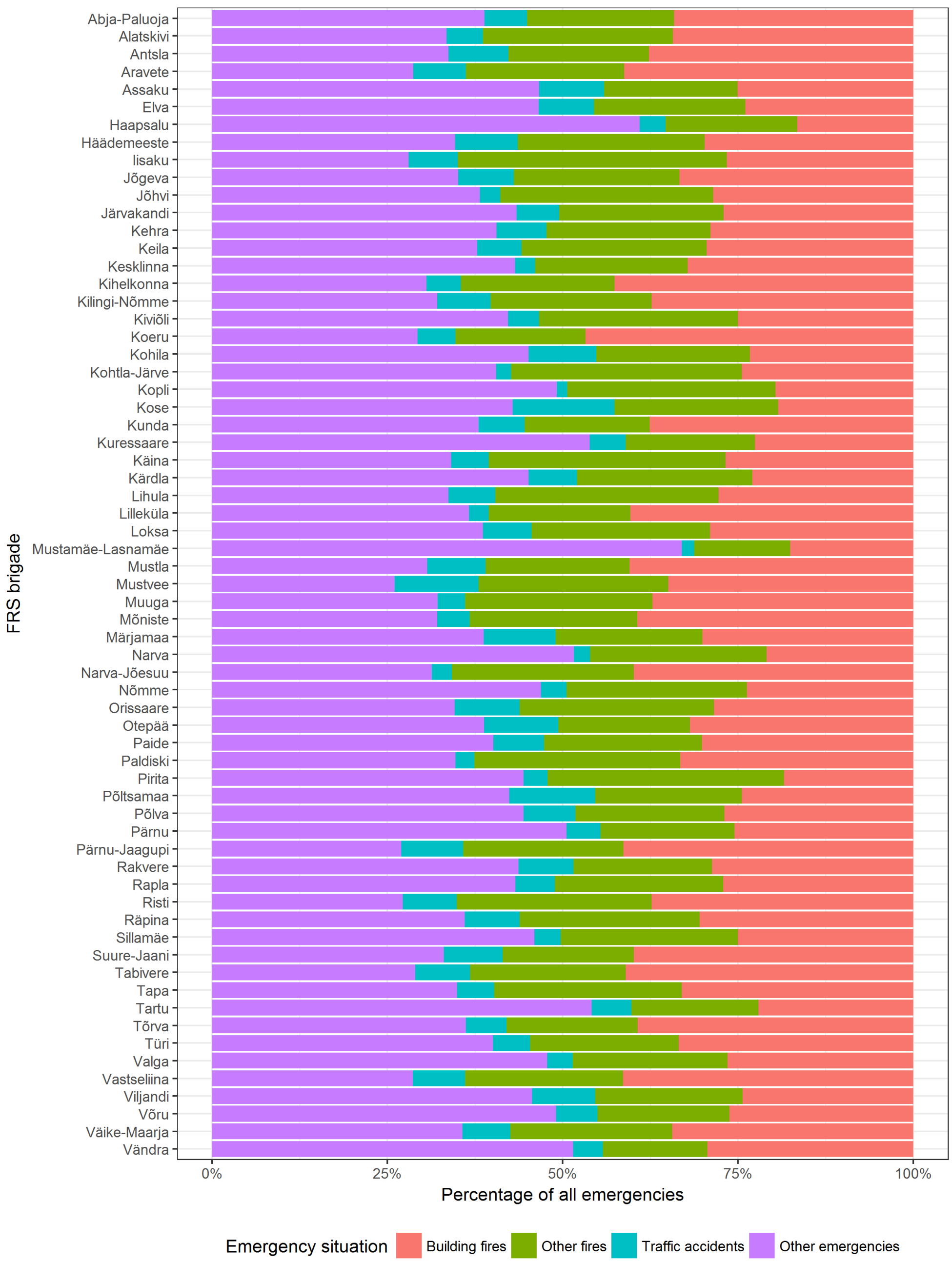


Appendix 5: The number of emergencies attended annually by different Estonian FRS brigades. (Source: Estonian Rescue Board 2016; Author's calculations)

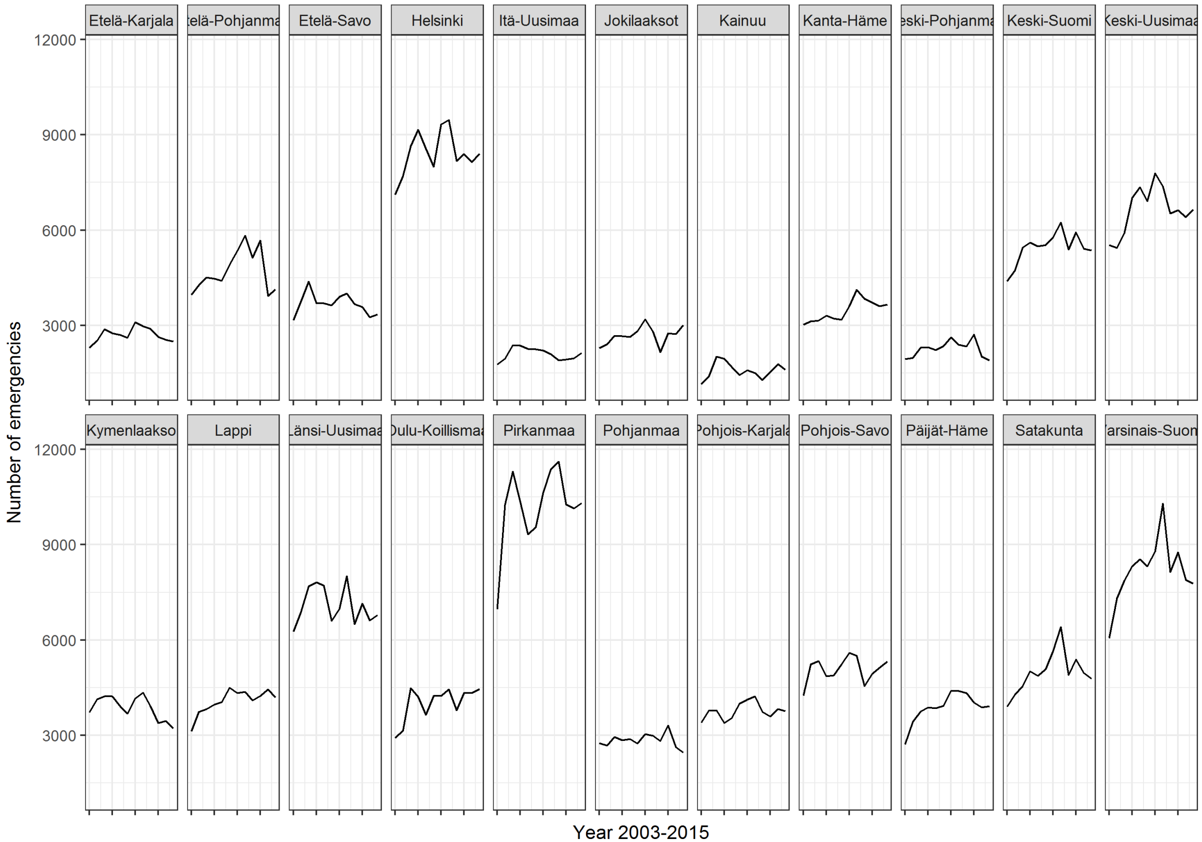


Appendix 6: The variation in number of emergencies attended annually (2004–2015) by different Estonian FRS brigades. (Source: Estonian Rescue Board 2016; Author's calculations)

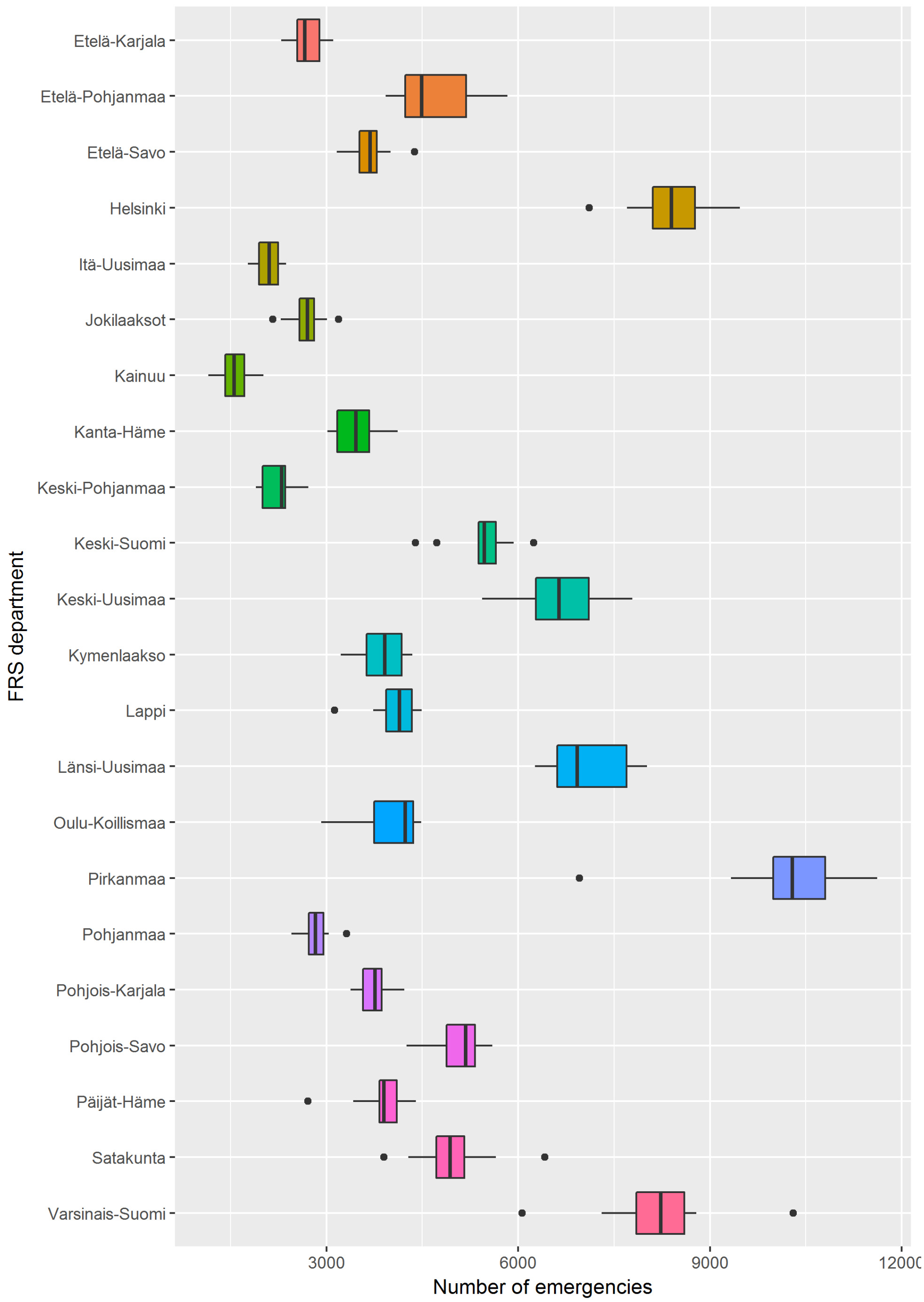
Note: Groups refer to ERB classification – Group 1 = small-sized FRS brigades, Group 2 = middle-sized FRS brigades, Group 3 = large-sized FRS brigades.



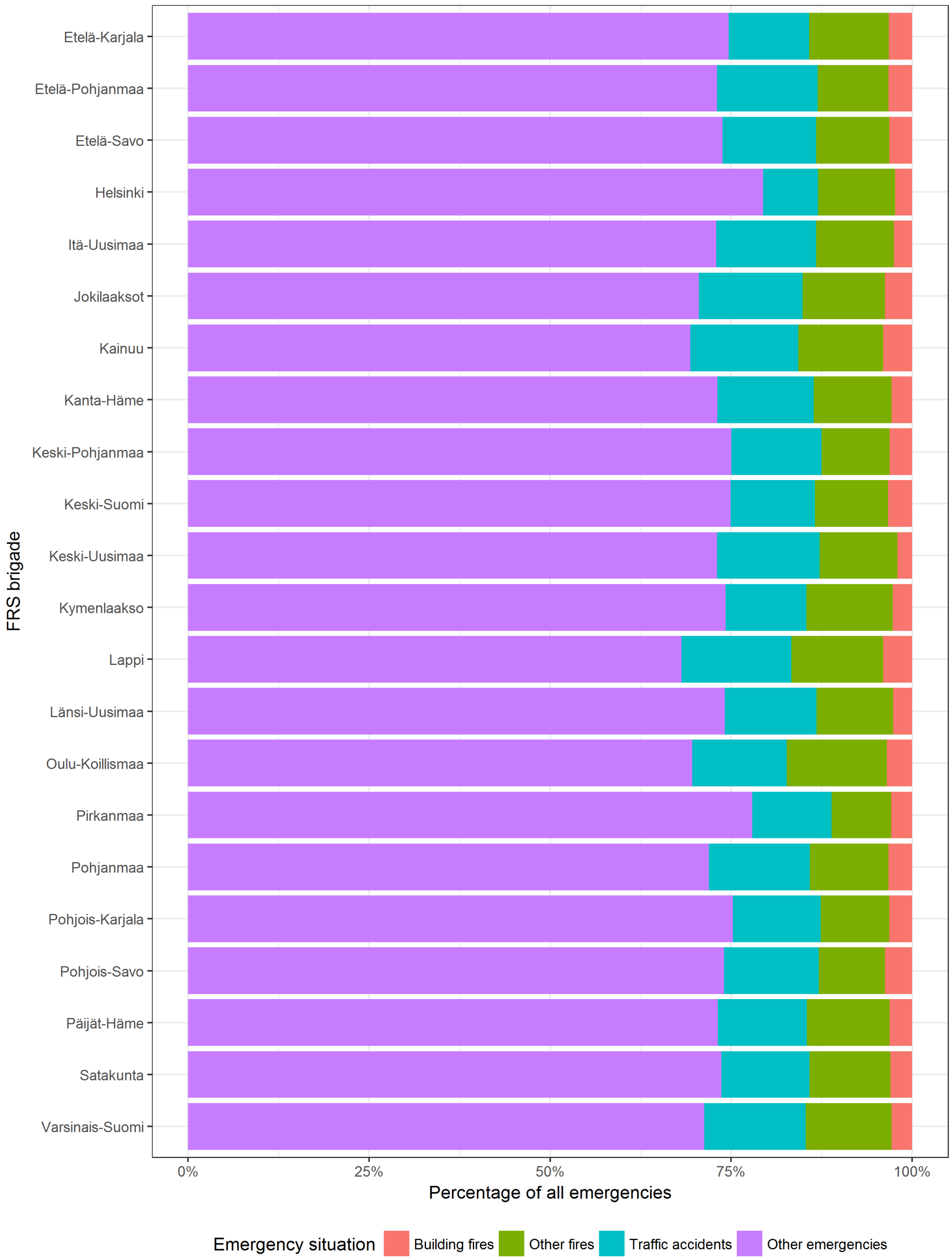
Appendix 7: The variation in shares of different emergencies attended by Estonian FRS brigades. (Source: Estonian Rescue Board 2016; Author's calculations)



Appendix 8: The number of emergencies attended annually by different Finnish fire departments. (Source: PRONTO 2016; Author's calculations)



Appendix 9: The variation in number of emergencies attended annually by different Finnish fire departments. (Source: PRONTO 2016; Author's calculations)



Appendix 10: The variation in shares of different emergencies attended by Finnish fire departments. (Source: PRONTO 2016; Author's calculations)

Appendix 11: Descriptive statistics on the number of departures to emergencies and share of fires in buildings, other fires, traffic accidents and other emergencies in Swedish municipalities

	Municipality	N	Mean	SD	Min	Max	%q1	%q2	%q3	%q4
1	Ale	11	229.545	26.871	193	275	11	25	20	44
2	Alingsås	10	421.600	47.796	359	492	9	13	18	60
3	Alvesta	11	201.182	17.435	169	227	15	17	23	45
4	Åmål	11	143.182	14.476	116	166	12	10	19	59
5	Aneby	8	87	36.387	2	114	13	9	20	58
6	Ånge	11	118.636	19.521	87	154	15	21	20	44
7	Arboga	10	153.100	14.541	119	166	11	13	16	60
8	Åre	11	163.636	40.463	115	249	10	9	21	59
9	Arjeplog	10	58.400	7.321	47	70	8	11	17	64
10	Årjäng	10	110.200	9.987	95	128	15	20	25	39
11	Arvidsjaur	11	64.909	16.961	32	98	17	18	25	40
12	Arvika	11	285.364	16.913	254	306	11	14	16	59
13	Åsele	10	41.100	5.466	33	50	18	21	25	36
14	Askersund	11	93.727	25.088	45	122	17	18	34	31
15	Åstorp	10	173	11.954	150	193	9	16	25	50
16	Åtvidaberg	10	101.300	12.553	86	125	13	13	18	56
17	Avesta	11	262.182	51.427	148	342	12	13	15	60
18	Båstad	10	172.100	20.994	137	199	9	11	20	60
19	Bengtsfors	11	113.727	16.426	94	142	18	10	18	54
20	Berg	11	101.818	17.798	74	122	18	19	23	40
21	Bjuv	11	194.455	55.518	36	244	8	16	14	62
22	Boden	11	335.455	51.259	244	430	11	15	11	63
23	Bollebygd	11	85.636	18.986	65	138	9	15	22	54
24	Bollnäs	11	247.636	32.654	197	296	14	18	16	51
25	Borås	11	1,108.727	85.062	997	1,241	10	15	16	60
26	Borgholm	11	142.818	16.345	116	175	12	20	17	51
27	Borlänge	11	552.818	30.792	487	601	13	18	18	51
28	Botkyrka	11	938.818	52.966	855	1,020	7	31	14	48
29	Boxholm	11	69.545	11.750	49	92	12	14	10	64
30	Bromölla	11	135.818	12.960	116	159	11	14	19	56
31	Bräcke	11	85.364	9.821	70	103	15	20	21	44
32	Burlöv	11	210.364	18.370	178	245	8	18	18	57
33	Dals-Ed	11	61.818	8.495	43	76	13	13	22	52
34	Degerfors	11	85.909	14.117	70	108	19	20	14	47
35	Dorotea	11	45.091	8.324	27	55	14	9	11	66
36	Eda	11	106.818	11.932	89	126	14	16	22	47
37	Ekerö	11	222.727	27.401	192	280	10	21	17	53
38	Eksjö	11	243.727	39.616	195	341	10	9	16	65
39	Emmaboda	11	118.091	15.404	99	154	16	13	22	49
40	Enköping	11	401.909	31.059	346	450	13	17	21	49
41	Eskilstuna	11	903.545	81.854	782	1,051	12	23	15	51
42	Eslöv	11	328	50.062	263	436	10	14	24	52
43	Essunga	11	49.364	9.500	32	63	17	14	22	46
44	Fagersta	11	151.364	15.214	124	168	12	13	12	63
45	Falkenberg	11	494.636	40.991	427	556	11	15	19	55

46	Falköping	11	292.909	25.450	252	338	11	11	19	58
47	Falun	11	567.818	44.708	490	634	12	16	16	56
48	Filipstad	10	186.200	25.841	126	221	15	14	13	58
49	Finspång	11	226.364	15.731	206	253	13	14	14	59
50	Flen	11	187.818	25.899	151	232	12	18	21	49
51	Forshaga	11	80.636	7.487	70	96	19	20	21	39
52	Färgelanda	11	77.273	12.459	54	96	13	16	23	48
53	Gagnef	11	96.455	9.395	78	108	17	17	20	45
54	Gislaved	11	346.909	48.743	231	395	13	10	21	57
55	Gnesta	11	81.364	13.261	65	101	14	21	21	44
56	Gnosjö	11	99.273	16.007	74	132	11	11	24	54
57	Gotland	11	563.455	55.237	482	650	15	15	15	55
58	Grums	11	115.818	12.040	98	141	14	18	18	50
59	Grästorp	7	63.429	11.858	46	84	12	12	27	49
60	Gullspång	11	66.455	23.636	39	127	21	18	21	39
61	Gällivare	11	255.455	17.598	227	283	9	11	15	65
62	Gävle	11	956.636	62.425	876	1,062	10	16	15	59
63	Göteborg	11	5,513.818	183.318	5,181	5,801	10	22	12	55
64	Götene	11	148.909	21.979	107	187	13	11	19	57
65	Habo	11	82.273	13.070	69	112	12	11	28	49
66	Håbo	11	195	23.160	164	237	11	19	16	55
67	Hagfors	11	123.727	36.417	67	187	15	16	18	51
68	Hallsberg	11	144.273	23.525	110	177	14	13	19	53
69	Hallstahammar	11	150.091	18.425	126	190	12	17	14	57
70	Halmstad	11	848.273	43.458	778	921	11	16	16	57
71	Hammarö	11	125.636	23.265	99	172	11	11	6	73
72	Haninge	11	661.909	43.572	573	721	11	24	18	48
73	Haparanda	11	111.909	16.483	84	144	14	15	15	56
74	Heby	11	127.455	12.011	105	151	17	20	23	41
75	Hedemora	11	169.727	46.401	54	223	13	17	20	50
76	Helsingborg	11	1,604.455	65.177	1,521	1,715	12	15	17	57
77	Herrljunga	11	116.727	21.541	91	156	16	13	16	55
78	Hjo	11	57	13.364	32	75	14	9	22	55
79	Hofors	11	99.364	23.170	67	141	13	16	17	55
80	Huddinge	11	716.818	78.325	591	840	10	22	18	50
81	Hudiksvall	11	369.273	28.082	337	436	13	15	21	50
82	Hultsfred	11	232.545	39.576	183	318	16	13	15	57
83	Hällefors	11	83.091	14.646	55	110	18	13	17	51
84	Härjedalen	11	166.455	14.828	150	197	13	18	26	42
85	Härnösand	11	253.364	34.840	200	307	14	15	17	53
86	Härryda	11	266.182	22.991	237	305	8	20	25	47
87	Hässleholm	11	540.364	36.327	476	605	11	12	23	54
88	Höganäs	11	263.727	50.652	198	389	14	12	11	63
89	Högsby	11	76.455	15.280	59	106	19	19	20	42
90	Hörby	11	157.545	18.430	126	181	11	12	31	46
91	Höör	11	166.273	25.605	132	208	11	10	26	53
92	Hylte	11	142.091	20.724	117	178	19	14	18	49
93	Jokkmokk	11	70.818	8.340	61	88	12	17	16	55
94	Järfälla	11	471.636	73.837	368	618	10	24	13	52
95	Jönköping	11	1,709.727	124.773	1,462	1,875	9	10	21	60
96	Kalix	11	175.818	23.004	121	201	15	15	19	51
97	Kalmar	11	542.182	42.951	488	647	11	12	17	60

98	Karlsborg	11	60.545	11.945	37	79	11	22	16	51
99	Karlshamn	11	384.636	29.924	346	426	9	11	13	67
100	Karlskoga	11	310	26.881	258	355	14	15	14	56
101	Karlskrona	11	709.455	37.861	661	775	11	11	14	64
102	Karlstad	11	722.455	55.189	619	802	10	14	19	57
103	Katrineholm	11	403.545	18.256	360	435	10	13	16	61
104	Kil	11	100	13.266	79	129	13	19	24	44
105	Kinda	11	126.727	18.499	94	156	9	10	14	67
106	Kiruna	11	281.818	68.693	227	471	13	9	12	66
107	Klippan	11	195.273	18.423	171	218	12	14	24	50
108	Knivsta	4	108	7.348	102	117	12	22	26	40
109	Kramfors	11	218.727	34.906	165	273	19	17	19	45
110	Kristianstad	11	801.364	47.473	735	872	10	14	20	57
111	Kristinehamn	11	270.455	28.144	225	334	12	19	19	50
112	Krokom	11	105.273	15.212	91	138	15	15	24	46
113	Kumla	11	190.455	25.727	139	220	14	14	18	54
114	Kungsbacka	11	504.818	39.625	445	570	10	18	17	55
115	Kungsör	10	89.500	18.082	67	121	18	12	20	49
116	Kungälv	11	435.727	31.856	351	476	7	13	19	61
117	Kävlinge	11	239.909	40.066	205	317	9	17	23	50
118	Köping	10	265.600	18.578	236	302	15	16	16	53
119	Laholm	11	251.182	19.067	228	297	12	14	20	54
120	Landskrona	11	523.545	30.775	477	581	11	15	16	57
121	Laxå	11	86.182	21.679	61	134	12	19	23	46
122	Lekeberg	11	55.091	10.454	33	67	17	17	27	38
123	Leksand	10	160.700	37.473	106	219	12	14	15	58
124	Lerum	11	305.818	34.426	246	341	8	20	17	55
125	Lessebo	11	81.727	9.296	70	100	14	17	16	53
126	Lidingö	11	172.909	30.905	99	213	10	18	10	62
127	Lidköping	11	299.636	23.161	258	345	14	14	19	53
128	Lilla Edet	11	218.727	22.218	188	272	14	13	18	56
129	Lindesberg	11	286.364	30.210	250	333	18	12	14	55
130	Linköping	11	1,373.455	105.397	1,194	1,580	8	10	12	70
131	Ljungby	11	365.455	27.201	314	400	9	12	25	54
132	Ljusdal	11	230.182	26.411	185	262	13	17	19	51
133	Ljusnarsberg	9	72.556	10.549	62	95	14	16	19	51
134	Lomma	11	164.636	16.274	142	192	6	14	30	50
135	Ludvika	11	354.364	34.398	280	401	12	11	16	61
136	Luleå	11	724.273	67.972	604	851	10	16	16	59
137	Lund	11	959.909	71.140	819	1,036	9	12	15	64
138	Lysekil	11	171.727	20.338	136	208	8	18	12	62
139	Malmö	11	2,946.455	377.905	2,479	3,493	12	23	16	49
140	Malung-Sälen	11	191.727	16.983	169	220	14	15	20	51
141	Mariestad	11	303.727	24.377	259	345	13	11	15	62
142	Mark	11	318.636	29.330	273	371	12	16	20	52
143	Markaryd	11	145.636	18.943	107	182	11	12	24	53
144	Mellerud	11	103.909	18.512	55	123	15	15	23	48
145	Mjölby	11	325.545	27.876	270	368	9	11	19	61
146	Mora	11	254.818	25.131	216	292	12	14	19	55
147	Motala	10	389.600	45.189	332	475	11	14	16	58
148	Mullsjö	11	69.364	14.821	47	90	13	12	23	52
149	Munkedal	11	171.545	41.495	117	247	9	13	20	57

150	Munkfors	11	47	10.305	35	66	20	15	16	49
151	Mölndal	11	559.364	36.484	476	632	8	17	15	60
152	Mönsterås	11	210	30.070	155	255	10	11	16	62
153	Mörbylånga	11	107.455	20.211	88	148	12	18	21	49
154	Nacka	11	554.727	57.303	451	637	9	20	10	61
155	Nora	11	82.818	9.988	69	99	16	14	22	48
156	Norberg	11	73.909	29.750	18	127	14	13	15	58
157	Nordanstig	10	94.700	16.364	68	117	22	14	34	30
158	Nordmaling	11	76.727	6.035	67	88	18	19	23	39
159	Norrköping	11	1, 459.636	80.110	1, 356	1, 665	12	12	14	63
160	Norrtälje	11	485	38.131	408	543	14	17	23	46
161	Norsjö	9	32.111	8.023	14	41	31	14	21	34
162	Nässjö	11	392.818	39.982	312	437	12	12	19	57
163	Nybro	11	227.455	30.962	179	273	12	14	19	54
164	Nyköping	11	595.455	54.430	524	683	9	15	21	55
165	Nynäshamn	11	265.727	39.289	210	341	10	24	18	47
166	Ockelbo	11	66.545	10.192	53	88	17	21	22	41
167	Olofström	11	110.545	17.963	83	139	15	15	23	47
168	Orsa	11	69.636	11.801	50	88	18	16	24	42
169	Orust	11	133	24.384	85	178	11	17	24	49
170	Osby	11	175.182	14.034	153	202	12	14	21	53
171	Oskarshamn	11	443.818	34.190	391	495	8	8	12	72
172	Ovanåker	11	108.727	12.166	90	137	19	17	15	49
173	Oxelösund	11	103.636	14.271	86	139	13	19	7	60
174	Pajala	8	53.750	8.828	38	70	30	27	25	18
175	Partille	11	241.091	25.450	213	278	9	19	11	61
176	Perstorp	11	163.182	16.827	132	199	10	7	11	72
177	Piteå	8	414.375	28.869	364	451	12	10	16	61
178	Ragunda	11	66.182	8.612	53	83	19	22	25	34
179	Robertsfors	8	64.625	12.592	43	84	15	15	24	46
180	Ronneby	11	359.182	50.750	295	470	10	11	16	63
181	Rättvik	11	135.182	18.054	102	166	14	15	19	53
182	Sala	8	268.750	28.177	223	318	12	13	19	56
183	Sandviken	11	396.091	47.103	342	481	12	16	14	57
184	Sigtuna	11	603.909	70.509	495	720	8	16	15	61
185	Simrishamn	11	184.273	22.010	145	207	13	12	19	55
186	Sjöbo	11	185.545	14.116	162	217	16	16	29	39
187	Skara	11	307.636	31.668	254	343	9	10	18	64
188	Skellefteå	11	550.545	53.283	487	638	16	17	19	48
189	Skinnskatteberg	8	71.250	15.673	49	96	9	14	22	54
190	Skurup	3	144.333	9.292	138	155	10	14	27	49
191	Skövde	11	593.545	44.704	514	641	8	8	13	71
192	Smedjebacken	8	140	18.408	117	170	13	15	16	56
193	Sollefteå	11	231.545	44.689	185	354	16	22	18	44
194	Sollentuna	11	467.727	52.443	391	541	9	18	18	54
195	Solna	10	601.600	43.108	539	668	8	13	10	69
196	Sorsele	11	33.182	6.210	26	43	22	21	22	35
197	Sotenäs	11	162	18.724	136	193	6	14	11	69
198	Staffanstorp	11	187.909	23.864	133	217	6	12	23	59
199	Stenungsund	11	278.545	33.625	216	344	9	9	20	62
200	Stockholm	8	6, 102.750	560.163	4, 756	6, 522	10	18	7	65
201	Storfors	11	39.273	6.828	27	48	22	16	19	44

202	Storuman	10	62.600	14.315	38	80	12	13	25	50
203	Strängnäs	11	419.545	34.558	348	470	8	18	16	58
204	Strömstad	8	227.125	31.343	170	256	9	11	21	59
205	Strömsund	11	125.364	21.393	80	158	18	19	23	41
206	Sundsvall	11	921.091	60.182	862	1,056	13	16	17	54
207	Sunne	8	140.750	22.417	119	187	19	15	21	46
208	Surahammar	11	85.909	12.818	70	114	11	16	19	54
209	Svalöv	11	144.091	31.810	107	212	11	16	22	51
210	Svedala	8	216.250	40.861	147	281	7	15	25	53
211	Svenljunga	11	120.636	10.661	97	138	14	13	22	51
212	Säffle	8	199	27.800	155	252	16	14	19	52
213	Säter	11	123	17.170	99	147	12	23	27	38
214	Sävsjö	11	128	16.125	96	151	14	13	19	54
215	Söderhamn	11	290	34.450	232	347	11	18	22	49
216	Söderköping	8	143.375	14.501	126	170	11	17	18	55
217	Södertälje	11	1,076.727	103.519	949	1,249	10	27	16	47
218	Sölvesborg	11	149.273	29.706	118	231	12	14	19	55
219	Tanum	8	235.375	22.219	203	267	7	12	25	56
220	Tibro	11	80	11.463	62	103	15	12	17	56
221	Tidaholm	9	132.222	14.464	110	156	13	12	19	55
222	Tierp	11	199.545	36.689	111	251	15	16	22	47
223	Timrå	11	197.364	24.586	170	253	12	21	16	50
224	Tingsryd	11	139	16.553	111	176	16	16	23	46
225	Tjörn	8	160.750	22.147	141	200	7	21	15	58
226	Tomelilla	11	128	12.985	113	153	13	17	27	42
227	Torsås	11	64.727	10.071	53	86	16	17	20	47
228	Torsby	11	193	18.050	168	224	16	15	21	47
229	Tranås	8	249	24.089	222	296	11	10	14	65
230	Tranemo	11	135.364	22.015	105	172	12	11	22	56
231	Trelleborg	8	413.500	44	371	507	9	12	18	60
232	Trollhättan	11	614.909	75.478	486	741	9	15	15	61
233	Trosa	8	108.875	12.597	93	126	10	13	25	53
234	Täby	7	317	37.139	267	381	8	16	12	63
235	Töreboda	11	86.818	15.420	60	103	16	15	17	52
236	Tyresö	11	232.818	20.760	201	266	10	28	12	50
237	Uddevalla	8	560.375	36.948	513	611	8	13	23	56
238	Ulricehamn	11	208.727	22.751	170	249	13	11	26	51
239	Umeå	8	791.375	54.962	708	888	12	11	15	62
240	Upplands-Bro	11	208.455	21.616	180	257	10	28	18	44
241	Upplands Väsby	11	321.818	40.393	256	367	9	24	18	49
242	Uppsala	8	1,445.875	123.727	1,242	1,646	13	20	13	53
243	Uppvidinge	11	101.818	15.165	74	121	16	15	29	40
244	Vadstena	10	81.300	11.066	60	97	13	10	11	65
245	Vaggeryd	11	180.818	29.849	137	230	11	13	26	50
246	Valdemarsvik	8	85.625	11.275	70	108	13	19	21	47
247	Vallentuna	10	180.200	28.003	138	224	9	24	18	48
248	Vansbro	8	109.375	13.158	90	130	22	14	24	41
249	Vara	10	183.900	17.629	160	210	15	14	28	43
250	Varberg	11	565.636	36.770	513	627	11	18	18	52
251	Vårgårda	8	139.500	34.339	70	167	11	11	27	51
252	Vaxholm	9	59.222	10.860	43	84	10	17	10	63
253	Vellinge	8	191.625	27.129	163	234	9	18	23	49

254	Vetlanda	11	320.091	43.413	239	385	15	11	21	54
255	Vilhelmina	10	60.800	11.063	43	83	20	20	19	41
256	Vimmerby	11	235.455	39.114	149	289	17	8	14	60
257	Vindeln	8	53.750	14.714	29	68	16	20	20	45
258	Vingåker	10	82.400	11.433	66	97	17	13	16	55
259	Vänersborg	11	344	63.559	247	492	11	14	16	58
260	Värmdö	10	242.500	70.407	170	430	11	30	15	44
261	Värnamo	8	391.500	25.467	339	413	12	11	28	49
262	Västerås	11	$1,240.182$	74.159	$1,093$	$1,340$	10	19	10	60
263	Västervik	8	434.250	40.732	380	507	12	12	17	60
264	Växjö	11	877.273	62.983	791	954	12	13	15	60
265	Älmhult	10	242.200	27.744	177	286	11	14	16	59
266	Älvdalen	7	122.143	12.903	106	145	13	15	22	50
267	Älvkarleby	10	185.200	35.007	129	230	10	11	7	72
268	Älvsbyn	10	79.600	10.596	67	101	14	21	20	45
269	Ängelholm	10	409	21.385	366	447	10	12	26	52
270	Öckerö	10	92	26.825	58	156	8	35	9	48
271	Ödeshög	7	108.143	11.127	94	123	10	10	29	50
272	Örebro	10	$1,109.300$	160.313	774	$1,320$	16	14	13	57
273	Örkelljunga	10	137.300	17.192	116	168	13	20	30	38
274	Örnsköldsvik	7	505	38.131	449	558	13	14	19	54
275	Österåker	6	240	42.005	189	302	10	26	17	48
276	Östersund	9	468.333	51.471	394	530	13	13	21	54
277	Östhammar	9	157.778	42.760	55	198	18	18	21	43
278	Östra Göinge	10	122	20.833	89	160	14	14	28	44
279	Överkalix	7	47.571	7.635	37	58	14	22	28	37
280	Övertorneå	9	34.889	7.656	23	48	21	20	20	39
281	Ydre	6	57.833	15.690	39	82	10	9	15	67
282	Ystad	11	364.364	61.583	282	459	11	9	17	63

Source: IDA 2016; Author's calculations.

Note: $\%q1$ - share of fires in buildings, $\%q2$ - share of other fires, $\%q3$ - share of traffic accidents, $\%q4$ - share of other emergencies.

SUMMARY IN ESTONIAN – KOKKUVÕTE

Riigiasutuse soorituse hindamine ebakindla nõudluse tingimustes Eesti, Soome ja Rootsi päästeteenuste näitel

Kõige üldisemalt on riigiasutuste ülesandeks täita talle administratiivsüsteemis seatud kohustusi rahvale teenuste pakkumisel. Eesmärgiks on pakkuda nõutavaid teenuseid parimal viisil mahtu ja kvaliteeti tasakaalustades. Avaliku sektori ja riigiasutuste tähtsus seisneb asjaolus, et need on suured tööandjad ja teenuseosutajad, kuid ühtlasi ka suured maksumaksja raha kasutajad ning läbi selle vastutavad ühiskonna toimimise eest. Seega on riigiasutustel suur mõju majandusele ja ühiskonnale üldisemalt, mistõttu on nende soorituse hindamine igati põhjendatud ja vajalik. Riigiasutuse sooritus on aktuaalne teema nii Eestis kui ka teistes riikides. Elanikkonna ootused avaliku sektori poolt osutatavate teenuste osas üha kasvavad, ent teenuse osutamiseks vajalikke ressursse ei ole võimalik suurendada vastavalt ootuste kasvule. Ühe võimaliku lahendusena nähakse riigiasutuste töö tulemuslikkuse, kui soorituse osa, pidevat parendamist. Seejuures vajavad lahendamist mitmed probleemid. Esiteks, teoreetilised probleemid avaliku sektori töö tulemuslikkuse määratlemisel ja nõudluse kõikumise mõju hindamisel tulemuslikkusele. Teiseks, metodoloogilised probleemid töö tulemuslikkuse kvantitatiivsel hindamisel ja võrdlemisel. Kolmandaks, rakenduslikud aspektid töö tulemuslikkuse väljatöötamisel infovarustuse spetsiifilistes tingimustes. Nende probleemide lahendamisele ongi käesolev uurimus suunatud.

Riigiasutuse soorituse hindamine hoogustus märgatavalt nn uue haldusjuhtimise (*New Public Management*, NPM) tõusuga. Paraku ei toonud ärijuhtimisest liiga otseselt üle võetud juhtimismeetodid avalikus sektoris oodatud lihtsat edu ning esialgne eufooria lahtus. Tulemuste mõõtmine on aga taas tõusu teel eelkõige seetõttu, et informatsiooni hulk on viimastel aastakümnetel hoomamatult kasvanud. Sellega paralleelselt on kaasnenud mõõtmismeetodite täiustumine. Käesolevas uurimuses süstematiseeritakse ning hinnatakse meetodeid riigiasutuse soorituse hindamisel lähtuvalt saadaoleva informatsiooni hulgast ja kvaliteedist.

Valdav enamus riigiasutustest toimib keerukas keskkonnas. Riigiasutuse töö planeerimisel on paljud tegurid riigiasutuse juhtidele teadmata ning ebakindlad. Selliseks ebakindlaks teguriks on näiteks oodatav nõudlus riigiasutuse poolt pakutavate teenuste järele. Ühe hea näitena riigiasutusest, mis tegutseb ebakindla nõudluse tingimustes, võib tuua päästeteenistuse. Päästeteenistus jaotab oma ressursid erinevate allüksuste vahel, kindlalt teadmata, kui palju neid ressursse vastavas asukohas rakendada tuleb. Seega, päästeteenistus peab tagama pideva valmisoleku, et päästesündmusele kiiresti reageerida. Tüüpiliselt nähakse valmisoleku tagamises üksnes kuluartiklit. Kulutõhusust, mis arvestaks valmisoleku tagamisega ebakindlale nõudlusele reageerimiseks, pole aga senini hinnatud. Käesolev uurimus hindab kulutõhususe, mis arvestab ebakindla nõudlusega kolme riigi päästevaldkonna näitel. Neis kolmes riigis (Eestis, Soomes ja Rootsis) erineb päästevaldkonna struktuur oluliselt. Uurimuses analüüsitakse võimalikke struktuurist tulenevaid erisusi soorituse hindamisele.

Käesoleva doktoritöö eesmärgiks oli välja töötada teoreetiline kontseptsioon ja rakendus, kuidas hinnata sooritust riigiasutustes, mis toimivad ebakindla nõudluse tingimustes. Väljapakutud mudelid oleksid aluseks ressursside jaotuse planeerimise parendamisel riigiasutustes. Soorituse hindamiseks kasutati mitmesuguseid piirianalüüsi meetodeid ning kontseptsiooni rakendati Eesti, Soome ja Rootsi päästevaldkonna näitel. Eesmärgi täitmiseks püstitati järgmised uurimisülesanded:

Teoreetiline raamistik:

- defineerida riigiasutuse sooritus ja soorituse hindamine ning sellega tihedalt seotud mõisted ja käsitlused;
- süstematiseerida riigiasutuse soorituse hindamise võimalikud kasutusalad, hinnates soorituse mõõtmiste võimalusi ja piiranguid;
- analüüsida ebakindla nõudluse mõju riigiasutusele teenuste osutamisel;

Meetodid:

- süstematiseerida riigiasutuse soorituse hindamisel rakendust leidnud meetodid;
- luua metoodika, mis võimaldaks integreerida ebakindla nõudluse riigiasutuse soorituse hindamisse;

Rakendus:

- näitlikustada väljapakutud kontseptsiooni toimimist teenuste osutamisel päästeteenuste näitel (Eestis, Soomes ja Rootsis), tuginedes varasematele uuringutele, seadustele ja arengukavadele;
- võrrelda erinevate hindamismeetodite karakteristikuid; põhjendada meetoditest tulenevate hinnangute erinevusi;
- hinnata empiiriliselt Eesti, Soome ja Rootsi päästeüksuste soorituse taset;

- hinnata väljapakutud ebakindla nõudluse kontseptsiooni rakendamise võimalusi otsuste tegemise parendamiseks.

Töö koosneb kolmest peatükist. Esimeses peatükis käsitletakse riigiasutuse soorituse hindamise teoreetilisi aspekte: defineeritakse riigiasutuse soorituse ja selle hindamisega seonduvad mõisted ja käsitlused, süstematiseeritakse soorituse hindamise võimalused ja lüngad ning hinnatakse ebakindla nõudluse mõju riigiasutuse soorituse hindamisele. Teine peatükk annab võrdlevhinnangu erinevate meetodite rakendamise võimalustele riigiasutuse soorituse hindamisel ning käsitleb ebakindla nõudluse mõju soorituse hinnangule metoodilisest aspektist. Kolmandas peatükis rakendatakse ebakindla nõudluse kontseptsiooni alusel piirianalüüsi meetodeid Eesti, Soome ja Rootsi päästeteenistuste soorituse hindamisel: kõrvutatakse erinevate piirianalüüsi meetoditega saadud tulemusi, võrreldakse päästevaldkonna sooritust neis kolmes riigis ning diskuteeritakse võimaluste üle rakendada empiirilise analüüsi tulemusi päästevaldkonna juhtimisotsuste tegemisel.

Töö uudsus seisneb mikroökonoomika teooria osana tuntud tootmisteooria ja praktilisema avaliku sektori juhtimise valdkondade tihedamal sidumisel ebakindla nõudluse tingimustes toimiva riigiasutuse soorituse hindamisega. Kuigi mõistetel 'tõhusus' ja 'tulemuslikkus' on tootmisteoorias kindel definitsioon, kasutatakse neid majanduse juhtimise praktikas tihti meelevaldselt ning järjekindlusetult. Teiseks on uudne ebakindla nõudluse kontseptsiooni sidumine piirianalüüsi meetodite rakendamisega. Kolmandaks on uudne uurimisobjekti süsteemne käsitlus. Kuigi päästevaldkonda kajastatakse põhjalikult valdkonna praktilisi probleeme käsitlevas kirjanduses, on seda akadeemilistes uurimustes käsitletud vaid üksikutel juhtudel ja fragmentaarselt. Seejuures on arvestamata jäetud ebakindla nõudluse mõju sooritusele. Töö kujutab endast seega esimest katset süsteemselt hinnata ja võrrelda päästevaldkonna töö tulemuslikkust kolmes riigis: Eestis, Soomes ja Rootsis. See omakorda avab uusi võimalusi päästevaldkonna töö tulemuslikkuse parendamise abinõude väljatöötamisel.

Riigiasutuse sooritus on mitme valdkonna uurimisobjektiks. Kui sooritust on võimalik hinnata, saab kogutud informatsiooni kasutada järgnevate otsuste ettevalmistamisel, et soorituse tulemuslikkust tõsta. Seega, soorituse hindamist tuleb käsitleda süsteemse tegevusena, mille tulemusena saadakse hinnang, kui hästi üks riigiasutus võrreldes teistega toimib ning kuivõrd täidab sellele asutusele pandud kohustusi. Riigiasutuste sooritus võib erineda mitmetel asjaoludel: neil võivad olla erinevad ülesanded, vajadused, teenuse osutamise viisid, koostöö partneritega, töö tõhusus, aruandluse põhimõtted jne. Lisaks kõigele esineb asutustes ka juhuslik soorituse taseme kõikumine, millele polegi võimalik ratsionaalset selgitust anda. Niisiis, soorituse hindamine võib aidata tööd paremini planeerida ja tulemuslikkust tõsta. Samas võib soorituse tulemuslikkuse võrdlevhinnang teiste riigiasutuste suhtes luua laiema ja tugevama baasi soorituse parendamise abinõude väljatöötamiseks. Soorituse täiustamise planeerimine sooritus hindamise eesmärgina on eelkõige suunatud organisatsioonisiseseks kasutamiseks, samas kui võrdlevhindamine teenib eelkõige avalikke huvisid, mis omakorda võib mõjutada organisatsioonis tehtavaid otsuseid.

Soorituse hindamisel on ka mitmeid puudusi, mis võib jagada kolme suuremasse gruppi: 1) tehnilised tõrked (mõõdetakse-hinnatakse valesti), mis seonduvad tulemusindikaatorite valiku, aga samuti info kogumise, analüüsi ja tõlgendamisega; 2) süsteemsed või kontseptuaalsed probleemid (mõõdetakse valesid aspekte), mis seondub soorituse hindamise tulemuste integreerimisega otsustusprotsessidesse, strateegilise juhtimise puudustega, eesmärkide laialivalguvusega jms; ning 3) kaasamise küsimustega (mõõtmistulemusi rakendatakse viisil, mis ei ole soovitav), mis seondub inimfaktoriga – kuivõrd huvitatud ollakse tulemuste rakendamisest ning kas seda tehakse eesmärgipäraselt.

Soorituse hindamiseks tuleb esiteks mõista, kuidas riigiasutuses tööprotsessid toimuvad. Kontseptuaalselt on siinjuures abiks sisend-väljund mudel. See mudel näitlikustab riigiasutuse toimemehhanisme, kuidas eraldatud ressursid transformeeritakse teenusteks. Sisendnäitajatena võib siinkohal käsitleda tööjõu ja kapitali mahtu, struktuuri ja kvaliteeti, aga samuti ajaressurssi iseloomustavaid andmeid jms, mis osalevad väljundite loomisel. Väljunditeks on osutatud teenused. Et sisendite rakendamise tulemusena tekiks väljund, on vaja teha teatud toiminguid (tegevusi). Antud lähenemisest võib tuletada tulemusindikaatorid, mille abil riigiasutuse sooritust iseloomustada ja hinnata. Päästeteenuste puhul on sisendeid iseloomustavateks näitajateks näiteks päästjate, päästemasinate ja muude tehniliste vahendite arv. Väljundeid iseloomustavate näitajatena võib käsitleda väljakutsete arvu erinevatele päästesündmustele (tulekahjud, liiklusõnnetused, muud õnnetused). Ühtlasi tuleb rõhutada, et päästeteenuseid osutatakse muutuvates keskkonnatingimustes, st riigiasutuse sooritus sõltub teeninduspiirkonna karakteristikutest, nagu näiteks hoonestuse kvaliteet ja struktuur, elanikkonna sotsiaalmajanduslik olukord ja riskikäitumine ning ilmastikutingimused.

Erinevate tulemusindikaatorite omavaheliseks seostamiseks on välja pakutud mitmeid hindamiskriteeriume. Hindaja võib olenevalt hindamise eesmärgist huvituda soorituse tõhususest, tulemuslikkusest, ökonoomsusest, võrdsusest, eetilisusest või teistest aspektidest. Käesoleva uurimuse fookuses on soorituse tõhususe hindamine, sest sellel on teoreetiliselt põhistatud määratlus, kuigi praktikas kasutatakse seda mõistet laialivalguvalt ja meelevaldselt. Juhul, kui kasutatakse erinevaid tulemusindikaatoreid ja hindamiskriteeriume, tuleb luua soorituse hindamise süsteem, mille alusel saab anda riigiasutuse sooritusele tervikliku hinnangu.

Tõhusus (sünonüümina võib kasutada efektiivsust) on mõistena kasutusel tootmisteoorias. Tehniline tõhusus on mõõdik, mis defineeritakse kui vaadeldava üksuse sisend-väljund komplekti (suhteline) kaugus tootmisvõimaluste piirist, seega on tegemist suhtnäitajaga (-mõõdikuga). Võrdluse võib esitada kahes vormis: üksuse vaadeldava ja optimaalse väljundkomplekti suhe juhul, kui sisendkomplekt on fikseeritud või üksuse vaadeldava ja optimaalse sisendkomplekti suhe juhul, kui väljundkomplekt on fikseeritud. Sisend-väljundvektor on tehniliselt efektiivne siis ja ainult siis, kui ükskõik millise väljundi koguse suurendamine või ükskõik millise sisendi koguse vähendamine on võimalik ainult juhul, kui sellega kaasneb vastavalt teise väljundi koguse vähendamine või teise sisendi koguse suurendamine.

Kui riigiasutus planeerib oma ressursside jaotamist, on tihti nõudlus nende teenuste järele kindlalt teadmata. Omamata täielikku teavet võimalikust nõudlusest, kindlustab otsustaja ennast ootamatute nõudlussšokkide vastu täiendavate ressurssidega (suurendab valmisolekut). Kulu- ja tõhususanalüüsid eeldavad üldjuhul, et nõudlust iseloomustavad parameetrid on teada (või ei pöörata sellele lihtsalt tähelepanu), mis on aga ebarealistlik ning alahindab teenuse osutamise keerukust riigiasutustes. Paljudel juhtudel toob see kaasa soorituse tõhususe liiga madala hinnangu. Üleliigset ressursivarustatust (valmisolekut) teenuste osutamiseks võib osaliselt selgitada otsustajate riskikartliku käitumisega ning seega peaks riskihinnang olema osa soorituse tulemuslikkuse hinnangust, et vältida sellise ressursivarustatuse käsitlemist ebatõhususena. *Ex ante*, võttes arvesse oodatud nõudlust, võib ressursside jaotus olla optimaalne, kuid mitte *ex post*, kui on teada reaalne nõudluse tase. Standardne (traditsiooniline) soorituse tulemuslikkuse hinnang sellist valmisolekureservi mahtu arvesse ei võta ja seetõttu hinnatakse riske arvestav käitumine ebatõhususeks. Tegelik väljakutse seisneb aga sellise vajaliku valmisoleku taseme eristamises ülemäärasest ressurssidega varustatusest. Et hinnata mitme üksusega ja mitmel tasandil toimivat teenuseid osutavat riigiasutust ebakindla nõudluse tingimustes, on hindaja ülesandeks luua süsteem, milles hinnatakse (a) keskse riigiasutuse (nn tööstusharu tasandil) kulutõhusust, (b) allüksuste alavarustatust ja (c) allüksuste väljundtõhusust, mis iseloomustab nende võimekust pakkuda suuremat hulka teenuseid.

Riigiasutuse soorituse hindamiseks on mitmeid erineva keerukusega meetodeid. Käesoleva doktoritöö teine peatükk keskendus meetodite süstematiseerimisele ja võrdlevanalüüsile soorituse tulemuslikkuse hindamise aspektist. Sobivaima meetodi valikul tuleb lähtuda sellest, milleks analüüsi teostatakse ning kes selle tulemusi rakendama hakkab. Kui hinnangute kasutajaks on poliitikakujundajad, siis peaks analüüs olema selgelt fokusseeritud ning pärinema pigem avaliku sektori juhtimise valdkonnast. Kui fookuses on aga teoreetiline kontseptsioon ning soov mõista, kuidas teenuse osutamist kõige paremini kirjeldada, siis baseeruvad kasutatavad meetodid eelkõige tootlikkuse analüüsil. Võimalike hindamismeetodite hulk on ulatuslik: ühikukulu analüüs, suhteanalüüs, riskidega kohandamine, Nelja Kvadrandi meetod, komposiitindeksi analüüs, klasteranalüüs, Tasakaalus Tulemuskaart, lineaarregressioon, mitmetasandiline lineaarne regressioon, näiliselt seostamata regressioon, andmeraja analüüs ja stohhastiline piirianalüüs. Antud loetelu ei ole küll lõplik, ent sisaldab kõiki peamisi meetodeid, mida siiani riigiasutuse soorituse tulemuslikkuse hindamisel on rakendatud.

Piirianalüüsi meetodite (mis leiavad peamiselt kasutust tootlikkuse võrdlevhindamisel) ühiseks jooneks on hinnata/modelleerida saavutatava tulemuslikkuse ülemine piir. Piiri on võimalik hinnata erinevatel eeldustel ning hindamistehnikatega. Järgmise sammuna kõrvutatakse vaadeldud üksuse tulemusindikaatorid hinnatud piiri suhtes ja nii leitaksegi suhteline tõhusushinnang. Järelikult, piirianalüüsi meetodid hindavad tegeliku tulemuse kauguse piirist – see tähendab tõhususe või tulemuslikkuse maksimaalset-minimaalset proportsiooni saavutatavasse tasemesse kindlas tehnoloogiatasemes ja keskkonnas. Kuigi piirianalüüsi meetodid on populaarsed akadeemilistes uuringutes, leiavad need meetodid harva kasutust poliitika kujundamisel ja riigiasutuste praktiliste juhtimisotsuste ettevalmistamisel. Kaks kõige laialdasemalt kasutatavat meetodit on andmeraja analüüs (*data envelopment analysis*, DEA), mis on deterministlik ja mitteparameetriline, ning stohhastiline piirianalüüs (*stochastic frontier analysis*, SFA), mis on, nagu nimigi ütleb, stohhastiline ja parameetriline. Lisaks neile kahele meetodile, millel on oma koolkond, leidub ka mitmeid edasiarendusi, mis proovivad mõlema meetodi häid omadusi ühendada. Ebakindlat nõudlust kui nähtust ei ole tootlikkuse võrdlevhindamise valdkonnas põhjalikult käsitletud või on seda tehtud oodatava kasulikkuse (*expected utility*) mudeli alusel, aga samuti tingimusliku oleku (*state-contingent*) lähenemist kasutades. Lisaks sellele on vaid üksikud tööd käsitlenud päästeasutuste soorituse tulemuslikkuse hindamist ning vaid üksikutes uuringutes on rakendatud piirianalüüsi meetodeid. Seega on antud meetodite teoreetiliselt põhjendatud rakendamine päästevaldkonna asutuste soorituse tulemuslikkuse hindamisel uudne lähenemine, millel praktiline tähtsus tõhususe juhtimise parendamiseks.

Väljapakutud teoreetilise ja metodoloogilise raamistiku testimiseks riigiasutuse soorituse tulemuslikkuse hindamiseks ebakindla nõudluse tingimustes rakendati piirianalüüsi meetodeid päästeasutuste soorituse tulemuslikkuse hindamiseks kolmes riigis: Eestis, Soomes ja Rootsis. Päästeteenuseid pakutakse tavapäraselt mitmete allüksuste poolt erinevates piirkondades. Seega, teenuste osutaja otsustab esialgselt, kuidas ressursid (päästjad, päästemasinad jms) erinevates piirkondades jaotada, teadmata seejuures, kui palju päästesündmusi seal tegelikult aset leiab. Allüksused peavad hoidma teatud valmisoleku taset, et vajaduse korral päästesündmusele reageerida. Valmisoleku tagamine on aga kõige kulukam komponent eelarves, mistõttu on oluline, et ressursid oleks jaotatud nii, et ei tekiks liigseid kulusid, st minimeeritakse valmisoleku taset optimaalse mahu ja kvaliteediga päästeteenuse pakkumiseks.

Päästeteenuste osutamise viis erineb kolmes uuritud riigis oluliselt. Eestis on süsteem tsentraliseeritud ning teenust osutab üks riigiasutus (Päästeamet). Soomes on teenuse osutamine detsentraliseeritud 22 regionaalse päästekeskuse ülesandeks, millel on suur autonoomia otsuste langetamisel. Rootsis on päästeteenuste osutamine kohalike omavalitsuste kohustus. Päästeameti strateegia näeb oma visioonina 2025. aastaks, et jõutakse põhjamaade tasemele, mis peab kaasa tooma õnnetuste ja õnnetustes hukkunute arvu vähenemise. Käesolev uurimus on toetavaks materjaliks võrdluses põhjamaadega ning annab võimaluse teaduslikult põhjendatult näidata, millised on erinevused (kulu)efektiivsuse seisukohast. Päästeteenuste osutamine on kõige kulukam Rootsis ning kõige vähemkulukas Eestis. Samuti on kulu ühe õnnetuse kohta Rootsis kõrgeim. Kuigi teenuste osutamise eesmärk ja üldine viis on kõikides riikides sarnane, ei ole need riigid siiski otseselt võrreldavad kulutõhususe, alavarustatuse ja väljundtõhususe analüüsi objektina organisatsioonilise ülesehituse ja sellest tuleneva aruandluse (infobaasi) erinevuste tõttu. Seepärast viidi analüüs läbi igas riigis eraldi, kuna neid kõiki ei ole kvalitatiivsetest (organisatsioonilistest) eripäradest lähtuvalt võimalik ühendada ühte mudelisse. Võimaluse piires anti võrdlevaid hinnanguid erineva analüüside tulemuste alusel.

Töö empiiriline analüüs teostati, kasutades kolme piirianalüüsi meetodit: DEA, FDH ja DFA. Kuna erinevad meetodid toetuvad erinevatele eeldustele, siis erinevad ka nende rakendamise tulemused teatud määral teineteisest. Esmalt hinnati keskse üksuse (nn tööstusharu tasandil) soorituse kulutõhusust kahel viisil: ühel juhul võeti ebakindlat nõudlust arvesse, kasutades minimaalse teenuseosutamise taseme (oodatava väljundite taseme) kontseptsiooni; teisel juhul kasutati nn naiivset viisi, mis läheneb probleemile standardselt, st kasutab soorituse tulemuslikkust iseloomustava näitajana väljundite tegelikku taset. Analüüsi tulemused näitavad veenvalt, et ebakindlat nõudlust arvestavate mudelite alusel saadud soorituse kulutõhususe hinnangud on kõrgemad nn naiivsel viisil saad kulutõhususe hinnangutest, olenemata analüüsitud riigist või rakendatud meetodist. See tulemus näitab, et otsustaja peaks soorituse kulutõhususe suurendamiseks lähtuma ressursside jaotamisel vajadusest kindlustada päästeüksuste nõudluse ebakindlust (kõikumist) arvestav valmisoleku võimekus reageerida päästesündmustele. Senini ei pööratud nõudluse kõikumistele ja reageerimise võimekuse tagamise vajadusele soorituse kulutõhususe analüüsimisel tähelepanu ning võimekust ebakindlale nõudlusele reageerida käsitleti päästeüksuse soorituse ebatõhususena. Päästeüksuste soorituse kulutõhusus hinnati vaatlusaluste aastate lõikes üpris stabiilseks, olenemata rakendatud hindamismeetodist. Seega on madalama kulutõhususega päästeüksustel võimalik õppida kulutõhusamatelt allüksustelt ressursside paremat rakendamist., mis on üks soorituse kulutõhususe võrdleva hindamise eesmärkidest.

Teiseks hinnati päästeüksuste alavarustatuse taset. Kui päästeüksus varustati sisenditega, mida oli minimaalse teenuseosutamise taseme saavutamiseks ebapiisavas koguses, siis hinnatakse üksus alavarustatuks. Kuna alavarustatuse hinnangu saamine on võimatu kindlat nõudlust eeldaval standardsel juhul, siis hinnati käesolevas uurimuses päästeüksuste soorituse tõhusust üksnes nõudluse ebakindlust arvestavate mudelite alusel. Kõigis kolmes vaatlusaluses riigis tõid kõik hindamismeetodid välja alavarustatud päästeüksuste osakaalu langustendentsi. Paraku ei olnud tulemused piisavalt robustsed rakendatud hindamismeetodite lõikes: ühe meetodi alusel alavarustatuks hinnatud päästeüksused ei pruukinud saada sama hinnangut teise meetodi alusel. See tähendab, et päästeüksuste alavarustatuse hinnangute rakendamiseks praktilises ressursijaotuse juhtimises tuleb hindamismeetodite eelduste ja seega alavarustatuse hinnangu olemust detailsemalt uurida.

Viimasena hinnati päästeüksuste soorituse väljundtõhusust. Väljundtõhususe võrdlevhinnang toob välja, kas päästeüksusel oleks jätkunud potentsiaali, pakkumaks täiendava nõudluse korral suuremat teenuste mahtu (võrdluses kõige tõhusamate üksustega). Ebakindla nõudluse tingimustes ei ole väljundtõhususe hinnangul tähtsust juhul, kui kogu nõudlus kaeti (st päästeüksus suutis reageerida kõigile väljakutsetele), st kõik päästeüksused tuleks hinnata täielikult väljundtõhusateks. Standardse lähenemisviisi korral on väljundtõhususe hindamisel mõtet. Kuna päästeüksused on teenuste osutamise viisilt ja ka mahult erinevad, siis on väljundtõhususe hinnangud madalad ja väljundtõhususe tõstmise võimalused tuuakse selgemalt välja. Erinevate meetoditega saadud väljundtõhususe hinnangute mediaanid liikusid aga vaatlusalusel perioodil samas suunas, st tulemused osutusid suhteliselt robustseteks.

Käesoleval uurimusel on ka kindel praktiline väljund, näiteks saab väljatöötatud ebakindla nõudluse kontseptsiooni rakendada riigiasutuste töö planeerimisel: minimaalse teenuseosutamise taseme hinnangu rakendamine toetab ressursside jaotamise protsessi parendamist (nt teenuslepingute koostamisel), st hinnatakse ressursside vajadust kindla tasemega teenuse osutamiseks. Teiseks on võimalik loodud tulemusindikaatorite süsteemi rakendada soorituse parendamisel (ressursside jaotus, kulutõhususe tõstmine). Kolmandaks võimaldab tulemusindikaatorite süsteem järjestada päästeüksused vastavalt soorituse tõhususe hinnangutele, mis avab uusi võimalusi (paremate tunnustamine, nõrgemate järele aitamine) juhtimise täiustamiseks.

Kokkuvõttes pakub ebakindla nõudluse tingimustes soorituse tulemuslikkuse hindamine uusi võimalusi avaliku sektori asutuste juhtimise parendamiseks. Eelkõige tuleb arvestada, et otsusetegijad peavad langetama oma otsuse ressursside jaotuse osas omamata täielikku informatsiooni väljundite mahu kohta. Riskikartlikult käitudes peavad nad jaotama piisavalt ressursse, et toime tulla ootamatute hüpetega nõudluses. Empiiriline analüüs tõi selgelt välja, et ebakindla nõudluse kontseptsiooni rakendamine erinevate tsentraliseeritustasemetega riigiasutustes avab võimalusi juhtimisotsuste parendamiseks. Ebakindlat nõudlust arvestavad mudelid loodi kolme erinevat meetodit (DEA, FDH ja DFA) kasutades, mis kindlustab sooorituse tulemuslikkuse võrdlevhinnangute mitmekesisuse. Tulevikuperspektiivina on järgmiseks loogiliseks sammuks lisada mudelitele müra komponent, st rakendada SFA meetodit.

CURRICULUM VITAE

Name: TARMO PUOLOKAINEN
Place and date of birth: Estonia, 11 August 1988
Nationality: Estonian
Address: Uus 1-39, Tartu, Tartumaa, Estonia
Phone: (+372) 566 30 960
E-mail: tarmo.puolokainen@ut.ee

Language skills: Estonian, English, German, Finnish

Education:
2013 MA, Economics, University of Tartu

2010 BA, Economics, University of Tartu

Employment:
2012 – ... Centre for Applied Social Sciences, University of Tartu

2009 – 2013 Faculty of Economics and Business Administration, University of Tartu

Studies, research abroad:
2016 – 2017 Research Study Abroad Program, Centre for Efficiency and Productivity Analysis (CEPA), University of Queensland, Brisbane, Australia.

2016 Participating in project "Possibilities and limits, challenges and obstacles of transferring CEE EU pre-accession best practices and experience to Moldova's and Georgia's pre-accession process", Ivane Javakhishvili Tbilisi State University, Tbilisi, Georgia.

2015 Summer School on Productivity and Efficiency Analysis, The 14th European Workshop on Efficiency and Productivity Analysis, Aalto University, Helsinki, Finland.

2015 Summer Course on Theory and Practice of Efficiency & Productivity Measurement: Static & Dynamic Analysis, Universidad Santo Tomas, Santiago, Chile.

2014 I Summer Programme on Efficiency and Productivity Analysis, University of Rome "Tor Vergata", Rome, Italy.

2010 Exchange student (Economics), University of Helsinki, Helsinki, Finland.

Main research interests: Models in uncertainty, Data Envelopment Analysis, Stochastic Frontier Analysis, Index numbers, Performance measurement, Resource allocation, Decision-making, Policy evaluations, Data visualisation

ELULOOKIRJELDUS

Nimi: TARMO PUOLOKAINEN
Sünnikoht ja -aeg: Eesti, 11. August 1988
Kodakondsus: Eesti
Aadress: Uus 1-39, Tartu, Tartumaa, Eesti
Telefon: (+372) 566 30 960
E-mail: tarmo.puolokainen@ut.ee

Keeleoskus: Eesti, inglise, saksa, soome

Haridus:
2013 MA, majandusteadus, Tartu Ülikool

2010 BA, majandusteadus, Tartu Ülikool

Teenistuskäik:
2012 – ... Sotsiaalteaduslike rakendusuuringute keskus, Tartu Ülikool

2009 – 2013 Majandusteaduskond, Tartu Ülikool

Enesetäiendamine:
2016 – 2017 Külalisdoktorant, Centre for Efficiency and Productivity Analysis (CEPA), Queenslandi Ülikool, Brisbane, Austraalia.

2016 Osalemine projektis “Possibilities and limits, challenges and obstacles of transferring CEE EU pre-accession best practices and experience to Moldova’s and Georgia’s pre-accession process”, Ivane Javakhishvili Tbilisi Riiklik Ülikool, Tbilisi, Georgia.

2015 Tootlikkuse ja efektiivsuse suvekool, The 14th European Workshop on Efficiency and Productivity Analysis, Aalto Ülikool, Helsingi, Soome.

2015 Tootlikkuse ja efektiivsuse mõõtmise suvekursus, staatiline ja dünaamiline analüüs, Santo Tomase Ülikool, Santiago, Tšiili.

2014 Tootlikkuse ja efektiivsuse suvekool, Rooma “Tor Vergata” Ülikool, Rooma, Itaalia.

2010 Vahetustudeng (Majandusteadus), Helsingi Ülikool, Helsingi, Soome.

Peamised uurimisvaldkonnad: mudelid määramatuse tingimustes, andmeraja analüüs, stohhastiline piirianalüüs, indeksid, tulemuslikkuse mõõtmine, ressursside allokatsioon, otsustusprotsessid, mõjuhindamine, andmete visualiseerimine

DISSERTATIONES RERUM OECONOMICARUM UNIVERSITATIS TARTUENSIS

1. **Олев Раю.** Экономическая ответственность и ее использование в хозяйственном механизме. Tartu, 1994. Kaitstud 20.05.1991.
2. **Janno Reiljan.** Majanduslike otsuste analüütiline alus (teooria, metodoloogia, metoodika ja meetodid). Tartu, 1994. Kaitstud 18.06.1991.
3. **Robert W. McGee.** The theory and practice of public finance: some lessons from the USA experience with advice for former socialist countries. Tartu, 1994. Kaitstud 21.06.1994.
4. **Maaja Vadi.** Organisatsioonikultuur ja väärtused ning nende vahelised seosed (Eesti näitel). Tartu, 2000. Kaitstud 08.06.2000.
5. **Raul Eamets.** Reallocation of labour during transition disequilibrium and policy issues: The case of Estonia. Tartu, 2001. Kaitstud 27.06.2001.
6. **Kaia Philips.** The changes in valuation of human capital during the transition process in Estonia. Tartu, 2001. Kaitstud 10.01.2002.
7. **Tõnu Roolaht.** The internationalization of Estonian companies: an exploratory study of relationship aspects. Tartu, 2002. Kaitstud 18.11.2002.
8. **Tiia Vissak.** The internationalization of foreign-owned enterprises in Estonia: An extended network perspective. Tartu, 2003. Kaitstud 18.06.2003.
9. **Anneli Kaasa.** Sissetulekute ebavõrdsuse mõjurite analüüs struktuurse modelleerimise meetodil. Tartu, 2004. Kaitstud 15.09.2004.
10. **Ruth Alas.** Organisational changes during the transition in Estonia: Major influencing behavioural factors. Tartu, 2004. Kaitstud 22.12.2004.
11. **Ele Reiljan.** Reasons for de-internationalization: An analysis of Estonian manufacturing companies. Tartu, 2004. Kaitstud 25.01.2005.
12. **Janek Uiboupin.** Foreign banks in Central and Eastern European markets: their entry and influence on the banking sector, Tartu, 2005. Kaitstud 29.06.2005.
13. **Jaan Masso.** Labour Reallocation in Transition Countries: Efficiency, Restructuring and Institutions, Tartu, 2005. Kaitstud 7.11.2005.
14. **Katrin Männik.** The Impact of the Autonomy on the Performance in a Multinational Corporation's Subsidary in Transition Countries, Tartu, 2006. Kaitstud 29.03.2006.
15. **Andres Vesilind.** A methodology for earning excess returns in global debt and currency markets with a diversified portfolio of quantitative active investment models, Tartu, 2007. Kaitstud 13.06.2007.
16. **Rebekka Vedina.** The diversity of individual values and its role for organisations in the context of changes, Tartu, 2007. Kaitstud 16.11.2007.
17. **Priit Sander.** Essays on factors influencing financing decisions of companies: risk, corporate control and taxation aspects, Tartu, 2007. Kaitstud 19.12.2007.
18. **Kadri Ukrainski.** Sources of knowledge used in innovation: an example of Estonian wood industries. Tartu, 2008. Kaitstud 22.04.2008.

19. **Kristjan-Olari Leping.** Heterogeneity of human capital and its valuation in the labour market. Tartu, 2008. Kaitstud 14.05.2008.
20. **Kadri Männasoo.** Essays on financial fragility – evidence from the corporate and banking sectors in Central and Eastern Europe. Tartu, 2008. Kaitstud 26.05.2008.
21. **Made Torokoff.** Patterns of learning organisation – Estonian experiences. Tartu, 2008. Kaitstud 30.06.2008.
22. **Helena Rozeik.** Changes in ownership structures, their determinants and role in the restructuring of enterprises during transition: evidence from Estonia. Tartu, 2008. Kaitstud 31.10.2008.
23. **Jaanika Meriküll.** Technological change and labour demand. Tartu, 2009. Kaitstud 19.05.2009.
24. **Anne Aidla.** The impact of individual and organisational factors on academic performance in estonian general educational schools. Tartu, 2009. Kaitstud 18.06.2009.
25. **Alexander Gofman.** Experimentation-Based Product Development in Mature Food Categories: Advancing Conjoint Analysis Approach. Tartu, 2009. Kaitstud 21.09.2009.
26. **Anne Reino.** Manifestations of organizational culture based on the example of Estonian organizations. Tartu, 2009. Kaitstud 06.11.2009.
27. **Krista Jaakson.** Management by values: the analysis of influencing aspects and its theoretical and practical implications. Tartu, 2009. Kaitstud 12.11.2009.
28. **Eve Parts.** Social capital, its determinants and effects on economic growth: comparison of the Western European and Central-Eastern European countries. Tartu, 2009. Kaitstud 18.12.2009.
29. **Egle Tafenau.** Welfare effects of regional policy in the constructed capital model. Tartu, 2010. Kaitstud 22.03.2010.
30. **Epp Kallaste.** Employee workplace representation: an analysis of selected determinants. Tartu, 2010. Kaitstud 21.06.2010.
31. **Danel Tuusis.** Interest rate influence on the behavior of economic subjects. Tartu, 2010. Kaitstud 22.10.2010.
32. **Elina Kallas.** Emotional intelligence, organizational culture and their relationship based on the example of Estonian service organizations. Tartu, 2010. Kaitstud 17.11.2010.
33. **Dorel Tamm.** Alignment between the factors of the innovation process and public sector innovation support measures: an analysis of Estonian dairy processors and biotechnology enterprises. Tartu, 2010. Kaitstud 16.12.2010.
34. **Rasmus Kattai.** The links between private sector indebtedness and banking sector vulnerability: An Estonian case study. Tartu, 2010. Kaitstud 17.01.2011.
35. **Kurmet Kivipõld.** Organizational Leadership Capability and its evaluation based on the example of Estonian service organizations. Tartu, 2011. Kaitstud 4.05.2011.

36. **Janno Järve.** Downward Nominal Wage Rigidity in the Estonian Private Sector. Tartu, 2011. Kaitstud 21.06.2011.
37. **Kristina Toming.** The impact of integration with the European Union on the international competitiveness of the food processing industry in Estonia. Tartu, 2011. Kaitstud 21.06.2011.
38. **Andrus Kotri.** Customer experience evoking and management in services. Tartu, 2011. Kaitstud 26.08.2011.
39. **Andres Kuusik.** Segmentation of repeat visitors using passive mobile positioning data: customer loyalty based approach. Kaitstud 31.08.2011.
40. **Tuuli Pärenson.** Social impact evaluation in social enterprises in Estonia: need, readiness and practices. Kaitstud 21.09.2011.
41. **Indrek Saar.** Optimal alcohol taxation in Estonia. Kaitstud 25.11.2011.
42. **Kertu Lääts.** Management accounting change in a dynamic economic environment based on examples from business and public sector organizations. Tartu, 2011, 250 p.
43. **Reelika Irs.** Teacher performance appraisal and remuneration aspects of performance management on the example of Estonian general educational schools. Tartu, 2012, 322 p.
44. **Anne Lauringson.** The Impact of the Generosity of Unemployment Benefits on Estonian Labour Market Outcomes in a Period of Crisis. Tartu, 2012, 268 p.
45. **Peeter Peda.** The relationship between governance and performance in water services provision in Estonian municipalities. Tartu, 2012, 326 p.
46. **Andres Kuusk.** Financial contagion during times of crisis: a meta-analysis based approach with special emphasis on CEE economies. Tartu, 2012, 211 p.
47. **Kerly Espenberg.** Inequalities on the labour market in Estonia during the Great Recession. Tartu, 2013, 312 p.
48. **Xiaotian Zhang.** Internationalization processes of Chinese firms: The role of knowledge. Tartu, 2013, 274 p.
49. **Helen Poltimäe.** The distributional and behavioural effects of Estonian environmental taxes. Tartu, 2014, 141 p.
50. **Eneli Kindsiko.** Organisational Control in University Management: A Multiparadigm Approach on the Example of the University of Tartu. Tartu, 2014, 211 p.
51. **Diana Eerma**. A Bookkeeping approach to social accounting for a university faculty: The case of the University of Tartu. Tartu, 2014, 293 p.
52. **Kaia Kask.** Public sector real estate asset management models and their evaluation. Tartu, 2014, 264 p.
53. **Ott Pärna.** Managerial and contextual factors influencing innovation in information technology-based public sector services: an exploratory cross-national study. Tartu, 2014, 410 p.
54. **Merle Tambur.** Workplace bullying in Estonian organizations: The prevalence and causes. Tartu, 2015, 210 p.

55. **Sten Anspal.** Essays on gender wage inequality in the Estonian labour market. Tartu 2015, 195 p.
56. **Oliver Lukason**. Characteristics of firm failure processes in an international context. Tartu 2016, 180 p.
57. **Marko Viiding**. Role of electricity price in competitiveness of the manufacturing industry in liberalised electricity markets: the case of NordPool. Tartu 2016, 188 p.
58. **Bianka Plüschke-Altof.** Images of the Periphery Impeding Rural Development? Discursive Peripheralization of Rural Areas in Post-Socialist Estonia. Tartu 2017, 237 p.