Technische Universität Berlin

School VII Economics and Management

Efficiency Analysis in Public Service Provision: Addressing Characteristics and Specificities Related to the Public Sector and Regulation

vorgelegt von

Dipl.-Kffr.

Maria Nieswand

aus Dresden

Von der Fakultät VII - Wirtschaft und Management
der Technischen Universität Berlin
zur Erlangung des akademischen Grades
Doktor der Wirtschaftswissenschaften
- Dr. rer. oec. -

genehmigte Dissertation

Promotionsausschuss:
Vorsitzender: Prof. Dr. Georg Meran (TU Berlin)
Berichter: Prof. Dr. Christian von Hirschhausen (TU Berlin)
Berichter: Prof. Dr. Andreas Stephan (Jönköping Business School)
Tag der wissenschaftlichen Aussprache: 20. August 2012

Berlin 2012
D 83

Acknowledgements

My thesis-writing period yet has been one of the most meaningful periods of my life, both from an academic and personal perspective. I consider myself lucky to have had the opportunity to meet so many people who I have crossed paths with and who have influenced me in so many ways. I now have the opportunity to express my gratitude to some of them, which has always been a part of the thesis that I was looking forward to.

First, I would like to express my sincere gratitude to my supervisor Professor Christian von Hirschhausen, who has always supported my research and given significant impulses at just the right time. I am thankful for the opportunity to do independent research, for thoughtful comments, and for becoming part of a fabulous *cloud*. The last years have been full of wonderful experiences and I am aware of *wie gut wir es hatten*! *Grazie mille*!

I would like to thank Professor Andreas Stephan for giving me a deeper understanding of efficiency analysis during the InfraTRAIN workshop at TU Berlin at the very beginning of my research, for helpful suggestions and comments, and especially for agreeing to become my second supervisor during the last phase of my doctoral research.

I extend my thanks to my colleagues, who contributed in various ways to this thesis and made the environment I worked in an inspiring, lively and inimitable place. I especially owe my grateful thanks to Astrid Cullmann and Anne Neumann for tirelessly encouraging me to undertake this work, for endlessly supporting me throughout all stages of my thesis, for sharing their perspectives and insights, and for their friendship. Further, I would like to thank Pio Baake for each single question he raised, for sharing his knowledge, and for his never-ending patience. I thank my co-authors for fruitful collaborations and their contributions. Also, I gratefully acknowledge Felix Groba, Katharina Pijnenburg, Michael Weinhardt, Aleksandar Zaklan, Petra Zloczysti, and Michael Zschille.

Of course, one never travels alone. Without the support of my family and my dear friends, writing this thesis would have been a much more arduous path. My wholehearted gratitudes go especially to Beatrice, Birgit, Hanni, Marika, Miriam, Patrick, and Ralf. Thanks for all the things that cannot be captured in equations!

Abstract

The government has been and remains crucial in providing goods and services in modern-day economies. However, the manner of this provision is complex and touches upon numerous areas of economic and social policy. Government action plays an especially central role for services of general interest, as well as in the regulation of their provision by the private sector. Thus, the efficient and effective implementation of governmental economic activitiy, as well as of governmental intervention in the behavior of private-sector providers is of high importance. Against this background, there is both a growing interest in and need for an empirically sound assessment of public service provision. Scientific efficiency analysis provides a suitable analytical approach to gaining knowledge about production and cost structures, to evaluate the impact of their various determinants, as well as to quantitatively estimate the potential for efficiency gains. In order to achieve unbiased and meaningful results on the governmental, i.e. public, provision of goods and services it is imperative that the properties and specificities of public service provision are adequately considered in the analysis. This dissertation considers four selected properties and specificities of public service provision: (i) the multiplicity of public services; (ii) external factors that may affect service provision; (iii) subsidies where the provider of the subsidy is also the recipient; and finally (iv) the modeling of technologies in the context of regulation, where informational asymmetries and conflicting goals exist between the regulator and the regulated companies. In the dissertation we both suggest and implement appropriate procedures for the evaluation of public service provision by means of efficiency analysis, which allow for an adequate consideration of these properties and specificities.

Keywords: Efficiency Analysis, Data Envelopment Analysis (DEA), Order-m, Stochastic Frontier Analysis (SFA), Nonparametric Restriction Test, Public Service Provision, Local Transport, Natural Gas Transmission, Multiplicity of Outputs, Heterogeneity, Subsidies, Dimensionality

Zusammenfassung

Der Staat ist und bleibt entscheidend an der Bereistellung von Gütern und Dienstleistungen in Volkswirtschaften beteiligt. Dabei ist die Art der Beteiligung komplex und berührt zahlreiche Bereiche der Wirtschafts- und Sozialpolitik. Eine zentrale Bedeutung hat staatliches Handeln insobesondere bei der Bereitstellung von Gütern und Dienstleistungen der Daseinsvorsorge sowie für die Regulierung der privat organisierten Versorgung. Eine effiziente und effektive Umsetzung des wirtschaftlichen Handelns durch den Staat sowie der staatlichen Eingriffe in das wirtschaftliche Handeln privater Leistungsersteller ist daher von hoher Bedeutung. Vor diesem Hintergrund steigt das Interesse und die Notwendigkeit einer empirisch fundierten Bewertung der öffentlichen Leistungserstellung. Die Ansätze wissenschaftlicher Effizienzanalyse stellen ein geeignetes Analyseinstrumentarium dar, anhand dessen wichtige Kenntnisse über Produktions- und Kostenstrukturen gewonnen, sowie das Ausmaß des Verbesserungspotentials quantitativ geschätzt und verschiedene Einflussfaktoren evaluiert werden können. Um unverfälschte und aussagekräftige Ergebnisse über die staatliche, also öffentliche, Leistungserstellung zu erzielen, ist es notwendig deren Eigenschaften und Spezifitäten methodisch adäquat in die Analyse einzubeziehen. Die vorliegende Dissertation beschäftigt sich mit vier ausgewählten Eigenschaften und Spezifitäten der öffentlichen Leistungserstellung: (i) mit der Multiplizität öffentlicher Leistungen; (ii) mit den externen Einflussfaktoren auf die Leistungserstellung; (iii) mit Subventionen, bei denen der Subventionsgeber gleichzeitig der Subventionsempfänger ist; und schließlich (iv) mit der Technologiemodellierung im Regulierungskontext, bei der zwischen der Regulierungsbehörde und den regulierten Unternehmen Informationsasymmetrien und Zielkonflikte bestehen. Innerhalb der Dissertation werden geeignete Vorgehensweisen zur Evaluierung öffentlicher Leistungserstellung mittels Effizienzanalyse vorgeschlagen und durchfgeführt, die eine adäquate Berücksichtigung dieser Eigenschaften und Spezifitäten erlauben.

Schlüsselwörter: Effizienzanalyse, Data Envelopment Analysis (DEA), Order-m, Stochastic Frontier Analysis (SFA), Nicht-parametrischer Restriktionstest, Öffentliche Leistungsbereitstellung, Nahverkehr, Erdgasübertragung, Multiplizität von Output, Heterogenität, Subventionen, Dimensionalität

Contents

List of Tables

List of Figures

Abbreviations

AG .... *Aktiengesellschaft*
ASEAG ...... *Aachener Straï¿œenbahn und Energieversorgungs-AG*
bn .... billion
COM .... Commission of the European Communities
CRS .... constant returns to scale
DEA .... Data Envelopment Analysis
DGCL .... *Direction générale des collectivités locales*
DGP .... data generating process
DMU .... decision making unit
Dth .... Dekatherm
DTI .... Dominion Transmission, Inc.
EPNGC ...... El Paso Natural Gas Company
ESL .... efficiency stepladder
etc. .... *et cetera*
EWEPA ..... European Workshop for Efficiency and Productivity Analysis
FDH .... Free Disposal Hull
FEAR .... Frontier Efficiency Analysis in R
FERC .... Federal Energy Regulatory Commission
FTE .... full-time equivalents
GDP .... gross domestic product
GmbH .... *Gesellschaft mit beschränkter Haftung*
GRASP ...... Growth and Sustainability Policies for Europe
Hp .... horsepower
IEA .... International Energy Agency
iid .... independent and identically distributed
Inc. .... Incorporated
INFRADAY .. Conference for Applied Infrastructure Research
INSEE .... *Institut national de la statistique et des études économiques*
IPOC .... Iroquois Pipeline Operating Company as Agent/Iroquois Gas Transmission System, L.P.

IRDES .... *Institut de recherche et documentation en économie de la santé*
km .... kilometer
KVG .... *Kieler Verkehrsgesellschaft mbH, Kraftwagenverkehr Koblenz GmbH*
KViP .... *Kreisverkehrsgesellschaft in Pinneberg mbH*
KVS .... *Kreisverkehrsbetriebe Saarlouis GmbH*
L.L.C. .... Limited Liability Company
L.P. .... Limited Partnership
mbH .... *mit beschränkter Haftung*
mn .... million
NAPW .... North American Productivity Workshop
NBPC .... Northern Border Pipeline Company
np .... nonparametric kernel smoothing methods for mixed data types
OECD .... Organisation for Economic Co-operation and Development
OLS .... ordinary least squares
Opex .... operating and maintenance expenditures
PC .... principal component
PCA .... principal components analysis
RP .... random parameters
RPI-X .... Retail Price Index X
skm .... seat-kilometer
std.dev. ...... standard deviation
TGPC .... Transcontinental Gas Pipeline Corporation
thsd .... thousand
TRE .... true random effects
UK .... United Kingdom
US .... United States
USD .... United States Dollar
VRS .... variable returns to scale

Chapter 1

Overview

"From birth to death,
our lives are affected in countless ways
by the activities of government."
Stiglitz (2000, p. 3)

Growing up in a purely socialist country that later, in changed constitution, turned into social market economy, I experienced many versions and interpretations of what public service provision includes and how it is performed. However, things keep changing. One situation I still vividly remember is that, as a child, I had to use public buses everyday, to get to school and sport activities. These buses were made by Ikarus (and some still roam the streets today, more than 20 years later!). Unfortunately, the stop-request button that passengers had to press was installed out of reach for children; namely above the doors. The only way for me, as a child, to get off the bus was to ask other passengers to press the button for me. Happily, getting off buses today is less challenging, not only because I grew. Back then, I never could have imagined that buses would be the beginning of my research activities. During my course in public sector management at the Chair of Energy Economics and Public Sector Management in Dresden, I pondered my childhood memories as I spent countless hours investigating the capital costs of buses. Eventually, I dreamed of spreadsheets with these capital cost numbers. With this dissertation, I can confirm that I still like, and enjoy, spending time with data, spreadsheets and empirical research.

1.1 The Issue

Government is and will remain an important provider of goods and services. The means through which provision is delivered are generally organized along all tiers of (federal) governments and touch on significant aspects of social and economic policy. The governmental supply of goods and services is further referred to as public service provision and includes direct service provision accomplished by both (i) public bodies and (ii) state-owned entities, as well as indirect services provided through (iii) financing (private) service provision, and (iv) regulating private service provision by public authorities and institutions.[1]

The outlined scope of public activities indicate the great importance of achieving public service provision that is well performed. This is particularly relevant for services of general interest and for the sectors moving toward more private involvement and, therefore, regulation. Most recently, the European sovereign debt crises, beginning in late 2009, demonstrates and emphasizes the need for efficient governments, effective control mechanisms, and cost reductions for both public and private sector performance. Thereby, the crises also reignited an old discussion on shifting service provision to the private sector.

There is a profound skepticism regarding the efficiency of public service provision, where on the one hand a clear definition of efficiency is missing (Pestieau, 2009), and on the other hand empirical evidence is inconclusive[2] and by no means complete. The objective of this thesis is to provide empirical evidence on the efficiency of public service provision and on regulatory efficiency measurement by means of frontier methods in order to broaden the scientific and economic evidence. Frontier methods are suitable instruments in this context because they quantitatively determine relative efficiency measures of public services and regulated private firms while, at the same time, they are especially designed for situations where not all information is known or obtainable.

Frontier methods are experiencing increasing interest, in particular in the context of regulation, where they are applied by the regulating authority to implement incentive-based regulation schemes (Haney and Pollitt, 2009; Farsi et al., 2006a), thus helping to implement incentives for cost reductions by inducing efficient price decisions (Vogelsang, 2002). The basic idea of frontier analysis is to obtain unit-specific measures of efficiency by comparing each observation relative to a best practice frontier that is estimated from similar observed units. The efficiency measures can subsequently be used by the regulator to derive targets that units have to achieve in order to increase efficiency.

Frontier techniques are instruments by which competitive environments can be imitated and are, therefore, interrelated to the idea of yardstick competition,

[1] Throughout the dissertation, state, governmental and public service provision are used as synonyms.

[2] With respect to ownership see e.g., Lonti and Woods (2008)

a mechanism first introduced by Shleifer (1985). Due to its broad applicability, meaning that production units can include e.g., organizations, firms and decision making units (DMUs) in general (Bogetoft and Otto, 2011), frontier analysis is not limited to the consideration of regulated private firms. Applying them in the context of public service provision is hence feasible and useful for gaining insights into the production and cost structures through comparative analysis.

However, considering the public service provision in the framework of frontier analysis requires an attentive implementation because the public service, as such, and the related questions that may arise, introduce various characteristics and specificities. From a methodological point of view these need to be taken into account appropriately to obtain meaningful and accurate efficiency estimates.

This thesis focuses on four of these characteristics and specificities: (i) the multiplicity of outputs in governmental service provision; (ii) the introduction and statistical evaluation of external factors that are not under the control of the DMU, but may impact its efficiency and further its choice of inputs and outputs (heterogeneity of framework conditions); (iii) the subsidization of public sector services where the government is coincidently the source and the receiver of subsidies; and (iv) the information asymmetry existing between the regulating authority and the regulated firms.

A central contribution of this thesis is to propose approaches that appropriately incorporate these characteristics and specificities in frontier analysis and provide empirical evidence on the performance of public service provision. The approaches and results developed in this thesis can, therefore, strengthen the evaluation of public service providers, the conclusion of policy implications, the identification of efficient production and cost levels, and the improvement of instruments for incentive-regulation.

This thesis unfolds as follows: the subsequent subsections of Chapter 1 give an overview with Section 1.2 outlining the increasing importance and consequences, market-oriented policies experience in European and Latin American economies since the 1970s, as well as giving the theoretical arguments that justify privatization and presenting the inconclusive findings of empirical research with respect to the performance comparison of private and public service provision; Section 1.3 introducing the main ideas of the used methodological approaches to efficiency analysis; Section 1.4 exposing the characteristics and specificities of public service provision that the thesis addresses; Section 1.5 presenting the contributions of the thesis; and Section 1.6 briefly giving an outlook. Chapters 2 and 3 are devoted to the performance estimation and explanation of local public services provided by local governments, more precisely by French *départements* and German municipalities, respectively. In Chapter 4 the impact of deficit-balancing subsidies on the cost efficiency of German local public bus companies is evaluated whereby the public owners pay transfers to companies (partly) owned by themselves. Chapters 5 and 6 deal with the information asymmetry regarding the specification of the technology in the context of regulation. Each chapter separately presents the methodological details, the relevant literature and ends with concluding remarks.

1.2 The Return of Public (Local) Service Provision

In mixed economies both the private and the public sectors provide goods and services that affect daily routines. The economic role of governments has been the subject of intense debate among politicians, economists, social scientists and thinkers of all shades ever since. As of today, theory suggests that state intervention in existing markets is mainly justified by four categories of market failure, i.e. natural monopolies (indivisibility), external effects, information asymmetries and shortcomings in adjustment; in these cases public self-provision and the regulation of private firms are two instruments to achieve allocative efficiency, which is deteriorated by market failure; see e.g., Fritsch et al. (2006).

Since the 1970s, neoliberalism, in particular, has influenced economic policy and thinking (Harvey, 2007). Thus market-oriented policies involving extensive privatization of state assets has shaped the constitution of many economies worldwide. Neoliberal thinking postulates a minimalist role for the government in economic activities where state interventions are limited to instances of market failure and economies are fundamentally structured with free markets, free trade, property rights and the individual freedom. When governments are intensively involved in economic activities, economic history shows that the privatization of state enterprises has traditionally been a crucial ingredient in reducing the government's economic activities and in promoting the private sector. In particular the economic reforms undertaken by the administration of Margaret Thatcher, British prime minister from 1979 to 1990, involved massive privatization of state-owned assets.

Thatcher's privatization program shows how governmental activities are not only withdrawn from firms peripheral to public sector activities but also from major public utilities and social provision (Wolfe, 1991; Harvey, 2007). The program privatized e.g., the state-owned sectors of gas, oil, water and bus services, as well as British Telecom and British Airways. Eventually even health care and social housing was privatized. By 1991, about half the state-owned assets had been sold to private investors (Marsh, 1991). Thatcher's market-oriented reforms were then perceived as successful and, consequently, spurred similar other privatization programs during the late 1980s and 1990s by many European countries, including France, Germany, Spain and Italy (Netter and Megginson, 2001).

Outside of Europe, privatization has played also an important role in Latin America. Responding to the debt crisis in the 1980s, many Latin American countries implement economic reforms inspired by the propositions of the Washington Consensus.[3] The ten reform recommendations formulated by the Washington Consensus effectively involve a notable shift toward market-oriented policies (Gore, 2000) and intend to restore the economic growth of the countries in Latin America (Kuczynski and Williamson, 2003) by fostering e.g., deregulation and privatization.

The general reasons to justify the reduction of governmental economic activities are multiple. Margaret Thatcher uses privatization to realize the ideal of self-responsibility, to develop national capital markets and to reduce the power of trade unions (Harvey, 2007). Other reasons, such as income and wealth redistribution and public finance improvement, are summarized by e.g., Yarrow (1986). A central argument favoring the private provision of goods and services is the gain in efficiency. On the one hand, Stiglitz (1999) argues that a limited scope of governmental service provision sharpens the focus of governmental activities and, hence, increases their efficiency. On the other hand, it is commonly assumed that privately owned firms perform better, i.e. more efficient, than those operated by the government.

The latter argument, stating that private production is superior to public production with respect to efficiency, is supported across multiple fields of theoretical research, e.g., industrial organization, agency theory, public choice, organizational theories, welfare economics, corporate and public finance, law and economics and macroeconomics (Netter and Megginson, 2001; Villalonga, 2000). However, in this respect the empirical research does not provide consistent evidence for the general superiority of private ownership. Surveys of the empirical literature related to the performance of privatized firms and the performance comparison between private and public firms include e.g., Arcas and Bachiller (2010), Estrin et al. (2009), Megginson and Sutter (2006), Djankov and Murrell (2002), Willner (2001), Netter and Megginson (2001), Pestieau and Tulkens (1993), and Pommerehne (1990). Referring to the ownership of railroad sector in Canada, Caves and Christensen (1980)

[3]The Washington Consensus is first named as such by the economist John Williamson in 1989 (Williamson, 1990), who intends to reflect the common conception among the Washington D.C.-based policy advising institutions at that time with respect to economic policy instruments that are believed able to stabilize the Latin American economies.

conclude that if state-enterprises are subject to competition and performance-based control mechanisms, the performance differences between them and private firms shrink.

Investigating a total of 15 surveys, including a few of the aforementioned, Mühlenkamp (2012) emphasizes that efficiency, which is used to evaluate the impact of ownership on performance, is defined in manifold ways; sometimes even, with respect to the purpose, in meaningless ways. Further, Mühlenkamp argues that the surveys differ significantly with respect to e.g., the underlying composition of the population and the data characteristics. Particularly interesting in the context of the work at hand is the identified difference in methodology: earlier work predominantly uses simple regression models while more recent work applies modern approaches of efficiency analysis, i.e. frontier analysis. According to Mühlenkamp, private firms do no longer perform more efficient than public entities when newer research, which uses cross-sectional data (instead of longitudinal data) and deals with the concepts of cost and production efficiency, is considered.

In addition to the empirical findings, two relevant lessons can be learned from Thatcher's privatization program and the Latin American experience. Among other reasons, the failure of the privatization policy conducted in Latin America in the 1990s is commonly attributed to an insufficient preparation of the transfer process and to the missing regulatory and competition policies, which are ignored by the Washington Consensus (Stiglitz, 1999). Margaret Thatcher in contrast, adequately prepared the privatization of state-owned entities (Harvey, 2007). However, the case of Britain also demonstrates that markets in the absence of proper incentive schemes do not always yield policy-desired results. After privatization, the quality of services decreased, while prices increased and the whole process was at taxpayers expense (Marsh, 1991, and references therein). Stiglitz (1999) emphasizes that although powerful regulation could overcome such undesired outcomes, it does not necessarily guarantees the achievement of broader public objectives.

The concern of broader public objectives underpins arguments that justify the involvement of governments, at least and in particular, in the provision of the fundamental services, e.g., basic education, roads, health care etc. Further, Yarrow (1986) emphasizes that ownership is less important than the regulatory and competitive environment in which service providers are embedded. To summarize, the economic role of governments, e.g., either in terms of self-provision or regulation, remains significant.

Competitive environments appear to be a key ingredient for realizing governmental activities that yield efficient service provision. These can be implemented through appropriate mechanisms, as for instance the yardstick competition proposed by Shleifer (1985), making competition and regulation policies workable, and therefore, making state intervention on the one hand, and state activities on the other, effective, as claimed e.g., by Stiglitz (1999). Frontier analyses provide a versatile instrument for generating information in many contexts of public service provision that is useable for benchmarking and yardstick competition. Rather than comparing private and public service provision, the dissertation at hand provides empirical analysis on efficiency measurement in public service provision of different kinds.

1.3 Efficiency Analysis

Commonly, partial productivity ratios, e.g., the produced number of output quantities over the number of employees, are used to make a relative comparison of the firms' success in converting their resources into output. However, when multi-dimensional production processes are considered, such key performance indicators can not only result in misleading implications, known as the Fox theorem, but also assume that the investigated technology exhibits constant returns to scales (CRS, Bogetoft and Otto, 2011). Therefore, it is desirable and necessary to use a concept that is flexible enough e.g., to capture multi-dimensionality and various technology characteristics.

Frontier analysis, based on the seminal work of Debreu (1951), Koopmans (1951) and Farrell (1957), provides an alternative, complementary concept. According to Bogetoft and Otto (2011), each productive unit can be described by its employed production plan, i.e. the input-output-combination. The basic idea of frontier analysis is to derive a reference performance, i.e. a best-practice frontier, from a given set of different input-output-combinations, against which each observation is compared to. The distance to that frontier is then interpreted as the waste of resources or the omission of potential outputs and provides a measure of inefficiency or the degree of efficiency, respectively. Excellent introductions to the topic of efficiency analysis are e.g., given by Coelli et al. (2005) and Bogetoft and Otto (2011).

The production plan, in turn, can be represented by a vector of inputs and outputs; more formally $x^i \in \mathbb{R}^p_+$ and $y^i \in \mathbb{R}^q_+$ denote the vectors of p inputs and q outputs where the production possibility set, denoted by Ψ, contains all feasible input-output-combinations that are available to production unit i. The efficient boundary of Ψ, i.e. Ψ^δ, represents the frontier that collects the optimal production plans and is also referred to as the production or technology frontier (Simar and Wilson, 2008). One option to determine the unit-specific efficiency is to consider the Debreu-Farrell efficiency measure that gives the radial distance of a particular observation to its corresponding frontier (Daraio and Simar, 2007a). An alternative and intimately linked measure of efficiency is the so called Shepard efficiency measure proposed by Shepard (1970). Throughout this dissertation, the input-oriented version of the efficiency measures is used, assuming that for a given level of output, which is fixed, the inputs, either in monetary or physical terms, need to be minimized for being considered efficient.

Since Ψ is generally unknown, it needs to be estimated from a (random) sample of n observed productive units $\mathcal{X} = \{(x_i, y_i) \mid i = 1, \ldots, n\}$ in order to obtain Ψ^δ and to estimate efficiency scores; see e.g., Simar and Wilson (2011); Daraio and Simar (2007a). For this purpose parametric and nonparametric frontier models are available where the first one mentioned assumes a functional form to define Ψ while the latter does not.

1.3.1 Nonparametric Approaches

1.3.1.1 Free Disposal Hull (FDH) and Data Envelopment Analysis (DEA)

The two most well known representatives of nonparametric estimators are the Data Envelopment Analysis (DEA) estimator that was initially proposed by Farrell (1957), and the Free Disposal Hull (FDH) estimator, first proposed by Deprins et al. (1984). DEA is later operationalized by Charnes et al. (1978) for CRS technologies and extended by Banker et al. (1984) to variable returns to scale (VRS).

Both estimators are deterministic frontier models meaning that all observations are assumed to belong to the attainable production possibility set Ψ (Simar and Wilson, 2008). Further, they rely on few properties that Ψ needs to satisfy. The only assumption FDH imposes on Ψ is free disposability while DEA additionally assumes convexity of Ψ. To construct the frontier, FDH and DEA employ mixed integer and linear programming, respectively. As a results of their deterministic nature, both FDH and DEA estimators are sensitive toward extreme values and outliers (Simar, 2003) and, therefore, single observations can have a disproportionally large impact on the efficiency score of others.

Today, the statistical properties of FDH and DEA are well-known; for a recent review see Simar and Wilson (2008) and the references therein. Compared to parametric estimators being usually n-root consistent, the nonparametric estimators, including FDH and DEA, achieve typically lower convergence rates, i.e. the speed with which in probability the estimator converges to the true parameter. The convergence rates of FDH and DEA estimators depend on the number of inputs and outputs and shrink when the dimensionality of the problem is increased (Kneip et al., 1998; Park et al., 2000). This issue is commonly referred to as the 'curse of dimensionality'. Consequently, applying these estimators requires large data sets in order to obtain meaningful results.

1.3.1.2 Order-m Estimator

Aiming to overcome the sensitivity of FDH and DEA to outliers, Cazals et al. (2002) propose the order-m estimator, which is based on an alternative formulation of the production process, i.e. the probability distribution function (joint probability function). This approach uses the concept of the expected minimum input function, or the expected maximum output function, respectively, to define the benchmark against each observation of interest is compared to. Thereby, the expected minimum (maximum) value for the unit in question is derived by considering the inputs (outputs) used (produced) by m randomly drawn units that produce at least the same output (use equally much or less inputs). Unlike, FDH and DEA estimators that construct full frontiers, the order-m approach involves the concept of partial frontiers due to constructing the frontier using a subset of m observations (Simar and Wilson, 2008).

Since the expected value is an average of the considered m units, the order-m estimator is robust against outliers and extreme values and provides a less strict but as reasonable benchmark. Further, the order-m efficiency scores are no longer bounded to unity and, therefore, allow units to be considered as super-efficient when they, on average, perform better than their reference set, i.e. the m randomly drawn units (De Witte and Kortelainen, 2009). Cazals et al. (2002) show that although the order-m estimator converges to the FDH estimator when m increases, particularly in finite samples it remains more robust.

1.3.1.3 Truncated Bootstrap Regression and Conditional Efficiency Analysis

The means by which heterogeneous factors that influence the activity of the productive units are incorporated in nonparametric efficiency analysis have been extensively discussed and constitute a constant research theme; for a comprehensive overview see Daraio and Simar (2007a).

Frequently, the family of two-stage approaches is used to control for environmental factors and to explain the performance differences. Typically, these approaches regress the nonparametrically obtained efficiency scores on a set of environmental variables. One recently proposed two-stage methodology is the semi-parametric bootstrap truncated regression model developed by Simar and Wilson (2007). According to the authors, this techniques overcomes, simultaneously, some important drawbacks of previously applied estimation approaches (e.g., Tobit regression) and allows valid statistical inference of the environmental variables' coefficients. The estimation approach by Simar and Wilson (2007) is able to take into account that estimates obtained by linear programming are truncated (censored) and might be serially correlated. One important assumption of the method is the separability condition between explanatory variables and input-output-space, meaning that the environmental variables do influence the efficiency but not the attainable production possibility set Ψ and, therefore, not the frontier's position (Daraio and Simar, 2007a).

When introducing the probabilistic production formulation and the order-m estimator, Cazals et al. (2002) propose the conditional efficiency approach to incorporate exogenous factors, i.e. environmental variables, into the efficiency estimation procedure. Unlike second-stage regressions, the method neither requires making parametric assumptions regarding the relationship between efficiency score and the environmental variables nor does it rely on the separability condition (De Witte and Kortelainen, 2009; Daraio and Simar, 2007a). To control for the environmental factors this estimation technique involves conditioning the production process on a set of environmental variables, which requires smoothing techniques, e.g. Kernel estimators.

De Witte and Kortelainen (2009, 2013) suggest to use a truncated mixed kernel function combined with a data-driven bandwidths selection approach. This broadens the applicability of conditional order-m estimation by allowing the inclusion of ordered and unordered variables, in addition to continuous variables, while the bandwidths are estimated for each observation individually. Moreover, the bandwidths selection procedure introduced by De Witte and Kortelainen avoids that the separability condition implicitly enters the conditional estimation through the chosen bandwidth parameters that are a crucial part in Kernel estimation. Daraio and Simar (2005) and Daraio and Simar (2007b) extend the initial conditional order-m approach to the full multivariate, hence full frontier, case (conditional FDH efficiency measure) and to convex technologies (conditional DEA efficiency measure), respectively.

1.3.2 Parametric Approaches

Stochastic Frontier Analysis (SFA), also referred to as the composed error regression approach (Cooper et al., 2007), is a parametric and stochastic approach to efficiency measurement, which originates from the stochastic frontier production model independently proposed by Aigner et al. (1977) and Meeusen and van den Broeck (1977). Relying on the theoretical premises of production and cost functions, which represent an ideal, econometric frontier estimation makes the empirical implementation consistent with the underlying theoretical proposition that no observed DMU can exceed this ideal (Greene, 2008).

Beside imposing a functional form on the transformation process used for converting inputs into outputs, stochastic frontier models allow introducing stochastic elements, e.g., statistical noise, into the estimation. The regression error term in stochastic frontier models, therefore, contains two components, i.e. the inefficiency term and the random error term. For both, distributional assumptions are made; for the different possibilities of distributional assumption see e.g., Greene (2008). Since the components cannot be estimated independently, the extraction of inefficiency estimates requires disentangling both components, which is commonly achieved by the conditional mean estimator introduced by Jondrow et al. (1982).

Additionally to cross-sectional frontier models, comprehensively surveyed e.g. by Kumbhakar and Lovell (2000), SFA provides models for panel data that are able to capture information that change over time. Among the group of panel frontier models, one can identify (i) random and fixed effects models, where the inefficiency is either treated as a random variable or a fixed parameter;[4] (ii) time-invariant and time-varying inefficiency models; (iii) models that account for observable heterogeneity as a component of the technology and/or as a component of the inefficiency term (Coelli et al., 2005); (iv) models that account for unobserved heterogeneity;[5] and (v) models that assume either common or heterogeneous technologies. For overviews, theoretical details, and empirical comparisons of the respective models, further see e.g., Coelli et al. (2005), Greene (2008), Coelli et al. (1999), Farsi et al. (2005a) and Farsi et al. (2006a).

[4]Greene (2008) in addition mentions Bayesian approaches to econometric efficiency analysis.

[5]The term unobserved heterogeneity is usually associated with time-invariant characteristics that are possibly related to the already included variables but are distinct from unobservable inefficiency (Greene, 2008).

1.3.2.1 *True* Random Effects (TRE) and Random Parameters (RP) Model

As in the nonparametric case, controlling for heterogeneous operating conditions is essential for obtaining unbiased and meaningful performance measures and for deriving reliable policy implications. In addition to observable heterogeneity, unobservable heterogeneity might be present. If not accounted for, both types of heterogeneity are likely to deteriorate the estimated efficiency estimates.[6]

The *true* random effects (TRE) model, proposed by Greene (2004, 2005b), is a parametric stochastic frontier model that disentangles unobserved and time-invariant heterogeneity from the compound error term by introducing a third stochastic term. By simulated maximum likelihood estimation, the unobserved heterogeneity can be separated from inefficiency.

The model of Greene (2004, 2005b) assumes a technology commonly shared by all DMUs. Depending on the context, this assumption seems rather strict, since it would imply for example that all DMUs have the same optimal cost shares of inputs. The random parameters (RP) model permits randomizing other coefficients of the production or cost function, in addition to the error terms. Both econometric models, i.e. the TRE model and the RP model, are closely linked while the TRE can be interpreted as a special case of the RP model (Greene, 2005a) with only the constant randomly shifting the technology.

1.3.2.2 Introducing Heteroscedasticity

Heterogeneity can be introduced by allowing the statistical error and/or the inefficiency term to be heteroscedastic (Greene, 2008), i.e. the standard deviation (std.dev.) of inefficiency is allowed to be non-constant. Neglecting heteroscedasticity can yield biased and even inconsistent parameter estimates (Bhattacharyya et al., 1995; Hadri, 1999). Caudill et al. (1995) emphasizes that the issue of heteroscedasticity is even more serious in frontier than in mean regression models because an increasing std.dev. changes the frontier.

DMU-internal factors, i.e. factors the DMU can control, reflect the managerial characteristics and might cause inefficiency (Bhattacharyya et al., 1995). When considering the ability of public service providers, these are of particular interest because they provide insights into public decision making. According to e.g., Hadri et al. (2003) and Hadri (1999), nonstochastic firm management characteristics are associated to the std.dev. of the inefficiency term whereas the variance of the statistical error term is affected by size-related heteroscedasticity.

[6] Some variation in the data might also be captured by the statistical error term.

1.4 Characteristics and Specificities of Public Service Provision

Considering public sector performance and regulation of private firms by means of frontier techniques involves characteristics and specificities, which are introduced by the objective itself and the related research questions. The applied methodology (modeling approach) must be able to control for these characteristics and specificities, because only then is the empirical evidence reliable, comprehensive and able to answer the intended questions. The four aspects considered in this dissertation are described in the following.

1.4.1 Multiplicity of Outputs

Unlike the traditional analysis of one-product firms, analyzing the governmental service provision induces the consideration of a wide range of responsibilities and tasks, and therefore multiple outputs. In many countries, the accomplishment of services is allocated among the different administrative layers. The sub-national tiers are of particular concern, because, in general, the provided services are important matters of local interest[7] while at the same time the budgets of local governments are frequently stressed. Further, their decision making with respect to service provision is not entirely independent from higher administration. Local governments undertake, and potentially pursue, multiple objectives (Fletcher and Snee, 1985; Pestieau, 2009). Particularly, the latter dissociates public service providers from private firms in many cases.

The challenge of empirically evaluating public sector performance is threefold: First, the analytical framework must be able to capture the multiplicity of outputs in order to derive meaningful results. In general and emphasized among others, e.g., by Balaguer-Coll et al. (2007), the governmental production process is complex and, like its inputs and outputs, difficult to model. Therefore, the evaluation and estimation of public service performance requires an analytical framework that can appropriately represent the governmental transformation of inputs into outputs, while avoiding the unnecessary imposition of untenable or unexaminable assumptions.

Second, the outputs need to be identified. Following Bradford et al. (1969) one could distinguish outputs that are directly produced, and the outputs that

[7] For example, are Spanish municipalities responsible for public street lighting (Balaguer-Coll et al., 2007), French *départements* for road construction (Nieswand and Seifert, 2011) and Portuguese municipalities for waste-collection (Afonso and Fernandes, 2006).

citizen-consumers are primarily interested in. Similarly, De Witte and Geys (2011) consider the provision of many public services as a two-step process of production and discriminate between demand-independent service potential and demand-dependent observed outputs. The alternative interpretations and definitions of outputs (and inputs) underlines the general problems associated with representing the transformation process of administrative units.

Third, and closely related to the identification of outputs, is output measurement. Once identified, outputs can be represented either by direct measures or rough proxies, the latter e.g., used by De Borger and Kerstens (1996a). However, even if direct measures are available, e.g., for educational services the number of lessons taught (Loikkanen et al., 2011), the pupils enrolled (Geys et al., 2010), the pupil exam performance (Giménez and Prior, 2007; De Witte and Kortelainen, 2009), and the population in the relevant age group (Kriese, 2008) are used, the contained information of the alternatives can vary considerably.

1.4.2 Heterogeneity of Operating Conditions

The purpose of addressing the operating environments in performance measurement is at least twofold: First, whenever pure measures of inefficiency are required, which is for example the case in incentive-based regulation, the heterogeneous operating conditions need to be controlled for, since they are likely to influence the success in producing output, but cannot be directly controlled by the unit under consideration. Taking these into account is necessary to obtain justified, reasonable, attainable and enforceable improvement targets. Second, researchers, politicians and other interest groups are frequently concerned about what impact particular conditions and policies have on the efficiency of output production in order to explain performance variations and in order to derive corresponding policy implications.

In general, the heterogeneity of the operating environment can be represented by two categories of variables. The first category refers to factors that only affect the degree of efficiency, i.e. the estimated distance to the best-practice frontier, whereas the second category refers to factors that in addition might influence the set of inputs and outputs, i.e. the production possibilities, and hence the position of the frontier; see e.g., Daraio and Simar (2007a) and De Witte and Kortelainen (2009). The assumption that environmental variables influence the inefficiency but not the production possibilities is also referred to as the separability condition; see e.g., Daraio and Simar (2007a). Further, the data type through which the external conditions are represented varies and includes continuous, e.g. population density and unemployment rates, as well as ordered and unordered discrete variables, e.g., income class and the location in one of the Federal States, respectively.

Each of the mentioned aspects, i.e. the type of influence and the type of data, requires adequate modeling. Due to recent methodological advances, e.g., the bootstrap truncated regression and the conditional efficiency analysis, frontier analysis is able to capture these aspects and, therefore, allow pursuing the two described purposes of controlling for the environmental factors.

1.4.3 Subsidies

The subsidization of otherwise unprofitable sectors is an important policy instrument to ensure the provision of essential public services such as local public bus transportation. With respect to public services, the necessity of financially supporting these sectors emerges, beside the economic reasons, predominantly from the social responsibility and social objectives of governments; see e.g., Vickrey (1980) in the context of public transport. However, subsidies are also means of compensating for the denial of profit making and of fostering competitiveness.

Subsidies differ in their characteristics; some of them are earmarked while others are not. To discuss the effects of subsidies on the performance and, more precisely, on efficiency, it is worth clarifying that the recipients of financial support payments are not necessarily restricted to private firms providing public services, but rather additionally include state-owned entities. Beside the phenomenon of subsidy-maximizing behavior, which is independent from ownership and also related to the separability condition mentioned above, a distinct situation is given when the government is service provider and simultaneously the source of subsidization. For sectors in which the operating companies are mainly publicly owned, this situation becomes highly relevant because such subsidies might yield inefficiency that can be identified among state-owned firms without comparing them to private firms. Additionally, in this case, the subsidy appears to be influenceable by the company itself making the variable endogenous.

The analysis of endogenous variables requires a corresponding estimation approach since the endogeneity introduces biases and, therefore, affects the consistency of estimators. The parametric framework of efficiency analysis, for example, provides an adequate approach by introducing the endogenous variable as a heteroscedastic variable.

1.4.4 Uncertainty and the Curse of Dimensionality

One main goal of analyzing alternative technology specifications is to make incentive-based regulation relying on efficiency analysis a powerful, effective and trustworthy instrument. Beside actively supplying public goods and services, either through administrative bodies or state-owned entities, the government is, as the regulator, also involved in the provision of goods and services by private firms. Applying frontier methods allows the regulator to control private production and to pursue and implement regulatory objectives, e.g., reducing monopolistic power and promoting the efficient use of resources. Thereby, regulators frequently use nonparametric frontier models, e.g., DEA in order to establish benchmarks for target determination (Haney and Pollitt, 2009).

However, regulatory benchmarking involves two main challenges: (i) the small number of observable firms, which is due to the monopolistic market structure regulated sectors naturally have; and (ii) the information asymmetry existing between the regulator and the regulated firms. Both aspects are relevant when nonparametric frontier methods are applied because their empirical consequences can yield deteriorated improvement targets and, therefore, make efficiency analysis an ineffective regulatory tool. The methodological problem is that nonparametric estimators have lower convergence rates than their parametric alternatives; see e.g., Simar and Wilson (2008), meaning that the probability of overestimating the firm's performance increases when, for a given number of observations, the number of variables representing the technology of the firms is increased. This offers a problem for the regulator: the regulator aims to model the production process as closely as possible while each additional variable increases the risk of overestimating the firms' performance.

To overcome this issue of regulatory benchmarking, approaches can be employed that improve the modeling of the technology sets.

1.5 Contribution of this Thesis

The evaluation of public services provision and the improvement of regulatory control mechanisms are crucial ingredients for making both private service provision and governmental activities more efficient and effective. Efficient governmental service provision is particularly important, because even powerful regulation may not guarantee the achievement of broader public objectives (Stiglitz, 1999) and, hence, government remains an essential service provider that does not completely withdraw its economic involvement.

This thesis thereby focuses on four characteristics and specificities related to public service provision by developing approaches to take these into account when the performance of public and regulated private service providers is evaluated by frontier analysis. Empirical results are only meaningful, reliable and useful for policy implications if the method can account for these characteristics and specificities. First, the efficiency of public spending in France and Germany is analyzed taking output multiplicity and heterogeneity of public service provision into account (Chapters 2 and 3). Subsequently, Chapter 4 considers the performance of publicly-owned German local public bus companies in terms of cost efficiency. It focuses on the impact of subsidies that are paid by the public owners who are also the operators. The two last chapters (Chapters 5 and 6) are devoted to the improvement of model specification in regulatory benchmarking and, therefore, to the improvement of efficiency estimation used to implement incentive-based regulation schemes. Considering the service provision of natural gas transmission, both chapters address the issue of information incompleteness between regulator and regulated firms and the curse of dimensionality.

1.5.1 What Drives Intermediate Local Governments' Spending Efficiency: The Case of French *Départements*

In Chapter 2, we analyze the effects of non-manageable external framework conditions on the spending efficiency of French *départements* that are intermediate local governments and have not been investigated by means of frontier analysis before. France is an example of ongoing efforts undertaken by the government to restructure the allocation of governmental competencies toward its intermediate administrative tier. Starting in the 1980s, the importance of the *départements* have been increased in waves by transferring additional tasks to them.

In a first step, the spending efficiency of 96 *départements* in metropolitan France in 2008 is estimated using DEA. We use total expenditures as the measure of input and consider five output variables: the number of inhabitants, of beneficiaries, of beds, and of pupils as well as the municipal road kilometers (km). The results indicate spending inefficiencies among *départements* averaging between 10 and 22 percent, depending on model specification.

In a second step, the obtained DEA efficiency scores are used to test whether the performance partly depends on geographical and socio-economic factors that are not under the control of the *départements*. For this purpose we apply the bootstrapped truncated regression approach proposed by Simar and Wilson (2007).

The analytical framework imposes the separability condition implying that external variables have no impact on the production possibilities, i.e. the choice of production plans.

The second-stage regression reveals that the distance to the national capital, the median income of households and the share of elderly inhabitants significantly influence the efficiency of public service provision. In contrast we find no empirical evidence for the *départements*' size and the dummy controlling for seaside location to explain inefficiency. In general, our analysis demonstrates that the framework conditions can favor and deteriorate, respectively, the performance of local governments. Thus, depending on the context and details of the analysis' purpose, the efficiency estimation requires appropriately taking into account the external factors in order to obtain reliable performance measures and reliable information on the influence these factors have on the performance. The latter can be particularly relevant when policy instruments are planned to be implemented.

1.5.2 Conditional Efficiency Analysis of Municipal Service Provision in Germany

Chapter 3 examines public service provision by German municipalities. In contrast to the previous chapter, the included environmental variables are not only not under the control of local governments but additionally are allowed to influence the production possibilities; hence, the separability condition is dismissed. Furthermore, without leaving the nonparametric framework, the environmental variables are allowed to be of discrete data type and are no longer restricted to the continuous variables. Both, the enlarged data types and the removal of the separability condition are important for the evaluation of local government, because it broadens the insights into how and which external factors' influence the governmental performance and makes the efficiency estimation more substantial and accurate. To the best of our knowledge, this ensemble has not been considered before.

The empirical analysis employs a cross-sectional data set covering 1,060 German municipalities, located in nine of the 16 Federal States for the year 2005. The municipal efficiency is estimated by the conditional order-m approach according to the methodology proposed by De Witte and Kortelainen (2009, 2013), who extend the ideas of Cazals et al. (2002) and Daraio and Simar (2005). The enhanced conditional order-m estimation allows deriving robust efficiency measures while simultaneously taking the heterogeneous operating environments of municipalities into account. The framework conditions are expressed by continuous as well as ordered and unordered discrete factors. Further, separability between external variables and the attainable production set is neither imposed during efficiency estimation nor for the selection of bandwidths used to obtain the kernel functions.

We use the total current expenditures to measure input and the total surface area, the public thoroughfare, the population, and the number of school buildings, kindergarten places, and sport clubs to measure output. To evaluate the variation in the municipal performance, we condition the efficiency estimation on a set of environmental variables and subsequently test by means of nonparametric regression methods whether their impact on efficiency is of statistical significance. We find that the average performance increases about 5 percent when taking the heterogeneity in operating conditions into account. All considered environmental variables are statistically significant. Our results suggest that the effect of unemployment on efficiency is mixed. Municipalities perform better when commercial activities are higher while the efficiency decreases with higher population density and higher classes of income. Furthermore, we find that the performance variations can be partly explained by the location, i.e. the Federal State in which the municipality is located whereas the population movement has no influence.

1.5.3 Cost Efficiency and Subsidization in German Local Public Bus Transit

In Chapter 4, the analysis of public sector efficiency is conducted within the context of state-owned firms supplying services of local public bus transit. Since this sector is traditionally unprofitable, it requires subsidies and yet the empirical evidence implies that they can have a harming effect on costs and possibly also on the performance of operators. More precisely, the analysis focuses on the impact of deficit-balancing subsidies on the cost inefficiency of local public bus companies in Germany, where a complex system allocates ample financial support. The specificity of these subsidies is that they are not earmarked and payed by the owners, who are in our case predominantly local governments, to make sure that the operator does not make a deficit; therefore we assume these subsidies to be influenceable by the operator.

Our empirical analysis relies on a unique data set of 33 companies observed over a period of up to twelve years (1997-2008), for a total of 231 observations. We employ a stochastic frontier cost function for panel data that accounts for unobserved heterogeneity and provide firm-specific, time-varying inefficiency estimates. One of our model specifications further allows for the optimal technology to vary among the considered companies by randomizing some cost functions' coefficients. For taking the specificity of the discussed subsidy, the variable representing the subsidies directly enters the inefficiency function as a heteroscedastic variable.

We find a positive effect of these subsidies on the std.dev. of inefficiency, which implies that the range of companies' inefficiency increases with the level of subsidies relative to total costs. However, we also find that non-subsidized firms perform better in terms of cost efficiency. To our knowledge, this is the first analysis of endogenous subsidies and local public bus transport in the German context. The analysis demonstrates that the performance varies across state-owned companies and implementing a competitive environment by means of frontier analysis could help to increase the efficiency in this sector.

1.5.4 Overcoming Data Limitations in Nonparametric Benchmarking: Applying PCA-DEA to Natural Gas Transmission

In this chapter, efficiency analysis is applied to natural gas transmission, a classical network industry, which exhibits monopolistic market structures and is therefore traditionally regulated. Chapter 5 addresses the issue of limited observations, i.e. the number of firms, available to the regulator for conducting efficiency analysis by frontier techniques. The limited number of firms is unfavorable when regulators rely on nonparametric and deterministic estimators, e.g., DEA, since the estimators' convergence rates are low relative to parametric estimators, and hence, their discriminatory power diminishes with each additionally included variable needed to model the technology.

Combining a certain number of variables by principal components analysis (PCA) is one alternative of reducing the technology set's dimensionality while maintaining most of the variation in the original data (e.g., Adler and Golany, 2002, 2001). We apply the efficiency measurement approach of PCA-DEA initially proposed by Ueda and Hoshiai (1997) and Adler and Golany (2001) to a data set consisting of 34 US natural gas transmission pipeline companies in 2007. Our results show that the efficiency scores notably change in response to the combination of variables and on average decrease compared to the classical DEA model. Further, by conducting a case study we find that the companies serving as peers for the company under investigation, exhibit more structural similarities with the evaluated company under the PCA-DEA model specification.

However, the presented approach is not undisputed. One point of criticism is that the combined variables have no economic meaning and the obtained efficiency estimates can therefore hardly be interpreted. Another criticism is that the aggregation of variables might be inappropriate for estimating efficiency measures using

DEA due to their radial character (Simar and Wilson, 2001). This is addressed in the following chapter.

1.5.5 Estimating Alternative Technology Sets in the Context of Nonparametric Efficiency Analysis in Regulation: A Restriction Test for Pooled Data

Chapter 6 presents a method for identifying the appropriate specifications of the production possibilities based on statistical inference while being suitable for pooled data. As indicated before, data availability in the regulatory context is mostly limited and regulators may pool observable cross-sectional data across multiple time periods in order to achieve sufficiently large samples. The limitation of information refers not only to the number of firms, but also to the uncertainty about the input and outputs that properly represent the technology under investigation.

Due to the curse of dimensionality, it is desirable to exclude irrelevant inputs and outputs from the analysis while it is simultaneously favorable for the regulator to support the choice of variables by statistical inference. Simar and Wilson (2001, 2011) and Schubert and Simar (2011) propose restriction tests and test procedures that are suitable to derive statistical inference with respect to alternative technology set specifications. However, the yet existing implementations of these tests are restricted to cross-sectional data and are therefore not applicable for (unbalanced) panel, and therefore, pooled data. Consequently, the yet existing restriction tests require a methodological extension that is developed in this chapter.

We demonstrate the usefulness of the method, by applying it to a pooled sample of US natural gas transmission pipeline companies that includes five years (2003-2007), yielding 171 observations in total (after outlier correction). Based on the scientific and practical discussion, we develop two alternative specifications of the production possibilities that differ in their output vectors. As input we use the operating and maintenance expenditures. Concerning the choice of outputs, we begin with two variables capturing the production output, i.e. the total length of mains and the amount of natural gas delivered during peak hours. The aim is to test whether the third potential output measure, i.e. the total amount of natural gas delivered, is a relevant or a redundant variable with respect to the investigated technology. From that we develop the test hypothesis and conduct the proposed test procedure.

We find empirical evidence that, for our sample, the total quantity of natural gas delivered is a redundant output measure and, therefore, can be excluded for efficiency estimation. Consequently, the dimensionality of the technology set is reduced and the accuracy of company-individual performance measures is improved. Our analysis applies the proposed method for the first time and emphasizes its relevance. Further it shows how regulators can benefit from powerful and reliable instruments in terms of both, the identification of improvement potentials and acceptance on the part of the regulated companies.

1.5.6 Summary of Contributions

Table 1.1 summarizes the content of the remaining five chapters.

1.6 Outlook

As emphasized by Mühlenkamp (2012), unlike private firms, state-owned companies do not solely pursue the profit-maximizing goals. Rather, they also perform goals that are not evaluated by markets. This situation applies to public service provision in general. Further research may thus increasingly focus on including these non-market goals into the theoretical, methodological and empirical analysis. Thereby, additional and valuable insights into the economics of public service provision could be gained, which most likely would also shed new light on the comparison of public and private service providers.

Table 1.1: Summary of chapters

Chapter	Publication and author's contribution	Involved governmental level or body	Analyzed DMU	Characteristic or specificity	Frontier method	Main findings
Chapter II	based on Nieswand and Seifert (2011, revise and resubmit with *Local Government Studies*), author worked on model development and interpretation of results	intermediate local government	French *départements*	multiplicity of outputs, external framework conditions	DEA, bootstrap truncated regression	performance partly influenced by non-influenceable external variables; for distance to Paris, median income of households and share of elderly population significant and for size and seaside dummy no significant effects
Chapter III	author's independent work	lowest governmental tier	German municipalities	multiplicity of outputs, external framework conditions	conditional order-m, nonparametric regression	conditioning the efficiency estimation increases performance estimates; mixed results for unemployment rate, commercial activity increases performance, higher population density and income classes deteriorate performance, performance differ across Federal States, no effect for population movement
Chapter IV	based on Nieswand and Walter (2010, revise and resubmit with *Transportation Research Part E*), research was initiated by the author, responsible for developing the model, running calculations, writing the bulk of the analysis; data collection and interpreting the results were collaborative	lower governmental tiers, publicly owned entities	German publicly owned local bus transit public companies	subsidies	SFA, heteroscedastic inefficiency term	the variation of inefficiency increases with larger share of influenceable subsidies, non-subsidized companies perform better, marginal costs vary across firms
Chapter V	based on Nieswand et al. (2010), responsible for collecting the data, developing the models and running calculations; interpreting the results and writing the analysis were collaborative	public regulating authority	US natural gas transmission pipeline companies	uncertainty and limited data in regulatory benchmarking	PCA-DEA	combination of outputs yield lower average efficiency estimates and more similar peers for the firm under investigation
Chapter VI	research was initiated by the author, responsible for collecting data, writing the bulk of analysis; developing models, calculations and interpreting the results were collaborative	public regulating authority	US natural gas transmission pipeline companies	uncertainty and limited data in regulatory benchmarking	order-m and subsampling	when considering operating and maintenance costs as input and the amount of natural gas delivered during peak hours and length of mains as output, the total amount of natural gas delivered is a redundant output measure

Chapter 2

What Drives Intermediate Local Governments' Spending Efficiency: The Case of French *Départements*

2.1 Introduction

In the course of the financial crisis starting in 2007, the sustainability of public finances was put on the public agenda. However, this increasing pressure on public budgets is not new but is more pronounced and manifold than before. The European Commission (2010) lists the falling share of working age people in the population, lower (potential) economic growth and higher costs associated with providing services for the ageing population as drivers of stressed public budgets. The OECD (2010b) emphasizes the strong need for fiscal consolidation, whereby structural reforms remain an essential policy tool for its facilitation. In particular, reforms targeting the increase in public sectors' productivity and efficiency would improve the fiscal positions of many countries (OECD, 2010b). Efficiency improvement potentials do not seem to be fully exploited, most notably at the subnational government level (OECD, 2007, 2009a). Using 2008 data on the 96 European French *départements*, this chapter evaluates the spending efficiency of this intermediate level of governance using nonparametric efficiency analysis. Further, it aims to discuss factors that might explain parts of the existing inefficiency by second-stage regression.

In general, the public sector comprises of economic activities in which governments are engaged either in the production, the delivery or the allocation of public goods and services. These activities range from providing a legal system to purchasing goods and services, from government production to government redistribution of income. How public sector activities are pursued and its scope strongly differs among economies (Stiglitz, 2000). In many countries the public sector includes more than one level of government (Atkinson and van den Noord, 2001) and notably contributes to the economic outcomes. In 2008, the average share of general government expenditures[1] in gross domestic product (GDP) was about 41 percent (OECD, 2010a) for OECD countries, emphasizing its economic relevance. Particularly in multilayer systems, two issues are relevant for fiscal sustainability and public sector performance: the allocation of responsibilities and the management of public spending (OECD, 2003). With respect to the former, Atkinson and van den Noord (2001) note that decision-making authority is preferable where it can best be exercised.[2] With respect to the latter, exercising control over public spending is an important instrument strengthening the management of public spending (OECD, 2003, 2010b).

Benchmarking is the systematic comparison of the performance of one unit to other units (Bogetoft and Otto, 2011), making status quo evaluation and identification of areas that can be improved possible. Thus, it is a tool to exercise control over public expenditures, independently from the contributing level of government. Efficiency analysis provides benchmarking approaches that identify best practices (frontier) used in the transformation of inputs to outputs (technology). Relative to the determined best practice, unit-individual inefficiency then can be measured.[3] To define the frontier and measure the inefficiency of French *départements* we use DEA, which is a deterministic and nonparametric benchmarking method.[4] Compared to alternative parametric techniques, e.g., SFA, one distinguishing feature of DEA is that, except for convexity, it does not require any assumptions, such as a functional form, regarding the technology (Hjalmarsson et al., 1996). This is very useful since governmental activity, contrary to the example of firm activity, does not have a convenient, well-established equivalence in microeconomic theory. Thus, governmental behavior might not be adequately represented by a production function. Furthermore, similar to other efficiency measurement techniques as SFA,

[1]These include expenditures by central, state and local government plus social security.

[2]This argumentation is in line with public choice theory (effectiveness and knowledge about needs), e.g. Müller (2003) and Balaguer-Coll et al. (2010).

[3]It is common consensus that the public sector production exhibits inefficiencies that arise from numerous sources, e.g., organizational settings and personnel, procurement and budgeting restrictions. Therefore, the private sector serves as the standard of comparison. Alternatively, inefficiencies can be identified by comparing economic activities of government bodies among a homogeneous group.

[4]For a comparison and discussion of alternative efficiency analysis methods, see e.g., Coelli et al. (2005) and Hjalmarsson et al. (1996).

DEA allows us to consider multiple outputs, representing different governmental duties.

Efficiency analysis is applied to numerous European countries, and, although France is one of the biggest economies in the world and an important member of the European Union, to our knowledge, France has not been individually studied.[5] In addition, such an analysis is worthwhile because the repeated failure of authorities to meet medium-term spending objectives reinforces the need to improve the capacity of decision makers to control public spending (OECD, 2003).

France possesses a unique organization of its public sector, which roughly consists of the central government, the subnational governments, the social security funds, and large publicly owned enterprises (OECD, 2003). The country is a decentralized, unitary state meaning that the central state holds all legislative power and delegates responsibilities for public service provision to subnational administrative bodies. According to OECD (2010a), the overall proportion of general government expenditures[6] to national GDP in France is the highest ratio among the OECD countries, about 53 percent in 2008. These expenditures include those made by the central government (about 34 percent), by the sub-national governments (about 21 percent), and by the social security (about 45 percent; OECD, 2009b).

The French Constitution entitles three levels of subnational governments: the *rï¿œgions* (regions), the *départements* (departments), and the *communes* (municipalities); each with an elected council, autonomously financed, and possessing - to a limited extent - fiscal sovereignty (French Constitution - *Constitution*, *Article* 72, 72-2). While the *rï¿œgions* contribute with 13 percent to local governments' expenditures, the *secteur communal*[7] contributes with 55 percent and *départements* with 32 percent (DGCL, 2010).

We are particularly interested in analyzing *départements* for two reasons: First, they constitute the intermediate level of subnational government, for which a lack of analysis still exists and the potential for efficiency improvement does not yet seem exploited at this level. Second, *départements* hold an important role in the shifting of power from national to local authorities. France has a long history of attempts of decentralization, which can be interpreted as part of a broader effort by the French state to deal with the increasing complexity of its responsibilities and

[5]France is included in cross-country analyses considering OECD countries, e.g., Afonso et al. (2005) and Maudos et al. (2003).

[6]This excludes expenditures contributed by the large publicly owned enterprises.

[7]The communal sector includes single *communes* and groupings of *communes*, that may share tasks in general public services as waste management or water supply, but can also jointly levy taxes.

management (Cole, 2006). The power of the *départements* was already enhanced with the reforms of 1982-1983 that conceded larger budgets, more staff and more service-delivery responsibilities. The reforms of 2003-2004 intending to clarify the responsibilities shifted additional power toward subnational levels to support better and more efficient governance. As a result the share of general government services delivered by them increased.

However, inefficiency can be influenced by factors over which the *départements* cannot fully exercise control. Thus, such exogenous factors explain some aspects of the inefficiency. Depending on the considered system, these variables can relate to, e.g., political, geographical or fiscal characteristics. In the French case, the physical location of a subnational government relative to Paris, both as a leading global economic center and as the center of French political power, can be one of such factors. *Départements* that are part of or are located closer to the Paris agglomeration may benefit from that location; whether due to the close proximity to policy makers, due to a pool of highly skilled labor force, and/or due to economic strength. In addition to other factors, this effect needs to be taken into account when discussing spending efficiency.

The chapter is organized as follows: Section 2 gives an overview of the literature on public sector efficiency. Section 3 introduces the methodologies applied in this chapter. In Section 4, the model specifications and data are presented. The results are discussed in Section 5 and Section 6 concludes.[8]

2.2 Literature Review

A broad literature on measuring public sector performance by means of frontier methods is evolving. Kalb (2010b), Afonso and Fernandes (2008), De Borger and Kerstens (2000) and Worthington and Dollery (2000) provide comprehensive overviews of the empirical evidence derived from both, parametric and nonparametric methodologies. The existing literature concentrates on evaluating the performance of the public sector either in terms of publicly provided services or in terms of administrative units. For example, the work by Hauner and Kyobe (2010), Worthington (1999) and Gorman and Ruggiero (2008) all refer to particular ser-

[8]This chapter is based on Nieswand and Seifert (2011) and joint research with Stefan Seifert. This chapter is produced as part of the project Growth and Sustainability Policies for Europe (GRASP), a collaborative project funded by the European Commission's Seventh Framework Programme, contract number 244725. We thank Kristiaan Kerstens, David Saal, Christian von Hirschhausen, Astrid Cullmann, Petra Zloczysti, Michael Zschille, Anne Neumann, Ronny Freier and the participants of the European Workshop for Efficiency and Productivity Analysis (EWEPA) 2011 young researchers session for fruitful discussions and helpful comments, further, Adam Lederer for excellent editing.

vices including the health sector, education, libraries and police work. However, our focus is on the performance of administrative units. Within this context, spending efficiency is understood as a global measure of the administrative bodies' capability to provide and manage the tasks they are in charge of, with respect to the multiple inputs placed at their disposal.[9]

Concerning the representation of inputs, mainly financial rather than physical measures are used. Like this chapter, some studies, including Geys and Moesen (2009), Sampaio de Sousa and Stošić (2005), Vanden Eeckaut et al. (1993) and Arcelus et al. (2007), use one financial aggregate to describe the inputs, i.e. total or current expenditures. Whereas other studies further decompose the aggregate into capital related and labor expenditures, e.g., Balaguer-Coll et al. (2010). The advantage of using financial data is that all inputs are considered. However, it also implies that the administrative units face identical input factor prices if input factor prices and quantities cannot be implemented adequately in the estimation.

Concerning the representation of outputs, i.e. the goods and services administrative units are providing, analyses predominantly rely on the tasks that are obligatory to the units due to legal prescription. Although this approach excludes voluntary tasks, depending on the application, it covers the vast majority of costs and thus, allows comparing the units. To measure these outputs, the literature provides a wide range of means. For example, educational service is measured as the number of lessons taught (Loikkanen and Susiluoto, 2006), pupils enrolled (Geys et al., 2010), pupil exam performance (Giordano and Tommasino, 2011), the number of schools, or even the population in the relevant age group (Kriese, 2008). Each measure contains information on education in general, but delivers different detailed insights.

Following Bradford et al. (1969), one could distinguish direct outputs, and outputs of primary interest to the citizen-consumer. For instance, while the number of lessons taught tries to assess directly the actual service provided, student exam attainment is an outcome that is also a result of other socio-economic factors, which are not under control of the local government. However, citizens may be more concerned about the final outcome, rather than the amount of services delivered (Afonso and Fernandes, 2006). The on-going discussion on defining inputs and outputs underlines the general problems associated with representing the transformation process of administrative units. Among others, Balaguer-Coll et al. (2007) and Afonso and Fernandes (2006) point out that the complexity of the production

[9]This approach is along the same lines as Stiglitz (2000), who refers to the governmental management as a public good itself where everybody benefits from a better, more efficient and responsive management.

process and inputs and outputs are difficult to model. Furthermore, prices are rarely available (De Borger and Kerstens, 1996b).

With respect to the character of administrative units, efficiency analyses are conducted at the different tiers of governmental organization. The country level, i.e. state level, is the level of highest aggregation. Work by Afonso et al. (2005) provides empirical evidence for cross-country comparisons. Much attention is on local governments for which tasks can be identified more precisely. Municipalities are analyzed for various countries, e.g., Belgium (Vanden Eeckaut et al., 1993; De Borger et al., 1994), Spain (Benito et al., 2008; Balaguer-Coll et al., 2010), Germany (Kriese, 2008; Kalb, 2010b), Japan (Tanaka, 2006), and Finland (Loikkanen and Susiluoto, 2006; Loikkanen et al., 2011).

However, the empirical evidence for intermediate levels of government – to which the French *départements* belong to – is very limited. By nonparametric deterministic techniques, Hauner (2008) analyzes the spending efficiency for 89 Russian regions in terms of health care provision, education and social services. The author finds significant differences between the regions in all sectors. Likewise, Giordano and Tommasino (2011) find efficiency differences among the 103 Italian provinces that perform municipal, regional and national tasks. In addition, the authors identify rather low correlation of efficiency scores for different responsibilities. Applying a stochastic frontier approach, Kellermann (2007) evaluates the spending efficiency of the 26 Swiss Cantons between 1990 and 2002, finding fairly low inefficiency and increasing efficiencies over time.

Subsequent to the measurement of the performance rendered by particular public services and administrative units, the literature is also concerned with explaining (in-)efficiency. The purpose of these analyses is to explain performance differences that are due to exogenous factors (determinants) that are not (fully) under the control of the DMUs. Following Fried et al. (1999), a clearer understanding of the nature of inefficiency is important for designing policies that improve resource allocation. Such analyses are commonly conducted in a second stage, during which a set of explanatory factors are regressed on efficiency scores obtained by efficiency analysis techniques.

Table 2.1 gives an overview on second-stage analysis, outlines the approaches used, and summarizes the main findings.

Table 2.1: Overview of second-stage analysis on local governments' efficiency

Authors	Sample	Method	Main finding	
			Positive impact on efficiency	Negative impact on efficiency
De Borger et al. (1994)	589 Belgian municipalities	Tobit	high local tax rates, inhabitants' education level	higher inhabitants' income, per capita block grants, number of coalition parties
De Borger and Kerstens (1996b)	589 Belgian municipalities	Tobit	higher property taxes, inhabitants' education level	block grants, high inhabitants' income
Balaguer-Coll et al. (2002)	258 Valencian municipalities	Tobit	large population, level of commercial activity	higher per capita tax revenue, higher per capita grants
Loikkanen and Susiluoto (2006)	353 Finnish municipalities	OLS	higher inhabitants' education, dense urban structure, share of municipal workers aged between 30 and 49	large population, high inhabitants' income, peripheral location, diverse service structure, unemployment
Balaguer-Coll et al. (2007)	414 Valencian municipalities	non-parametric kernel smoothing	large population	tax revenues, self-generated revenues, deficit, grants
Giménez and Prior (2007)	258 Catalonian municipalities	Tobit	large population, inhabitants' income, commercial activity, tourism	distance to region's capital
Afonso and Fernandes (2008)	278 Portuguese municipalities	Tobit	inhabitants' education, inhabitants' purchasing power	
Hauner (2008)	89 Russian regions	truncated regression	inhabitants' income, good governance, democratic control	federal grants, higher spending
Loikkanen et al. (2011)	353 Finnish municipalities	OLS	dense urban structures, higher inhabitants' education, large share of municipal workers aged between 30 and 49, city manager's education, co-operation	high unemployment, larger population, peripheral location

Note: This overview extends the overview given by Afonso and Fernandes (2008).

The determinants can be contextually grouped into political, geographic, fiscal, and socio-economic factors. However, Table 2.1 shows that for some, the evidence is inconsistent, e.g., population size. While De Borger et al. (1994), Giménez and Prior (2007), and Balaguer-Coll et al. (2007) find a positive impact of population size on efficiency, the results of Loikkanen and Susiluoto (2006) and Loikkanen et al. (2011) indicate a negative relationship. Similarly, population density is found to be positively related to efficiency in some studies (Geys et al., 2010 and Loikkanen and Susiluoto, 2006) while Afonso and Fernandes (2008) do not find significant effects. Likewise, the results are ambiguous regarding the influence of inhabitants' economic situation, e.g., in terms of income or purchasing power, while some studies find significant negative impact (De Borger and Kerstens, 1996b and Loikkanen and Susiluoto, 2006), other authors find significant positive influence (Giménez and Prior, 2007 and Afonso and Fernandes, 2008). Concerning dependence on central government transfers, most studies find a negative relationship between central government grants and efficiency (De Borger and Kerstens, 1996b and Balaguer-Coll et al., 2007). Similarly, several studies find a negative impact of tourism and in-commuting on efficiency (Kellermann, 2007 and Sampaio de Sousa and Stošić, 2005), which might be due to the additional costs of public goods provided to non-residents. In contrast, increasing urbanization and commercial activity (Loikkanen and Susiluoto, 2006; Balaguer-Coll et al., 2002 and Giménez and Prior, 2007) and higher resident education levels (De Borger et al., 1994 and Loikkanen and Susiluoto, 2006) are generally found to be positively related to efficiency. The latter are also used as indicator for citizen political participation, which is also found to positively influence efficiency (De Borger and Kerstens, 2000; Giménez and Prior, 2007).

Second-stage analysis predominantly employs regression techniques such as ordinary least squares (OLS) and Tobit regression. While Tobit regression accounts for the limitation of efficiency scores at unity, it still imposes strong statistical assumptions and requires a correct model specification. Simar and Wilson (2007) show that this technique has several drawbacks and may lead to biased results. Recent analyses of government efficiency take this, to some extent, into account: Hauner (2008) uses a truncated rather than a censored regression model following the suggestion of Simar and Wilson (2007). Balaguer-Coll et al. (2007) aim to overcome the problems with a nonparametric smoothing approach, which demands no functional specification and avoids assumption violations, but allows only for bivariate analysis. The conditional efficiency framework proposed by Daraio and Simar (2007b) is an alternative approach that takes the operational environment directly into account when efficiency is measured. However, this approach demands a great number of observations to derive meaningful results. Therefore, in this chapter we use bootstrapped truncated regression, as proposed by Simar and Wilson (2007), which has, to the authors' knowledge, so far not been applied to analyze government efficiency.

2.3 Methodology

2.3.1 Performance Measurement with DEA

We use DEA to measure the spending efficiency of French *départements*. Thereby, the *départements* can be considered as DMUs transforming inputs to outputs. DEA determines the best practice technology (frontier) by piecewise linear programming whereby the frontier envelopes all observed input-output combinations. Thus, the frontier sets the benchmark against which each of the *départements* is compared to and any distance to the frontier is interpreted as inefficiency. Those *départements* lying on the frontier are considered to be relatively efficient and serve as peers for others. A *département* is fully efficient on the basis of available evidence if and only if the performance of other *départements* does not show that some of its inputs or outputs can be improved without worsening some of its other inputs or outputs (Cooper et al., 2011). More formally, we analyze a set of $i = 1, \ldots, n$ *départements* that transform an input vector $x^i \in \mathbb{R}^p_+$ including p inputs, into an output vector $y^i \in \mathbb{R}^q_+$ including q outputs. According to Simar and Wilson (2008), the production set Ψ can be understood as the set of physical available points (x, y), or in other words as a set of feasible input-output combinations, i.e.

$$\Psi = \left\{(x, y) \in \mathbb{R}^{p+q}_+ \mid x \, \text{can produce} \, y\right\}. \tag{2.1}$$

This production set constraints the production process. To describe the efficient boundary (frontier) of Ψ we assume input-orientation meaning that we identify the minimum amount of inputs required to produce a given amount of outputs. Hence, for every *département*, we obtain the maximum potential reduction of inputs for its observed level of outputs, which is available in the feasible production set. This is a reasonable assumption because the obligatory tasks of the French *départements* are determined by law and thus, choices related to outputs are limited. Furthermore, practical consolidation favors spending-based budget retrenchment OECD (2010b) for which the input-oriented boundary of Ψ provides useful information.

For a *département* with the input-output combination (x_0, y_0), the input-oriented efficiency measure θ is then defined by

$$\theta\,(x_0, y_0) = inf\,\{\theta \geq 0 \mid (\theta x_0, y_0) \in \Psi\} \tag{2.2}$$

where $\theta\,(x_0, y_0)$ gives the radial, i.e. proportional, reduction of inputs a unit could undertake to become efficient (Simar and Wilson, 2008). By construction, θ is equal or less than unity, but cannot take values smaller than zero. For $\theta = 1$, the *département* is efficient and cannot reduce its input. For $\theta < 1$, the *département* can produce the same level of output with only using $1 - \theta$ times its input. Thus, it could save θ percent of each input.

Based on the ideas of Farrell (1957), different linear programs have been developed to allow the technology, i.e. the frontier, to be of certain nature. Most frequently, the models proposed by Charnes et al. (1978) and Banker et al. (1984) are applied where the technology exhibits CRS and VRS, respectively. We assume VRS, which assures that local governments are benchmarked against units of similar structure. An efficiency estimate $\hat{\theta}$ for an observation operating at level (x_0, y_0) is then derived by solving the following program

$$\begin{aligned} \hat{\theta}\,(x_0, y_0) = \min_{\theta,\lambda_1,\ldots,\lambda_n} \{\theta > 0 \mid \;\; & \theta x_0 \geq \textstyle\sum_{i=1}^{n} \lambda^i x_k^i ; k = 1, ..., p \\ & y_0 \leq \textstyle\sum_{i=1}^{n} \lambda^i y_l^i ; l = 1, ..., q \\ & \textstyle\sum_{i=1}^{n} \lambda^i = 1;\ \lambda^i \geq 0 \,\forall\, i = 1, ..., n\}. \end{aligned} \tag{2.3}$$

with λ^i being a vector of unit-individual weights for inputs and outputs that are used to construct the efficient linear combination. VRS assumption is introduced by the constraint $\sum_{i=1}^{n} \lambda^i = 1$.

Nonparametric deterministic frontiers, such as those constructed by DEA, are appealing since they rely on only few assumptions. However, when applying DEA, particularly two aspects must be carefully considered: the convergence rate of the DEA estimator and extreme values or outliers in the data. The convergence rate measures how fast an estimator converges to the true and unknown parameter subject to the number of observations. Compared to alternative parametric approaches, the DEA estimator exhibits a slow degree of convergence. Hence, the validity of DEA estimates strongly depends on the number of variables used, i.e. the dimensionality of the model specification, relative to the observations included. To obtain a reasonable discriminative power and meaningful estimation results, an appropriate ratio of variables and observations is necessary. We address this issue by restricting ourselves to a single input and the most relevant outputs, i.e. the mandatory tasks.

2.3.2 Outlier Detection

As aforementioned, DEA frontiers are sensitive to extreme values and outliers (Simar, 2003). Extreme values and outliers can indicate either data errors, for which DEA cannot correct, or indicate observations that are outside the normal range but nevertheless valid. Because DEA relies on envelopment, extreme values and outliers belong to the attainable set with certainty. Thus, when identified as peers, they can directly influence the efficiency measures of other observations. To overcome this issue, we use two methodologies, first the super-efficiency analysis proposed by Banker and Gifford (1988) and Andersen and Petersen (1993) to detect outliers and then the efficiency stepladder (ESL) proposed by Edvardsen (2004) to test the frontier's robustness.

The concept of super-efficiency constructs efficiency measures by avoiding that the evaluated unit can help span the frontier (Bogetoft and Otto, 2011). Consequently, super-efficient observations obtain efficiency scores larger than unity and can be subject to an individual inspection. We use the results of this analysis to identify observations with a super-efficiency score[10] greater than 1.2. These are further investigated using the ESL approach that indicates the sensitivity of the individual efficiency scores to measurement errors (Edvardsen, 2004). Thus, efficiency estimates can be investigated in terms of robustness. For every observation, the first step of this iterative approach is to identify its most influential peer, i.e. the peer whose exclusion leads to the greatest efficiency increase. The detected peer is removed and DEA is conducted again. This procedure is done repeatedly until the given observation becomes fully efficient. The changes of the measured efficiency occurring in these steps indicate the sensitivity of the measured efficiency scores against the other observations in the data set. This allows us to evaluate the influence of the elimination of those observations that are found to be potentially super-efficient.

2.3.3 Bootstrapped Truncated Second-stage Regression

To investigate which, and whether, exogenous variables have explanatory power on inefficiency, we conduct the bootstrapped truncated regression proposed by Simar and Wilson (2007). This approach allows for valid inference in the second stage and is therefore superior to others. Previous studies on local government efficiency predominantly use OLS or Tobit regression. However, Simar and Wilson (2007) note that due to serial correlation, Tobit regression yield inappropriate and biased estimation results. Basically, two sources of errors cause biases: on the one hand, the observations are empirically obtained and not independently distributed, but underlie serial correlation. On the other hand, since only a sample is used and the

[10]Based on Monte Carlo simulation, Banker and Chang (2006) propose to define observations as outliers that exceed an efficiency level of 1.2.

most efficient observations are not captured, the efficiency scores are likely to be biased upwards. Even though our sample covers the whole population of French *départements*, the true frontier remains unknown. Furthermore, inefficiencies may still exist for the efficient observations.

To evaluate the influence of exogenous factors on the spending efficiency of French *départements*, we investigate the following relationship

$$\theta^i = \alpha + \beta Z^i + \varepsilon^i, \tag{2.4}$$

with θ^i representing the unknown true efficiency of the ith observation, α being a constant term (intercept) and β being the vector of coefficients to be estimated. For each variable, β is the same for all observations and indicates the relationship between Z^i, a vector of exogenous factors, to the efficiency score. ε^i is the statistical noise term of the ith observation, which is restricted by the condition $\varepsilon^i \leq 1 - \alpha - \beta Z^i$. Following Simar and Wilson (2007) this term is assumed to follow a truncated normal distribution with zero mean (before truncation), unknown variance and a truncation point determined by this condition. Since the true θ is unknown, it is replaced with the Debreu-Farrell efficiency scores obtained in the first stage ($\hat{\theta}^i$, bounded between zero and unity). The econometric problem becomes

$$\hat{\theta}^i = \alpha + \beta Z^i + \varepsilon^i \text{ with } \varepsilon \sim N\left(0, \sigma^2\right) \tag{2.5}$$

such that $\varepsilon^i \leq 1 - \alpha - \beta Z^i$, which has to be solved by maximum likelihood estimation with respect to β and σ. By using bootstrapping methods with b replications, b estimates for these coefficients are calculated. Confidence intervals for those estimators can be constructed following Simar and Wilson (2000). A positive sign of the second-stage estimation coefficient indicates a positive relationship between spending efficiency and the respective explanatory variable.

2.4 Model Specification and Data

2.4.1 Specification of Inputs and Outputs

We consider the French *départements* as units that contribute expenditures (input) in order to provide a certain bundle of publicly provided services (outputs) without assuming a functional form of this process.

We use total expenditures (TOTEX) as a single input employed by the *départements* to provide public services for that they are in charge of. Using TOTEX as input measure, on the one hand, allows incorporating all relevant input information. On the other hand, it implicitly assumes that input factor prices are the same for all *départements*. This assumption appears to be reasonable in the case of France: With respect to labor it is justified since wages of civil servants are mainly regulated by the government. With respect to capital expenditures, De Borger and Kerstens (1996b) argue that Belgian local governments have access to the same capital markets and thus, face similar capital related inputs prices, which can also be assumed for the French *départements*. A further issue related to capital input is the issue of the dynamic character of investments. However, our data show that investment expenditures remain fairly steady over time.

To specify the outputs, we follow the work done by Vanden Eeckaut et al. (1993) and De Borger et al. (1994) and concentrate on the *départements*' legal obligations (mandatory tasks) in the fields of: (i) social services: care for elderly and provision of minimum subsistence grants; (ii) secondary education; (iii) road construction and maintenance; and (iv) general administration.

These categories are represented by five output indicators: the number of beneficiaries of minimal subsistence grants (BENEF) and the number of beds in private and public retirement and nursing homes (NURSING)[11] are used to measure social services. The road network km (ROAD) are used as an indicator for efforts undertaken concerning road construction and maintenance and the number of pupils on public schools (PUBPUPILS)[12] approximates education services provided. General administrative services can be approximated with the total population; see e.g., Sampaio de Sousa and Stošić (2005); De Borger and Kerstens (1996b). However, this variable is highly correlated with PUBPUPILS (Pearson correlation of over 96 percent). Therefore, using the number of pupils on public schools also as the indicator for general administration reduces the number of dimensions in our DEA problem without losing too much information.

Although, these outputs do not comprise the entire array of services provided, the restriction is rational. The selected outputs cover the most relevant competencies of the French *départements*, both, in terms of responsibility and in terms of the share in expenditures.[13] Furthermore, it prevents us from having a poor ratio of variables to observations, which would deteriorate the meaning of our estimation results.

[11]Contrary to the pure number of elderly, e.g., the population over 65, this variable contains more information on the number of dependent elderly.

[12]This variable is chosen to measure the services regarding the provision of education infrastructure. In our opinion, this is a more appropriate measure then the number of schools, since it also takes different school sizes into account.

[13]In 2008 the current expenditures for this four sectors sum up to about 78 percent of total current expenditures.

2.4.2 Specification of Exogenous Variables

In the second stage, we aim to identify the impact of some selected exogenous variables on the *départements*' spending efficiency. For this purpose a set of variables is regressed on the efficiency scores obtained from the DEA analysis in the first stage. The literature provides a wide range of possible variables. However, their exogeneity is neither always absolutely certain nor applicable in our context. For example, the dependence on grants as a fiscal variable influences the transformation process of the *départements* itself and thus, would introduce endogeneity to the estimation. As a result the obtained coefficients would be biased. Hence, political or fiscal variables are omitted in the analysis at hand. Rather, we choose three geographical and two socio-economic factors that are assumed to have some impact on the spending efficiency of French *départements*.

First, we test how efficiency is influenced by the *département* size (SIZE). The territorial size of the *départements* is predetermined and we hypothesize that larger *départements* face disadvantages due to the lack of positive agglomeration economies. The effort devoted to general coordination may be higher and provision of certain services, e.g., safety and fire fighters, might be relatively more costly. Thus, size may affect spending efficiency.

Second, we test the influence of the distance to Paris (DISTANCE). In the French context this variable is of particular interest, since it captures the peripheral character associated with centralized states. Being spatially closer to the economic and political capital seems to be advantageously to local governments. The *départements* further from Paris, for example, may experience greater migration of highly skilled workers and may have limitations in exercising political influence. We therefore expect a negative relationship between DISTANCE and spending efficiency.

Third, we test a variable that contains information on the coastal location (SEASIDE). Due to special circumstances, the *départements* could be forced to have additional expenditures, e.g. for flood control or road and port construction and maintenance. Thus, coastal regions should have spending efficiency negatively affected.

Fourth, following De Borger and Kerstens (1996b), we include the median income of households (MED_INCOME) to the set of explanatory variables. The authors argue that the households' income may influence the efficiency in two ways. First, local governments' higher fiscal capacity may facilitate featherbedding and on-the-job-leisure. This is not necessarily the case for the French *départements*, since their tax revenues are mainly independent from income levels. Nevertheless, a negative relationship between income and monitoring of the government by the society may exist: due to higher opportunity costs households decide to spend less of their time on monitoring their government, which facilitates inefficiencies. Moreover, as Geys et al. (2010) argue, income may influence the preferences of the inhabitants. Due to additional income, the demand for public goods of higher quality might increase.[14] Based on these arguments, we expect a negative relationship between median income and efficiency.

Lastly, we investigate what effect the population composition, especially the old-age dependency ratio (SHARE_ELDERLY), has. The structure of the population can significantly influence public sector efficiency and budgets as shown for example by Geys et al. (2008) and Seitz (2008). Even though the French population is expected to grow, ageing will significantly change the structure leading to a higher share of dependent elderly persons relative to total population, which will also lead to a change in the demand for public goods and services. Nevertheless, this demographic change is already present and leads to demand for additional public services for elderly, whereas at the moment especially rural counties are affected.

2.4.3 Data

Our data consists of the 96 French *départements* exclusively located in Europe.[15] For 2008 we gather monetary and physical data from the French Ministry of the Interior, the French National Institute of Statistics and Economic Studies (INSEE) and the Institute for Research and Information in Health Economics (IRDES).

Table 2.2 presents the main characteristics of our data. We restrict our analysis to 2008, since the *départements* obligations were extended considerably in the previous years: In 2004 and 2005, responsibilities in the social sector, concerning especially social welfare, care for elderly, as well as youth work, were extended. Similarly, in 2006 competences for the care of disabled were extended and responsibility for more than 17,000 km of roads was transferred from national to local governments. Finally, in 2007, the technical staff in secondary schools, in total more than 95,000 employees, was transferred to the local government. However, our output variables are not able to capture the additional competencies assigned to the *départements* during this decentralization process. Hence, the changes in

[14] Loikkanen et al. (2011) point out that resident income might also be an indicator of regional input price differentials. They argue that capital cost and especially land prices will be higher in areas with higher income.

[15] Overseas *départements* have to be excluded to maintain comparability of observations. They differ strongly from the *départements* located in Europe in socio-economic, but also geographic terms. Furthermore, input prices may diverge.

Table 2.2: Descriptive statistics for French *départements*

Variable	Mean	Min	Max	Std.dev.
Inputs				
TOTEX [mn Euro]	653.8	118.2	2,648.3	483.9
Outputs				
BENEF [number of beneficiaries]	10,471	727	71,813	12,003
NURSING [number of beds]	4,799	383	12,694	2,680
PUBPUPILS [number of pupils]	24,699	2,373	92,604	19,035
ROAD [km]	3,931	0	7,762	1,562
Exogenous variables				
DISTANCE [km]	352	0	917	205
SIZE [square km]	5,666	105	83,534	1,924
SEASIDE [dummy]	1	0	1	0
MED_INCOME [Euro]	26,589	20,944	39,671	3,216
SHARE_ELDERLY [percent]	0.18	0.11	0.37	0.04

Source: French Ministry of the Interior, French National Institute of Statistics and Economic Studies, Institute for Research and Information in Health Economics. Notes: observations=96, year=2008.

technology, i.e. the additional responsibilities, prevent us from pooling data for more years. For the same reason, results from comparing the year considered to previous ones, give only very limited information on the dynamics of spending efficiency.

Input is measured as total expenditure (TOTEX), which contains all operating expenditures, including personnel expenditures, interest payments, general expenditures and other expenditures, and all investment expenditures, including investment costs, debt amortization, and granted subsidies. The *départements* spent on average 654 million (mn) Euro, with a minimum of 118 mn and a maximum of 2.6 billion (bn) Euro. This spread indicates the large variety between the government in terms of size and services provided. This is also reflected by the different output measures, which vary strongly. However, the road network km for Paris are set to zero since this task is carried out at the municipal level. For the *départements* of Corsica, which are, contrarily to the other *départements* not responsible for secondary schools, the non-zero number of pupils is used to measure the output concerning general administrative services.

Concerning exogenous factors, the following variables are chosen: distance to the capital (DISTANCE) is measured as linear distance between Paris and the capital of the considered *département*. The size of each *département* (SIZE) is measured as territory in square km and ranges from between 105 and 83,534 squared km indicating the substantial differences between the jurisdictions concerning the service area. Coastal location is represented by a dummy, SEASIDE that equals one if the *département* has seashore. For 26 out of 96 *départements* this is the case. Inhabitants' income is measured as median household income in 2008

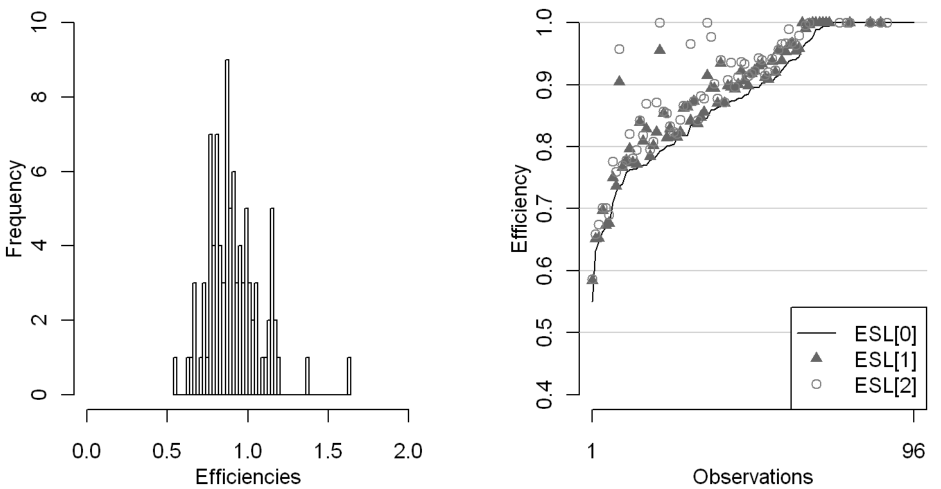
(MED_INCOME) and its wide spread (between about 21 and 40 thsd Euro) shows that notable economic differences between the territorial units exist. Finally, the old-age dependency ratio (SHARE_ELDERLY) is the share of over 65 years old in the total population. This variable ranges from 11 to 37 percent and indicates that population composition varies significantly across *départements*.

2.5 Results

2.5.1 Identifying Outliers and Extreme Values

In order to obtain meaningful and robust efficiency estimates using DEA, we check the data for extreme values.[16] The histogram in Figure 2.1 shows the frequency distribution of efficiency measures from super-efficiency analysis. Two observations, *Lozère* (1.4) and *Loire-Atlantique* (1.6), have efficiency scores exceeding the critical value of 1.2 proposed by Banker and Chang (2006). Thus, they are candidates to be excluded. We assume that these results are not driven by data errors,[17] and review the observations with respect to their characteristics. *Lozère* is the smallest unit in terms of both input and most outputs. Hence, we conclude that *Lozère* is an extreme, but still valid, observation outside the normal range. In contrast, *Loire-Atlantique* is one of the most populous *départements* and is among the jurisdictions with the lowest per capita expenditure while achieving high scores in several outputs. In order to decide which observation to exclude, we further test their impact on other *départements*' performance using the ESL.

Figure 2.1: Histogram of super-efficiency analysis and ESL



[16]Calculations are conducted using the statistical software *R* with the additional package "FEAR" version 1.12 by Wilson (2008).

[17]However, one cannot rule out the possibility of measurement errors.

The first step (ESL[1]) already increases the efficiency estimates of numerous observations considerably, but few observations are rated as fully efficient after excluding only one observation. A closer look at the procedure reveals that the potential outliers, flagged by the super-efficiency approach, are serving as most influential peers only for two observations in this first stage. Hence, in the first phase of the ESL approach the frontier seems to be rather robust against these potential outliers.

The same applies to the second stage (ESL[2]), in which efficiency scores increase again considerably for certain observations. However, the potential efficiency gains by excluding the most influential observations are caused only in two cases by the potential outliers. The results show that the potential outliers' impact on the frontier, and thus, on the efficiency scores of the other observations, is limited. Therefore, we do not exclude any observations and further analysis is conducted for the 96 *départements*.

2.5.2 Spending Efficiency Measurement

After carefully checking the data for outliers, efficiency measurement for the 96 *départements* is carried out. The DEA estimation results for the 96 observations are summarized in Table 2.3. The mean spending efficiency of French *départements* is about 88 percent. This implies an average improvement potential of about 12 percent meaning that the *départements* could save this amount of inputs while providing the observed level of output. The maximum value of spending efficiency is 100 percent by definition. 25 of the 96 observations are rated as efficient and shape the best practice frontier. These observations may serve as references for identifying improvement potentials. The lowest efficiency score achieved is about 56 percent and signalizes enormous saving potentials for the most inefficient unit.

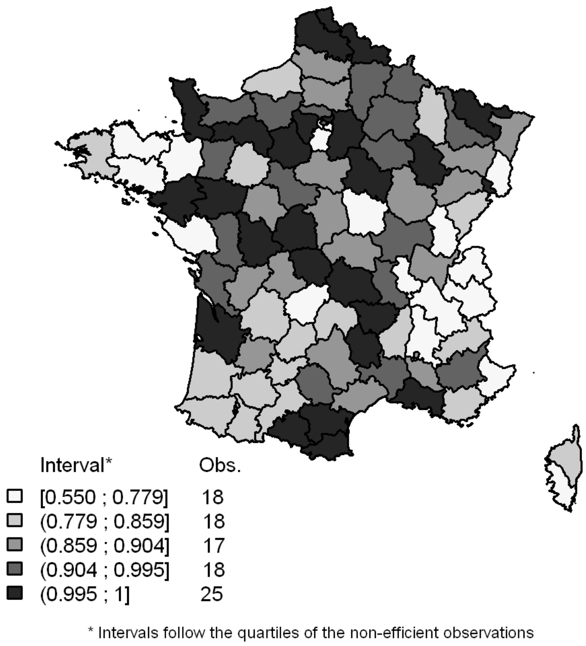
Table 2.3: Spending efficiency estimates for French *départements*

Statistic	Efficiency/ number
Min	0.555
Mean	0.883
Median	0.895
Max	1
Std.dev.	0.108
Efficient units	25
Share of efficient units	26

Figure 2.2 maps the efficiency scores of the individual *départements*. This illustration shows some interesting patterns and efficient and inefficient areas seem to concentrate in certain areas. First, a large number of highly efficient observations

can be found in the north of France and the area surrounding Paris. A second area of highly efficient *départements* is located in the south at the Spanish boarder. Finally, a long belt of efficient units includes *départements* from the north-west to the south of France. On the contrary, inefficient *départements* are especially located in the south-east, in the south-west and in the north-west of the country. However, among the efficient observations, sparsely populated *départements* can be found as well as urban areas. Likewise, we find the north, which is a rather industrialized area, as highly efficient, while the also economically important *Rhône-Alpes* region is found to possess notable improvement potential. In a second stage, these efficiency differences are assessed and bootstrapped truncated regression is used to explore exogenous determinants of efficiency.

Figure 2.2: Map of spending efficiency of French *départements*



2.5.3 Explaining Efficiency

To explain the performance of French *départements*, a set of exogenous factors is regressed on the efficiency scores obtained from the DEA program. The estimation results of the second-stage regression are summarized in Table 2.4.[18] Due to the truncation at unity the fully efficient units are omitted in this regression and the number of observations reduces from 96 to 71.

[18]Note that regression results may vary when bootstrapping is applied. Nevertheless, our results are robust.

Table 2.4: Second-stage regression results for French *départements*

Exogenous variable	β	p-value	CI LB[a]	CI UB[b]
DISTANCE	-0.362**	0.048	-0.072	0
SIZE	0.009	0.602	-0.024	0.042
SEASIDE	0.039	0.339	-0.121	0.042
MED_INCOME	-0.028***	0.002	-0.045	-0.009
SHARE_ELDERLY	-1.716**	0.040	-3.356	-0.077
Constant	1.957***	0	1.204	2.711
σ	0.098***	0	0.073	0.123
Log-likelihood	80.659			
Observations	71			

Note: ***,** indicate statistical significance at the 0.01 and 0.05 level. [a,b] Lower and upper bound of the confidence interval, respectively.

For the variable DISTANCE, we find a significant negative impact on the spending efficiency, meaning being located closer to Paris fosters performance.[19] This is in line with previous analyses on other European unitary states, such as Portugal (Afonso and Fernandes, 2008) and Finland (Loikkanen and Susiluoto, 2006 and Loikkanen et al., 2011). Distance to policymakers might influence efficiency in several ways: First, remote *départements* might face migration of highly skilled workers to the capital. This is possibly even more relevant for France because of the exceptional economic and political position of Paris. The capital city attracts an especially young and highly skilled population, which also improves the pool of candidates for the public sector. Second, this finding might be interpreted as the ability of local governments to exercise direct influence on national politics to their advantage. Since Paris hosts the major political institutions at national, regional and departmental levels, closeness to the political decision-makers can be beneficial, e.g. when the redistribution of subnational tasks during the process of decentralization is discussed. Regarding this point, further analysis of the influence of political variables would be beneficial as far as they can be represented by exogenous measures.

As indicated by Figure 2.2 we find no significant impact on the French *départements*' spending efficiency for the variables SIZE and SEASIDE. Concerning the first, modern communication technology possibly simplifies coordination and thus, reduces transaction costs. Concerning the latter, the effect of coastal location seems to be negligible when analyzing public sector efficiency for French *départements*.

Similar to other studies, e.g., De Borger and Kerstens (1996b), our results show

[19]We also test a variable that contains information on the topography, i.e. the highest elevation in the *départements*. Such a variable is highly correlated with DISTANCE. Therefore, our estimator might also include effects of the land-form on efficiency.

a significant negative relationship between spending efficiency and median income (MED_INCOME). As previously noted, there are two explanations for this finding: on the one hand, high-income households probably sacrifice less time monitoring their government due to higher opportunity costs, which facilitates inefficiency. On the other hand, demand for public goods of higher quality might increase in high income areas, driving up the costs for the local government. This question needs to be considered when quality indicators are available.

The coefficient of the share of elderly population (SHARE_ELDERLY) has a negative sign and is highly significant. Thus, demographic structure seems to affect spending efficiency. An explanation for this could be that costs of service provision are higher for the elderly segments of the population, as shown by Seitz (2008) for Germany. Since population projections for France forecast a significant increase in the elderly population, local government budgets will be especially affected at the *département* level due to the allocation of responsibilities among the layers of government. In light of this demographic challenge, analyzing and reducing public sector inefficiency becomes even more important.[20]

Overall, our results suggest that efficiency is partly driven by exogenous factors. Peripheral location and greater residents' income are negatively related to spending efficiency. Likewise, a higher share of elderly population is found to negatively influence efficiency. Contrarily, *départements*' size and coastal location are not found to significantly influence *départements*' performance.

2.6 Conclusion

The French government has reallocated responsibilities for service deliveries in the first years of the 2000s. These developments have led to a considerable increase of the responsibilities of the French *départements*. Structural reforms, such as this decentralization process, are considered as important means of fiscal consolidation that appear to be necessary in order to overcome the increasing pressure on public budgets. Furthermore, the restructuring aims to help improving the public sector's productivity and efficiency.

To identify the efficiency of public spending and potential improvements, we use DEA, a nonparametric deterministic approach of efficiency analysis, to the 96 *départements* in metropolitan France in 2008. This approach is particularly suitable since the behavior of public sectors might not be adequately represented

[20] A higher share of elderly population is also related to a rural structure of a *département*. Therefore our estimator might also include negative effects from this factor, e.g. by allowing for agglomeration and scale economies.

by production or cost functions relying on microeconomic assumptions. We define total expenditures as the single input administrative units employ. For the representation of the responsibilities, we focus on the obligatory tasks. Hence, the best practice displays the minimal amount of expenditures required to provide the given level of obligatory tasks provided.

Similar to analyses on the spending efficiency at municipal, e.g., De Borger and Kerstens (1996b) and national levels, e.g., Afonso et al. (2005), we find significant inefficiencies in public service provision at the intermediate level of government. More precisely, we identify a mean spending efficiency of the French *départements* of about 88 percent. Hence, one average the expenditures could be reduced by 12 percent, while providing the same amount of public services. The range of efficiency varies significantly among the *départements*, which is in line with previous analysis, e.g., Afonso and Fernandes (2008). Based on our results, a number of *départements* with different socio-economic and economic characteristics can serve as reference points to identify possible improvements. However, our results also indicate that inefficiency is not only due to inefficient usage of resources.

In fact, exogenous factors can contribute to inefficiency and thus, must be taken into account when evaluating the potential improvement. We are interested in identifying those factors that impact the *départements*' performance but are not under their control. For this purpose we conduct a bootstrapped truncated regression as proposed by Simar and Wilson (2007). Our results suggest that the population structure, the households' median income and the distance to Paris negatively affect the spending efficiency.

Our analysis shows that the efficiency of French subnational governments could be increased to improve the fiscal position. Against the background of the increasing importance of subnational tiers, this is particularly relevant for the public provision of goods and services and for the public budget from a global perspective. In 2008 France established a committee for the reform of the *collectiviti¿œs territoriales* (*Comiti¿œ pour la ri¿œforme des collectiviti¿œs locales*) with the objective of reviewing the territorial organization and local administration. Following the arguments by OECD (2003), this process should address the allocation of responsibilities and implementation of proper control mechanisms. Efficiency analysis could contribute to this discussion being a useful tool not only for the performance evaluation, but also for assessing the consolidation of budgets and territories.

Chapter 3

Conditional Efficiency Analysis of Municipal Service Provision in Germany

3.1 Introduction

3.1.1 Motivation

This chapter examines the performance of 1,060 German municipalities by means of nonparametric frontier and regression models. The analytical framework allows deriving robust measures of municipal efficiency while immediately accounting for the heterogeneous framework conditions. Further, it is able to identify how these operating conditions effect the performance and whether the effects are of statistical significance.

Similar to other countries, e.g., Finland, Belgium and Spain, local governments in Germany shape daily routines by providing important local public sector services, e.g., educational institutions, waste collection, recreation parks, and public welfare. Although these services are important, their provision is, in many cases, difficult. German municipalities and cities are facing a tense financial situation and are subject to substantial changes that either increase the pressure or aim to improve their positions. For example, the demographic change and the devolution of responsibilities from higher to lower governmental tiers directly affect the financial endowment and enhance the range of responsibilities. Thus, the accomplishment of tasks is further complicated while enormous deficits already frequently exist; see e.g., Statistisches Bundesamt (2008, 2010, 2011). Due to these trends, municipal capabilities are sometimes exceeded, which compromises service provision and, consequently, endangers the exercise of self-governance. Reinforcing self-governance as well as improving the efficiency and productivity of local public sector activities, in turn, are essential purposes toward motivating structural reforms. Against this background, systematic analysis of the public service provision can contribute valuable insights.

Both parametric and nonparametric frontier models are frequently used to evaluate the performance and variations in performance of and among local public service provision; see e.g., Worthington and Dollery (2000). These methods provide a quantitative measure of efficiency whereby local governments are considered as DMUs that transform resources (inputs) into public services (outputs) using a common technology. From a given set of observations a best-practice frontier is constructed and serves as a reference against which each unit is compared to. The performance is determined relative to this frontier and expressed by the resulting efficiency score.

Surprisingly, there is little empirical evidence for German municipalities and cities, including e.g., Kalb (2010b); Geys et al. (2010); Montén and Thater (2011); Hitzschke (2011). Except for the latter, dealing with 112 German cities, to our best knowledge, there is no analysis addressing local governments across Germany states. In addition to using a unique data set, the current chapter differs from previous work on municipal efficiency, in particular with respect to the methodology, i.e. the estimation approach proposed by De Witte and Kortelainen (2009, 2013), which extends the ideas of Cazals et al. (2002) and Daraio and Simar (2005, 2007b). Thereby, this chapter aims to make contributions to previous findings that are relevant for both, the German case and the efficiency measurement of local governments in general.

The literature clearly emphasizes that local governments not only undertake multiple and heterogeneous activities, but also potentially pursue multiple objectives; see e.g., Fletcher and Snee (1985) and Pestieau (2009). The latter in particular dissociates public service provision from private firms. This, and the existing uncertainty regarding the exact structure of the municipal transformation process, give reason to rely on nonparametric approaches when considering local governments. In contrast to parametric alternatives that impose a functional structure on the transformation process, nonparametric techniques rely on a few formal properties that the production set needs to satisfy; see e.g., Simar and Wilson (2008) and Daraio and Simar (2007a). Further, nonparametric techniques allow for the consideration of multi-input-multi-output technologies without the requirement of predefined weights; see e.g., Afonso and Fernandes (2008).

Applications that estimate the unknown frontier and the efficiency of local governments nonparametrically, commonly apply the FDH estimator proposed by Deprins et al. (1984) or its convex version, the DEA estimator proposed by Charnes et al. (1978). Both are deterministic envelopment estimators with appealing characteristics, e.g., great flexibility and easy computability (Simar and Wilson, 2008). However, they are sensitive toward outliers and extreme values. The method applied in this chapter is based on the order-m estimator proposed by Cazals et al. (2002), which obtains a less restrictive benchmark. Unlike FDH and DEA, this estimator is robust against outliers and extreme values since it constructs a partial frontier that depends on a randomly drawn subset of the given sample.

Another advantage of the applied approach is that it facilitates inclusion of operating environments by conditioning the order-m efficiency estimation on a set of environmental variables, also referred to as z-variables. Obviously, municipalities act under very heterogeneous conditions, e.g., different unemployment rates and legal regulations, which are likely to influence the success in providing public services, but cannot be controlled by the municipality itself. Therefore, inefficiency might partly be due to these factors. Previous research mainly relies on parametric methods to control for the exogenous environment, i.e. Tobit regressions and the truncated bootstrap regression proposed by Daraio and Simar (2007b). Compared to these second-stage regressions, the conditional efficiency approach has three main distinctions: Firstly, it includes the environment immediately in the efficiency estimation rather than controlling for it in a second-stage. Secondly, it does not impose a functional relationship between the performance measure and the z-variables. Thirdly, it avoids the separability condition, i.e. z-variables do influence the efficiency but neither the attainable technology set nor the frontier (Daraio and Simar, 2007a). Although, there might be variables for which the separability condition holds, it is unlikely to be true for many socio-economic, financial and political variables researchers and politicians are particularly interested in.

De Witte and Kortelainen (2009) further extend the conditional efficiency approach. Most importantly for the application at hand, the method allows for simultaneous consideration of (unordered and ordered) discrete and continuous external factors, while avoiding the potential drawbacks of sample-splitting alternatives (frequency-based methods).[1] These drawbacks include, e.g., small sample

[1]Zelenyuk and Simar (2011) argue that the smoothing approach by Racine and Li (2004) involved in this methodology for smoothing discrete regressors might yield over- and/or undersmoothing of some groups of the discrete variables and therefore introduces biases into the estimates of the true regression relationship. In their simulations the authors compare the “simple smoothing” approach (i.e. the smoothing approach proposed by Racine and Li (2004)) with the “no-smoothing” approach (frequency-based approach) and the “complete-smoothing” approach. They find that the “simple smoothing” approach performs poorer relative to the other approaches. As noted by Zelenyuk and Simar, the true regression relationship in real data, how-

sizes of the sub-groups and challenging tests of significance of variables with multiple categories. Thus, we can estimate the performance of the German municipalities and control not only for, e.g., unemployment rates but also for different income classes and Federal States. Furthermore, the enhanced method allow visualizing and statistically testing the impact of z-variables on the municipal efficiency without leaving the nonparametric estimation framework.

The remainder of the chapter is a follows. The next subsection briefly outlines the institutional background of the German local governments and Section 2 reviews the relevant literature. Afterward, Section 3 introduces the methodology, beginning with the robust and conditional frontier estimation and then proceeding with the approaches of evaluating the impact of exogenous variables on the performance of municipalities. Section 4 discusses the choice of variables and describes the data. The estimation results are presented and discussed in Section 4. Finally, Section 6 concludes.[2]

3.1.2 German Municipalities

The German constitutional structure comprises of three levels of administration, i.e. the Federal government (*Bund*), the Federal States (*Länder*), and local governments (*Kommunen*). The latter, in turn, unifies the lowest territorial entities, namely the municipalities (*Gemeinden*) and the counties (*Landkreise and Gemeineverbände*). Counties can further be distinguished into rural counties (*Landkreise*), which are the association of a determined number of municipalities, and urban counties consisting of only one municipality (county-independent city). Although municipalities and counties are endowed with the right of self-governance, they have no national jurisdiction and are constitutional parts of the *Länder* (Nierhaus, 2006). Hence, they are subject to surveillance and decisional authority.[3]

The right of self-governance is assured by the German constitution (*Grundgesetz - GG*, *Artikel* 28, *Absatz* 2 *GG*) and implies that local governments can execute all matters of local interest on their own.[4] Rather than describing a fixed

ever, remains unknown and, consequently, a categorical denial would be hasty. We leave this question for further research.

[2] We especially thank Mika Kortelainen for methodological insights and for kindly sharing parts of his code. Further, we thank Pio Baake, Astrid Cullmann, Anne Neumann, and Christian von Hirschhausen for fruitful and inspiring discussions. All remaining errors are ours, the usual disclaimer applies.

[3] One consequence is that municipalities are not entirely immune against consolidating or disbanding intentions (Nierhaus, 2006).

[4] The counties' self-governance right is mitigated and restricted to duties assigned by the legislature (Nierhaus, 2006). The traditional duties of counties are of super-municipal, auxiliary or countervailing character.

stock of duties, the term self-responsibility relates to different fields of municipal sovereignty, e.g., the sovereignty of territory, of organization, of personnel and finance. In general, the specific tasks accomplished by the municipalities, can be distinguished into four groups: (i) voluntary self-governmental tasks, e.g., green space, theater and libraries; (ii) mandatory self-governmental tasks, e.g., waste disposal and land-use planning; (iii) mandatory tasks with instruction;[5] e.g., fire brigade and public order office; and (iv) mandatory transferred state tasks, e.g., conducting population censuses and federal elections. The first two groups refer to the matters of local interest where the municipalities can independently decide whether, when and how they fulfill the voluntary tasks. With respect to the mandatory tasks they are obliged to perform them while being subject to legal and/or functional supervision. However, the scope of municipal duties varies considerably across Federal States, because each Federal State has individual municipal enactments.

Due to the sovereignty of finance, the municipalities can autonomously control their revenues and expenses within the framework of the legal orderly budget (Nierhaus, 2006). Basically there are three sources of income available, each of which roughly accounting for one third of total revenue (Wehling, 2006). The main sources are (i) tax revenues; (ii) grants and fund allocation; and (iii) charges and financial contributions. The local business tax and personal income tax, both depending on the economic power of the individual municipality, are particularly important since they can be spent at the municipality's own discretion. On the contrary, user charges and financial contributions are only raised for specific return services, e.g., swimming pool use. Because tax revenues and user charges alone do not cover the expenses that are needed to accomplish the municipal tasks, the *Bund*, principally, and *Länder* additionally assign grants and fund allocations, where the latter can be either of general or specific nature.

However, the general financial position of the local governments in Germany remains stressed. Since 1992, the revenues are frequently exceeded by the expenses (Statistisches Bundesamt, 2008, 2010, 2011). One reason might be that additional duties are transferred to the municipal level. Thereby, it remains unclear whether the municipalities have a claim against *Bund* and *Länder* related to (additional) obliged and transferred duties (Nierhaus, 2006). In general, the *Bund* is not committed to financially compensate local governments due to the two-tier structure of the state; recall that local governments are constitutional parts of the *Länder*. Only a few *Länder* guarantee a full coverage of the additional burdens. In other Federal States, the legislature determines rules for cost coverage, and in some cases

[5]Mandatory tasks with instruction can also relate to tasks with local interest characteristics.

refers only to general financial constitutional rules.

3.2 Literature Overview

Existing research concerned with the measurement of public sector efficiency by means of frontier techniques consists of an extensive body of literature whose scope has multiple dimensions. For example, the analyses deal with different tiers of governments,[6] apply different methodologies and use different data. As noted by Balaguer-Coll et al. (2007), their aims and conclusions vary notably. Worthington and Dollery (2000), De Borger and Kerstens (2000) and more recently Kalb (2010b) give comprehensive surveys of the related literature.

In order to measure the performance of units performing public sector activities, e.g., subnational governments and public entities, the empirical literature relies on both parametric, i.e. SFA, and nonparametric, i.e. DEA and FDH, frontier methods. Comparing the alternative approaches, e.g., De Borger and Kerstens (1996a), Worthington (2000) and Geys and Moesen (2009) discuss their underlying assumptions and show that results differ significantly. Going beyond the pure estimation of efficiency, researchers are interested in explaining the variations in the estimated performance; using both frontier approaches. For this purpose, factors are considered that are commonly assumed to be out of the units' control. Those factors are mainly of socio-economic, demographic, geographic, financial and political nature, including e.g., population density, share of unemployed to total population, measures of decentralization, geographical distance to capital, education of population etc. For only few of them, e.g., governmental transfers, the empirical evidence is predominantly consistent; see e.g., Nieswand and Seifert (2011).

With respect to parametric approaches, e.g., Kalb (2010a) investigates the impact of grants as well as other financial and socio-economic control variables on the cost efficiency of 1,111 German municipalities in the state of Baden-Wuerttemberg. The author uses the model proposed by Battese and Coelli (1995), which makes the mean of the inefficiency term a function of the exogenous (non-discretionary) variables. Kalb finds that the mean of cost efficiency increases with additional grants and is higher for abundant and financially weak municipalities. Among other findings, the unemployment rate and the population density negatively impacts the inefficiency's mean, whereas accommodation facilities, political concentration and monopolization, as well as the share of left-wing parties have increasing effects. For a similar sample, Geys et al. (2008) investigate the implications of demographic

[6] We exclude efficiency analyses at the country level from further discussion.

change for the cost efficiency of municipalities. Relying on the same estimation approach, they estimate the cost efficiency and examine the municipalities' economies of scale, expressed as the cost elasticity of population size, in more detail. A rolling regression reveals that particularly small municipalities would suffer from a shrinking population. Further, Geys et al. (2010) show that greater voter involvement increases the average efficiency.

More closely related to this chapter are nonparametric approaches of efficiency estimation. Explaining variations in nonparametrically estimated inefficiency, predominantly relies on different parametric techniques that are used in a second stage to regress a set of (exogenous) variables on the obtained performance estimates (Nieswand and Seifert, 2011). To our best knowledge, Hitzschke (2011) presents the only analysis addressing German local governments, an analysis of 112 cities. The evaluation relies on the average values of data for the years 2004 to 2008. To identify the optimal city size, Hitzschke regresses the population on the scale efficiency estimates obtained by DEA using quadratic and cubical model specifications and OLS regression. The results indicate that the optimal size of cities is roughly 220,000 inhabitants, which roughly coincides with the observed sample mean size.

De Borger et al. (1994) and De Borger and Kerstens (1996b) estimate the productive and technical efficiency assuming FDH technologies, for 589 Belgian municipalities in 1985. Using Tobit regression, they find that larger municipalities are more efficient, with higher tax prices and better inhabitant education levels increasing efficiency. In contrast, block grants and lower average personal income negatively affects the performance. The coefficients of the political variables, e.g., the number of coalition members, are statistically insignificant.

Afonso and Fernandes (2008) evaluate Portuguese municipalities for 2001, regressing socio-economic explanatory variables on DEA efficiency scores by means of Tobit regression. The results of the analysis indicate that the percentage of population with higher education is positively related with spending efficiency in most regions. Similar, the purchasing power effects the performance. For other variables, i.e. population density, population growth and geographic distance to the capitol, the findings are mixed.

Spanish municipalities are investigated by Balaguer-Coll et al. (2002) and Giménez and Prior (2007). Using the same approach as Afonso and Fernandes (2008), Balaguer-Coll et al. (2002) consider 258 Valencian municipalities for the years 1992-1995 and find negative effects of per capita tax revenues and grants, whereas commercial activities increase technical efficiency. Other included socio-economic and financial variables (per capita financial liabilities, unemployment rate, tourism index, per capita disposal household income) do no significantly impact performance.

Giménez and Prior (2007) assess a sample of 258 Catalonian municipalities in 1996 with more than 2,000 inhabitants. The authors examine the municipal cost efficiency and three proposed decompositions of it, i.e. short-term variable cost efficiency, capacity utilization and scale efficiency. Giménez and Prior using linear programming to conduct a non-convex frontier analysis and Tobit regression in the second step. With respect to the operational environment, municipalities with high per capita income perform worst compared to median and low per-capita income. The existence of libraries, an additional service provided by larger municipalities (more than 5,000 inhabitants) also reduced cost efficiency. Like population indicating the size of municipalities, commercial and tourism activity increase the performance. Specific population characteristics, share of children and retired persons, are not statistically significant.

Hauner (2008) applies a truncated regression to overcome the drawbacks related to Tobit regressions, conducting an analysis on four public sectors (health, education, social protection and social sectors) for 89 Russian regions. The main findings include that populations income, good governance and domestic control positively while e.g., federal grants negatively impacts the performance.

The bootstrap truncated regression proposed by Simar and Wilson (2007) is used by Nieswand and Seifert (2011), who investigate the spending efficiency of 96 subnational governments in France during 2008. The results suggest that higher median income, higher shares of elderly and the distance to the capital decrease performance. The coefficients of the local governments' size and the dummy accounting for locations at seaside are not significantly different from zero.

Balaguer-Coll et al. (2007) are the first to assess the impact of explanatory variables on municipal inefficiencies in a fully nonparametric framework. The first stage incorporates estimating the performance of 414 Spanish municipalities located in the Valencian region during 1995 using DEA and FDH technologies. Unlike previous studies, the second stage incorporates nonparametric (bivariate) joint density estimation (kernel smoothing) in order to investigate how selected political and fiscal influence the municipal performance. Similar to previous findings, grants and tax revenues decrease efficiency, as do self-generated revenues and deficits. For the governing party share of votes and loans and issued securities, the effect cannot clearly be determined.

To our best knowledge, the conditional efficiency framework that allows the incorporation the operational environment directly in the efficiency estimation is not yet applied in the context of local government performance. However, the performance of particular public services, e.g., water supply, libraries and education, are empirically investigated separately using conditional efficiency measurement.

Recently, Zschille (2012) assesses the potential gains from horizontal mergers for the municipal service of water supply in Germany using the approach by Daraio and Simar (2005, 2007b). For the sample of 651 observations in 2006, Zschille finds that the conditional DEA efficiency measures are positively effected when accounting for output density and the share of water losses, whereas the share of groundwater usage the impact appears to be weakly negative. Other applications of this approach include De Witte and Marques (2011) and Marques and De Witte (2011) for the water sector, and Van Klaveren and De Witte (2010) for education.

De Witte and Kortelainen (2009) further develop the ideas of the conditional efficiency measurement. Empirically, they estimate the performance of students using the robust and conditional order-m approach while accounting for continuous and discrete background variables. The 2006 sample comprises of 3,992 Dutch 15-year old student test scores. The performance considerably changes when the control variables are included. Further, by nonparametric regression, De Witte and Kortelainen show which of these variables exhibit statistical significance. For example, the results indicate that the parent's education level significantly and the number of books at home increase the test outcomes of students. Additionally, having an own room at home and the family structure are examples for unordered discrete variables with significant impact.

The same technique is applied by De Witte and Geys (2011), who reconsider the provision of most public services, e.g., libraries and education, as a two-step production process. They provide empirical evidence for 290 municipal public libraries in Flanders. De Witte and Geys argue that the provided public service potential, e.g., opening hours of libraries, is not identical to the observable outcomes and therefore dependent on public demand which needs to be considered in performance measurement. The authors find that, e.g., having a female mayor, a wealthier population, and a higher share of public service revenues coming from local resources, positively impact the library productive efficiency.

To summarize, the service provision of local governments has been a continual subject for empirical research. However, conditional efficiency analysis has not been applied in this context. This chapter aims to bridge this gap and thereby contribute to existing evidence on the efficiency of public service provision.

3.3 Methodology

3.3.1 Conditional Order-m Efficiency Estimator

Evaluating the efficiency of municipalities by means of nonparametric frontier techniques, requires estimating the underlying technology set. For this purpose, municipalities can be considered as DMUs that transform p inputs into q outputs using a common technology. The set of technically feasible combinations of inputs and outputs describe the production set Ψ, which can formally be defined by

$$\Psi = \left\{(x,y) \in \mathbb{R}_+^{p+q} \,|\, x \text{ can produce } y \right\}, \tag{3.1}$$

where x represents the vector of inputs and y the vector of outputs. Usually, only a few assumptions are imposed on Ψ, such as free disposability of inputs and output; for details see e.g., Shepard (1970) and Daraio and Simar (2007b). The boundary of Ψ, denoted by Ψ^δ, represents the efficient production frontier and can be formally expressed as

$$\Psi^\delta = \left\{(x,y) \in \Psi \,|\, \left(\gamma x, \gamma^{-1} y\right) \notin \Psi \text{ for any } \gamma < 1 \right\}. \tag{3.2}$$

A DMU using a production plan that belongs to Ψ^δ is regarded as efficient and its input-output-combination cannot be improved. Therefore, DMUs operating at points that are in the interior of Ψ^δ exhibit inefficiencies (Simar and Wilson, 2001). We use the input-oriented Debreu-Farrell[7] measure of efficiency to determine the extend of inefficiency, i.e. the distance from a given observation (x_0, y_0) to that frontier. This measure is defined as

$$\theta\left(x_0, y_0\right) = inf\left\{\theta \,|\, \left(\theta x_0, y_0\right) \in \Psi\right\} \tag{3.3}$$

with θ denoting the DMU individual efficiency score. The DMU is considered efficient, if $\theta = 1$. Whereas, $\theta < 1$ indicates inefficiency and corresponds to the proportionate reduction of inputs to achieve the efficient frontier. In practice, the production set Ψ, its frontier Ψ^δ, and hence, the efficiency scores θ are unknown, and need to be estimated from a random sample of production units $\mathcal{X} = \{(x_i, y_i) \,|\, i = 1, \ldots, n\}$ (Simar and Wilson, 2007). Alternatively to stochastic and deterministic frontier models, Cazals et al. (2002) propose the conditional efficiency approach that relies on the probabilistic formulation of the production

[7] This measure is based on the work of Debreu (1951) and Farrell (1957). Alternatively, the concept proposed by Shepard (1970) could be used.

process while allowing to immediately incorporate external variables, also referred to as z-variables.[8] The idea of this concept is to use additional information captured by z-variables that potentially affect the production process, but are neither inputs nor outputs under the control of the considered units (Simar and Wilson, 2007). According to Daraio and Simar (2007b), in this framework, the production process is fully described by the joint probability function, which gives the probability that a unit operating at input-output-levels (x, y) is dominated. Being dominated means that another unit produces at least as much output while using no more of any input than the considered unit (Simar and Wilson, 2008). Conditioning the joint probability function on given levels of a set of z-variables, i.e. $Z = z$, allows incorporating the operating conditions without assuming the separability condition. The conditional joint probability function, denoted by $H_{XY|Z}$, is given by:

$$H_{XY|Z}\left(x, y \,|\, z\right) = Prob\left(X \leq x, Y \geq y \,|\, Z = z\right). \tag{3.4}$$

defining Ψ^z, the feasible production set when $Z = z$. Therefore, $H_{XY|Z}$ represents the probability of a unit operating at the level (x, y) to be dominated given that $Z = z$. For the input-oriented measure of efficiency, the expression in Equation 3.4 can be further decomposed as follows

$$\begin{aligned} H_{XY|Z}\left(x, y \,|\, z\right) &= Prob\left(X \leq x \,|\, Y \geq y, Z = z\right) \,|\, Prob\left(Y \geq y \,|\, Z = z\right) \\ &= F_{X|Y,Z}\left(x \,|\, y, z\right) S_{Y|Z}\left(y \,|\, z\right). \end{aligned} \tag{3.5}$$

where $F_{X|Y,Z}$ denotes the conditional distribution of X, and $S_{Y|Z}$ the conditional survivor function of Y.[9] Supposing that the conditional probabilities exist, i.e. $S_{Y|Z}\left(y|z\right) > 0$, Ψ^z can be defined by the support of $F_{X|Y,Z}$ for all y (Daraio and Simar, 2007b). Then, similar to the unconditional case, under the assumption of free disposability, the lower boundary of $F_{X|Y,Z}$ can then be considered as the input-oriented Farrell efficiency frontier. For an unit producing the output level y with operating conditions described by z, the corresponding input-oriented efficiency measures are

$$\begin{aligned} \theta\left(x, y|z\right) &= inf\left\{\theta \,|\, \left(\theta x_0, y_0\right) \in \Psi^z\right\} \\ &= inf\left\{\theta \,|\, F_{X|Y,Z}\left(\theta \,|\, y, z\right) > 0\right\}. \end{aligned} \tag{3.6}$$

Defining a nonparametric estimator of $\theta\left(x, y|z\right)$, demands an empirical analog of the unknown conditional distribution function $F_{X|Y,Z}$. This requires smoothing in z since the conditioning on a respective operating environment is described by

[8]The ideas of Cazals et al. (2002) are further extended by Daraio and Simar (2005) to the multivariate case and by Daraio and Simar (2007b) to the convex technologies.

[9]The conditional distribution is nonstandard due to the event describing the condition, i.e. $Y \geq y$ instead of $Y = y$ (Daraio and Simar, 2007b).

the equality constraint $Z = z$ (De Witte and Kortelainen, 2009). As shown by Cazals et al. (2002), an estimator of $\theta(x, y|z)$ can be obtained by plugging the nonparametric kernel estimator of $F_{X|Y,Z}$ defined as:

$$\hat{F}_{X|Y,Z} = \frac{\sum_{i=1}^{n} I(X_i \le x, Y_i \ge y) K_{\hat{h}}(z, z_i)}{\sum_{i=1}^{n} I(Y_i \ge y) K_{\hat{h}}(z, z_i)} \tag{3.7}$$

into Equation 3.6, where $I(\cdot)$ denotes the indicator function, $K_{\hat{h}}(\cdot)$ the kernel function and $\hat{h}$ the bandwidths parameter that need to be estimated using an appropriate bandwidths choice algorithm. The corresponding efficiency estimator is then given by

$$\hat{\theta}_n(x, y|z) = inf\left(\theta \,|\, \hat{F}_{X|Y,Z_n}(\theta \,|\, y, z) > 0\right). \tag{3.8}$$

Similar to the unconditional case, this yield an estimator that asymptotically converges to the deterministic FDH estimator (Cazals et al., 2002) and therefore, could be termed as the 'conditional FDH efficiency measure' (Daraio and Simar, 2005). Formally the FDH estimator assumes that all observations belong to the feasible production set causing sensitivity toward outliers and extreme values; see e.g., Daraio and Simar (2007a). To overcome this drawback, Cazals et al. (2002) propose the order-m estimator, which uses the concept of the expected minimum input function to derive the input-oriented efficiency scores.[10] Rather than building the frontier by enveloping all observations (full frontier), a partial frontier is constructed based on a randomly drawn subset of m out of n units that at least produce the output level y, given that $Z = z$. Among those m units, the expected minimum achievable input level, i.e. the average of the minimum values of inputs, represent the alternative reasonable benchmark against which an observation is compared to (Daraio and Simar, 2007a). The corresponding feasible production set $\Psi_m^z(y)$ can be defined as

$$\Psi_m^z(y) = \left\{(x, y') \in \mathbb{R}_+^{p+q} \,|\, x \ge X_i, y' > y,\, i = 1, \ldots m\right\} \tag{3.9}$$

Daraio and Simar (2005) note that this set depends on z since the m *iid* random variables X_i, $i = 1, \ldots, m$ are generated through the conditional p-variate distribution function $F_{X|Y,Z}$. When defining

$$\tilde{\theta}_m^z(x, y) = inf\left\{\theta \,|\, (\theta x_0, y_0) \in \Psi_m^z(y)\right\} \tag{3.10}$$

[10]The approach also includes estimating output-oriented efficiency relying on the expected maximum output function.

the conditional order-m input-oriented efficiency measure can then be obtained by

$$\begin{aligned}\theta_m(x,y\,|\,z) \quad &= E\left(\tilde{\theta}_m^z(x,y),|\,Y \geq y, Z = z\right) \\ &\int_0^\infty \left[1 - F_{X|Y,n}(ux,y)\right]^m du.\end{aligned} \tag{3.11}$$

Using the plug-in-principle, the empirical version of Equation 3.11 is then defined as

$$\hat{\theta}_{m,n}(x,y|z) = \int_0^\infty \left[1 - \hat{F}_{X|Y,Z_n}(ux\,|\,y,z)\right]^m du. \tag{3.12}$$

Note, that the order-m estimator is no longer bounded to unity and estimates can take values larger than one. A value of $\hat{\theta}_{m,n}(x,y)$ smaller than 1 indicates the extent to which a unit could reduce its inputs relative to the average performance of the randomly drawn reference set. If $\hat{\theta}_{m,n}(x,y)$ is greater than 1, the unit is on average performing superior than its m randomly drawn reference units (De Witte and Kortelainen, 2009). Unlike for the unconditional order-m estimator[11], the curse of dimensionality cannot be avoided in the conditional case due to smoothing in z; see e.g., Cazals et al. (2002); Park et al. (2006); Jeong et al. (2010). However, based on Li and Racine (2008), De Witte and Kortelainen (2009) show that the convergence rates of the nonparametric estimators for conditional density and distribution functions do not depend on the number of discrete but only on the number of continuous variables in Z. Therefore, the curse of dimensionality cannot completely overcome by conditioning when continuous variables are considered.

For the kernel estimation we follow De Witte and Kortelainen (2009) and apply a methodology, which generalizes existing approaches of conditional efficiency estimation such that continuous and (ordered and unordered) discrete z-variables can be incorporated. Furthermore, De Witte and Kortelainen show the practical implementation of the optimal bandwidth selection method proposed by Bădin et al. (2010). This method provides municipality-specific bandwidths, but more importantly, avoids the separability condition also when selecting bandwidths (this is not the case when applying, e.g., the k-nearest neighbor approach suggested by Daraio and Simar, 2005). As mentioned by De Witte and Geys (2011), estimating kernel densities around the evaluated environment provides the empirical probabilities for a unit of the full sample to be drawn as one of the m units that construct the partial frontier; the more similar the greater the probability.

In order to estimate the kernel function, De Witte and Kortelainen (2009) separate the components of the multivariate Z according to the three groups of z-variables and define a vector of r continuous variables ($z^c \in \mathbb{R}^r$), a vector of

[11] Cazals et al. (2002) show that the unconditional order-m estimator is more robust against outliers than the deterministic, nonparametric estimator FDH.

v ordered discrete ($z^o \in \mathbb{R}^v$), and a vector of w unordered discrete ($z^u \in \mathbb{R}^w$). Hence, the vector of observed z-variables is redefined by $z_i = (z^c, z^o, z^u)$, $i = 1, \ldots, n$. Further, z^o_{is} and z^u_{is} represent the sth component of z^o_i and z^u_i. Both types of discrete variables can take more than two different values. For each of the components, standard multivariate product kernel functions are used to smooth the z-variables. By multiplying these functions, the generalized product kernel function is obtained, defined by

$$K_h(z, z_i) = \prod_{s=1}^{r} \frac{1}{h^c_s} l^c \left(\frac{z^c_s - z^c_{is}}{h^c_s} \right) \prod_{s=r+1}^{r+v} l^o(z^o_s, z^o_{is}, h^o_s) \prod_{s=r+v+1}^{r+v+w} l^u(z^u_s, z^u_{is}, h^u_s) \quad (3.13)$$

where $l^c(\cdot)$, $l^u(\cdot)$ and $l^o(\cdot)$ are univariate kernel functions and h^c_s, h^o_s and h^u_s are bandwidths for the continuous, ordered and unordered discrete z-variables, respectively. Following De Witte and Kortelainen, we use the Epanechnikov kernel function having compact support, i.e. kernels for which $k(z) = 0$ if $\mid z \mid \geq 1$, for the continuous variables, the Aitchison and Aitken (1976) discrete univariate kernel function for the unordered discrete variables, and the Li and Racine (2007) discrete kernel function for the ordered discrete variables, respectively.

Finally, the method of bandwidth selection needs to be specified. This is a crucial part of nonparametric kernel estimation, since different bandwidths can generate radically different impressions of the underlying distributions (Pagan and Ullah, 1999; Li and Racine, 2007). De Witte and Kortelainen (2009) suggest to use the cross-validation method proposed by Li and Racine (2008) to estimate optimal bandwidths where the choice criterion is based on the weighted integrated squared errors. By this approach, municipality-specific optimal bandwidths for each of the z-variables can be obtained while they are estimated considering a limited reference set which consists of other municipalities that produce at least the output level the considered one, i.e. $y_i > y$.

3.3.2 Nonparametric Significance Test

Testing the significance of variables requires estimating the unknown relationship between the dependent and explanatory variables. For this purpose, De Witte and Kortelainen (2009) suggest using the nonparametric regression method developed by Li and Racine (2004) and Racine and Li (2004). Conveniently in our context, this method avoids imposing any parametric assumption on the relationship and is able to incorporate both continuous and discrete variables. The developed estimation technique uses kernel weighted local linear methods that smooths both types of regressors with smoothing parameters selected by least square cross-validation.

In general, an advantage of local linear estimators is that they can be analyzed using standard regression techniques (Pagan and Ullah, 1999).

To formulate the unknown regression function by which the influence of external variables on the production process is evaluated, define the ratio of the conditional and the unconditional efficiency scores for municipality i as $\hat{Q}_i^z = \frac{\hat{\theta}_{m,n}(x,y|z)}{\hat{\theta}_{m,n}(x,y)}$ and consider the nonparametric regression model:

$$\hat{Q}_i^z = \tilde{f}\left(z_i\right) + \epsilon_i \tag{3.14}$$

where $\tilde{f}$ represents the conditional mean function of the estimated ratio with $\tilde{f} = \tilde{\alpha} - \left(z_i^c - z^c\right)\tilde{\beta}$, z_i the vector of continuous, ordered and unordered external variables for municipality i, and ϵ_i the individual error term. The external variables and the error term are assumed to be uncorrelated, i.e. $E\left(\epsilon_i \mid z_i\right) = 0$. The parameters α and β can be interpreted as the conditional mean of Q_i^z for $Z_i = z$, which is equivalent to the local linear regression problem of estimating an intercept, and its partial derivative, which is equivalent to the local linear regression problem of estimating a slope, respectively (Racine, 2008). Performing a weighted least square regression gives the estimators of α and β. By solving the minimization problem

$$\min_{\alpha,\beta} \sum_{i=1}^{n} \left(\hat{Q}_i^z - \tilde{\alpha} - \left(z_i^c - z^c\right)\tilde{\beta}\right)^2 K_h\left(z, z_i\right), \tag{3.15}$$

the local linear estimators of α and β, i.e. $\tilde{\alpha} = \tilde{\alpha}\left(z\right)$ and $\tilde{\beta} = \beta\left(z^c\right)$, are obtained where K_h is the generalized product kernel function as specified in Equation 3.13 and h the vector of bandwidths estimated by least-square cross-validation (Li and Racine, 2004) . $\tilde{\alpha}\left(z\right)$ and $\tilde{\beta} = \left(z^c\right)$ are consistent estimators for the true conditional mean function $f\left(z\right) = E\left(Q^z \mid z\right)$ and the gradient $\beta\left(z^c\right) = \frac{\delta E(Q^z \mid z)}{\delta z^c}$. Racine (1997) emphasizes that nonparametric regression techniques allow the gradient vary over its domain whereas parametric approaches typically assumes the gradient to be constant over its entire domain affecting the type of test statistic.

Regarding the statistical inference on the discrete and continuous external variables, we follow De Witte and Kortelainen (2009) and apply the significance tests proposed by Racine (1997) and Racine et al. (2006), respectively. Based on the distribution-free nature of the regression methods, both tests derive statistical inference in a fully nonparametric and robust framework, and therefore, do not rely on parametric assumptions themselves. Further, they allow for the accounting of the different types of external variables. According to Racine et al. (2006) significance tests in nonparametric kernel frameworks are generally consistent under less restrictive assumptions than those required for parametric approaches.

The basic idea of the tests is evaluating whether the continuous or discrete z-variable is an irrelevant regressor in the regression model. Irrelevant here means that the conditional mean remains unchanged whatever value the discrete variable takes and whenever the continuous variable is included or not.

First, turning to the significance test for continuous variables, the test hypothesis for a specific regressor can be formulated as

$$H_0 : \; E\left(Q^z \mid \tilde{Z}, Z^c\right) = E\left(Q^z \mid \tilde{Z}\right) \text{ almost everywhere, and} \tag{3.16}$$

where Z^c denotes the sth component of the continuous variables under consideration while $\tilde{Z}$ represent all other continuous and discrete variables. The alternative hypothesis is the negotiation of the null, i.e. $H_1 : E\left(Q^z \mid \tilde{Z}, Z^c\right) \neq E\left(Q^z \mid \tilde{Z}\right)$. In order to derive a test statistic, Equation 3.16 can be reformulated as follows

$$H_0 : \frac{\delta E\left(Q^z \mid \tilde{Z}, Z^c_s\right)}{\delta Z^c_s} = \beta\left(Z^c_s\right) = 0 \tag{3.17}$$

declaring that the partial derivative of $f(Z)$ with respect to the z-variable under consideration, i.e. Z^c_s, equals zero. Hence, the respective z-variable has no influence on the conditional mean. Since the partial derivative varies over its domain, some aggregated measure is involved in stating the hypotheses, for details see Racine (1997). The test statistic is

$$I^c = E\left\{\beta\left(Z^c_s\right)^2\right\} \tag{3.18}$$

and the average of the squared local linear estimator of the partial derivative, $\hat{\beta}\left(z^c\right)$, provides the consistent empirical analog of the test statistic, i.e.

$$I^c_n = \frac{1}{n}\sum_{i=1}^{n} \hat{\beta}\left(z^c_{is}\right). \tag{3.19}$$

I^c_n is a consistent estimator of I^c since $I^c \to 0$ under the null, and $I^c \to 1$ under the alternative (Racine, 2008). In order to approximate the null distribution and deduce the critical value of the test statistic, the naive bootstrap (resampling algorithm) with 1,000 replications is applied. According to Racine (1997) this technique yield in a test that has correct size, possesses high power and is notably insensitive to bandwidth choices.

Turning to the significance test on discrete variables, the null hypothesis can be written as

$$H_0 : \; E\left(Q^z \mid \tilde{Z}, Z^d\right) = E\left(Q^z \mid \tilde{Z}\right) \text{ almost everywhere} \tag{3.20}$$

and the alternative hypothesis is again determined by its negotiation, i.e. H_1 : $E\left(Q^z \,|\, \tilde{Z}, Z^d\right) \neq E\left(Q^z \,|\, \tilde{Z}\right)$. To constitute the respective test statistic, it is assumed that the discrete variable Z_s^d can take different g values: $\{1, 2, \ldots, g-1\}$. Again, the null hypothesis can be redefined and represented by

$$f\left(\tilde{Z}, Z_s^d = l\right) = f\left(\tilde{Z}, Z_s^d = 0\right) \text{ for all } \tilde{Z} \text{ and for all } l = 1, 2 \ldots, g-1. \quad (3.21)$$

Hence, irrespective of the value the (ordered or unordered) discrete variable Z_s^d takes, the conditional mean function equals to the case where $Z_s^d = 0$. The test statistic is given by an estimator of

$$I^d = \sum_{g=1}^{g-1} E\left\{\left[f\left(\tilde{Z}, Z_s^d = l\right) - f\left(\tilde{Z}, Z_s^d = 0\right)\right]^2\right\}. \quad (3.22)$$

The estimator for Equation 3.22 can be obtained by replacing the unknown conditional mean functions by its local linear estimators $\hat{f}(\cdot)$ at the given values of the variable. Hence, the feasible test statistic is

$$I_n^d = \frac{1}{n} \sum_{i=1}^{n} \sum_{g=1}^{g-1} \left[\hat{f}\left(\tilde{Z}, Z_s^d = l\right) - \hat{f}\left(\tilde{Z}, Z_s^d = 0\right)\right]^2. \quad (3.23)$$

Only if H_0 is true, $I_n^d = 0$, and positive otherwise, i.e. $I_n^d \geq 0$; therefore, I_n^d is a consistent estimator of I^d and proper measure for testing H_0 (Park et al., 2006). To derive the null distribution of the test statistic and its critical values, we apply the bootstrap method I, from Racine et al. (2006), with 1,000 replications. Similar to the significance test for continuous variables, the test for discrete variables has correct size and power increasing with deviations from the null hypothesis. Furthermore, the test appears to be more powerful in finite samples compared to frequency-based alternatives, i.e. tests that rely on splitting the sample.

3.3.3 Partial Regression Plots

Although, the significance test provide inference for the z-variables, it does not provide useful information about the direction of influence. In order to explore the effect of the exogenous variables, e.g., Daraio and Simar (2005, 2007a) suggest plotting the ratio $\hat{Q}^z$ against the continuous variable and its nonparametric smoothing regression line. For the input-orientated efficiency assessment, the influence of the univariate and continuous Z can be considered unfavorable if the smoothed regression line is increasing. In this case, the respective variable represents an operating environment that reflects an undesired output to be produced requiring additional resources. In contrast, a decreasing line indicates a favorable operating environment that influences the production process as if it would be an substitutive input. The z-variable has no impact when the smoothed regression line is flat.

To interpret both, continuous and discrete z-variables in a multivariate setting, partial regression plots; see e.g., Daraio and Simar (2007a), can be used to detect their impacts. In these plots, the ratio $\hat{Q}^z$ is plotted against the variable in question while all other exogenous variables are set to a fixed value, e.g., the median. The interpretation of the smoothed regression line is analogue. Note, that discrete variables are evaluated at their particular categories (levels) and an average effect cannot be deduced for unordered discrete variables.

3.4 Model Specification and Data

The analysis is based on data coming from two sources, i.e. the Statistisches Bundesamt (2007a) and the Deutscher Städtetag (2008). The sample includes municipalities with more than 10,000 inhabitants[12] and consists of 1,060 observations for the year 2005. This covers about 67 percent of all municipalities and urban counties with more than 10,000 inhabitants in 2005.[13] To maintain as much comparability between the units as possible, we solely consider municipalities, omitting rural and urban counties. This reduced the original sample from 1,366 to 1,265 observations and further reductions are due to missing data. The data set is unique and outstanding as it comprises municipalities of nine Federal States[14] and is rich in terms of variables and observations.

Table 3.1 gives the characteristics of the data we use. Similar to other applications, like e.g., De Borger and Kerstens (1996b) and Geys et al. (2008), we use a single monetary aggregate to measure the input, namely the total current expenditures. These encompass expenditures that mainly arise on a regular basis for executing the administration and commissioning facilities and institutions, i.e. personnel expenditures, current goods and service expenditures, transfers to third parties, and other financial transfers.[15]

The chosen output measures relate to the most important municipal responsi-

[12]Unfortunately, the tables provided by the German Association of Cities do not consider the municipalities and counties (urban and rural) with fewer inhabitants.

[13]In 2005, there were 12,340 municipalities and urban counties in total of which 12.9 percent have more than 10,000 inhabitants. Roughly 73 percent of the total population lived in these areas (Statistisches Bundesamt, 2007b).

[14]We excluded the three city states Berlin, Hamburg and Bremen. Further, important financial data is not provided for Brandenburg, Mecklenburg-Hither Pomerania and Saxony-Anhalt.

[15]Not included are payments from the same passing through payments and clearance positions.

Table 3.1: Descriptive statistics for German municipalities

Variable	Min	Mean	Median	Max	Std.dev.
Input					
Total current expenditures [mn Euro]	0.52	37.87	20.85	1503.40	72.63
Outputs					
Total surface area [hectare]	440	6,909	6,110	35,750	4,556
Public thoroughfare[hectare]	28	431	376	3,267	286
School buildings [number]	1	7	6	85	6
Kindergarten places [number]	0	836	575	23,338	1,054
Sport clubs [number]	0	32	25	335	27
Population [thsd]	0.60	25.89	18.03	515.73	27.92
Continuous external variables					
Unemployment rate [percent]	4.32	11.25	10.20	32.49	4.27
Population density [inhabitants/hectare]	0.26	5.05	3.65	25.28	4.20
Unordered discrete external variables					
Federal State [levels$^{(1)}$]	2	5.35	6	11	2.61
Migration [levels$^{(2)}$]	1	1.54	2	2	0.50
Commute [levels$^{(3)}$]	1	1.37	1	2	0.48
Ordered discrete external variables					
Income class [levels$^{(4)}$]	1	2.50	2.50	4	1.12

Source: *Statistisches Bundesamt, Deutscher Stï¿œdetag.* Notes: observations=1060, year=2005. $^{(1)}$Federal State; 2: Baden-Wuerttemberg, 3: Bavaria, 4: Hesse, 6: Lower Saxony, 7: North Rhine-Westphalia, 8: Rhineland-Palatinate, 9: Schleswig-Holstein, 10: Saarland, 11: Saxony. $^{(2)}$Migration; 1: emigration $\leq$ migration, 2: emigration $>$ migration. $^{(3)}$Commute; 1: number of socially insured employees at place of work $\leq$ number of socially insured employees at place of living, 2: number of socially insured employees at place of work $>$ number of socially insured employees at place of living. $^{(4)}$Income class; 1: first quartile, 2: second quartile, 3: third quartile, 4: fourth quartile.

bilities and include both, voluntary and mandatory tasks. Thereby, we select the variables according to the objective of this chapter which is to compare municipalities with respect to their service provision, i.e. their efficiency in supplying a set of tasks. This corresponds to that particular part of the municipal production process, De Witte and Geys (2011) refer to as the first-stage. This is worthwhile to mention since some variables may indeed relate to service provision, but depend on the demand and, therefore, may reflect some other type of output. An example for this is the provision of sport clubs as the municipal tasks and the registered members of sport clubs. The six output measures address particularly the following five tasks: (i) land-use planning; (ii) municipal road construction and transportation; (iii) education and school administration; (iv) youth and health care; and (v) transferred state tasks. Whenever possible, we chose direct measures of output rather than (rough) proxies.

The total surface area in hectare represents the municipal tasks of urban land-use planning, a central instrument of territorial organization. Although, the planning should generally be for the public good and promote a sustainable development, local governments are basically free in their eventual decisions with respect to allocating land. The reason for including this variable is that greater areas imply greater complexity and require greater coordination of interests.

Further, we use the space dedicated as public thoroughfares in hectare to measure the municipal task of road construction and transportation. Although, current expenditures do not consider investments, many expenses are devoted to the maintenance of roads.

Turning to education and school administration, two output measures are considered. The municipal responsibility refers to the outer school administrations (school buildings and the equipment of schools) while education authorities are responsible for the inner school administration. To capture this task, we use the number of basic and intermediate schools, which are predominantly public.[16] The number of kindergarten places accounts for education. Historically, the provision of kindergarten places is much more developed in regions located in former East Germany. However, the German government intends to significantly expand the supply of child care facilities, with municipalities responsible for providing, and baring most of the costs, of the supply, thus, making this variable particularly important.

The number of sport clubs represents tasks related to youth and health care. Like other tasks, e.g., museums and theaters (cultural services) and recovering areas (recovery services), providing sport facilities and promoting associations are voluntary responsibilities and crucial to communal life organization. Due to limited data availability and to keep the dimensionality of the model slim, we foresee including more measures.

Lastly, the population measured by the total number of inhabitants (in thousand, thsd) is included, particularly to reflect the mandatory state transferred tasks such as issuing identification cards.[17] Additionally, it captures the genuine municipal tasks of administration, and presumably those services that municipalities provide beyond the legal minimum at their own expense (Balaguer-Coll et al., 2007). Hence, museums etc. might enter through this variable.

Regarding the contextual variables that incorporate the operational environment into the model, we select two continuous, three unordered and one ordered discrete variables. The continuous variables are the unemployment rate, measured as the ratio of unemployed persons to the total population aged between fifteen and 75 years, and the population density, defined as the total population per hectare of municipal surface area. The unemployment rate affects the economic power of

[16]Other studies, e.g., Geys et al. (2008) alternatively use the number of enrolled students in public schools. However, the required information is not provided for all *Länder*.

[17]Although, e.g., De Borger and Kerstens (1996a) note that population is rather a proxy than a direct output of municipal production, it might be more than that in our case due to the characteristics of the transferred state tasks in Germany.

municipalities, certainly narrowing the financial scope at all. Hence, it impacts the choice of public services provided by the municipality according to the preferences or requirements of the inhabitants. For example, the more social services are required and the less people are willing or able to pay for voluntary tasks, e.g., swimming pools and theaters, the more public budgets might be stressed. The population density clearly accounts for rural versus urban structures. Urban areas might possess positive effects on the costs of service provision due to the municipality's ability of raising potential agglomeration gains (De Borger and Kerstens, 1996a). However, the overall effect remains unclear since urban areas might induce reverse cost effects, e.g., due to higher property tax etc. (Kalb, 2010a).

The first unordered discrete variable describes the Federal State in which the municipality is located. Obviously, this is an important aspect when comparing the municipalities across federal borders. Particularly the different regulations for local governments are captured, but also all other systematical differences that might exist. For example, differences in input factor prices, if existing, can be controlled for. This aspects is probably one of the most striking (methodological) barriers which prevent previous nonparametric efficiency analysis from conducting nation-wide comparisons.

The second unordered discrete variable addresses effects occurring due to the residential mobility. We distinguish between the two cases where in the considered municipality emigration is greater (level 1) than migration and its counterpart (level 2). Tibout (1956) argues that the (allocative) efficiency local governments is increased when the population, more precise consumer-voters, is mobile and able to move to jurisdictions offering a mix of goods they prefer. We assume that in Germany mobility is generally given and formulate the according hypothesis that efficiency is positively effected when emigration is dominant (level 1). However, we are aware of the fact, that we neither can capture dynamics of the mobility nor can we identify the characteristics of moving people, e.g., whether they tax paying voters or not. The commuter variable distinguishes between municipalities in which the number of socially insured employees at place of work is smaller (level 1), and larger (level 2) respectively, than the number of socially insured employees at place of living. By this, we aim to capture the appearance of industry in the particular area,[18] or put it differently, whether people rather work or live in the respective municipality. Similar to unemployment, existing industry is expected to influences the provision of public services in one way or another. For example, we assume that it effects in particular outputs related to infrastructure, such as

[18]Alternatively, we could use the trade tax revenues. However, these are sensitive to economic cycles and therefore not meaningful in cross-section analysis.

streets or the supply of childcare facilities. Moreover, the related tax revenues increase the sovereignty of finance since they are not earmarked.

Finally, we consider the income of citizens measured by the municipal revenues of personal income tax per socially insured persons at place of living. Commonly, personal income is presented group-wise in order to make statements about income distribution. To incorporate the citizen's wealth, we define an ordered discrete variable that takes one of four levels according to the sample's quartiles of the measure. For example, level 1 denotes municipalities for which the personal income per socially insured persons does not exceed the samples' 25 percent quantile of this variable. Revenues related to personal income tax strengthens the radius of municipal operations since the local governments can spend them at their own discretion. Hence, it is likely that input and output choices differ among the defined income groups. For example, Geys et al. (2010) argue that higher income implies preferences toward services of higher quality and De Borger and Kerstens (1996a) point out that citizens of high-income municipalities may be less motivated to monitor the expenditures. Further, the population's wealth increases the bureaucratic slacks (technical inefficiency) and favors inefficient operations since fiscal capacity is sufficient. These arguments would suggest a negative relationship between citizens' income and efficiency. Note, this expectation implies that all outputs are addressed in the analysis and that, e.g., bureaucratic slacks involve no utility.

3.5 Results

Table 3.2 presents the summary statistics of both, the unconditional and the conditional order-m efficiency scores for the 1,060 German municipalities considered.[19] The individual scores are robust (against outliers and extreme values) and obtained by comparing the respective observation against a partial frontier that is constructed by 220 randomly drawn units, i.e. $m = 220$.[20] The estimation results indicate that the median performance of the municipalities notably increase from about 89 percent ($\hat{\theta}_m\,(x, y)_{median} = 0.8924$) to roughly 100 percent ($\hat{\theta}_m\,(x, y \mid z)_{median} = 0.9975$) when the heterogeneity in the framework conditions is taken into account. Although particularly conditional median performance in-

[19]All calculations are conducted using the statistical software *R*. For the conditional efficiency estimation and the nonparametric significance test, the additional package “np” version 0.40-12 by Hayfield and Racine (2008) is used.

[20]To select m, we recalculated the conditional order-m algorithm for the a given interval, i.e. $m = [20, 40, 60, \ldots, 800]$, and chose m such that the number of super-efficient observations stabilize (De Witte and Kortelainen, 2009).

Table 3.2: Order-m efficiency estimates and bandwidths

	Min	Mean	Median	Max	Std.dev.
Efficiency estimates					
Unconditional order-m estimates	0.1778	0.8726	0.8924	8.4413	0.3065
Conditional order-m estimates	0.2211	0.9243	0.9975	1.0002	0.1231
Bandwidths for z-variables					
Unemployment rate	0	1.06E+05	0.0324	2.55E+06	2.71E+05
Population density	0	4.09E+04	1.1627	2.39E+07	9.41E+05
Federal State	0	0.4314	0.5189	0.6379	0.1951
Migration	0	0.2681	0.3505	0.3588	0.1259
Commuter	0	0.1331	0.1002	0.3588	0.1200
Income class	0	0.5328	0.7174	0.7177	0.2724

dicates very efficient public service provision, the municipalities could provide the given level of services using on average 8 percent ($\hat{\theta}_m\,(x, y\,|\,z)_{mean} = 0.9243$) and 13 percent ($\hat{\theta}_m\,(x, y)_{mean} = 0.8726$) less resources. The other statistics of the efficiency scores reveal the difference between the models more clearer. The minimum efficiency score is larger in the conditional model and the maximum estimate heavily decreased compared to the unconditional case. Interestingly, the maximum conditional efficiency score exceeds only slightly unity, i.e. $\hat{\theta}_m\,(x, y\,|\,z)_{max} = 1.0002$, meaning that there are no municipalities that perform much better than the partial frontier average. Furthermore, the std.dev. diminishes from $\hat{\theta}_m\,(x, y)_{sd} = 0.3065$ to $\hat{\theta}_m\,(x, y\,|\,z)_{sd} = 0.1231$ clearly resulting from more similar conditions of the comparators in the conditional model.

In addition, Table 3.2 shows the descriptive statistics of the municipality-specific bandwidths used to smooth the kernel function in the conditional estimation approach. Notably, the average bandwidths for the continuous variables unemployment rate and population density are very large, which might be, according to De Witte and Kortelainen (2009), due to outlying values. Large maximum values can be a result of effectively smoothing out the insignificant variable for particular evaluated observations. The other bandwidths show no abnormalities.

Since the bandwidths themselves provide no formal evidence on the statistical significance of the corresponding z-variables, we conduct a fully nonparametric significance test based on nonparametric regression techniques. Table 3.3 shows the p-values obtained by the test. The results indicate that each of the included z-variable is statistically significant, and hence, relevant for explaining variations in the performance of the municipalities. Further, Table 3.3 summarizes the average effects of the continuous and ordered discrete variables[21] and indicates the overall implication of each z-variable for the municipal performance measured by

[21] Note that an average effect for unordered variables is meaningless. Instead, the effect is interpreted among the levels.

Table 3.3: Nonparametric significance test and performance implications

Variable	p-value	Average effect	Performance ...
Unemployment rate	0.013**	mixed	in- and decreases depending on evaluation point
Population density	2E-16***	unfavorable	decreases with increasing population density
Federal State	2E-16***	na	$3 > 8 > 6 > 4 = 9 = 11 > 10 > 2 > 7$
Migration	0.006***	na	$1 = 2$
Commuter	2E-16***	na	$1 > 2$ (higher commercial activity is favorable)
Income class	0.003***	unfavorable	decreases with increasing income class

Note: ***,** denote statistical significance at the 0.01 and 0.05 level. The average effects and performance implications are obtained by the visual inspection of the partial regression plots.

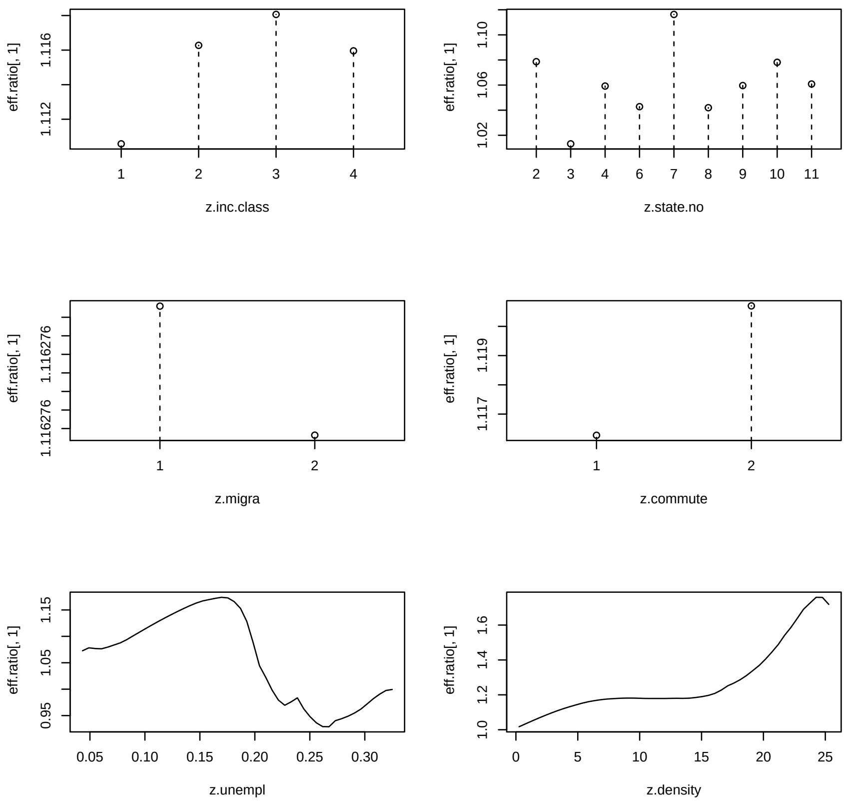
the conditional order-m estimator. Because the direction of influence cannot be derived by the statistical significance test, we inspect the partial regressions plots for interpretation.

Figure 3.1 show the partial regression plots for each of the six environmental variables. The plots illustrate the impact of the framework conditions on the ratio of the conditional and unconditional efficiency scores, i.e. $\hat{Q}^z$, by allowing the respective z-variable to change, while all other variables are kept at their median value. The upper left panel (z.inc.class) shows that higher income classes negatively influence the performance of municipalities compared to the lowest income class. Thereby, the third income class is the most unfavorable. This is in line with most previous research; see e.g., Giménez and Prior (2007) and Nieswand and Seifert (2011). Although, the present analysis includes the voluntary task of promoting sport clubs, there mights be additional voluntary tasks, e.g., swimming pools and theaters, that are financed by the free disposable revenues (coming e.g., from personal income tax) but are not accounted here.

The upper right panel in Figure 3.1 (z.state.no) visualizes how the location of municipalities affect their performance. Obviously, the Federal State that a municipality belongs to plays a role. Compared to Bavaria (level 3), all other Federal States seem to provide less favorable environments. Since the z-variable for Federal States (z.state.no) captures all related differences, the effect on the performance can have multiple reasons. For example, the price levels and geographic characteristics differ significantly and municipal regulations might allocate responsibilities more effectively. This results supports the common perception that particularly nonparametric efficiency analysis need increased attention if municipalities are assessed that act under different regulations.

The effect of migration is shown by the middle left panel (z.migra). Although, the variable is statistically significant, virtually no difference between municipalities where more people migrate than emigrate (level 1) and vice versa (level 2) can be found. Note, the scale of the y-axis does not change apparently. As indicated

Figure 3.1: Partial regression plots for external variables



before, our hypothesis is formulated referring to the mobility of consumer-voters and the variable neglects dynamics. The middle right panel (z.commute) gives the results for the commuter variable. Similar to the results of Giménez and Prior (2007), it demonstrates that a higher commercial activity (level 1) provides a more favorable environment for public service provision.

Turning to the continuous variables, the lower left panel (z.unempl) shows that the effect of the unemployment rate on the municipal performance is mixed depending on the point of evaluation. About up to 17 percent, the unemployment rate negatively affects the performance and has positive effects favorable thereafter. For higher unemployment rates an unfavorable effect can be observed. The literature proves mixed evidence: Kalb (2010a) identifies a negative impact for the municipalities in Baden-Wuerttemberg, whereas it is insignificant for the Spanish municipalities Balaguer-Coll et al. (2002). The mainly unfavorable impact of population density is visualized by the lower right panel (z.density). Only the most densely populated municipalities of our sample appear to benefit from agglomeration effects. As aforementioned, also for population density, the literature provides mixed empirical evidence.

3.6 Conclusion

This chapter investigates the performance of German municipalities using a fully nonparametric framework of frontier analysis. The contribution of the chapter is twofold: first, it applies a recently proposed methodology to the efficiency measurement of local governments, and thereby secondly, provides new empirical evidence on the public service provision in Germany. Methodologically, we follow De Witte and Kortelainen (2009), who extend the conditional order-m estimator initially developed by Cazals et al. (2002). Therefore, without requiring any functional structure on the production process and predefined weights, we are able to estimate robust efficiency measures for the municipalities that typically supply multiple services. By conditioning the efficiency estimation on environmental variables, the approach further allows for the heterogeneous operating conditions to be directly accounted for, either represented by continuous or (unordered and ordered) discrete variables. Note that, due to the algorithm proposed by De Witte and Kortelainen, neither the conditional efficiency estimates nor the bandwidths used for kernel function smoothing are subject to the separability condition between exogenous variables and inputs and outputs.

For the year 2005, we collect a unique data set of 1,060 German municipalities that are located in nine of 16 Federal States. To describe the municipal service provision, we use the total current expenditures as input and six output measures that represent mandatory and voluntary responsibilities, i.e. total surface area, public thoroughfare, school buildings, kindergarten places, sport clubs and total population. Further, six environmental variables are included to control for different operating environments, i.e. the unemployment rate, the population density, the Federal State, the migration status, the commuting status, and finally, the inhabitants' income class. For our sample, we find a mean inefficiency of 13 percent which decreases to 8 percent when the heterogeneous operating conditions are taken into account.

As suggested by De Witte and Kortelainen (2009), we apply nonparametric regression and partial regression plots to identify the effect and the statistical significance of the exogenous variables. According to our estimation results, each of the included variable significantly influences the municipal performance. Similar to previous findings, the performance of the local governments is negatively affected by high income classes. Further, notable variations in the success of public service provision can be identified depending on the Federal State in which the municipalities are located. Although statistically significant, the migration status reflecting whether more people migrate than emigrate and vice versa, barely influences the efficiency. Differently, the performance benefits from higher commercial activity, which is captured by whether more socially insured employees work or alternatively live in the respective area. Further, we find that the municipal efficiency decreases with increasing population density whereas the empirical evidence for unemployment is mixed.

Although our data set is rich in variables and observations, the results should carefully be interpreted. Particularly, the individual efficiency measures need to be viewed in light of the included inputs, outputs, and exogenous variables. The interpretation might also be limited due to dynamic effects that are omitted in cross-sectional analysis. However, the analysis benefits from its nonparametric nature and provides useful insights of the current situation where the municipalities are subject to tense financial positions and ongoing structural reforms.

Chapter 4

Cost Efficiency and Subsidization in German Local Public Bus Transit

4.1 Introduction

Currently, Germany's local public transit sector in total receives about 13 bn Euros of public financing per year.[1] The funding is distributed in ample and sometimes parallel ways which might offer even contrary incentives for the recipients (Umweltbundesamt, 2003b). An evaluation of the effect of subsidies on the performance of operators can help identify improvements of the subsidies' guiding function and make appropriate recommendations on potential reductions in funding. In general, subsidies are considered crucial for a suitable provision of public transit that is an important component of population mobility. Hereby, the bus sector is of particular interest since it supplies more than two-thirds of the demand for general local public transportation[2] and it is notably employed in rural areas (Verband Deutscher Verkehrsunternehmen, 2009a). The reality that bus transit demand in non-urban areas is twice that in cities demonstrates the necessity of publicly providing mobility services, particularly in less densely populated areas. Several hundred bus companies, which are predominantly publicly owned, serve the market and constitute local and temporary monopolies.

Even though the average degree of cost coverage of the German public tran-

[1] The largest proportion of this amount is dedicated to infrastructure-intensive and rural public transport provided with light trains.

[2] General local public transit refers to modes including buses, trams, light railway and metros. It is particularly distinguished from rail-bound local public traffic provided with light trains. All modes together build the local public transit in total.

sit sector grew over the past decade (Verband Deutscher Verkehrsunternehmen, 2009b), it is commonly assumed that most local public transit is unprofitable, i.e. the overall costs exceed the revenues from fares, and, therefore, depends heavily on public financial support. Historically, the most significant reason for the sector's fiscal deficit originates in the growth of private vehicle usage during the second half of the twentieth century (Goeverden et al., 2006). In response, Germany established a complex financing system whereby all governmental units, i.e. the federal government, federal states, and lower-level government bodies, act as financiers. The individual financing instruments can be roughly divided into investment-related and non-investment-related groups (Umweltbundesamt, 2003a). The first group concerns investments in infrastructure and vehicles, while the second group can be subdivided further into: (i) grants for operating costs; (ii) compensation for target group-related traffic; (iii) tax reductions and other benefits; and (iv) deficit balancing (Umweltbundesamt, 2003b). The annual level of subsidies is partially determined by the profitability of the operators and some types might affect the performance of companies negatively.

This chapter focuses on deficit balancing, which we theorize can be influenced by firm management. We investigate the impact of this subsidy type on the performance of local bus companies by conducting a parametric form of efficiency analysis. To account for the character of deficit balancing in our econometric analysis, we link it directly to the cost inefficiency distribution. To our knowledge, this approach has not been pursued in the literature on public bus operations, and we believe it promises to broaden our understanding about subsidizing the sector.

The deficits that occur when profits from ordinary activities become negative can be addressed by various accounting treatments, e.g., loss forwarding or loss absorption.[3] Loss absorption can be characterized as an internal payment provided by the owners (shareholders) of a bus company to equalize the annual deficit that is not balanced by amounts from retained earnings. These additional payments differ from other types of subsidies because they depend on the extent and treatment of the losses reported by a company. In other words, the payments are influenced by the firms. Vickrey (1980) outlines three rationales to justify subsidies to local transit. First, because transit operates under conditions of substantial economies of scale,[4] marginal costs are lower than average costs, and under-pricing average costs, e.g., for social reasons, produces a gap in cost coverage. Second, competing modes of transit receive substantial subsidies. Third, special requirements

[3]The German transport accounting standards denote these payments as "*Verlustübernahme durch Eigentümer*".

[4]This has also been shown in a variety of empirical studies, among them Cambini et al. (2007) and Farsi et al. (2006b).

for the underprivileged or disabled, e.g., an inability to use alternate forms of transportation, justify public financial support. Essentially, Vickrey (1980) states the conclusions later reached by Karlaftis and McCarthy (1998), who note that with the exception of the economies of scale rationale, public transit subsidies are based upon non-economic arguments, i.e. social objectives. In addition, public subsidies are a second-best instrument to address the urban externalities such as noise, congestion and pollution, in order to shift demand from private to public transportation (Button, 1993). However, a large body of literature provides empirical evidence for cost-increasing effects of subsidies in public transit; for a review, see e.g., Karlaftis and McCarthy (1998). Thus, financial support might extend the failure to cover costs instead of compensating for exogenously caused cost increases.

The literature analyzing cost structure and performance of public bus transportation dates to the 1950s and divides into the two strands: regression analysis and frontier analysis; Piacenza (2001) surveys theoretical and empirical issues associated with both approaches. Early work, including Johnston (1956), Miller (1970), Viton (1981) and Berechman and Guiliano (1985), is chiefly concerned with establishing concepts of cost models and cost functions' properties within the context of public bus transportation,[5] but also with appropriate regression estimation techniques. The regression analyses concerning subsidies highlight further aspects of public transportation, e.g., fares, unit costs, and demand. Bly et al. (1980) and Bly and Oldfield (1986) find reduced effects on fares and increased effects on demand as well as increased unit costs and reduced labor productivity because of subsidies. The data they use comprises multiple countries, and therefore the findings appear to reflect general trends. Pickrell (1985) examines the relationship between deficits and subsidies in the US transit sector and concludes that government subsidy programs would be more effective if transit operators could gain a measure of control over operating costs, adapt their services to changes in demand, and reconstruct fares to recognize the variations in supply costs. In addition, Pickrell proposes that a revision of the subsidy mechanism could also contribute to improving the situation. Thereby, a major effort in revising state and federal programs is reestablishing incentives for operators. The success of subsidies appears to be closely related to the level of government awarding the financial support. Andersen (1983), Pucher (1988) and Filippini et al. (1992) find that subsidies by low-level government bodies cause fewer cost increases than subsidies funded by high-level government bodies. In other words, the impacts of subsidies on costs are less harmful when close relationships exist between funding

[5]See Berechman (1983) for a general survey of public transport.

bodies and companies.

During the early 1980s, performance measurement using frontier analysis entered the discussion; for a review, see e.g., De Borger et al. (2002) and De Borger and Kerstens (2008). Based on the idea of Farrell (1957), frontier methods determine the best practice behavior in an industry (or a sample) and estimate the unit-specific degree of inefficiency relative to the best-practice benchmark. Frontier approaches mainly estimate the efficient frontier either by nonparametric linear programming, or by parametric techniques which assume a functional form representing the underlying input-output-transformation. The advantages of parametric efficiency analysis are its accountability for statistical noise, applications to panel data, and incorporation of the time horizon. This chapter applies SFA, a widely used parametric technique, which yields estimation residuals that are interpreted as measures of inefficiency.

Even though there is continued interest in performance measurement focusing on public bus operators, the empirical evidence on subsidies derived from frontier analysis is limited. In a nonparametric analysis, Obeng (1994) investigates the technical efficiency of 73 US single mode bus systems in 1988 by comparing the efficiency scores from a base model to its re-estimation including subsidies (measured as total operating and capital subsidies from all sources) as an additional variable. The author finds higher technical efficiencies when subsidies are considered. However, it is unclear whether Obeng's results are truly subsidy-related, or are driven by the curse of dimensionality.[6]

Kerstens (1996) uses nonparametric technology references to evaluate the technical efficiency of 114 French bus operators in 1990. Conducting a Tobit regression in a second stage, the author shows that subsidies (measured as the share of subsidies in total operating costs) subvert technical efficiency.

Filippini et al. (1992) estimate the cost efficiency of a panel of 62 Swiss bus operators in 1988 by displaced OLS. The subsequent OLS regression reveals that cost efficiency is positively influenced by the low-level government share in deficit subsidies and the amount of compensatory payments.

Sakano and Obeng (1995) examine the technical and allocative efficiency of 134 US single mode bus firms in 1988 using a stochastic frontier approach developed by Schmidt and Lovell (1979, 1980). Using OLS regression, they find that firm size rather than operating and capital subsidies affects the allocative efficiency between labor and capital.

Sakano et al. (1997) extend Sakano and Obeng (1995) by incorporating the

[6]For theoretical considerations of the curse of dimensionality, see Simar and Wilson (2008) and Adler and Yazhemsky (2010); for an empirical investigation, see e.g., Nieswand et al. (2010).

operating and capital subsidies in the cost minimization problem such that firms minimize costs net of subsidies subject to the production function constraints. This specification allows them to distinguish allocative inefficiency due to subsidies, or to internal factors. They pool data on US urban bus companies from 1983 to 1992 and find that allocative inefficiency mainly originates in factors internal to the firms, not the subsidies. Further, Sakano et al. (1997) indicate that subsidies cause notable deviations from optimal input factor proportions, i.e. the excess use of labor relative to capital and the excess use of fuel relative to capital and labor.

Unlike previous research, we directly incorporate the firm-influenced subsidy as a heteroscedastic variable in the std.dev. of inefficiency term, i.e. the half-normal error term. This approach is proposed by Bhattacharyya et al. (1995) and Hadri et al. (2003) who suggest among others, to assign factors which are under the control of firms (managerial determinants) to the inefficiency term.[7] Our approach accounts for endogeneity of inefficiency and deficit-balancing subsidies, enabling us to capture a potential bias due to heteroscedasticity. Caudill et al. (1995) argue that especially the residuals in frontier estimation are sensitive to heteroscedasticity, because the frontier changes when the error dispersion increases.[8] This sensitivity is likely to carry over to the inefficiency measures and therefore must be considered.[9]

This chapter contributes to the existing literature by investigating a sample of German local public bus operators and applying recent panel data model specifications of SFA that account for unobserved heterogeneity and provide time-varying and firm-specific efficiency estimates. Moreover, to allow for variations in the optimal (reference) production technology, one of our two model specifications relaxes the strong assumption of equal output and price parameters by randomizing them. The remainder of the chapter is structured as follows. Section 2 discusses the applied methodology and introduces the model specifications and data. Section 3 shows the results of our regressions and discusses in depth our analysis of firm-specific cost efficiencies. Conclusions and suggestions for policy-makers are given in Section 4.[10]

[7]The association of factors under the control of firms (managerial determinants) and inefficiency is particularly distinguished from exogenous factors that are instead associated with the noise term.

[8]In regression estimation this is a minor problem, because average cost functions are usually estimated by least squares and estimators based on means are no longer efficient but still unbiased when symmetric error dispersion is present (Caudill et al., 1995).

[9]Using a Monte Carlo study for the estimation of a cross-sectional cost frontier of banking institutions, Caudill et al. (1995) find overestimation of inefficiency for small firms and underestimation of inefficiency for large firms when heteroscedasticity is ignored.

[10]This chapter is based on Nieswand and Walter (2010) and joint research with Matthias Walter. We thank Pio Baake for fruitful discussions and Arne Beck, Anne Neumann, Astrid

4.2 Methodology and Data

4.2.1 Cost Function

Public transit can be considered a production process whereby inputs, e.g., labor and capital, are transformed into one or multiple outputs, e.g., seat-km (skm). The production process is well-known and the corresponding cost function of public bus operators has been discussed at length. Kumbhakar (1997) notes that independent from the output produced, it is important to use inputs in order to minimize the cost of producing a given level of output. Cost-minimizing behavior is required when a cost function is applied (Coelli et al., 2005). Further, output quantities are predetermined by public (government) entities that make decisions about the public transport services to be supplied. Therefore, we apply an input-oriented approach and the total cost function can be written as

$$C = f\left(Y, p_L, p_K, di, D_{east}, t\right) \tag{4.1}$$

where total costs (C) depend on the level of output (Y), two input factor prices for labor (p_L) and capital (p_K), and structural variables that are beyond the control of companies. These structural variables are the density index (di) and a dummy variable (D_{east}), which obtains the value of one if a company operates in one of the East German states. A linear time trend (t) captures a neutral technological change.

We opt for a flexible functional form, i.e. the translog cost function.[11] We choose the mean to be the local point around which the function is approximated. Hence, the variables for output, factor prices, and density index are divided by their respective mean. This transformation allows interpreting the estimated coefficients as elasticities. After imposing linear homogeneity of costs in input prices of degree one by dividing cost-related measures by the input factor price for labor, the translog cost function is

Cullmann, Christian von Hirschhausen and the participants of the North American Productivity Workshop (NAPW) VI and the Graduate Center seminar at DIW for helpful comments. Further, we thank Marika Geissler, Stefan Seifert, and Samuel Schneider for help with the data collection and Ann Stewart for excellent proof reading.

[11] For previous applications, see e.g., Bhattacharyya et al. (1995), Farsi et al. (2006b) and Filippini and Prioni (2003).

$$
\begin{aligned}
\ln\left(\frac{C_{it}}{p_{L_{it}}}\right) = \ln C^*_{it} &= \beta_0 + \beta_Y \ln Y_{it} + \beta_{p_K} \ln\left(\frac{p_{K_{it}}}{p_{L_{it}}}\right) \\
&+ \frac{1}{2}\left[\beta_{YY}\left(\ln Y_{it}\right)^2 + \beta_{p_K p_K}\left(\ln\left(\frac{p_{K_{it}}}{p_{L_{it}}}\right)\right)^2\right] \\
&+ \beta_{Y_{p_K}} \ln Y_{it} \ln\left(\frac{p_{K_{it}}}{p_{L_{it}}}\right) + \beta_{di} \ln di_{it} + \beta_{D_{east}} D_{east} + t
\end{aligned} \tag{4.2}
$$

where β_0 represents the intercept, and all other $\beta's$ represent the variables' coefficients to be estimated. The indices i and t indicate the unbalanced panel structure of our data where $i = 1, 2, \ldots, 33$ denote the companies, and $t = 1, 2, \ldots 12$ the time period of the specific observation.

4.2.2 Econometric Model

To estimate the translog cost function we employ stochastic frontier models for panel data. The advantage of using panel data models is that they allow accounting for both unobserved heterogeneity between firms and dynamics. The first panel data models for SFA were proposed by Pitt and Lee (1981) and Schmidt and Sickles (1984). Both models allow for firm-specific inefficiency estimation but regard only time-invariant inefficiency. Thus, they are no longer considered here. Numerous approaches include time-varying inefficiency, such as Kumbhakar (1990) and Battese and Coelli (1992, 1995). This chapter uses the TRE model proposed by Greene (2005b), who extends conventional models by including an additional random intercept which captures unobserved heterogeneity. This model can be illustrated as

$$\ln\left(C^*_{it}\right) = \alpha_0 + \alpha_i + x'_{it}\beta + v_{it} + u_{it} \tag{4.3}$$

with C^* depicting the transformed cost variable and $x'\beta$ collecting the explanatory variables and the respective parameters. α_0 is a common intercept and $\alpha_i \sim iid\ N\left(0, \sigma^2_\alpha\right)$ a firm-specific random intercept which captures unobserved time-invariant heterogeneity. Noise is captured by a two-sided error term $v_{it} \sim iid\ N\left(0, \sigma^2_v\right)$, while $u_{it} \sim iid\ N\left(0, \sigma^2_u\right)$ denotes a one-sided, non-negative random variable which represents the firm-specific inefficiency. Since we wish to include a managerial determinant as heteroscedastic variable z in the inefficiency function, we parameterize the std.dev. of the one-sided inefficiency term such that $\sigma_{u_{it}} = \exp\left(\delta z_{it}\right)$, following Bhattacharyya et al. (1995). z collects an intercept (z) and the heteroscedastic variable which represents our measure of deficit-balancing subsidies (z_1). This variable is devided by its std.dev. to improve the estimation. δ denotes the vector of coefficients to be estimated in the heteroscedastically specified inefficiency function. Introducing heteroscedasticity in the half-normal model implies an individual-varying mean of the inefficiency since $E[u_i] = \sigma_{u,i}\phi(0)/\Phi(0) = 0.79788\sigma_{u,i}$ where ϕ denotes the probability density function of the inefficiency function of the normal distribution and Φ is its cumulative distribution function (Greene, 2007).

Several extensions of heteroscedastic models have been proposed: Hadri (1999) introduces double heteroscedasticity (heteroscedasticity in both the one-sided and the two-sided error terms) for cost frontiers; Hadri et al. (2003) extend this approach to the cases of production frontiers and panel data. We concentrate on the heteroscedasticity of the one-sided error term. According to Kumbhakar (1997), from an economic view, it makes more sense to model heterogeneity in the variances of firm-specific components, especially when there are unobserved firm-specific components.

The TRE model assumes that the explanatory variables are uncorrelated with the firm-specific effect. Farsi et al. (2005b) point out that at least time-variant efficiency measures are not very sensitive to such correlations because the correlations may be captured by the coefficients of the cost function and do not affect residuals. The TRE model is a special case of the RP model which additionally allows other coefficients to be randomized. We let the coefficients of output (β_Y) and the price ratio (β_{p_K}) vary between companies. Hence, the frontier estimated by this RP model does not assume the same optimal technology for every firm. Justifications for assuming a different technology may origin first, in different bus types, e.g., diesel versus hybrid, or low floor versus conventional, and second, in different optimal input factor ratios according to a company's environment. The heteroscedastic formulation of the inefficiency term and all other assumptions are the same as before.

4.2.3 Data

The data consists of an unbalanced panel of 33 German bus operators in urban and rural areas. The time period covers twelve years (1997-2008) for a total of 231 observations. The panel structure is such that 50 percent of the companies are observed seven years or less, and 25 percent are observed ten years or more. Table 4.1 presents the data characteristics. The data were derived from multiple sources, i.e. the physical data, e.g., skm, is from the annual statistics of the German Association of Transportation Companies (*Verband Deutscher Verkehrsunternehmen*), and the

Table 4.1: Descriptive statistics for German local public bus companies

Variable	Min	Mean	Median	Max	Std.dev.
Total costs (Y) [mn Euro[(a)]]	3.82	39.47	33.47	95.04	24.22
Skm (*skm*) [mn km]	55	750	719	1,870	423
Labor price (p_{L}) [Euro/FTE[(a)]]	10,693	46,896	46,689	86,243	11,566
Capital price (p_{K}) [Euro/seat[(a)]]	568	1,360	1,237	3,517	590
Density index (*di*) [inhabitants/km[(b)]]	61	412	344	2,460	333
Dummy east (D_{east}) [level[(c)]]	0.00	0.26	0.00	1.00	0.44
Subsidy ratio (z_1) [percent[(a,d)]]	0.00	0.14	0.14	0.55	0.14

Source: German Association of Transportation Companies, Federal Gazette. Notes: observations=231, companies=33, years=1997-2008. [(a)] Base year 2003. [(b)] Population in operating area per km of network length. [(c)] Dummy east; 1: company operates in Eastern Federal States (59 observations), 0: company operates in Western Federal States (173 observations). [(d)] Loss absorption in Euro divided by total costs in Euro.

monetary data, e.g., personnel expenditures and loss absorption, is from the balance sheets published in annual reports and the Federal Gazette (*Bundesanzeiger*). Total costs (C) include personnel expenditures, material costs, other operating expenses, depreciation, interest on borrowed capital,[12] and opportunity costs of equity. The latter is measured by multiplying the individual equity base of each observation with the corresponding interest rates of corporate bonds (Deutsche Bundesbank, 2010) plus a 2 percent risk premium. We note that this approach treats the companies equally and is justified by the fact that our dataset includes operators that are predominantly publicly owned. Only five companies[13] have a mixed ownership structure (public and private), and none are purely privately owned.

Dividing personnel expenditures by the number of full-time equivalents (FTE) provides the input factor price for labor (p_L). To approximate capital costs we use the residual from subtracting personnel expenditures from total costs, and thus consider all non-labor costs as capital costs. This approach is frequently used when companies do not report capital costs directly or it is not possible to apply the capital inventory method; see e.g., Farsi et al. (2006b) and Filippini and Prioni (2003). We then calculate the input factor price for capital (p_K) as the ratio of capital costs to the number of seats.[14] Seats are our preferred unit measurement, because unlike the number of buses, the number of seats accounts for different bus sizes. Both input factor prices vary notably. Walter (2011) argues that labor and capital cost shares are significantly related to outsourcing, because outsourcing moves internal labor costs into purchased services which are part of material

[12]We use the account "interest paid and similar costs" reported in the financial reports.

[13]These companies are ASEAG in Aachen, KVG in Kiel, KVG in Coblenz, KVS in Saarlouis, and KViP in Uetersen.

[14]The number of seats is calculated by the number of skm multiplied by the number of buses divided by the number of vehicle-km. This approach assumes a similar deployment of all buses in the fleet, which should be the usual case.

costs. The large variation in labor prices furthermore depicts the interregional wage differentials, particularly for the distinction between wage levels in Eastern and Western parts of the country.[15] The differences in capital prices seem to be due to rural and non-rural characteristics of the operating environments.[16] The capital price is lower for rural operating areas where companies tend to employ older buses with less comfort devices. Cost reductions due to lower depreciation costs appear to outweigh the higher maintenance costs associated with old buses. All cost data is inflation-adjusted to 2008 using the German producer price index (Statistisches Bundesamt, 2009).

Skm is the supply-oriented measurement of output. De Borger and Kerstens (2008) note that objectives and heterogeneity of public bus transit imply that both supply- and demand-oriented approaches are relevant. We use the former approach since local public transport is a public service obligation with pre-determined service levels which, and at least in the short-run, are not open to companies' influence.

For comparability between operators, we use a density index (di) capturing the network characteristics beyond firm's control. We define di as the ratio of population living in the operating area over the km of network length gathering, e.g., differences in the service accessibility for customers, in speed, and in network complexity. A dummy variable (D_{east}) addresses the cost differences between companies operating either in newly formed (D_{east} equals 1) or in old West German States. A substantial restructuring of public transport in the newly formed German States, supported by state aid, followed Germany's reunification and hence affects cost structures.

To determine the firm-influenced subsidies, we use the amount of loss absorption directly paid by the firm owners (shareholders) to balance negative revenues from ordinary activities. Deficits can also be recovered by depleting accumulated retained earnings, carrying forward a loss, appropriating reserves, etc., all of which depict a firm's ability to handle losses without demanding additional money.[17] Such accounting treatments support the assumption of firm influence on loss absorption. We assume that the amount of required loss absorption can be assessed by firm's management, since exogenous cost disadvantages are addressed by other

[15]The average labor factor price is 50,616 Euro/FTE in the old West German States, and 36,050 Euro/FTE in the newly formed German States.

[16]The average capital price is 1,133 Euro/seat in rural and 1,563 Euro/seat in non-rural operating areas, respectively.

[17]We make only one exemption from this treatment and consider the appropriation of reserves as loss absorption if shareholders obviously add the amount of the pending loss to the reserves. In this case only the accounting practice differs, but these accounts mirror the behavior we are studying.

subsidies mentioned in Section 1. The subsidy ratio (z_1) is then constructed by the ratio of loss absorption over total costs.

4.3 Empirical Evidence and Interpretation

4.3.1 Regression Results

Table 4.2 provides the regression results for the TRE and the RP models.[18] The obtained results are robust and show significant coefficients with small standard errors. The first order coefficients, β_Y and β_{p_K}, have the expected, positive signs and are statistically significant at the 1 percent level. Given that all variables of the cost function are in logarithmic form, we can interpret the estimates as cost elasticities. The TRE model shows an output cost elasticity of 0.457 for the mean company, indicating an under-proportional increase of costs when output enlarges. With the same implication of existing economies of scale, β_Y is substantially higher in the RP model (0.622) and exhibits a significant std.dev. of 0.263. Based on the significant std.dev. of the output coefficients, our results indicate that marginal costs variations are present within the same transportation system, i.e. motor-buses. These differences might be due to differences between urban and rural operating systems. The capital price coefficients are similar across models (0.413 and 0.415).

However, the randomized capital price coefficient in the RP model has a large std.dev. of 0.320. Since the price coefficient can be interpreted as the optimal cost share of the individual input factor, σ_{p_K} indicates an optimal input mix that varies across companies. This is likely to be related to firm size and emphasizes the necessity to account for different production structures and operating frameworks.[19] Moreover, varying prices and marginal costs can be explained by the diversity of input virtues, i.e. regarding capital diversity; De Borger and Kerstens (2008) mention that bus fleets are heterogeneous in terms of vintages therefore lead to diverse depreciation patterns. They also mention that different fuel power technologies are applied. Even though hybrid power technologies might not have an important impact over the sample period, low floor technologies, quality improving devices, e.g., air conditioning, and different types of buses, e.g., standard and articulated buses, are relevant. While hybrid technologies might not have important impacts during our sample period, we note that low-floor technologies, quality improvements, e.g., air conditioning, etc. are relevant.

[18] We conduct the estimation with LIMDEP 9.0 using 1,000 Halton draws for each model.

[19] The mean share of capital costs in total costs is 55 percent with a std.dev. of 13 percent.

Table 4.2: Regression results

		TRE model		RP model	
		Coefficient	Std.dev.	Coefficient	Std.dev.
Parameters of the cost function					
Constant	α_i	-7.042***	0.014	-6.970***	0.011
Std.dev. of α_i	σ_α	0.246***	0.009	0.190***	0.006
Output (skm)	β_Y	0.457***	0.019	0.622***	0.013
Std.dev. of output	σ_Y			0.263***	0.009
Capital price	β_{p_K}	0.413***	0.012	0.415***	0.011
Std.dev. of capital price	σ_{p_K}			0.320***	0.013
Output^2	β_{YY}	-0.363***	0.020	-0.215***	0.015
Capital price^2	$\beta_{p_K p_K}$	0.074***	0.019	-0.066***	0.023
Output capital price	$\beta_{Y p_K}$	-0.285***	0.021	-0.185***	0.018
Density index	β_{di}	0.042***	0.007	0.024***	0.006
Dummy east	$\beta_{D_{east}}$	-0.235***	0.015	-0.238***	0.014
Linear time trend	β_t	-0.006***	0.001	-0.011***	0.001
Parameters of the inefficiency function					
Constant of σ_u	δ_0	-4.348*	2.223	-4.567***	1.675
Subsidy ratio	δ_{z_1}	1.681**	0.799	1.906***	0.625
Std.dev. of v	σ_v	0.066***	0.002	0.043***	0.001
Lambda	σ_u/σ_v	0.326		0.483	
Wald test H_0: δ_0=δ_{z_1}=0			5.533$^{(a)}$		13.879$^{(b)}$
Log-likelihood function			218		263

Notes: (a) p-value=0.063. (b) p-value=0.001. ***,**,* indicate statistical significance at the 0.01, 0.05 and 0.1 level, respectively.

The second order coefficients, i.e. β_{YY} and $\beta_{p_K p_K}$, and the interaction coefficients, $\beta_{Y p_K}$, are statistically significant but do not always show the expected negative sign. The positive coefficient $\beta_{p_K p_K}$ in the TRE model violates the concavity property of cost functions in input prices and suggests a non-cost-minimizing behavior of firms in response to changes in prices. The same result found in other regulated industries; see e.g., Karlaftis and McCarthy (2002), Farsi et al. (2005b) and Farsi and Filippini (2009), is explained by the considerable barriers of cost-minimizing strategies.[20] Plausibly, these constraints could also apply to the German public local bus transport as a highly state-influenced sector. However, the more flexible technology shows a negative sign of the second-order coefficient $\beta_{p_K p_K}$ and thus, the RP model satisfies the theoretical requirements of cost functions. Applying the Wald test we cannot confirm the hypothesis of Cobb-Douglas-typed technologies at the 1 percent level.[21]

For the coefficients of the structural variables, β_{di} and D_{east}, both models show consistent implications of the estimates. Commonly, urban transportation systems are characterized by lower average speeds and higher network complexity, which explains the positive sign of the density index coefficient β_{di}. In addition, the

[20] For example, input prices that are constrained by regulation and a less-distinctive sensitivity to price changes in public sectors can be considered barriers.

[21] The test statistics take value of 481 (TRE model) and 217 (RP model) which clearly exceed the critical value of 12.84 at the 0.5 percent significance level.

coefficient contains some costs associated with network length, e.g., costs for bus stops. Thus, operating areas with a higher population density yield higher costs. The dummy variable's coefficient D_{east} implies lower costs for companies operating in eastern Germany. Apparently, restructuring after German reunification shows a significant cost-reducing impact. The expected negative coefficient value of the linear time is small (0.006 and 0.011) which implies only minor technological advances associated with cost reductions. De Borger and Kerstens (2008) explain the small time trend with the established technology of bus driving, increasing congestion levels impeding performance improvements, and improvements in technical efficiency rather than technological progress.

The focus of this chapter is on the heteroscedastic variable, i.e. on the effect of the subsidy ratio on the cost inefficiency's variance. The two models reveal positive and statistically significant coefficients for the subsidy ratio. δ_{z_1} is 1.681 in the TRE model and 1.906 in the RP model which implies an increasing std.dev. in cost inefficiencies for larger z_1. Conducting a Wald test on the heteroscedasticity of the inefficiency's std.dev. fails to confirm the hypothesis of zero values for δ_0 and δ_{z_1} at the 10 percent significance level in the TRE model and at the 1 percent level in the RP model. Since efficiency is half-normally distributed, the distribution's probability function flattens with increasing z_1 and the probability mass shifts towards the tail. Therefore, we conclude that the performance range increases when the proportion of subsidies to total costs increases.

4.3.2 Cost Efficiencies

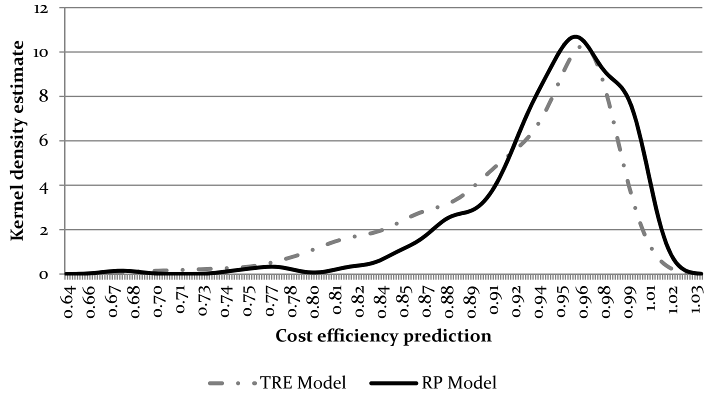
Table 4.3 shows the characteristics of the predicted cost efficiencies. For the 231 observations considered, the econometric models show an overall mean cost efficiency of 92 percent and 93 percent in the TRE model and RP model, respectively. The minimum value of cost efficiency is 69 percent in the TRE model and 66 percent in the RP model while the highest is close to 99 percent in both.

Table 4.3: Cost efficiency estimates

Model	Min	Mean	Median	Max	Std.dev.
TRE model	0.6908	0.9214	0.9398	0.9886	0.0582
RP model	0.6599	0.9280	0.9368	0.9873	0.0475

Figure 4.1 depicts the distribution of the cost efficiency predictions. Both curves support our assumed half-normal distribution of efficiency and so we conclude that the underlying models are appropriate for the given data. However, the probability mass in the RP model is closer to the efficient tail of the distribution

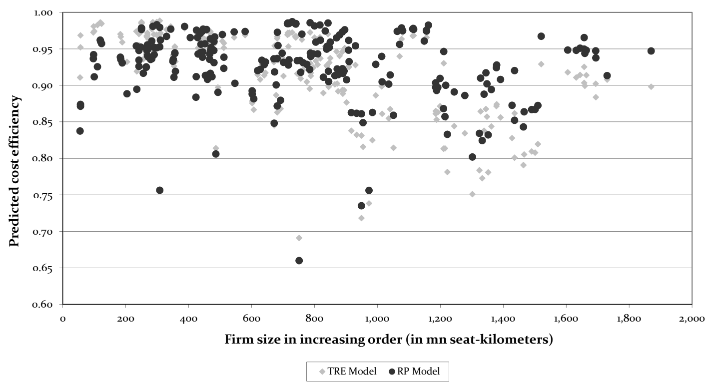
Figure 4.1: Kernel density of cost efficiency predictions



representing the model's characteristic of allowing more heterogeneity between companies without attributing the cost differences to inefficiency.

A detailed look at firm-specific efficiency estimates with respect to firm size in Figure 4.2 reveals that larger firms particularly benefit from the RP model with a more flexible underlying technology. This has two implications: First, there is no clear indication for size-related differences in performance between smaller and larger local public bus operators; second, the TRE model appears to miss important information on the technology characteristics.

Figure 4.2: Cost efficiency and firm size



Since we are interested in whether less-subsidized operators perform better, we conduct a Welch test which compares the mean cost efficiency of two groups while allowing for any underlying distribution of the std.dev. We divide the companies in two groups according to the subsidy ratio z_1; group 1 comprises all observations recording zero deficit balancing (95 observations) and group 2 comprises all others (136 observations). The test first calculates the mean cost efficiency of all group members for each group and then tests whether the means differ significantly from each other. The obtained test results are illustrated by Table 4.4.

Table 4.4: Welch test on group mean cost efficiency

	Group size	TRE model Mean	TRE model Std.dev.	RP model Mean	RP model Std.dev.
Overall mean cost efficiency	231	0.9214	0.0582	0.9280	0.0475
Mean cost efficiency group 1[(a)]	95	0.9322	0.0050	0.9423	0.0369
Mean cost efficiency group 2[(b)]	136	0.9139	0.0054	0.9181	0.0044
t-value		2.484		4.171	
p-value		0.014		0.000	

Notes: [(a)] indicates that subsidy ratio z_1 equals zero and [(b)] indicates that the subsidy ratio z_1 is greater than zero.

Both models consistently show that group 1 performs better in terms of cost efficiency. Those companies with a subsidy ratio of zero achieve a mean cost efficiency of 93.22 percent and 94.23 percent in the TRE model and the RP model, respectively, while companies with a positive subsidy ratio achieve mean cost efficiencies of 91.39 percent (TRE model) and 91.81 percent (RP model). We cannot confirm the null hypothesis of non-differing mean cost efficiency values, since the respective average values are different at the 5 percent (TRE model) and at the 1 percent significance levels (RP model). From this we conclude that operators demanding no subsidies in the form of loss absorption, on average, perform better. This coincides with De Borger and Kerstens (2008), who conclude from their empirical evidence that subsidies have cost-increasing and performance-worsening effects. Our results extend the empirical evidence of efficiency decreasing effects to subsidies which are firm-influenced, target-unspecific, and unlimited.

4.4 Conclusion

Subsidies are commonly allocated to public bus transportation to compensate exogenously caused cost increases. However, the empirical evidence implies reversing effects of financial supports on costs, i.e. cost increases due to subsidies. Interest in curtailing Germany's generous public budgets and previous empirical findings spurred our examination of the effect of subsidies on operator performance. We considered loss absorption, i.e. a payment by a firm's owner to balance negative revenues from ordinary activities, as firm-influenced subsidies for two reasons. First, a wide range of subsidies exists to compensate for exogenously caused cost disadvantages, and second, losses can be balanced via different accounting treatments.

We hypothesized that bus operators with higher subsidies would perform worse and thus exhibit reduced cost efficiencies. Using a heteroscedastic stochastic frontier cost function, we analyzed an unbalanced panel of 33 German bus companies observed over a time period of twelve years (1997-2008) for a total of 231 observations. To estimate the translog cost function, we used two stochastic cost frontier models (TRE model and RP model), which differ in their ability to allow for varying optimal cost structures among companies. The PR model is preferable to the TRE model in three respects: first, it achieves a higher Log-likelihood function; second, it satisfies the concavity property of the cost function; and third, it shows significant std.dev. coefficients for output and prices. To ignore the latter leaves important information unexploited.

The finding of a positive effect of subsidies on the std.dev. of cost inefficiency showed that inefficiency is not equally distributed across subsidy levels. Relative to total costs, the larger the subsidies the wider the range of companies' efficiency. We also find that German bus companies are more cost efficient when they have lower ratios of subsidies to total costs. Our results suggest that the performance of publicly owned providers of local public bus transit can be improved by comparative analysis. Further, they suggest that reconsidering the allocation of subsidies, which operators can control, may help leaving inefficient cost structures.

Chapter 5

Overcoming Data Limitations in Nonparametric Benchmarking: Applying PCA-DEA to Natural Gas Transmission

5.1 Introduction

Natural gas transmission is a typical network industry. Theoretical, e.g., Sharkey (1982), and empirical evidence, e.g., Gordon et al. (2003), underline the subadditivity in the cost structure and therefore gas transmission companies remain regulated. The purpose of this chapter is to provide empirical evidence of a robust benchmarking technique for regulation when the number of regulated companies and/or data observations is small.

Since the late 1980s a substantial reform process was undertaken with the objectives of cost reductions and efficiency increases in regulated network industries. The transition from cost-plus regulation, where companies recover their costs with a fixed rate of return (Joskow, 2007; Farsi et al., 2007) to incentive-based regulation is the latest development towards more efficient production and cost reduction. In an incentive-based regulatory framework, price and revenue caps are set based on the RPI-X formula (Littlechild, 1983; Beesley and Littlechild, 1989) where the determination of the expected efficiency savings (X) is usually based on empirical results obtained from sophisticated efficiency analysis approaches (also called benchmarking analyses).

This framework, where the efficiency performance of the companies is evaluated against a reference performance (Farsi et al., 2005a), has mainly been favored by European regulators and played a crucial role in the regulatory processes in the UK and the Nordic countries. Using benchmarking methods in regulatory practice has been widely criticized (Shuttleworth, 2003, 2005). One of the major criticisms is the low number of observations in this sector, for a robust and consistent benchmarking. As shown in Table 5.1 the low number of observations is caused by strong concentration and absence of competition in natural gas transmission; for Germany see e.g., Hirschhausen et al. (2007). In fact, in most of the European countries, e.g., Finland and Belgium, a single transmission company is operating. In others, e.g., Spain, Sweden and Austria, several independent companies are operating. In Germany, for the first round of determining efficiency scores data on only eight companies were considered in the benchmarking procedure. Moreover, the regulator often collects data on a yearly basis, thus additionally restricting sample size. Hence, both the low number of companies and the yearly data basis severely limit sample size for an application of traditional benchmarking methods. Regulators of natural gas transmission system operators require guidance in adapting their models to the empirical challenges.

A possible solution to expand the number of observations is to use data from other countries found in international benchmarking exercises. Jamasb et al. (2007) analyze relative efficiencies of European natural gas transmission operators. The sample consists of four European countries (one company each, covering different time spans (3-5 years)) and 43 US companies over 9 years. However, two major problems with international comparisons are the strong heterogeneity of firms and the differences in data definitions across countries. In addition, Europe has the problems that data collection still remains a responsibility of national regulators and a harmonized consistent European data pool is not yet implemented. Hence, efforts are predominantly undertaken to establish national efficiency standards with a limited data sample that consolidate theoretical requirements and practical applicability. A wide range of benchmarking approaches and frameworks exist in the literature (Jamasb and Pollitt, 2001, 2003; Farsi et al., 2007) and the approaches can be separated into two main streams: nonparametric and parametric methods with DEA and SFA being the most commonly used respective approaches. The nonparametric methods determine the reference technology by means of linear programming methods whereas the parametric SFA assumes a functional relationship for the production process and determines the reference technology based on econometric methods.

From a practical regulatory point of view both approaches have been useful to regulators: directly as part of the regulation process or as an additional control instrument for decision-making (Farsi et al., 2007). Both methods differ in

Table 5.1: European regulated natural gas transmission system operators

Country	Number	Country	Number
Austria	7	Latvia	1
Belgium	1	Lithuania	1
Czech Republic	1	Luxembourg	1
Denmark	1	Netherlands	1
Estonia	1	Poland	1
Finland	1	Portugal	1
France	2	Romania	1
Germany	20	Slovakia	1
Greece	1	Slovania	1
Hungary	1	Spain	1
Ireland	1	Sweden	1
Iceland	1	UK	1

Source: COM(2009)115, Technical Annex.

their requirements for the underlying data volume in order to derive meaningful results.[1] Even if DEA, in terms of statistical properties, is more inefficient practical experience shows that DEA is used more frequently than SFA in the practical applications of efficiency analysis in the energy sectors; see e.g., Haney and Pollitt (2009); Jamasb et al. (2007).

A further empirical challenge is that in regulatory practice a detailed benchmarking model, describing the production process by means of exact input and output variables of the firms is indispensable. Hence, the model should include as much relevant information as possible. This requires a reasonable number of observations to distinguish companies and derive meaningful results. However, given a pre-determined sample size, an increase in dimensions, i.e. more explanatory variables—which might contribute to more appropriate modeling of reality—leads to more observations determining the efficiency frontier. This subsequently affects efficiency scores in nonparametric efficiency analysis. For example, utility regulation is often conducted on a yearly basis, making it impossible to increase sample size when all possible installations are already included in the sample. Hence, this practical obstacle often constrains the regulator's ability to meet the statistical requirements. However, reducing dimensions and conserving all available information at the same time improves the estimation of technical efficiency in a DEA framework.

A feasible solution is the application of PCA in DEA that reduces dimensions of the original set of variables whilst maintaining the information on variation of data (Härdle and Simar, 2003). The combination of DEA and PCA was proposed by Ueda and Hoshiai (1997), and Adler and Golany (2001, 2002), who aim to

[1]Simar and Wilson (2008) prove that the theoretical foundations of DEA are based on large datasets to produce meaningful results. By contrast, parametric approaches reveal a desirable feature in terms of consistency of the estimator, i.e. its convergence to the unknown parameter at a certain rate when sample size increases to infinity.

overcome the issue of over-estimation of relative efficiency due to large numbers of variables in DEA. They show that PCA can improve discriminatory power in DEA and give more reliable efficiency measurement in small samples. Fields of application refer mainly to network industries. Whereas Ueda and Hoshiai (1997) apply their approach to the telecommunication sector, Adler and Golany (2001) and Adler and Berechman (2001) refer to the airline industry, and Adler and Golany (2002) to university departments. Adler and Yazhemsky (2010) provide further theoretical developments and show the applicability of PCA to radial DEA models when only additive DEA models[2] were previously considered.

There are also other discrimination-improving approaches related to DEA. For example, Adler and Yazhemsky (2010) compare PCA with the approach of variable reduction based on partial covariance and find better performance of PCA. Podinovski and Thanassoulis (2007) controvert simple approaches, i.e. increasing the number of units and reducing the number of variables by means of aggregation or reduction, and more sophisticated approaches, where the latter can be grouped using additional information and additional measurements data have an advantage over additional information. They do not require information that is not directly given by the data and that is often difficult to determine. Frequently, regulators are unable to identify more realistic profiles of an optimal mix of inputs and outputs that could be implemented in DEA by weight restrictions (Podinovski and Thanassoulis, 2007). Weight restrictions based on trade-offs modify the efficient boundary of the production possibility set such that unrealistic input-output-compositions are no longer used as reference. However, the PCA-DEA formulation causes similar effects without the need of additional information (Adler and Yazhemsky, 2010). Although the weights imposed by PCA-DEA may not necessarily reflect those economic weights proposed by DEA, Ueda and Hoshiai (1997) prefer summarizing the variables parsimoniously. Alternatively, the presence of correlated variables selection requires (industry) expertise. In contrast, PCA based weights are objective based constraints (Adler and Golany, 2002).

This chapter provides the first PCA-DEA (in terms of radial efficiency measurement) in the context of natural gas transmission regulation. Since European natural gas companies are not easily comparable, we use the US natural gas market as our reference model. The US natural gas market often serves as a reference model given the long and good record of regulatory experience and publicly available company data over the last three decades. Rather than potentially including US data in a European benchmarking exercise we use data on US natural gas transmission companies to illustrate how data limitations affect radial efficiency

[2]For the difference between radial and additive models; see e.g., Cooper et al. (2006).

measurement and how PCA-DEA improves it. For a discussion comparing the US and European natural gas market see Jamasb et al. (2008). Our contribution to the literature and practical application is to support a pragmatic approach for European regulators who predominantly undertake efforts for national benchmarking and therefore face problems of limited data.

The remainder of the chapter is structured as follows. Section 2 introduces traditional DEA methodology and describes the issue of small samples in nonparametric benchmarking. DEA is extended by means of PCA following Adler and Yazhemsky (2010). The model specifications are outlined in Section 3, which also presents the data we use. Within this section outlier detection is reviewed. Our results are presented in Section 4 and Section 5 concludes.[3]

5.2 Methodology

DEA is a nonparametric method frequently used in regulatory practice to evaluate relative efficiency and to set company-individual efficiency targets subsequently. The reference technology is not determined by imposing a functional form that describes the production process or cost structure, but by piecewise linear DEA models that consider mainly two types of technology: CRS proposed by Charnes et al. (1978), and VRS suggested by Banker et al. (1984). The first translates into strict regulation practice assuming one optimal firm size whereas the latter allows for scale inefficiencies. We limit ourselves to assume VRS technology because it seems to be more reasonable in small samples (Adler and Yazhemsky, 2010). We also impose input-orientation, meaning that input is minimized while output remains fixed. This is a reasonable and common assumption in network industries because firms are generally required to supply service to a fixed geographical area, and hence, the output vector is essentially fixed (Coelli and Walding, 2006).[4] The standard radial DEA environment incorporating VRS technology and minimizing individual relative efficiency θ can be written as the following linear program:

[3]This chapter is based on Nieswand et al. (2010), joint work with Anne Neumann and Astrid Cullmann, and resulting from the research program on 'Efficiency Analysis,' run jointly by DIW Berlin and the Chair of Energy Economics and Public Sector Management at TU Dresden. Earlier versions have been presented at the XI EWEPA in Pisa, the 8$^{\text{th}}$ INFRADAY in Berlin, and the Workshop on Exploring Research Frontiers in Contemporary Statistics and Econometrics in *Louvain-la-Neuve*. We thank an anonymous referee, Nicole Adler, Victor Podinovski, David Saal, Christian von Hirschhausen and the workshop participants for valuable input and fruitful discussions.

[4]Input-orientation can be implemented in parametric and nonparametric approaches. For a parametric application see e.g., Farsi et al. (2005a)

$$
\begin{aligned}
\min_{\theta,\lambda} \quad & \theta \\
s.t. \quad & Y\lambda - s_Y = Y_j \\
& -X\lambda - s_X = \theta X_j \\
& e\lambda = 1 \\
& \theta, \lambda, s_Y, s_X \geq 0
\end{aligned}
\tag{5.1}
$$

where θ represents the relative efficiency (that is the absolute efficiency of the unit under consideration relative to a maximum value of obtained efficiency by any of the units considered) of each company contained in the set $J = \{1, 2, ..., n\}$. X_j and Y_j are column vectors of k inputs and l outputs of unit j. Collecting the column vectors yields in a $k \times n$ matrix for inputs X and a $l \times n$ matrix for outputs Y respectively. The input and output weights are given by the column vector λ. The constraint $\lambda = 1$ ensures that the VRS restriction is taken into account.[5] The slack variables s_X and s_Y allow the constraints to be stated as equalities. Furthermore, θ, λ, s_X and s_Y are supposed to be nonnegative.

To obtain meaningful results with DEA the number of relevant input and output variables should be in proportion to the number of observations. Regulatory practice demands a sophisticated model with a high number of inputs and outputs to describe the production process or cost structure realistically. How well the method is able to sufficiently discriminate between utilities becomes an issue particularly when the data are limited, which is a known issue in real regulatory practice. This is addressed by PCA, which can be used to reduce dimensions (number of variables) of the optimization problem by means of constructing linear combinations of the original data (Adler and Golany, 2001, 2002). This conversion alters the original coordinate system (Adler and Yazhemsky, 2010). Selecting the number of linear combinations then can reduce the dimensions of the new coordinate system.

The number of dimensions comprising this new coordinate system depends on satisfying a selection criterion, e.g., the Kaiser-Guttmann-criterion or the Jolliffe-criterion. We follow the study by Adler and Golany (2002) who select two as the number of principal components (PC) that satisfy discrimination purposes. However, we exclusively consider the limitation of dimensions in terms of outputs since there is no problem with a single input. Thus, for the purpose of translating output data, the correlation matrix C is obtained from the output matrix Y with

[5] Relaxing this constraint yields CRS technology, i.e. $\lambda \geq 0$.

$Y = [Y_1, Y_2, ..., Y_l]$. The (normalized) eigenvectors $v_1, v_2, \ldots, v_l$ given by C are used to create linear combinations of the form $PC_{Y_i} = v_i^t Y = v_{1i} * Y_1 + v_{2i} * Y_2 + \ldots + v_{li} * Y_l$ where superscript t denotes the transpose operator and i represents the i-th element of the eigenvectors. These linear combinations are also known as PCs each of which explains a certain ratio of the original variables variance, whereby this ratio corresponds to the eigenvalues $\eta_1 \geq \eta_2 \geq \ldots \eta_l$ of C. Commonly, eigenvalues are in descending order, and so are therefore PCs, i.e. PC_1 covers most of the variation in the data, PC_2 covers less of it, and PC_l covers the lowest proportion.

Here we consider the combination of PCA and radial DEA models according to Adler and Yazhemsky (2010). However, one drawback of PCA-DEA is its requirement of data transformation. In PCA-DEA data are transformed initially by PCA and have to be remodeled to the original form after optimization. It appears that only some radial DEA settings are tolerant towards data transformation. Pastor (1996) proves output translation invariance for input-oriented DEA models under the VRS assumption. Hence, in general, the optimal solution using original data does not change when data are transformed. Although translation invariance is not supported by all DEA models, their general properties are not affected by PCA-DEA (Adler and Yazhemsky, 2010).

For one unmodified input and all outputs to be transformed into PCs, the dual linear program under VRS assumption can be written as follows:

$$\begin{aligned}
&\min_{\theta,\lambda} \quad \theta \\
&s.t. \qquad Y_{PC}\lambda - L_Y s_{PC} = Y_{PC,j} \\
&\qquad\qquad -X\lambda - s_X = \theta X_j \\
&\qquad\qquad L_Y^{-1} Y_{PC} \geq s_{PC} \\
&\qquad\qquad e\lambda = 1 \\
&\qquad\qquad \theta, \lambda, s_Y, s_X \geq 0
\end{aligned} \tag{5.2}$$

where $Y = [y_1, y_2, \ldots, y_p]$ is the matrix of p outputs and x the single input vector we use. L_Y is the matrix collecting the output weights obtained by PCA. The original data are weighted and enter through PCs Y_{PC} where $Y_{PC} = l_i^t Y = l_{1iy1} + l_{2iy2} + l_{3iy3} + l_{4iy4}$ and l_i are the normalized eigenvectors from the correlation matrix of Y. Because all outputs are transformed into PCs, the minimization problem does not include separate output vectors. Both the slack variable s_{PC} and the original output data are weighted by the linear coefficients obtained by PCA.[6] As

[6]Due to data transformation a new constraint enters the linear problem, which ensures the

stated in Equation 5.1 VRS technology and nonnegativity of parameters and slack variables are assumed. If and only if all *PC*s are included, i.e. *PC*s explain 100 percent of the original data variation, the solutions of Equations 5.1 and 5.2 are equivalent (Adler and Yazhemsky, 2010).

5.3 Model Specification and Data

5.3.1 Model Specification

We want to determine the pipelines' relative ability (pipelines refer to companies operating such facilities) to provide services at least cost where we consider the demand as fixed in the short-term. Hence, the model set up is based on the idea of a cost driver analysis, meaning that costs are explained by output variables that are relevant to costs of the pipelines under consideration. This approach deviates from the purely technical representation of the production process by physical data but is often applied in regulatory practice; see e.g., Jamasb et al. (2007); Bundesnetzagentur (2006).

An important issue that arises almost immediately when applying benchmarking in regulatory practice, is cost comparability. There are essentially two ways of constructing the benchmarking basis, i.e. the short-run maintenance model and the long-run service model; for a broad discussion see Burns et al. (2005). The first model incorporates operating expenditures while the second model incorporates total expenditures (operating expenditures plus capital costs). Although the total cost approach offers some advantages, the evaluation of capital costs still must be conducted carefully and in a reliable manner. However, in practice regulators more often rely on the first model (Haney and Pollitt, 2009), and therefore, we conduct our analysis of efficiency on the basis of the short-run maintenance model. The determination of variables to be included is discussed broadly in the literature. A comprehensive investigation of the variables to use as cost measures and cost drivers for international benchmarking and regulation purposes is presented by Jamasb et al. (2007, 2008) examine the productivity development of US natural gas transmission companies and review the literature with respect to variables. We note that most of the studies presented in the latter paper rely exclusively on parametric approaches.

We develop two model settings (Model 1 and Model 2), each containing the same cost measurement but differ in their number of cost drivers. We select total

slack variable to be equal or smaller than the product of inverse weighting matrix and weighted output data.

Table 5.2: Model specification under VRS assumption

	Model 1		Model 2	
	DEA	PCA-DEA	DEA	PCA-DEA
Opex	x	x	x	x
TotDeliv	x	x	x	x
TransSys	x	x	x	x
PeakDeliv	x	x	x	x
HorPow	x	x	x	x
TransLos			x	x

operating and maintenance expenses (Opex) as the input to be minimized.[7] Although there are arguments in favor of total expenses including capital costs, we do not consider them here. However, Jamasb et al. (2007) shows high correlation between these two measurements. The basic model (Model 1) treats total amount of natural gas delivered (TotDeliv), transmission system (TransSys), peak deliveries (PeakDeliv), and total installed horsepower (Hp) of compressor stations (HorPow) as Opex determinants and therefore outputs. The second model (Model 2) adds transmission system losses (TransLos), which is an undesired output[8] and, therefore, must be treated differently. It is not our aim to present the particular effect of this undesired output itself; rather, we wish to demonstrate how an additional output will affect the empirical analysis and therefore regulatory consequences. For the purpose of demonstration and comparison, each of the two models is specified under traditional DEA and PCA-DEA methodology, both assuming VRS technology. The resulting four model specifications are listed in Table 5.2.

5.3.2 Data

We use data on US natural gas transmission companies. The US natural gas industry offers a comprehensive record of publicly available data and regulatory history, making it ideal for our analysis. We compile data from the US Federal Energy Regulatory Commission's (FERC) database of the major interstate natural gas pipelines. This covers each natural gas company whose combined gas transported or stored for a fee exceed 50 mn Dekatherms (Dth) in each of the previous three calendar years (FERC, 2008). In total our original sample contains 37 US natural gas transmission companies in 2007 operating only onshore pipelines.[9] However, these companies are either stand alone units or units covering a broader business portfolio (holdings). Table 5.3 summarizes all variables we use.

The sample includes natural gas transmission pipelines that spend about 2,860

[7]This is known as Opex-benchmarking; Haney and Pollitt (2009) list international regulators who in fact conduct Opex regulation.

[8]All resources devoted to the production of natural gas lost in the system are captured by

Table 5.3: Descritive statistics of US natural gas transmission companies

Variable	Sum	Min	Mean	Median	Max	Std.dev.
Opex [mn USD]	2,860.32	1.25	77.31	31.50	402.67	99.61
Total deliveries [mn Dth]	34,191.24	49.93	924.0	403.89	6,046.71	1,255.53
Transmission system [miles]	127,783.20	59.00	3,453.60	1,680	14,463.20	3,703.33
Peak deliveries [mn Dth]	86.81	0.19	2.35	1.68	8.44	2.12
Installed horsepower [thsd Hp]	11,003.22	9.00	125.95	297.38	1,434.27	371.72
Transmission system losses [thsd Dth]	38,677.68	0.00	1,045.34	615.66	6,684	1,399.32

Source: US FERC. Notes: observations=37, year=2007, onshore pipeline companies included only.

mn USD on operating and maintenance for approximately 127,783 miles of onshore facilities. This covers about 66.5 percent of total US interstate pipeline mileage. Pipelines differ in transmission system[10] and total deliveries[11], ranging from 49.93 mn Dth to over 6,046.71 mn Dth. The data indicates that some deliver low amounts of gas in peak times[12] with a minimum of 0.19 mn Dth, while others deliver up to the maximum 8.44 mn Dth. Another output is compressor stations' total installed Hp, an important characteristic of gas transmission. Installed Hp is calculated as the product of the number of stations and their certified Hp. This enables us to incorporate a capacity measurement. In fact, the data show significant differences in installed Hp ranging from a minimum of 9 thsd Hp to a maximum of nearly 1,435 thsd Hp. The std.dev. of 371.72 thsd Hp indicates the strong variation in the data.

An additional output variable is transmission system losses. In total nearly 39.7 mn Dth of natural gas are lost that would not occur in total deliveries. Pipelines report data ranging from no losses to 6,685 thsd Dth. A record of zero losses is technically very unlikely. Therefore, we suspect measurement errors, which we try to overcome with the subsequent outlier detection. The variable TransLos must be treated differently from the others because of the inverse interpretation of undesirable outputs. To ensure a correct representation, we translate this variable such that more losses are disadvantageous to companies' performance. Thus, we subtract from a large number[13] and choose 10,000,000 as the large number; the results are insensitive to a variation of the large number to 8,000,000 instead.

Opex.

[9] We omit companies that also operate offshore pipelines since the technology differs.

[10] Petal Gas Storage, L.L.C. operates the smallest pipeline system, 59 miles, and Tennessee Gas Pipeline Company operates the largest, 14,463 miles.

[11] Natural gas delivered does not only account for own sales but also for interactions with others.

[12] Natural gas delivered in peak times refers to single day peak deliveries summing deliveries to interstate pipelines and "others".

[13] Other ways to implement undesirable outputs in the DEA framework are discussed in Dyson et al. (2001).

5.3.3 Outlier Detection based on Super-efficiency

Because nonparametric methods are sensitive to outliers (Simar, 2003), we conduct an outlier detection based on the concept of super-efficiency proposed by Banker and Gifford (1988) and Andersen and Petersen (1993). Following Banker and Chang (2006), we choose the selection criterion of 1.2: companies achieving an efficiency score equal to or smaller than 1.2 are accepted for the sample and those exceeding this criterion are excluded from further analysis. We find that three of the 37 utilities are super-efficient: Columbia Gas Transmission Corporation with 5.42, Petal Gas Storage, L.L.C. with 2.97, and Vector Pipeline L.P. with 2.02. In addition, this outlier detection confirms doubts from reporting non-transmission system losses for two of the three.[14] Hence, our final sample size is 34 pipelines.

5.4 Results

This section presents our results.[15] First, we deal with the results of the PCA, followed by the efficiency estimation for the two models (Model 1 without TransLos and Model 2 with TransLos) and methodologies (DEA and PCA-DEA). We then discuss the results for our model specifications for a particular pipeline to illustrate the relevance of PCA-DEA for real-world regulatory practice.

5.4.1 PCA

The PCA enables us to reduce the dimensions of the linear program and thus to increase discrimination between the pipelines of interest. Table 5.4 shows the results of our separate PCA analysis for both models. In terms of output, the first PC ($PC1$) captures at least 82 percent of data variation in both models. Taking also $PC2$ into account results in a cumulative explanation of more than 95 percent in Model 1 and 91 percent in Model 2 of total data variation. Using only these two output PCs does not cause much loss of information for either Model 1 (5 percent) or Model 2 (9 percent). Since we consider only one input PC, we capture all information. Hence, it exactly represents the single input and does not affect efficiency measurement.

[14] The other two of the four pipelines which report zero transmission losses do not determine the frontier in the super-efficiency analysis.

[15] Calculations are conducted using the statistical software R with the additional package "FEAR" version 1.12 by Wilson (2008) and the computer program PCA_DEA developed by Adler and Yazhemsky (2006).

Table 5.4: PCA for Models 1 and 2

	Variance explained by PCs			
	Model 1		Model 2	
PC	Input	Output	Input	Output
1	1	0.8776	1	0.8219
2		0.0752		0.0834
3		0.0335		0.0571
4		0.0138		0.0268
5		-		0.0108

5.4.2 Efficiency of Pipelines

Descriptive statistics of the pipelines' individual efficiencies given by DEA and PCA-DEA for each model are shown in Table 5.5. A company is radially efficient if it achieves 100 percent. The lower the efficiency score the worse the company has performed relative to its peers.

We find two general results. First, compared to the traditional DEA approach, PCA-DEA yields lower efficiency across both model specifications. For example, pipelines in Model 1 (without TransLos) achieve 67 percent on average but 47 percent under the PCA-DEA specification. This empirically reflects the argument of Adler and Yazhemsky (2010) by which PCA-DEA has effects similar to the imposition of weight restrictions, which renders parts of the efficient boundary of the production possibility set no longer efficient. In other words, companies that are really specialists in one of the original dimensions would be considered efficient performers due to linear programming. In fact, only specialization in this particular dimension would lead to efficiency, whereas the overall performance of the affected company does not. The single feature criterion (specialist in one dimension) is a particular problem for nonparametric approaches, while the weights of the variables by the coefficients attenuate the empirical problem in parametric SFA frameworks (Riechmann and Rodgarkia-Dara, 2006). This overestimation of efficiency occurs especially when only a few observations are present relative to the number of variables. By means of PCA-DEA we reduce the output space to only two dimensions and thus improve the efficiency determination.

Second, comparing the particular specifications of Model 1 with their counterparts in Model 2 (including TransLos), we observe higher efficiency in the latter model. This observation is almost true for every statistic except for the minimum values, e.g., DEA specification in Model 1 reveals a mean of 67 percent and 78 percent in Model 2, and PCA-DEA specification reveals a mean of 47 percent in Model 1 and 60 percent in Model 2.[16] So far, both models appear to differ in

[16]Again, DEA estimates are lower due to the substantially reduced number of outputs included

Table 5.5: Efficiency of US natural gas transmission pipeline companies

	Model 1		Model 2	
Statistic	DEA	PCA-DEA	DEA	PCA-DEA
Min	0.2702	0.1923	0.3065	0.1910
1st quartile	0.4478	0.3146	0.5383	0.3952
Mean	0.6689	0.4654	0.7755	0.6004
Median	0.6386	0.3951	0.9345	0.4839
3rd quartile	0.9553	0.5708	1	0.9850
Max	1	1	1	1

some respects, e.g., to median or 3rd quartile scores, which is highly relevant to regulatory practice. However, the robustness of PCA-DEA analysis is supported when considering pipeline-specific efficiency scores.

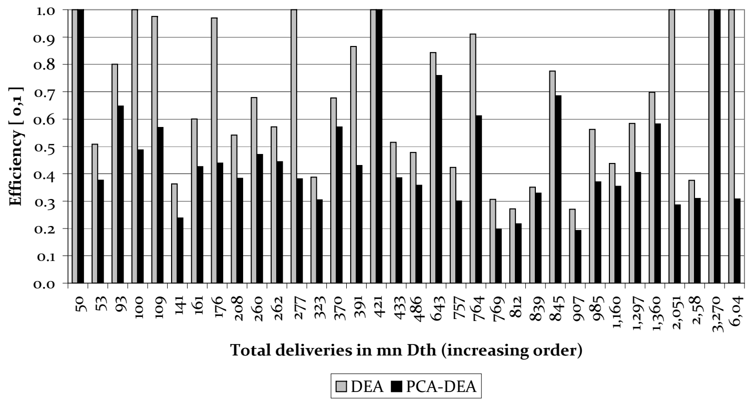
Figure 5.1 shows how company-specific efficiency scores change with DEA and PCA-DEA, and with our two model specifications. In addition to the findings already discussed—also retraceable here—other noticeable findings occur. In both graphs the pipelines are arranged in increasing order of total deliveries (TotDeliv), indicating their size. For DEA specification, none of the plots suggests an identifiable trend of better performance depending on pipelines' size. This can be explained by the VRS approach. However, for PCA-DEA specification, the larger pipelines seem to be better performers. Intuitively, the impact of single features, which make companies efficient in the range of smaller companies when VRS technology is assumed, is attenuated.

The number of pipelines that are part of the efficiency frontier is clearly higher when DEA applies. In this case, Model 1 depicts seven efficient utilities, and Model 2 even defines half of the sample as efficient due to the additional output variable TransLos. It is for technical reasons that the more variables are included in traditional DEA, the more units are considered to be efficient. This has particular importance in small samples. Moreover, Adler and Yazhemsky (2010) show by means of Monte Carlo simulation, that a trade-off occurs between incorrect classification of (in-)efficient DMUs under traditional DEA and PCA-DEA. If technology and salient variables are correctly specified, traditional DEA never defines truly efficient units incorrectly as inefficient, i.e. the probability of error type 1 is zero. But at the same time, the probability of incorrectly defining inefficient units as efficient (error type 2) is high in DEA under VRS. Thus, we can expect a remarkable proportion of pipelines to be overestimated in terms of efficiency, and potential cost reduction would remain uncovered. Therefore, the aim of regulation is not achieved.

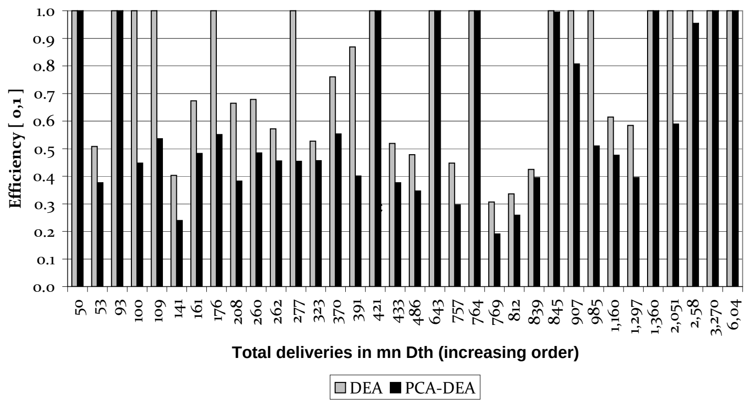
Even though PCA-DEA can improve benchmarking activities while notably

in Model 2.

Figure 5.1: Individual efficiency scores for US natural gas transmission pipeline companies



(a) Model 1 (without TransLos)



(b) Model 2 (with TransLos)

lowering the level of over-estimation, there is a cost. PCA-DEA causes a certain level of under-estimation. However, in radial efficiency measurement, this effect is minor. Adler and Yazhemsky (2010) demonstrate that with PCA-DEA the probability of under-estimation (error type 1) is very small while the probability of over-estimation (error type 2) significantly improves. Empirically PCA-DEA in our analysis defines three (Model 1) and nine (Model 2) pipelines as efficient. Note that in both cases only two output PCs are included in the analysis and thus, the ratio of variables and observations is acceptable. Hence, PCA-DEA offers methodological features that are preferable to those of traditional DEA.

However, for both models we observe that most of the pipelines suffer from introducing PCA-DEA. In Model 1, the second-smallest pipeline delivering about 53 mn Dth of natural gas achieves 51 percent under the DEA specification and decreases to 38 percent under PCA-DEA; a larger pipeline delivering 1,360 mn Dth of natural gas achieves 70 percent under DEA and decreases to 58 percent under PCA-DEA. But in Model 1 there are also companies that do not suffer from introducing PCA-DEA, i.e. those delivering 50, 421, and 3,270 mn Dth. We note that only peers (fully efficient companies) remain at the same level as before. It seems that their respective efficiency score is not distorted from unique characteristics[17] and full efficiency is justified.

According to Adler and Yazhemsky (2010) it is preferable to avoid the omission of relevant variables because it leads to under-estimation of the mean efficiency. For regulatory practice including operating characteristics, quality variables, etc., in a sophisticated model can be important. However, the request for a realistic representation of company structures easily increases the number of variables substantially and hence, harms the ratio between observations and variables. The known consequence is a deteriorated discrimination capability of DEA. In fact, including TransLos in place of the mentioned variables yields significantly changed efficiency scores in both model specifications of Model 2, i.e. DEA and PCA-DEA. Still, the methodological difference induces a reduction of dimensions when PCA-DEA is applied; thus, using PCA-DEA does not affect the discriminatory capability although more variables are considered before.

When we compare the PCA-DEA results of Model 1 (without TransLos) and Model 2, 29 percent of the companies (10 out of 34) exhibit lower efficiency under Model 2. Other companies improve or remain as good as before. The maximum individual deterioration of about 4 percent is experienced by the company delivering 100 mn Dth in total. Note that because dimensions are equal in both models, changes seem to be associated with new information. At the same time, the PCA-DEA specification in Model 2 discloses the ability of PCA to account for specialists which we explain by one specific pipeline in more detail in the following section.

5.4.3 Case Study: Northern Border Pipeline Company (NBPC)

Northern Border Pipeline Company (NBPC) delivers 907 mn Dth in total and achieves very low efficiency scores in Model 1 (27 percent with DEA and 19 percent with PCA-DEA), but the efficiency scores increase significantly when including TransLos in Model 2. Under traditional DEA, the pipeline achieves 100 percent efficiency. This indicates specialization in the particular variable TransLos which accounts for roughly 78 thsd Dth (so it seems unlikely to be an error in reporting). In contrast, when applying PCA-DEA, the efficiency score falls to 81 percent.

[17] Riechmann and Rodgarkia-Dara (2006) point out that statistical fuzziness and unique characteristics are sources of distortion.

Table 5.6: Case study: peers of NBPC in PCA-DEA model specification

		Peers in Model 1		Peers in Model 2	
Variable	NBPC	IPOC	TGPC	DTI	EPNGC
Opex [mn USD]	165.3	9.3	117.3	70.7	373.4
Total deliveries [mn Dth]	907.0	420.6	3,270.0	1,360.1	6,046.7
Transmission system [miles]	1,399	414	10,325	3,344	10,240
Peak deliveries [mn Dth]	2.6	1.4	8.4	4.0	5.1
Installed horsepower [thsd Hp]	536.6	78.3	1,434.3	350.2	1,136.4
Transmission system losses [thsd Dth]	77.9	489.4	6,684.6	398.5	3,038.8

What cannot be seen from this graph directly is how the reference set of NBPC changes between Model 1 and Model 2 with respect to PCA-DEA specification. Table 5.6 provides more insight on the relevance of this reference set (peers) on the efficiency of our example. In Model 1 NBPC is compared to the efficient utilities Iroquois Pipeline Operating Company as Agent/Iroquois Gas Transmission System, L.P. (IPOC) and Transcontinental Gas Pipeline Corporation (TGPC), whereas in Model 2 Dominion Transmission, Inc. (DTI) and El Paso Natural Gas Company (EPNGC) appear to be its peers.[18]

Obviously, IPOC is a much smaller company, e.g., Opex are only about 9 mn USD, and total deliveries account for roughly 421 mn Dth. Peers in the reference set of Model 2 are structurally more alike than the peers in Model 1. This can also be observed in Figure 5.1, where in Model 1 (PCA-DEA specification) the peers of NBPC are the efficient companies delivering 421 and 3,270 mn Dth, and in Model 2 the peers are those efficient pipelines delivering 1,360 and 6,047 mn Dth. This finding confirms the idea of DEA in the regulatory context. Burns et al. (2005) relate benchmarking techniques to yardstick competition and point out that one key feature of DEA is that it identifies “local” conditions, i.e. analyzes the efficiency of a firm with reference to other firms that are similar in their combinations of outputs, for example. If regulators want benchmarking to fulfill this prerequisite, our results support its fulfillment when relevant variables are part of the analysis and discrimination power is given by applying PCA-DEA.

5.5 Conclusion

The purpose of this chapter is to empirically demonstrate how improving discriminatory power in nonparametric efficiency analysis affects the efficiency scores of natural gas transmission companies. Moreover, we want to support a pragmatic approach of efficiency evaluation for (European) regulatory authorities that accounts

[18]The peers in Model 1 with DEA specification are TGPC and Guardian Pipeline, L.L.C. In Model 2 with DEA specification NBPC serves as the peer, because of its specialization.

for a poor ratio between the number of variables and the number of observations.

Over the last decades network industries with natural monopoly character have experienced extensive restructuring towards incentive-based regulation schemes. Restructuring aims to motivate more efficient production and cost structures. Benchmarking has become an established tool in regulatory practice to identify company-individual targets for achieving these goals. Although there is an increasing interest in parametric benchmarking methods, e.g., SFA, practical experience show frequent application of nonparametric approaches such as DEA. For meaningful efficiency measurement, DEA requires a sufficient amount of data. However, due to the former monopolistic market structures and yearly conducted efficiency evaluation, this cannot always be guaranteed in reality. Limited data negatively affects DEA's discriminatory power, and thus increases the probability of efficiency over-estimation. This issue amplifies when a large number of variables are considered to describe the production process or cost structure of companies.

To address this issue, DEA can be combined with PCA. By means of linear combinations of the original variables PCA reduces the dimensions while maintaining a large proportion of the variation in the original data. Consequently, discriminatory power in PCA-DEA improves and results in more robust efficiency scores. If regulators want benchmarking to fulfill this prerequisite, our results support its fulfillment when relevant variables are part of the analysis and discrimination power is given, i.e. by applying PCA-DEA. We test our hypotheses by applying PCA-DEA to a large sample of US natural gas transmission pipelines. We chose to employ US data because it is publicly available and the industry has a significant regulatory record. We defined two models, one with four output variables and a second with five; both models had a single input.

Our results suggest that PCA-DEA improves nonparametric efficiency analysis. Models applying traditional DEA display a high proportion of fully efficient pipelines (up to 50 percent), where we can suspect many are over-estimated. Because over-estimation decreases, pipelines on average perform less well under PCA-DEA than under DEA, which we trace back to more realistic efficiency measurement. We then show that additional outputs significantly change the results and, in PCA-DEA models, improve the evaluation of pipelines. Efficiency score changes between the different PCA-DEA model specifications appear to be not due to higher model dimensions, but due to worthwhile information and structurally similar reference companies. We conclude that these findings support current regulatory practice by mitigating the conflict between too few observations, and the demand for many variables to produce an appropriate representation of the relevant structures.

Chapter 6

Estimating Alternative Technology Sets in the Context of Nonparametric Efficiency Analysis in Regulation: A Restriction Test for Pooled Data

6.1 Introduction

We demonstrate an approach that provides statistical inference for testing hypotheses about the specification of alternative technology sets in nonparametric efficiency analysis when pooled data is used. This is, for example, of particular relevance in regulatory benchmarking where there is incomplete information about the production process and there are only a limited number of observations.

It is well known that the private sector draws on comparative analyses, such as activity analysis, to improve its performance. Starting in the 1990s, regulatory authorities also have made increasing use of benchmarking techniques in order to facilitate incentive regulation of network utilities; see e.g., Jamasb and Pollitt (2003). In particular, electricity and natural gas transmission and distribution utilities are involved in regulatory activities; see e.g., Jamasb et al. (2004); Cullmann (2012); Farsi et al. (2007); Sickles and Streitwieser (1998); Hollas et al. (2002). Applying benchmarking methods allows the regulator to simulate competitive market structures (quasi-competition), thus helping to pursue and implement regulatory objectives, e.g., reducing monopolistic power and promoting the efficient use of resources. Beside process and activity analysis and parametric frontier models, e.g., SFA, regulators frequently rely on DEA in order to establish benchmarks for target determination (Haney and Pollitt, 2009).

DEA is a nonparametric method of frontier analysis, i.e. efficiency analysis, and closely related to the classical models of activity analysis.[1] It offers an alternative way to evaluate the performance of any kind of production entities.[2] Different from classical activity analysis, the concept of efficiency analysis intends to express productive efficiency in a multiple-input-multiple-output framework while avoiding the index number problem (Farrell, 1957; Cooper et al., 2011). In efficiency analysis, the production unit's performance is determined by comparing it to a group of production entities that have access to the same transformation process (technology) by which they convert the same type of its resources (inputs) into the same type of products (outputs). From the observed input-output-combinations a best practice (frontier) is constructed against which each entity is assessed individually. The distance to that frontier, if any is present, reflects the production unit's ability to transform inputs into outputs, relative to what empirically is found and therefore assumed to be feasible. Hence, efficiency analysis provides a quantitative measure of the existing potential of improvement. As pointed out by Bogetoft and Otto (2011), the scope of application of DEA method is rich, since conceivable production entities include firms, organizations, divisions, industries, projects, DMUs, or individuals; either pursuing for or non-profit targets. Hence, in addition to regulated network industries, empirical analysis investigates, for example, warehouses (Schefczyk, 1993) and coal mines (Thompson et al., 1995).

As a nonparametric method, DEA has, on the one hand, appealing characteristics (Simar and Wilson, 2008); beside its great flexibility and easy computability, it requires only few assumptions on the technology set and its frontier. Particularly, it does neither assume a distributional expectation about the inefficiency term nor does it impose a functional form to express the production process generating the observed input-output-combinations (Haney and Pollitt, 2009; Simar and Wilson, 2008). On the other hand, the DEA estimator has drawbacks that are highly relevant not just for regulatory but also other types of benchmarking. Due to its deterministic nature, the DEA estimator is outlier sensitive and has a lower convergence rate compared e.g., to parametric alternatives. Therefore, in order to obtain reasonable estimation results in finite samples, it is preferable describing the technology set of the firms under investigation by only a small number of inputs and outputs. However, in situations where there is uncertainty about

[1]For more details about the methodological linkages of activity and efficiency analysis, the reader is referred to e.g., Färe and Grosskopf (2005).

[2]Homburg (2001) gives detailed insights on how nonparametric efficiency analysis can contribute to activity-based management.

the correct specification of the technology set, e.g., due to existing information asymmetries between the regulator and the regulated company, statistical inference about alternative specifications is desirable in order to make decisions about the reasonable choice of variables.

Simar and Wilson (2001) and Simar and Wilson (2011) propose different restriction tests for nonparametric efficiency analysis that allow to investigate whether certain variables are relevant and whether aggregates of variables can be used. Schubert and Simar (2011) further extend these tests and demonstrate how to examine the specificity of innovation input compared to other inputs, i.e. labor and capital. Although the benefits are obvious, restriction tests on production process formulations yet receive at least empirically only scant attention. These approaches notably improve nonparametric benchmarking, because they increase the confidence in the chosen representation of the production process by providing statistical inference. Hence, they notably reduce the risk of overestimating the performance due to the 'curse of dimensionality' when variables are identified as irrelevant and consequently are excluded from further investigation.[3] This is exacerbated by the fact that the existing implementation of the proposed tests is restricted to cross-sectional data and are therefore not applicable to (unbalanced) panel data. Due to the market structure in network industries, many regulatory benchmarking applications rely on a small number of observations; see e.g., Jamasb et al. (2008). Two obvious options exist in order to obtain a larger sample size: First, using cross-country analysis, and second, pooling cross-sectional data across time periods. For pooling across countries principally the simple cross-section tests can be used as long as we guarantee that all countries have access to the same technology. However, for the often more relevant comparisons across time the additional problem emerges that a firm's present and past observations are generally not independent. So pure cross-section methods will lead to false inference, even if the technology did not change over the respective period.

Therefore, the contribution of the chapter at hand is twofold: First, we further develop the theoretical underpinnings of the tests in order to enhance their applicability to (unbalanced) panel data. This requires accounting for intra-observational dependencies. Second, we demonstrate the relevance of the proposed test procedure for benchmarking by applying the method to a data set of US natural gas transmission companies. Clearly, the main benefits of this approach are im-

[3]Alternatively, variables could be omitted or aggregated. Omitting variables based on correlations should be avoided for translation invariant DEA models (Dyson et al., 2001) and aggregating variables based on PCs might be inappropriate for radial efficiency measurement (Simar and Wilson, 2001). However, the restriction tests proposed by Simar and Wilson (2001) and Schubert and Simar (2011) provide statistical inference procedures for the investigation of aggregates.

proving the efficiency estimation and overcoming lacks of information regarding the production process (information asymmetries or uncertainty). Although, our demonstration relates to the regulatory framework, it is straightforward to apply the technique to any other setting where the mentioned problems arise.

The chapter is organized as follows: Section 2 briefly introduces the nonparametric performance estimator DEA and formalizes the hypotheses we aim to test. The hypotheses are constructed accordingly to the described challenge of technology set modeling. Furthermore, the test arrangement and the derivation of test decisions are explained. Additionally, Section 2 deals with testing for technical change, which needs to be ruled out in order to accomplish the requirement for pooling cross-section observations across time in nonparametric efficiency analysis. Section 3 describes our empirical framework and presents the results. Finally, Section 4 concludes.[4]

6.2 Methodology

6.2.1 Technology Estimation using the DEA Estimator

In order to make sound decisions regarding the relevant components, i.e. inputs and outputs, of a particular production process, we begin by defining the estimator that represents the production possibility set and which approximates the unknown technology.

Let $x^i \in \mathbb{R}_+^p$ and $y^i \in \mathbb{R}_+^q$ denote the vectors of p inputs and q outputs. The production possibility set Ψ represents the feasible input-output-combinations available to firm i (Bogetoft and Otto, 2011) and can be defined as

$$\Psi = \left\{(x, y) \in \mathbb{R}_+^{p+q} \mid x \,\text{can produce}\, y\right\}. \tag{6.1}$$

For Ψ we assume free disposability and convexity. The boundary of Ψ, denoted by Ψ^δ, describes the efficient production frontier, i.e. the technology, and can be defined as

$$\Psi^\delta = \left\{(x, y) \in T \mid \left(\gamma x, \gamma^{-1} y\right) \notin \Psi \,\text{for any}\, \gamma < 1\right\}. \tag{6.2}$$

According to Equation 6.2, a firm that employs a production plan which belongs to Ψ^δ, is regarded as efficient and its input-output combination cannot be improved. Companies that operate at points in the interior of Ψ exhibit inefficiencies (Simar and Wilson, 2001), which can be diminished by moving toward the efficient

[4]This chapter is joint research with Anne Neumann and Torben Schubert. We thank the participants of the 10^{th} INFRADAY in Berlin for valuable discussions.

frontier. Being suitable for multi-input and multi-output frameworks, the Debreu-Farrell measure[5] quantifies the respective firm-individual degree of efficiency. For any particular coordinate $(x_0, y_0) \in \Psi$, the Debreu-Farrell efficiency score is determined by the radial distance from (x_0, y_0) to the efficient frontier Ψ^δ. It expresses the maximal proportional contraction of all inputs x that allows to produce output level y for input-orientation, and the maximum proportional expansion of all outputs y that is feasible with the given inputs x, for output-orientation, respectively.

We restrict ourselves to the input-orientated firm-specific efficiency measure, which can formally be expressed as

$$\theta\left(x_0, y_0\right) = inf\left\{\theta \geq 0 \mid \left(\theta x_0, y_0\right) \in \Psi\right\}. \tag{6.3}$$

Hence, if $\theta\left(x_0, y_0\right) = 1$, the company is efficient and operates along the frontier Ψ^δ. If $\theta\left(x_0, y_0\right) \leq 1$, the company can improve its performance by reducing its input quantities proportionally. Together with the imposed assumptions, Equations 6.1 and 6.2 set up the *true* economic production model, and hence, characterize the data generating process $\mathbb{P}$ (DGP).[6] However, the *true* technology set Ψ, and hence, the *true* efficient technology Ψ^δ against which observations are compared to, are unknown and both need to be estimated from the observed input-output-combinations.

To approximate Ψ, we apply the DEA estimator proposed by Banker et al. (1984). This model incorporates the mentioned assumptions and in addition, imposes the assumption of VRS. Thus, the linear program estimating the unknown input-oriented efficiency score θ becomes:

$$\begin{aligned} \hat{\theta}\left(x_0, y_0\right) = \min_{\theta, \lambda_1, \ldots, \lambda_n} \{\theta > 0 \mid \; & \theta x_0 \geq \textstyle\sum_{i=1}^{n} \lambda^i x_k^i; k = 1, \ldots, p \\ & y_0 \leq \textstyle\sum_{i=1}^{n} \lambda^i y_l^i; l = 1, \ldots, q \\ & \textstyle\sum_{i=1}^{n} \lambda^i = 1;\ \lambda^i \geq 0 \, \forall \, i = 1, \ldots, n\}. \end{aligned} \tag{6.4}$$

It is well known that the rate of convergence for nonparametric estimators, such as DEA, is small compared to parametric estimators (Simar and Wilson, 2008). The consistency of this estimator is proven by Kneip et al. (1998). But like most nonparametric estimators it suffers from the curse of dimensionality, which implies

[5]This measure is based on the work of Debreu (1951) and Farrell (1957). Alternatively, the concept proposed by Shepard (1970) could be used.

[6]To comprehensively define the DGP, assumptions on the statistical model are necessary. Due to space limitations, we omit the discussion and refer the reader to e.g., Simar and Wilson (2001).

that the rate of convergence (i.e. the speed by which the estimation errors goes is reduced in sample size) goes down with increasing the number of inputs and outputs. Additionally, the DEA estimates are upward biased. This implies that the true efficiency is lower than the estimated one. Hence, the precision of the estimation results is significantly affected and a considerable interest arises to test for the relevance of particular inputs and outputs because reducing the dimensionality of the technology set Ψ can offer substantial gains in estimation efficiency and decrease finite sample biases.

6.2.2 Testing Restrictions

Having specified the estimation approach, we need to formulate and formalize the restrictions on the technology set we aim to test. It is the chapter's objective to test whether particular outputs are relevant for modeling the technology set appropriately. Although we focus on the relevance of outputs in this chapter, note, that the method is broader. Alternatively, the relevance of input variables can be considered. Further, it can be tested whether inputs and outputs are individual relevant contributors in the production or can be included as an aggregate. To implement this, we extend a test procedure suggested by Simar and Wilson (2001) to panel data while, following Schubert and Simar (2011), using subsampling procedures.

The basic idea of the original approach is to compare efficiency estimates obtained from a technology set including all potential outputs with efficiency estimates obtained from a restricted technology set that excludes at least one output. The rationale behind assigning a particular output as possibly irrelevant, is the uncertainty about its relationship to the considered input(s). An output is identified as irrelevant, if the difference between the estimates of both technology sets, where the restricted is nested in the unrestricted, do not differ significantly. Hence, conceptionally, this means that the irrelevant output is not produced by the firm, or putting it differently, the considered inputs do not contribute to the irrelevant output. The main benefit of this approach is twofold: First, selecting outputs can be based on statistical tests which improves the technology specification's quality and second, when outputs can be excluded yielding fewer dimensions, the estimation's quality improves leading to an increase in the speed of convergence and a reduction in the finite sample upward bias.

To formalize this reasoning, we respecify the output vector y into two subsets of outputs, i.e. $y = (y^1, y^2)$, where $y^1 \in \mathbb{R}^{q-r}$ denotes the vector of $q-r$ outputs that are assumed to be relevant outputs of the production process under consideration, and $y^2 \in \mathbb{R}^r$ denotes the vector of r outputs which are possibly irrelevant. The hypothesis then is that x influences the level of y^1 but not of y^2. The null and alternative hypothesis can therefore be written as

$$\begin{aligned} &H_0\colon x\,\text{influences the level of}\,y^1\;\;(y^2\,\text{is irrelevant})\\ &H_1\colon x\,\text{influences the level of}\,y^1\text{and}\,y^2\;\;(y^2\,\text{is relevant})\,. \end{aligned} \tag{6.5}$$

For any given input-output-combination $(x, y) = (x, y^1, y^2) \in \Psi$, the corresponding reformulated input-oriented Farrell efficiency scores in Equation 6.3 are:

$$\begin{aligned} \theta_U\,(x,y) &= inf\,\{\theta \mid (x, y^1, y^2) \in \Psi\}\\ \theta_R\,(x,y) &= inf\,\{\theta \mid (x, y^1) \in \Psi\} \end{aligned} \tag{6.6}$$

where θ_U and θ_R represent the efficiency for the unrestricted and the restricted technology set. If the outputs in y^2 is truly irrelevant, θ_R equals θ_U. Opposing, if outputs in y^2 contains relevant outputs, then θ_R would be smaller than θ_U. From that we can derive the following inequalities:

$$\begin{aligned} &\text{if } H_0 \text{ is true: } 1 \geq \theta_U\,(x,y) = \theta_R\,(x,y)\,,\ \text{for all }\ (x,y) \in \Psi\\ &\text{if } H_1 \text{ is true: } 1 \geq \theta_U\,(x,y) > \theta_R\,(x,y)\,,\ \text{for some }\ (x,y) \in \Psi \end{aligned} \tag{6.7}$$

According to Equation 6.4, θ_U and θ_R can be estimated from the sample, denoted by $\mathcal{X}_n$, as follows:

$$\begin{aligned} \widehat{\theta_U}\,(x,y) = \min_{\theta,\lambda_1,\ldots,\lambda_n} \{\theta > 0 \mid \;\; & \theta x \geq \textstyle\sum_{i=1}^n \lambda^i x_k^i; k = 1, ..., p\\ & y^1 \leq \textstyle\sum_{i=1}^n \lambda^i y_l^{1,i}; l = 1, ..., (q-r)\\ & y^2 \leq \textstyle\sum_{i=1}^n \lambda^i y_l^{2,i}; l = 1, ..., r\\ & \textstyle\sum_{i=1}^n \lambda^i = 1;\ \lambda^i \geq 0\,\forall\, i = 1, ..., n\}. \end{aligned} \tag{6.8}$$

and

$$\begin{aligned} \widehat{\theta_R}\,(x,y) = \min_{\theta,\lambda_1,\ldots,\lambda_n} \{\theta > 0 \mid \;\; & \theta x \geq \textstyle\sum_{i=1}^n \lambda^i x_k^i; k = 1, ..., p;\\ & y^1 \leq \textstyle\sum_{i=1}^n \lambda^i y_l^{1,i}; l = 1, ..., (q-r)\\ & \textstyle\sum_{i=1}^n \lambda^i = 1;\ \lambda^i \geq 0\,\forall\, i = 1, ..., n\}. \end{aligned} \tag{6.9}$$

where the relationship $1 \geq \widehat{\theta_U}(x, y) \geq \widehat{\theta_R}(x, y)$ holds by construction.

In order to test H_0, we have to find a valid test statistic that appropriately compares the estimated efficiencies under both production possibility sets. The quantity depending on the generic DGP $\mathbb{P}$ that has been proposed by the literature is the following (Simar and Wilson, 2001):

$$t(\mathrm{P}) = \mathrm{E}\left(\frac{\theta_U(X, Y)}{\theta_R(X, Y)} - 1\right). \tag{6.10}$$

From Equation 6.7 we know that the ratio is equal to zero, i.e. $t(\mathrm{P}) = 0$, if H_0 is true, whereas it is strictly positive otherwise, i.e. $t(\mathrm{P}) > 0$. Empirically, the ratio can easily be obtained by the sample empirical mean that is a consistent estimator (Simar and Wilson, 2001; Schubert and Simar, 2011). The empirical equivalent of $t(\mathrm{P})$ is therefore:

$$t_n(\mathcal{X}_n) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{\theta}_U(X_i, Y_i)}{\hat{\theta}_R(X_i, Y_i)} - 1\right). \tag{6.11}$$

As mentioned before, by construction this quantity is always greater than or equal to zero. Thus, the important question is, how big it should be to be reasonable sure that H_0 is not true, i.e. y^2 is likely to be a relevant output of x. The usual approach is to use critical values corresponding to the distribution of the term in Equation 6.11. However, although this distribution can be shown to be non-degerate, it is complicated and depends on local parameters. So far the only way to determine critical values is by bootstrap-based simulation techniques. A particularly comfortable as well as flexible way is to use the subsampling approach as argued by Schubert and Simar (2011). This approach is described and extended to panel data in the next subsection.

In order to answer the question of how large the test ratio must to be to reject the null hypothesis, we need to compute a p-value or a critical value. This purpose requires the approximation of the unknown (asymptotic) sampling distribution of $\tau_n(t_n(\mathcal{X}_n) - t(\mathrm{P}))$, i.e. the convergence of the test statistic $t_n(\mathcal{X}_n)$ against the true population parameter $t(\mathrm{P})$ at rate τ_n where t_n is a function of the sample size. Note that $t_n(\mathcal{X}_n)$ is the estimate of $t(\mathrm{P})$ that discriminates between the H_0 and H_1.

The subsampling approach is a special kind of bootstrap. It differs from the normal procedure of generating pseudo samples of the original size n in that the samples here are of size $m < n$ such that $m/n \to 0$ when $n \to \infty$. This easy adjustment makes the subsampling approach robust to deviations from assumptions necessary for the consistency of the bootstrap. In particular with DEA and related estimators the frontier problem occurs that renders bootstrapping inconsistent, while the subsampling is not.

To derive an approximation of the sampling distribution of $\tau_n t_n (\mathcal{X}_n)$, we follow Schubert and Simar (2011) and use the algorithm based on subsampling proposed by Politis et al. (2001).[7] According to the algorithm, a sufficiently large number of subsets $b = 1, \ldots, B$, denoted by $\mathcal{X}^*_{m,b}$, are constructed,[8] each of which is producing a test statistic $t_{m,b} \left(\mathcal{X}^*_{m,b}\right)$ as defined in Equation 6.11. The large number of estimated test statistics approximate the sampling distribution for which a *critical value* $\hat{t}^c_m$ can be derived. The *critical value* depends on m and the $(1-\alpha)$ quantile. At the significance level α, the test rejects H_0 if and only if $\tau_n t_n (\mathcal{X}_n) \geq \hat{t}^c_m (1-\alpha)$ where τ_n equals $\sqrt{n} n^{2/(p+q+1)}$; for details see Schubert and Simar (2011).

In this chapter we further develop the work by Schubert and Simar (2011) in the sense that we extend the applicability of the algorithm by Politis et al. (2001) to panel data. The panel is allowed to be unbalanced but year-wise missing observations are assumed to be completely random. So we assume away (non-random) panel selection, such as attrition. Let n be the total number of observations and n_p be the number of different companies in the panel. Comparably, we define m_p as the subsample size with respect to n_p. Obviously, then $n_p \leq n$. Furthermore, if the panel is balanced and L is the time length of the panel, then $n_p = L_n$. Instead, if the panel is unbalanced, then the number of observations per company is a random integer, say Z_i such that it has support on $0, 1, \ldots L$. For distinction between the cases, we rename the parameters as n_p and m_p when referring to the pooled data case. To implement the algorithm using pooled data, the companies are clustered across time and subsampled block-wise (Davison and Hinkley, 1997).

For company i, the test statistic in Equation 6.11 is then expressed as the intra-observational mean of the company-individual yearly estimates and can be rewritten as:

$$t_{n_p} (\mathcal{X}_n \mid Z) = \frac{1}{n} \sum_{i=1}^{n_p} \sum_{1}^{L} t = \left(\frac{\hat{\theta}_{U,it} (X_i, Y_i, Z_i)}{\hat{\theta}_{R,it} (X_i, Y_i, Z_i)} - 1 \right). \tag{6.12}$$

[7]Other bootstrap methods, e.g. the homogeneous bootstrap proposed by Simar and Wilson (1998) and further developed by Simar and Wilson (2001) or the double smooth bootstrap proposed by Kneip et al. (2008) are not applicable in our setting. Because we need a method that is able to allow for heteroscedasticity and that is valid for all data points considered simultaneously (Schubert and Simar, 2011), which excludes both mentioned alternatives.

[8]A large number of subsets, and hence, of subsampling replications is required in order to reconstruct the behavior of the unknown parameter. Usually, the number of replications B is set to 2,000; see e.g., Daraio and Simar, 2007a and Simar and Wilson, 2000.

This equation is quite comparable to Equation 6.11. However, two thing are different. First, because observations belonging to the same unit are likely to be correlated, the subsampling has to account for the dependence among the observations. This problem can easily be solved using block-wise subsampling as suggested by Davison and Hinkley (1997). Second, there is a nuisance variable capturing the random panel response. So, for the subsampling procedure to be consistent at all, we have to prove that this nuisance parameter does not affect the central consistency condition, i.e. that $\tau_{n_p} t_{n_p} (\mathcal{X}_n, Z)$ converges to a non-degenerate distribution; compare Schubert and Simar (2011). This proof is presented in the appendix of this chapter.

The test procedure, irrespective of the cross-sectional or panel data case, is sensitive to the choice of m_p, which implies a trade-off between too small and too large values. On the one hand, too much information is lost, if m_p is too small. On the other hand, if m_p is too large, the subsample size almost corresponds to the sample size n_p, yielding inconsistent inference (Daraio and Simar, 2007a). Therefore, an intermediate level of m_p is supposed to balance the costs of both extremes. The algorithm by Politis et al. (2001) involves a data-driven approach through which m_p is chosen such that the volatility of the resulting measure of interest is the smallest. As volatility index we calculate the std.dev. of the 95 percent quantile of the test statistic on a running window from $m_p - 2$ to $m_p + 2$.[9] Simar and Wilson (2011) show that this data-driven approach allows for tests on m_p and on desirable power properties, e.g. rejecting H_0 with high probability when H_0 does not hold (Schubert and Simar, 2011). In order to evaluate the test statistic's volatility with respect to the choice of m_p, a grid of values m_p can reasonably take, is defined. These values belong to the interval $[m_{p,min}, m_{p,max}]$. For each of these values $\hat{t}^c_{m_p} (1 - \alpha)$ can be calculated and investigated with respect to their volatility. Therefore, a plot of the critical values $\hat{t}^c_{m_p} (1 - \alpha)$ against the possible values of m_p reveals a first impression of where the interval's region that exhibits stable results (smallest volatilities) lies.

6.2.3 Outlier Detection

Since the DEA estimator is deterministic, i.e. it envelops all observed data points to construct the full frontier, it is not robust against extreme values and data errors, further referred to as outliers; see e.g., Simar (2003); Simar and Wilson (2008). Before testing the restrictions on the technology set, we therefore, perform

[9]This corresponds to the selection rule proposed by Simar and Wilson (2008) that selects a value of m for which the resulting sample distribution and some of its features, e.g., relevant moments, are stable with respect to deviations from this particular value.

an outlier detection procedure and use the approach suggested by Pastor et al. (1999) to identify suspicious observations. To evaluate the influence of a particular observation (say DMU_j) on the performance measure of other observations, two steps are involved: In the first step, DMU_j is removed from the sample and the efficiency estimates for all other observations are obtained as usual. For the second step, an artificial sample is constructed that contains the observations identified as efficient in the first step plus the efficient projections of observations identified as inefficient in the first step. In the second step, the efficiency measurement program is conducted using the artificial sample and the DMU_j excluded before. If DMU_j has no impact on the performance measurement of the remaining observations, the two efficiency estimates obtained for each of the remaining observations in the first and second step are equivalent. If both estimates differ, the remaining observations (or their efficient projections) can reduce their inputs (in the case of input-orientation) to the extend of the efficiency score obtained in the second step. We use the standard test assumptions proposed by Pastor et al. (1999), where DMU_j is considered as influential if it makes more than 5 percent of the remaining observations reduce their efficiency to less than 95 percent. Based on these parameters, a p-value can be derived, which indicates whether DMU_j should be excluded from further analysis.

6.3 Application to US Natural Gas Transmission Companies

6.3.1 Technology Specification and Variable Selection

The introduced method is applied to the sector of natural gas transmission, which is frequently involved in regulatory benchmarking activities worldwide. As pointed out by Jamasb et al. (2008), the regulation schemes vary among countries, with the most obvious differences between European countries and the US. Regulating the gas transmission industry traditionally relies on cost-of-service or rate-of-return in the US; overviews of the implemented scheme is given, e.g., by Sickles and Streitwieser (1992, 1998) more recently by O'Neill (2005). In contrast, the European regulators increasingly shift toward incentive regulation, an approach that is in general discussed e.g., by Vogelsang (2002). As shown by Haney and Pollitt (2009), European regulators frequently use DEA for incentive-based regulation of the natural gas transmission companies.

A crucial part of regulatory benchmarking is to specify the technology set and, consequently, extensive attention is usually devoted to the choice of variables; see e.g., Jamasb et al. (2008). The conflict in real life applications arises from the opposing interests of regulating authorities and regulated firms: firms, on the one hand, intend to increase the number of the considered variables in order to make the model as detailed as possible and therefore increase the dimensions of the technology set. In the case of high dimensionality nonparametric efficiency analysis as an regulatory instrument is compromised because no meaningful efficiency estimates can be obtained due to the curse of dimensionality. Hence, the regulators, on the other hand, focus on only a few variables that appropriately model the technology set. We draw on the discussions found in the literature in order to establish alternative specifications of the technology set that we use to perform the proposed restriction test.

The primary task of natural gas transmission companies is to transfer natural gas from other upstream facilities[10] to city gates, storage facilities and some large industrial customers. From the city gates on, the commodity is distributed to all other customers via local distribution systems that do not belong to the transmission system. To accomplish the task, the companies essentially employ pipelines, compressor stations, natural gas as fuel and personnel.

In order to conduct a benchmarking analysis it is necessary to specify the variables that represent the inputs and outputs involved in this production process.[11] We begin with input variable selection. Similar to firms in other sectors, commonly considered input factors for natural gas transmission are labor, capital, and "other inputs" containing e.g., fuel, materials, and power (Coelli et al., 2003). The expenses on labor and "other inputs" basically constitute the operating expenses, whereas investment spending relates to capital expenses. Since compressor stations require a notable amount of fuel and maintenance, the relative share of "other inputs" is large in natural gas transmission compared to other technologies. The crucial contributors to the pipeline operating costs are therefore the number of compressor stations and labor expenses (IEA, 2003). With unknown factor prices, we use operating and maintenance expenses ($O\&M$) as an aggregated input measure, which sufficiently covers expenses for labor and "other inputs". The aggregated measure implies that factor prices are identical for all firms. Although this is a strong assumption, it seems reasonable in our context. However, it needs to be carefully considered in each application. An advantage

[10]These mainly include gas storage facilities, gas processing and treatment plants, and liquefied natural gas storage and processing plants.

[11]For a general overview of commonly considered inputs and output of network industries, the reader is referred to Coelli et al. (2003); a comprehensive discussion on the variable selection in the context of gas transmission is given by e.g., Jamasb et al. (2008).

of the monetary aggregate is that it ensures to account for all employed inputs. In addition, from an analyst's perspective, it overcomes information asymmetries; authorities find it difficult to obtain accurate input factor prices and physical input quantities (Jamasb et al., 2008). Note that the legitimacy of input (or output) aggregation should also be tested, e.g. by means of restriction tests; however, this it outside the focus of the present work.

Excluding capital has three reasons: First, the data on capital costs or capital stock are often very limited or hardly comparable. Second, regulators frequently rely on model specifications excluding capital input related measures; see e.g., Haney and Pollitt (2009). Third, capital (or infrastructure) could alternatively be considered as a factor that enters the equation through determining the amount of labor input and "other inputs" rather than being a separate input factor. This means in our application that the pipeline networks' characteristics determine how much personnel and maintenance is required to run the business. Therefore, the pipeline network does not necessarily constitute an individual input.[12]

There is a broad consensus about the plurality of outputs in network industries. The most obvious and frequently used measure to include is the natural gas delivered (*deliv*) Coelli et al. (2003). Additionally, we consider the amount of natural gas delivered in peak times (*peak*) since the difference across firms is relevant in particular when regional characteristics vary. The provision of infrastructure (or the service supplied by using this infrastructure) itself can be considered a distinct output. Unlike other studies in which length of mains (*length*) is incorporated as some capital measure, e.g., Jamasb et al. (2008), we use it as proxy for transportation service. In addition, including *length* improves the comparability among the investigated pipeline companies. Typically, larger (existing) networks are associated with higher operational costs: Compressor stations, installed to maintain the network pressure,[13] determine a large part of personnel expenditures and maintenance costs (including fuel consumption). Not considering this technical aspect leaves companies with high *O&M* due to large networks at a disadvantage, *per se*. The network length appears to be a suitable proxy for the number of installed compressor stations since they occur in rather regular intervals of 150-200 km, corresponding to about 93-124 miles (Natgas.info, 2011).[14] Another frequently considered measure is the number of customers supplied, which accounts for the

[12]This approach however, requires additional methodological implementations that are beyond the scope of this chapter.

[13]The transport of natural gas is based on a pressure differential at the inlet and outlet.

[14]However, we are aware of the fact that the length of mains cannot fully explain the differences of compressor station's total operational costs since these also depend on the engineering characteristics.

multiplicity of output. However, the number of connections seems to be of minor importance in gas transmission networks. We therefore exclude it from consideration. Furthermore, pollution (as a bad output) is sometimes taken into account but not considered here.

We derive two model specifications, i.e. Model 1 and Model 2, each of which uses *O&M* as input but different sets of outputs. Model 1 is the base specification of Ψ we aim to estimate and takes *deliv*, *peak* and *length* as outputs into account. According to the line of argumentation of the test, we consider *deliv* as potentially redundant for modeling the technology set of natural gas transmission. The reasons for considering the relevance of *deliv* as uncertain is twofold: (i) much of the resources approximated by *O&M* are devoted to the compressor stations responsible for the network's capacity provision. With *peak*, we therefore already capture this. Similarly, the German regulator argues, with respect to the technology set of electricity distributors, that the energy delivered is not a primary cost driver as long as capacity provision is encountered for (Bogetoft and Agrell, 2007); (ii) because *peak* and *deliv* are highly correlated, it is uncertain whether *deliv* contains distinct information about the production process. The dimensionality in that case would unnecessarily be inflated. Thus, Model 2 uses a reduced set of outputs, i.e. *peak* and *length*, and corresponds to H_0 in Equation 6.5. Not rejecting H_0 would imply that *deliv* is not required to approximate the technology set under investigation. This is exactly, what we aim to investigate for demonstrating the approach with this application.

6.3.2 Data

We employ data on US natural gas transmission companies provided by the FERC. FERC Form No. 2, includes all natural gas companies whose combined gas transported or stored for a fee exceed 50 mn Dth. Given we assume that the technologies of onshore and offshore pipelines differ, we consider companies operating onshore facilities only. Some missing values and data irregularities were excluded from the data set. The remaining sample contains information on 43 natural gas transmission pipeline companies that are observed with unequal frequency over a five-year time period (2003-2007).[15] In total, the unbalanced panel includes 191 observations.

By pooling cross-sections, we assume that all observations have access to the same technology, meaning that technical change is absent during the considered

[15]Note that we want to empirically apply our proposed method and are therefore not concerned about the exact period under consideration.

time span.[16] Hence, changes in productivity are rather driven by productivity and technical efficiency change.

Table 6.1 presents the characteristics of the data. All variables are related to the companies' transmission branch. In general, all variables exhibit high std.dev.s, indicating notable differences between the sample companies. Median values are consistently below corresponding mean values suggesting that the sample consists of relatively more large-size firms. For *O&M* we use the reported sum of transmission expenses for operation and maintenance. The monetary values are inflation adjusted to 2003 dollars for comparability purposes. On average the pipeline companies spend 42 mn USD on *O&M*. *Deliv* represents the account for the total quantity of natural gas delivered by the respective company and ranges from about 20 mn to 3 bn Dth. In order to ensure comparability with peak period information, we transformed this variable into Dth per day. The corresponding measure of supplied quantity then has a minimum and maximum value of 0.06 to 8.6 mn Dth a day, respectively. For *peak*, we use the single day account of the amount of natural gas delivered during system peak period. The sample companies report peak deliveries between 0.1 and 7 mn Dth (per day). *Length* represents the total length of transmission mains which highly varies among the companies. The smallest pipeline network has 80 miles of pipelines and the largest has over 9,000 miles.

Table 6.1: Descriptive statistics for US natural gas transmission companies

Variable	Min	Mean	Median	Max	Std.dev.
Opex (O&M) [thsd USD[a]]	268	42,421	20,593	244,284	50,632
Total deliveries (deliv) [thsd Dth[b]]	55	1,389	994	8,597	1,381
Peak deliveries (peak) [thsd Dth]	122	1,614	1,303	7,124	1,328
Length of mains (length) [miles]	80	2,379	1,402	9,627	2,505

Source: US FERC. Notes: observations=191, n=43, years=2003-2007, onshore pipeline companies included only. [a] Yearly operating and maintenance expenses are deflated to 2003. [b] Per day measures derived by dividing the total amount of natural gas delivered by 365 days.

[16] We test for technical change using the Malmquist approach proposed by Färe et al. (1992). We find no empirical evidence for technical change at the 1 percent level of significance. Therefore, the observations have access to the same technology and pooling is valid in our case. Further, a similar data set on US natural gas transmission companies for the years 1998 to 2004, is evaluated by Jamasb et al. (2008). The results indicate that findings on technical change are sensitive to the model specification where the magnitude is very low in most models and years.

6.3.3 Results

First, we present the results of the outlier detection based on *Model 1*.[17] The routine for detecting outliers is performed on a yearly base with results shown in Table 6.2. For each year, the table list those companies that causes, if included, a loss in efficiency larger than 5 percent for at least one other company. Further, the respective number of influenced observations and the corresponding p-values are given. If the p-value deceeds the 10 percent significance level, we consider the candidate to be an outlier and exclude it from further analysis. Since for 20 observations, the p-values is smaller than 10 percent, our remaining sample used for the subsequent analysis is reduced to 171 observations.

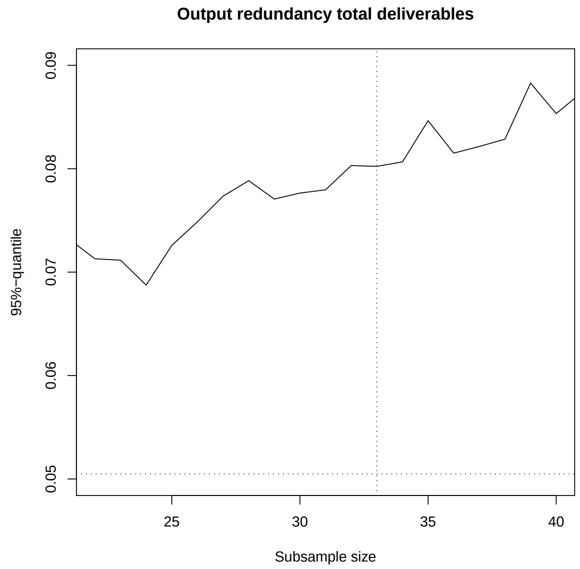
Table 6.2: Results of outlier detection

Year	ID	Influenced companies	p-value	Year	ID	Influenced companies	p-value
2003	22	8	0.000	2005	175	15	0.000
2003	37	22	0.000	2006	11	4	0.061
2003	78	9	0.000	2006	22	12	0.000
2003	172	11	0.000	2006	24	13	0.000
2003	175	1	0.774	2006	48	1	0.785
2004	22	25	0.000	2006	53	3	0.188
2004	53	2	0.447	2006	78	13	0.000
2004	78	12	0.000	2006	175	9	0.000
2004	175	10	0.000	2007	11	1	0.708
2005	7	1	0.774	2007	22	4	0.030
2005	22	12	0.000	2007	24	4	0.030
2005	24	14	0.000	2007	37	1	0.708
2005	43	1	0.774	2007	76	2	0.339
2005	53	3	0.175	2007	97	7	0.000
2005	78	11	0.000	2007	175	18	0.000

Turning to the main interest of this chapter, Figure 6.1 illustrates the test results of the restriction test based on subsampling for pooled data. In order to compare the technology sets of Models 1 and 2, we calculate for each subsample size m_p, 2,000 replications of the test statistic $t_{m_p,i}\left(\mathcal{X}^*_{m,b} \mid Z\right)$ using randomly drawn subsamples according to the presented method. From that the empirical approximation of the sampling distribution of $t_{m_p,i}\left(\mathcal{X}^*_{m,b} \mid Z\right)$ can be derived. The solid line illustrates the corresponding estimated 95 percent quantiles of the test statistic $t_{m_p,i}\left(\mathcal{X}^*_{m,b} \mid Z\right)$, i.e. the critical values $t^c_{m_p}(1-\alpha)$ at the preferred level of significance $\alpha = 5$ percent, as a function of the subsample size m_p. The vertical dashed line indicates the optimal subsample size of 33 companies, determined by the smallest measured volatility index. This corresponds to a region where the test statistic graphically appears to remain stable when slightly deviating from the identified optimal value of m_p.

[17] Calculations are conducted using the statistical software *R* with the additional package "FEAR" version 1.12 by Wilson (2008).

Figure 6.1: Results of restriction test



Note that in our applied approach of subsampling for pooled data, m_p refers to the number of cross-section covered by the sample, not to the total number of observations. For the subsample size of 33, we obtain a 95 percent quantile of 0.0798 and the corresponding p-value of 0.36. Since, the observed value of $t_{n_p}\left(\mathcal{X}^*_{n_p} \mid Z\right)$ obtained from Equation 6.12 and represented by the horizontal dashed line, is clearly smaller than the critical value of 0.0798 and the p-value of 0.36 clearly exceeds the preferred significance level of 5 percent, we do not reject our null hypothesis. Note that the presented results may vary slightly when replicating because of the stochastic nature of subsampling.

For our sample, we conclude that the reduced technology set of Model 2 is preferred to the unrestricted Model 1. Consequently, the technology set of the considered natural gas transmission companies, is determined by the single input *O&M* and the two output measures *peak* and *length,* while the variable *deliv* can be neglected as output. Note, that this result depends on the sample and does not provide general evidence for the sector. With the reduced set of outputs, we improve the efficiency estimation by reducing the dimensionality and hence, the risk of overestimating the performance.

Table 6.3 shows how the company-individual efficiency scores obtained for Model 1 and 2 differ using the DEA programs defined by Equations 6.8 and 6.9. By definition, none of the statistics becomes negative, meaning that the unrestricted technology set provide a more favorable performance evaluation. Based on our restriction test, we can interpret this difference as the extent that Model 1 overestimates company efficiency. On average, the efficiency scores estimated

Table 6.3: Differences of efficiency estimates for Model 1 and Model 2

Difference between models	Min	Mean	Median	Max	Std.dev.
$\hat{\theta}_U - \hat{\theta}_R$	0.0000	0.0215	0.0000	0.1994	0.0415

using the restricted output set of Model 2 differ by 2.15 percent from the scores estimated using Model 1. A detailed analysis of the results reveals that 62 out of 171 observations are not affected by omitting *deliv* from the output set, i.e. the efficiency scores do not change (minimum difference equals zero). For 82 observations the change is positive and smaller than 5 percent while 15 observations exhibit difference between 5 percent and 10 percent. Only 12 observations have differences great than 10 percent, with a maximum of roughly 20 percent. This underlines the importance and necessity of defining technology set specifications based on reliable approaches as the one herein proposed.

6.4 Conclusions

This chapter develops an approach to determine a correct specification of a production possibility set in nonparametric efficiency analysis based on statistical inference for pooled data. A typical application is production units converting the same type of inputs to produce identical outputs using the same technology. This has a tradition in benchmarking exercises for industrial entities and, in particular, regulated network industries. DEA relies on few assumptions when determining the efficient technology. We provide a tool that allows statistical interference on the derived production possibility set. The basic idea is that from a number of potential outputs (or inputs) for the technology the analyst (even with technical knowledge) cannot be certain that the technology set is defined correctly.

This chapter uses three outputs, two of which are certain variables of the technology set. The third potential output is subject to some elements of uncertainty in terms of relevance for defining the technology. The estimated efficiency scores from nonparametric analyses are thus not reliable and there is a need to test the relevance of all output parameters. For this, we propose to estimate, for simplicity, two distinct models defining the technology set: Model 1 including all potential output variables and Model 2 testing an output-restricted technology. Following, nonparametric efficiency estimations are carried out for both models delivering an efficiency score each. We then test the difference between both statistically in order to decide on the correct technology specification. The expected value of the efficiency scores is the lagged ratio of both. Using the sample empirical mean of the ratio determines a test statistic for our testable hypothesis. Next and in order to derive an approximation of the sampling distribution of this test statistic, we advance the subsampling algorithm of Simar and Wilson (2001) and Schubert and Simar (2011) to an application to pooled data. For each model a critical value for any chosen level of significance is derived that allows a decision towards Model 1 or Model 2. Hence, the technology defined by DEA is specified correctly using statistical interference rather than expert knowledge, which will always be put under scrutiny by interest groups in real world applications.

Finally, we apply the proposed approach to US natural gas transmission pipelines where an unbalanced panel data set is available for 2003 through 2007. Having identified *O&M* as input for transporting natural gas, there are three potential output variables: line length, natural gas delivered and peak deliveries. First, we want to specify the production technology and are certain on the relevance of line length and peak delivery as output variables. With respect to the variable natural gas delivered, it is less clear as to whether the inclusion provides any value for determining the true technology. After outlier detection, 171 observations from an unbalanced panel from 191 companies over five periods remain. At a significance level of 95 percent we cannot reject the null hypothesis that only the two initial outputs determine the technology. The reliability of the estimated efficiency scores are improved significantly by reducing the number of output variables whilst leaving natural gas delivered out. In essence, the information asymmetry between the analyst and the production entity delivering the data and possibly being subject to regulatory benchmarking is removed using a sound and reproducible method.

Appendix

Consider the test-statistic proposed in Equation 6.12. Under H_0 the restricted model is consistent such that efficiency measures $\theta_{U_{it}}(X_i, Y_i, Z_i) \xrightarrow{p} \theta_{R_{it}}(X_i, Y_i, Z_i)$, where X and Y are random vectors of inputs and outputs and Z defines the random number of times an unit is observed in time. However, at the same time $\theta_{U_{it}}(X_i, Y_i, Z_i) \geq \theta_{R_{it}}(X_i, Y_i, Z_i)$ by construction. It follows that the term in Equation 6.12 asymptotes to 0, if and only if H_0 is true. But it remains positive for any finite sample size. This follows directly from the arguments made Schubert and Simar (2011, eq. 20), who propose consistent restriction tests in the case of ordinary cross-section data.

What complicates the subsampling procedure in this setting is that the observations come from a panel setting and are thus unlikely to be independently across the time dimension; even though they will be along the cross-section dimension. A robust approach is to subsample block-wise; compare Davison and Hinkley (1997). This allows for arbitrary dependence between the observations belonging to the same cross-section unit. We will now show that this procedure meets the essential consistency requirements set out in Politis et al. (2001). Let sample size n_p be defined by the number different cross-section observations.

Proposition: Let $n(Z) = \sum_{i=1}^{n_p} Z_i$ where are *iid* random variables Z_i with distribution function F_Z defined on the support $S_Z = 1, \ldots, L$ and expectation $c \in [1, L]$, then for the test-statistic $t_{n_p}(X, Y, Z)$ the asymptotic distribution of $\sqrt{n_p} n_p^{2/(p+q+1)} t_{n_p}(X, Y, Z)$ is non-degenerate with expectation zero.

Proof: Consider the term $t_i(X, Y \mid Z = z) = \frac{1}{z_i} \sum_{t=1}^{z_i} \left(\frac{\theta_{U_{it}}(X_i, Y_i, \mid Z_i = z_i)}{\theta_{R_{it}}(X_i, Y_i, \mid Z_i = z_i)} - 1 \right)$. It follows from the results of Kneip et al. (2008) that $n^{2/(p+q+1)} \left(\frac{\theta_{U_{it}}(X_i, Y_i, \mid Z_i = z_i)}{\theta_{R_{it}}(X_i, Y_i, \mid Z_i = z_i)} - 1 \right) \xrightarrow{d} H_n$, where H_n is a random variable with an asymptotic distribution function Q that is non-degenerate and has mean 0 under H_0. Furthermore we can rewrite $n = n_p \overline{z}$, where $\overline{z}$ is the empirical mean of Z. Replacing and rearranging yields $n_p^{2/(p+q+1)} \left(\frac{\theta_{U_{it}}(X_i, Y_i, \mid Z_i = z_i)}{\theta_{R_{it}}(X_i, Y_i, \mid Z_i = z_i)} - 1 \right) \xrightarrow{d} \frac{1}{\overline{z}^{2/(p+q+1)}} H_n \equiv D_n$ because $\overline{z}$ is, given $Z = z$, just a constant. Since D_n is a scaled version of H_n, also $n_p^{2/(p+q+1)} \left(\frac{\theta_{U_{it}}(X_i, Y_i, \mid Z_i = z)}{\theta_{R_{it}}(X_i, Y_i, \mid Z_i = z)} - 1 \right)$ has a non-degenerate distribution. This implies that the conditional distribution of $n_p^{2/(p+q+1)} t_i(X, Y, Z)$ is non-degenerate. Furthermore, we obtain $t_i(X, Y, Z)$ by marginalizing out Z as follows $t_i(X, Y, Z) = \int_{z \in S_Z} t_i(X, Y \mid Z = z)\, dF_Z$. This does not change anything about the convergence rate. Thus we conject that $n_p^{2/(p+q+1)} t_i(X, Y, Z)$ itself follows a non-degenerate distribution function. In order to complete the proof, since $t_{n_p}(X, Y, Z)$ is an empirical mean of the $t_i(X, Y, Z)$, it follows that $\tau_{n_p} t_{n_p(X,Y,Z)}$ with $\tau_{n_p} = \sqrt{n_p} n_p^{2/(p+q+1)}$ is non-degenerate and additionally has an asymptotic expectation equal to zero under H_0, because the mean associated with the asymptotic distribution Q is zero. As a consequence of this result, the subsampling methods proposed by Politis et al. (2001) are consistent, when subsampling is conducted block-wise along the cross-section dimension. The sub-sampling size m_p is as usual defined as the integer part of n_p^k for $0 < k < 1$. It should be noted that these results include the case of ordinary cross-section data and a balanced panel setting. In the former case $z_i = 1$ and $n = n_p$ yielding just the formulae in Schubert and Simar (2011). In the latter case $z_i = L$ implying $t_i(X, Y \mid Z) = t_i(X, Y, Z)$.

Bibliography

Adler, N. and Berechman, J. (2001). Measuring airport quality from the airlines' viewpoint: An application of Data Envelopment Analysis. *Transport Policy*, 8(3):171–181.

Adler, N. and Golany, B. (2001). Evaluation of deregulated airline networks using Data Envelopment Analysis combined with Principal Component Analysis with an application to Western Europe. *European Journal of Operational Research*, 132(2):260–273.

Adler, N. and Golany, B. (2002). Including principal component weights to improve discrimination in Data Envelopment Analysis. *Journal of the Operational Research Society*, 53(9):985–991.

Adler, N. and Yazhemsky, E. (2006). PCA_DEA. Website. `http://pluto.huji.ac.il/\~msnic/PCADEA.htm`, retrieved 15 June 2010.

Adler, N. and Yazhemsky, E. (2010). Improving discrimination in Data Envelopment Analysis: PCA-DEA or variable reduction. *European Journal of Operational Research*, 202(1):273–284.

Afonso, A. and Fernandes, S. (2006). Measuring local government spending efficiency: Evidence for the Lisbon region. *Regional Studies*, 40(1):39–53.

Afonso, A. and Fernandes, S. (2008). Assessing and explaining the relative efficiency of local government. *Journal of Socio-Economics*, 37(5):1946–1979.

Afonso, A., Schuknecht, L., and Tanzi, V. (2005). Public sector efficiency: An international comparison. *Public Choice*, 123(4-5):321–347.

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

Aitchison, J. and Aitken, C. G. G. (1976). Multivariate binary discrimination by the kernel method. *Biometrika*, 63(3):413–420.

Andersen, P. and Petersen, N. C. (1993). A procedure for ranking efficient units in Data Envelopment Analysis. *Management Science*, 39(10):1261–1664.

Andersen, S. C. (1983). The effect of government ownership and subsidy on performance: Evidence from the bus transit industry. *Transportation Research Part A*, 17(3):191–200.

Arcas, M.-J. and Bachiller, P. (2010). Operating performance of privatized firms in Europe: Organizational and environmental factors. *International Journal of Public Administration*, 33(10):487–498.

Arcelus, F. J., Arocena, P., Cabasés, F., and Pascual, P. (2007). On the efficiency of the delivery of municipal services. Working Paper Series DT 92/07, Departemento de Gestión de empresas, Universidad Pública de Navarra, Tudela.

Atkinson, P. and van den Noord, P. (2001). Managing public expenditure. Some emerging policy issues and a framework for anlaysis. OECD Economics Department Working Papers 285, OECD, Paris.

Balaguer-Coll, M. T., Prior, D., and Tortosa-Ausina, E. (2007). On the determinants of local government performance: A two-stage nonparametric approach. *European Economic Review*, 51(2):425–451.

Balaguer-Coll, M. T., Prior, D., and Tortosa-Ausina, E. (2010). Decentralization and efficiency of local government. *The Annals of Regional Science*, 45(3):571–601.

Balaguer-Coll, M. T., Prior-Jiménez, D., and Vela-Bargues, J. M. (2002). Efficiency and quality in local government management. The case of Spanish local authorities. Working Paper 2002/02, Department of Business Economics, Universitat Autònoma de Barcelona, Barcelona.

Banker, R. D. and Chang, H. (2006). The super-efficiency procedure for outlier identification, not for ranking efficient units. *European Journal of Operational Research*, 175(1):1311–1320.

Banker, R. D., Charnes, A., and Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis. *Management Science*, 30(9):1078–1092.

Banker, R. D. and Gifford, J. L. (1988). *A relative efficiency model for the evaluation of public health nurse productivity*. Carnegie Mellon University, Pitsburgh.

Battese, G. E. and Coelli, T. J. (1992). Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India. *Journal of Productivity Analysis*, 3(1-2):153–169.

Battese, G. E. and Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*, 20(2):325–332.

Beesley, M. and Littlechild, S. (1989). The regulation of privatized monopolies in the United Kingdom. *RAND Journal of Economics*, 20(3):454–472.

Benito, B., Bastida, F., and García, J. (2008). Explaining differences in efficiency: An application to Spanish municipalities. *Applied Economics*, 42(4):515–528.

Berechman, J. (1983). Costs, economies of scale and factor demand in bus transport. *Journal of Transport Economics and Policy*, 17(1):7–24.

Berechman, J. and Guiliano, G. (1985). Economies of scale in bus transit: A review of concepts and evidence. *Transportation*, 12(4):313–332.

Bhattacharyya, A., Kumbhakar, S. C., and Bhattacharyya, A. (1995). Ownership structure and cost efficiency: A study of publicly owned passenger-bus transportation companies in India. *Journal of Productivity Analysis*, 6(1):47–61.

Bly, P. H. and Oldfield, R. H. (1986). The effects of public transport subsidies on demand and supply. *Transportation Research Part A: General*, 20(6):415–427.

Bly, P. H., Webster, F. V., and Pounds, S. (1980). Effects of subsidies on urban public transport. *Transportation*, 9(4):311–331.

Bogetoft, P. and Agrell, P. (2007). Development of benchmarking models for German electricity and gas distribution. Sumicsid AB, Final Report, Project No. 730.

Bogetoft, P. and Otto, L. (2011). *Benchmarking with DEA, SFA, and R*. International Series in Operations Research & Management Science. Springer, New York.

Bradford, D. F., Malt, R. A., and Oates, W. E. (1969). The rising cost of local public services: Some evidence and reflections. *National Tax Journal*, 22(2):185–202.

Bădin, L., Daraio, C., and Simar, L. (2010). Optimal bandwidth selection for conditional efficiency measures: A data-driven approach. *European Journal of Operational Research*, 201(2):633–640.

Bundesnetzagentur (2006). Bericht der Bundesnetzagentur nach §112a EnWG zur Einführung der Anreizregulierung nach §21a ENWG. Website. `http://www.zfk.de/cms/Infothek/Regulierung/Regulierung_/Anreiz-Endbericht_30_06_06.pdf`, retrieved 02 November 2009.

Burns, P., Jenkins, C., and Riechmann, C. (2005). The role of benchmarking for yardstick competition. *Utilities Policy*, 13(4):302–309.

Button, K. J. (1993). *Transport Economics*. Edward Elgar Publishing, Aldershot.

Cambini, C., Piacenza, M., and Vanonni, D. (2007). Restructuring public transit systems: Evidence on cost properties from medium and large-sized companies. *Review of Industrial Organization*, 31(3):183–203.

Caudill, S. B., Ford, J. M., and Gropper, D. M. (1995). Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity. *Journal of Business & Economic Statistics*, 13(1):105–111.

Caves, D. W. and Christensen, L. R. (1980). The relative efficiency of public and private firms in a competitive environment: The case of Canadian railroads. *Journal of Political Economy*, 88(5):958–976.

Cazals, C., Florens, J.-P., and Simar, L. (2002). Nonparametric frontier estimation: A robust approach. *Journal of Econometrics*, 106(1):1 – 25.

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.

Coelli, T., Estache, A., Perelman, S., and Trujillo, L. (2003). A primer on efficiency measurement for utilities and transport regulators. World Bank Institute, Development Studies.

Coelli, T., Perelman, S., and Romano, E. (1999). Accounting for environmental influences in stochastic frontier models: With application to international airlines. *Journal of Productivity Analysis*, 11(3):251–273.

Coelli, T. J., Rao, D. P., O'Donnell, C. J., and Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*. Springer, New York, 2nd edition.

Coelli, T. J. and Walding, S. (2006). Performance measurement in the Australian water supply industry: A preliminary analyis. In Coelli, T. J. and Lawrence, D., editors, *Performance measurement and regulation of network industries*, pages 29–62. Edward Elgar Publishing, Cheltenham.

Cole, A. (2006). Decentralization in France: Central steering, capacity building and identity construction. *French Politics*, 4(1):31–57.

Cooper, W. W., Seiford, L. M., and Tone, K. (2006). *Introduction to Data Envelopment Analysis and its uses: With DEA solver software and references.* Springer, New York.

Cooper, W. W., Seiford, L. M., and Tone, K. (2007). *Data Envelopment Analysis - A comprehensive text with models, applications, references and DEA-solver software.* Springer, New York, 2$^{\text{nd}}$ edition.

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*, International Series in Operations Research & Management Science, Volume 164, pages 1–39. Springer, New York, 2$^{\text{nd}}$ edition.

Cullmann, A. (2012). Benchmarking and firm heterogeneity: A latent class analysis for German electricity distribution companies. *Empirical Economics*, 42(1):147–169.

Daraio, C. and Simar, L. (2005). Introducing environmental variables in nonparametric frontier models: A probabilistic approach. *Journal of Productivity Analysis*, 24(1):93–121.

Daraio, C. and Simar, L. (2007a). *Advanced robust and nonparametric methods in efficiency analysis: Methodology and applications.* Studies in productivity and efficiency. Springer, New York.

Daraio, C. and Simar, L. (2007b). Conditional nonparametric frontier models for convex and nonconvex technologies: A unifying approach. *Journal of Productivity Analysis*, 28(1):13–32.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap methods and their application.* Cambridge University Press, Cambridge.

De Borger, B. and Kerstens, K. (1996a). Cost efficiency of Belgian local governments: A comparative analysis of FDH, DEA, and econometric approaches. *Regional Science and Urban Economics*, 26(2):145–170.

De Borger, B. and Kerstens, K. (1996b). Radial and nonradial measures of technical efficiency: An empirical illustration for Belgian local governments using an FDH reference technology. *Journal of Productivity Analysis*, 7(1):41–62.

De Borger, B. and Kerstens, K. (2000). What is known about municipal efficiency? the Belgian case and beyond. In Blank, J. L. T., editor, *Public provision and performance*, pages 299–330. North-Holland, Amsterdam.

De Borger, B. and Kerstens, K. (2008). The performance of bus-transit operators. In Hensher, D. A. and Button, K. J., editors, *Handbook of transport modelling*, pages 693–714. Elsevier, Amsterdam, 2nd edition.

De Borger, B., Kerstens, K., and Costa, A. (2002). Public transport performance: What does one learn from frontier studies? *Transport Reviews*, 22(1):1–38.

De Borger, B., Kerstens, K., Moesen, W., and Vanneste, J. (1994). Explaining differences in productive efficiency: An application to Belgian municipalities. *Public Choice*, 80(3-4):339–358.

De Witte, K. and Geys, B. (2011). Evaluating efficient public good provision: Theory and evidence from a generalised conditional efficiency model for public libraries. *Journal of Urban Economics*, 69(3):319–327.

De Witte, K. and Kortelainen, M. (2009). Blaming the exogenous environment? Conditional efficiency estimation with continuous and discrete exogenous variables. MPRA Paper No. 14034, University Library of Munich, Munich.

De Witte, K. and Kortelainen, M. (2013). What explains the performance of students in a heterogeneous environment? Conditional efficiency estimation with continuous and discrete environmental variables. *Applied Economics*, 45(17):2401–2412.

De Witte, K. and Marques, R. C. (2011). Big and beautiful? On non-parametrically measuring scale economies in non-convex technologies. *Journal of Productivity Analysis*, 35(3):213–226.

Debreu, G. (1951). The coefficient of resource utilization. *Econometrica*, 19(3):273–292.

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 measurements*, pages 243–267. North-Holland, Amsterdam.

Deutsche Bundesbank (2010). Monatsbericht Januar 2010. Website. `http://www.bundesbank.de/Redaktion/DE/Downloads/Veroeffentlichungen/Monatsberichte/2010/2010_01_monatsbericht.pdf?__blob=publicationFile`, retrieved 10 February 2010.

Deutscher Städtetag (2008). Statistisches Jahrbuch der Gemeinden. 93. Jahrgang 2006, Deutscher Städtetag, Berlin.

DGCL (2010). Les collectivités locales en chiffres 2010. Direction Générale des Collectivités Locales, Paris.

Djankov, S. and Murrell, P. (2002). Enterprise restructuring in transition: A quantitative survey. *Journal of Economic Literature*, 40(3):739–792.

Dyson, R., Allen, R., Camanho, A., Podinovski, V., Sarrico, C., and Shale, E. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research*, 132(2):245–259.

Edvardsen, D. F. (2004). *Four essays on the measurement of productive efficiency.* PhD thesis, University of Göteborg, Department of Economics, School of Economics and Law.

Estrin, S., Hanousek, J., Kočenda, E., and Svejynar, J. (2009). The effects of privatization and ownership in transition economies. *Journal of Economic Literature*, 47(3):699–728.

European Commission (2010). Public finances in EMU 2010. Technical report, Directorate-General for Economic and Financial Affairs, European Commission, Brussels.

Färe, R. and Grosskopf, S. (2005). *New directions: efficiency and productivity.* Studies in productivity and efficiency. Kluwer Academic Publishers, Boston.

Färe, R., Grosskopf, S., Lindgren, B., and Roos, P. (1992). Productivity changes in Swedish pharmacies 1980-1989: A non-parametric Malmquist approach. *Journal of Productivity Analysis*, 3(1-2):85–101.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120(3):253–290.

Farsi, M., Fetz, A., and Filippini, M. (2007). Benchmarking and regulation in the electricity distribution sector. CEPE Working paper series 07-54, CEPE Center for Energy Policy and Economics, ETH Zurich, Zurich.

Farsi, M. and Filippini, M. (2009). An analysis of cost efficiency in Swiss multi-utilities. *Energy Economics*, 31(2):306–315.

Farsi, M., Filippini, M., and Greene, W. H. (2005a). Efficiency measurement in network industries: Application to the Swiss railway companies. *Journal of Regulatory Economics*, 28(1):69–90.

Farsi, M., Filippini, M., and Greene, W. H. (2006a). Application of panel data models in benchmarking analysis of the electricity distribution sector. *Annals of Public and Cooperative Economics*, 77(3):271–290.

Farsi, M., Filippini, M., and Kuenzle, M. (2005b). Unobserved heterogeneity in stochastic frontier models: An application to Swiss nursing homes. *Applied Economics*, 37(18):2127–2141.

Farsi, M., Filippini, M., and Kuenzle, M. (2006b). Cost efficiency in regional bus companies: An application of alternative stochastic frontier models. *Journal of Transport Economics and Policy*, 40(1):95–118.

FERC (2008). FERC financial report. FERC Form No. 2: Annual report of major natural gas companies and supplemental form 3-Q: Quarterly financial report. Website. `http://www.ferc.gov/docs-filing/forms/form-2/form-2.pdf`, retrieved 12 May 2008.

Filippini, M., Maggi, R., and Prioni, P. (1992). Inefficiency in a regulated industry - The case of the Swiss regional bus companies. *Annals of Public and Cooperative Economics*, 63(3):437–454.

Filippini, M. and Prioni, P. (2003). The influence of ownership on the cost of bus service provision in Switzerland - An empirical illustration. *Applied Economics*, 35(6):683–690.

Fletcher, J. and Snee, H. (1985). The need for output measurements in the service industries: A comment. *The Service Industries Journal*, 5(1):73–78.

Fried, H., Schmidt, S. S., and Yaisawarng, S. (1999). Incorporating the operating environment into a nonparametric measure of technical efficiency. *Journal of Productivity Analysis*, 12(3):249–267.

Fritsch, M., Wein, T., and Ewers, H.-J. (2006). *Marktversagen und Wirtschaftspolitik*. Verlag Vahlen, München, 6th edition.

Geys, B., Heinemann, F., and Kalb, A. (2008). Local governments in the wake of demographic change: Evidence from German municipalities. *FinanzArchiv: Public Finance Analysis*, 64(4):434–457.

Geys, B., Heinemann, F., and Kalb, A. (2010). Voter involvement, fiscal autonomy and public sector efficiency: Evidence from German municipalities. *European Journal of Political Economy*, 26(2):265–278.

Geys, B. and Moesen, W. (2009). Measuring local government technical (in)efficiency: An application and comparison of FDH, DEA and econometric approaches. *Public Performance and Management Review*, 32(4):489–504.

Giménez, V. M. and Prior, D. (2007). Long- and short-term cost efficiency frontier evaluation: Evidence from Spanish local governments. *Fiscal Studies*, 28(1):121–139.

Giordano, R. and Tommasino, P. (2011). Public sector efficiency and political culture. Temi di discussione 786, Bank of Italy, Roma.

Goeverden, C. v., Rietveld, P., Koelemeijer, J., and Peeters, P. (2006). Subsidies in public transport. *European Transport*, 32(1):5–25.

Gordon, D. V., Gunsch, K., and Pawluk, C. V. (2003). A natural monopoly in natural gas transmission. *Energy Economics*, 25(5):473–485.

Gore, C. (2000). The rise and fall of the Washington Consensus as a paradigm for developing countries. *World Development*, 28(5):789–804.

Gorman, M. F. and Ruggiero, J. (2008). Evaluating US state police performance using Data Envelopment Analysis. *International Journal of Production Economics*, 113(2):1031–1037.

Greene, W. H. (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. H. (2005a). Fixed and random effects in stochastic frontier models. *Journal of Productivity Analysis*, 23(1):7–32.

Greene, W. H. (2005b). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics*, 126(2):269–303.

Greene, W. H. (2007). *LIMDEP Version 9.0 Econometric Modeling Guide Vol. 2.* Econometric Software Inc., Plainview.

Greene, W. H. (2008). The econometric approach to efficiency analysis. In Fried, H. O., Lovell, C. K., and Schmidt, S. S., editors, *The measurement of productive efficiency and productivity growth*, pages 421–521. Oxford University Press, Oxford.

Hadri, K. (1999). Estimation of a double heteroscedastic stochastic frontier cost function. *Journal of Business & Economic Statistics*, 17(3):359–363.

Hadri, K., Guermat, C., and Whittaker, J. M. (2003). Estimating farm efficiency in the presence of double heteroscedasticity using panel data. *Journal of Applied Economics*, 11(2):255–268.

Haney, A. B. and Pollitt, M. G. (2009). Efficiency analysis of energy networks: An international survey of regulators. *Energy Policy*, 37(12):5814–5830.

Härdle, W. and Simar, L. (2003). *Applied multivariate statistical analysis.* Springer, Berlin.

Harvey, D. (2007). *A brief history of neoliberalism.* Oxford University Press, Oxford.

Hauner, D. (2008). Explaining differences in public sector efficiency: Evidence from Russia's regions. *World Development*, 36(10):1745–1765.

Hauner, D. and Kyobe, A. (2010). Determinants of government efficiency. *World Development*, 38(11):1527–1542.

Hayfield, T. and Racine, J. S. (2008). Nonparametric econometrics: The np package. *Journal of Statistical Software*, 27(5):1–32.

Hirschhausen, C. v., Neumann, A., and Rüster, S. (2007). Wettbewerb im Ferntransport von Erdgas? Technisch-ökonomische Grundlagen und Anwendung auf Deutschland. *Zeitschrift für Energiewirtschaft*, 31(3):183–193.

Hitzschke, S. (2011). The optimal size of German cities an efficiency analysis perspective. Darmstadt Discussion Papers in Economics 202, Darmstadt Technical University, Darmstadt.

Hjalmarsson, L., Kumbhakar, S. C., and Heshmati, A. (1996). DEA, DFA and SFA: A comparison. *Journal of Productivity Analysis*, 7(2-3):303–327.

Hollas, D. R., Macloed, K. R., and Stansell, S. R. (2002). A Data Envelopment Analysis of gas utilities' efficiency. *Journal of Economics and Finance*, 26(2):123–137.

Homburg, C. (2001). Using Data Envelopment Analysis to benchmark activities. *International Journal of Production Economics*, 73(1):51 – 58.

IEA (2003). The challenges of future cost reductions for new supply options (pipelines, LNG, GTL). 22nd World Gas Congress Tokyo, website. `http://www.dma.dk/themes/LNGinfrastructureproject/Documents/Infrastructure/IEA-The\%20challenges\%20of\%20further\%20cost\%20red\%20new\%20supply\%20options.pdf`, retrieved 26 September 2011.

Jamasb, T., Newbery, D., Pollitt, M., and Triebs, T. (2007). International benchmarking and regulation of European gas transmission utilities. Report prepared for the Council of European Energy Regulators (CEER), Cambridge.

Jamasb, T., Nillesen, P., and Pollitt, M. (2004). Strategic behaviour under regulatory benchmarking. *Energy Economics*, 26(5):825–843.

Jamasb, T. and Pollitt, M. G. (2001). Benchmarking and regulation: International electricity experience. *Utilities Policy*, 9(3):107–130.

Jamasb, T. and Pollitt, M. G. (2003). International benchmarking and yardstick regulation: An application to European electricity distribution utilities. *Energy Policy*, 31(15):1609–1622.

Jamasb, T., Pollitt, M. G., and Triebs, T. (2008). Productivity and efficiency of US gas transmission companies: A European regulatory perspective. *Energy Policy*, 36(9):3398–3412.

Jeong, S.-O., Park, B., and Simar, L. (2010). Nonparametric conditional efficiency measures: Asymptotic properties. *Annals of Operations Research*, 173(1):105–122.

Johnston, J. (1956). Scale, costs and profitability in road passenger transport. *Journal of Industrial Economics*, 4(3):207–223.

Jondrow, J., Lovell, K. C. A., 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. J. (2007). Regulation of natural monopolies. In Polinsky, A. M. and Shavell, S., editors, *Handbook of law and economics*, volume 2, pages 1227–1348. North-Holland, Amsterdam.

Kalb, A. (2010a). The impact of intergovernmental grants on cost efficiency: Theory and evidence from German municipalities. *Economic Analysis and Policy*, 40(1):23–48.

Kalb, A. (2010b). *Public sector efficiency - Applications to local governments in Germany*. Gabler Verlag, Wiesbaden.

Karlaftis, M. G. and McCarthy, P. (1998). Operating subsidies and performance in public transit: An empirical study. *Transportation Research Part A: Policy and Practice*, 32(5):359–375.

Karlaftis, M. G. and McCarthy, P. (2002). Cost structures of public transit systems: A panel data analysis. *Transportation Research Part E*, 38(1):1–18.

Kellermann, K. (2007). Die öffentlichen Ausgaben der Kantone und ihrer Gemeinden im Quervergleich. Strukturberichterstattung No. 37, Eidgenössisches Volkswirtschaftsdepartement EVD, Staatssekretariat für Wirtschaft SECO, Direktion für Wirtschaftspolitik, Schweizerische Eidgenossenschaft, Bern.

Kerstens, K. (1996). Technical efficiency measurement and explanation of French urban transit companies. *Transportation Research Part A*, 30(6):431–452.

Kneip, A., Park, B. U., and Simar, L. (1998). A note on the convergence of nonparametric DEA estimators for production efficiency scores. *Econometric Theory*, 14(6):783–793.

Kneip, A., Simar, L., and Wilson, P. W. (2008). Asymptotics and consistent bootstraps for DEA estimators in nonparametric frontier models. *Econometric Theory*, 24(06):1663–1697.

Koopmans, T. C. (1951). An analysis of of production as an efficient combination of activities. In Koopmans, T. C., editor, *Activity analysis of production* *and allocation, cowles commission for research in economics*, pages 33–97. Monograph No. 13, Wiley, New York.

Kriese, M. (2008). Effizienzanalyse der Sächsischen Gemeinden. ifo Dresden berichtet 15 (5), CES ifo Group, Dresden.

Kuczynski, P.-P. and Williamson, J. (2003). *After the Washington Consensus: Restarting Growth and Reform in Latin America*. Number 350 in Peterson Institute Press: All Books. Peterson Institute for International Economics.

Kumbhakar, S. C. (1990). Production frontiers, panel data, and time-varying technical inefficiency. *Journal of Econometrics*, 46(1-2):201–211.

Kumbhakar, S. C. (1997). Efficiency estimation with heteroscedasticity in a panel data model. *Applied Economics*, 29(3):379–386.

Kumbhakar, S. C. and Lovell, K. C. A. (2000). *Stochastic Frontier Analysis*. Cambridge University Press, Cambridge, 1^{st} edition.

Li, Q. and Racine, J. S. (2004). Cross-validated local linear nonparametric regression. *Statistica Sinica*, 14(3):485–510.

Li, Q. and Racine, J. S. (2007). *Nonparametric econometrics: Theory and practice*. Princeton University Press, Princeton.

Li, Q. and Racine, J. S. (2008). Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. *Journal of Business and Economic Statistics*, 26(4):423–434.

Littlechild, S. C. (1983). Regulation of British telecommunications' profitability. Report to the Secretary of State, Department of Industry in London, London.

Loikkanen, H. A. and Susiluoto, I. (2006). Cost efficiency of Finnish municipalities in basic service provision. HECER Discussion Paper 96, HECER - Helsinki Center of Economic Research, Helsinki.

Loikkanen, H. A., Susiluoto, I., and Funk, M. (2011). The role of city managers and external variables in explaining efficiency differences of Finnish municipalities. HECER Discussion Paper 312, HECER - Helsinki Center of Economic Research, Helsinki.

Lonti, Z. and Woods, M. (2008). Towards government at a glance: Identification of core data and issues related to public sector efficiency. OECD Working Papers on Public Governance 7, OECD Publishing, Paris.

Marques, R. C. and De Witte, K. (2011). Is big better? On scale and scope economies in the Portuguese water sector. *Economic Modelling*, 28(3):1009–1016.

Marsh, D. (1991). Privatization under Mrs. Thatcher: A review of the literature. *Public Administration*, 69(4):459–480.

Maudos, J., Pastor, J. M., and Serrano, L. (2003). Human capital in OECD countries: Technical change, efficiency and productivity. *International Review of Applied Economics*, 17(4):419–435.

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.

Megginson, W. L. and Sutter, N. L. (2006). Privatisation in developing countries. *Corporate Governance: An International Review*, 14(4):234–265.

Miller, D. R. (1970). Differences among cities, differences among firms, and costs of urban bus transport. *The Journal of Industrial Economics*, 19(1):22–32.

Montén, A. and Thater, C. (2011). Determinants of efficiency in child-care provision. *FinanzArchiv: Public Finance Analysis*, 67(4):378–403.

Mühlenkamp, H. (2012). Zur relativen (In-)Effizienz öffentlicher (und privater) Unternehmen - Untersuchungsziele, Effizienzmaße und empirische Befunde. In *Renaissance öffentlicher Unternehmen - Bestandsaufnahme - Kontexte - Perspektiven*. Nomos, Baden-Baden.

Müller, D. C. (2003). *Public Choice III*. Cambridge University Press, Cambridge, 3rd edition.

Natgas.info (2011). Gas pipelines. Website. `http://natgas.info/html/gaspipelines.html`, retrieved 26 September 2011.

Netter, J. M. and Megginson, W. L. (2001). From state to market: A survey of empirical studies on privatization. *Journal of Economic Literature*, 39(2):321–389.

Nierhaus, M. (2006). Die kommunale Selbstverwaltung. In Büchner, C. and Franzke, J., editors, *Komunale Politik und Verwaltung*. Universitäsverlag Potsdam.

Nieswand, M., Neumann, A., and Cullmann, A. (2010). Overcoming data limitations in nonparametric benchmarking: Applying PCA-DEA to natural gas transmission. *Review of Network Economics*, 9(2):1–24.

Nieswand, M. and Seifert, S. (2011). Some determinants of intermediate local governments' spending efficiency: The case of French départements. DIW Berlin Discussion Paper No. 1130, DIW Berlin, Berlin.

Nieswand, M. and Walter, M. (2010). Cost efficiency and subsidization in German local public bus transit. Discussion Paper 1071, DIW Berlin, Berlin.

Obeng, K. (1994). The economic cost of subsidy-induced technical inefficiency. *International Journal of Transportation Economics*, 21(1):3–20.

OECD (2003). *Economic Surveys France*, volume 2003/11. OECD Publishing, Paris.

OECD (2007). *Economic Surveys France*, volume 2007/13. OECD Publishing, Paris.

OECD (2009a). *Economic Surveys France*, volume 2009/5. OECD Publishing, Paris.

OECD (2009b). Governance at a glance 2009, country note: France. Website. `http://www.oecd.org/dataoecd/35/42/43925776.pdf`, retrieved 02 May 2011.

OECD (2010a). *Economic Factbook 2010*. OECD Publishing, Paris.

OECD (2010b). *Economic Outlook 2010*. OECD Publishing, Paris.

O'Neill, R. P. (2005). Natural gas pipelines. In L.Moss, D., editor, *Network access, regulation and antitrust*, pages 107–120. Routledge, London.

Pagan, A. and Ullah, A. (1999). *Nonparametric Econometrics*. Cambridge University Press, Cambridge.

Park, B., Simar, L., and Weiner, C. (2000). The FDH estimator for productivity efficiency scores: Asymptotic properties. *Econometric Theory*, 16(6):855–877.

Park, B. U., Jeong, S.-O., and Lee, Y. K. (2006). Nonparametric estimation of production efficiency. In Fan, J. and Koul, H. L., editors, *Frontiers in statistics*, pages 407–428. Imperial College Press, London.

Pastor, J. T. (1996). Translation invariance in Data Envelopment Analysis: A generalization. *Annals of Operations Research*, 66(2):91–102.

Pastor, J. T., Ruiz, J. L., and Sirvent, I. (1999). A statistical test for detecting influential observations in DEA. *European Journal of Operational Research*, 115(3):542–554.

Pestieau, P. (2009). Assessing the performance of the public sector. *Annals of Public and Cooperative Economics*, 80(1):133–161.

Pestieau, P. and Tulkens, H. (1993). Assessing and explaining the performance of public enterprises. *Finanzarchiv*, 50(2):293–323.

Piacenza, M. (2001). Productive structure, cost efficiency and incentives in the local public transport: A survey of theoretical and empirical issues. Working Paper 1, HERMES, Moncalieri.

Pickrell, D. H. (1985). Rising deficits and the uses of transit subsidies in the United States. *Journal of Transport Economics and Policy*, 19(3):281–298.

Pitt, M. M. and Lee, L.-F. (1981). The measurement of sources of technical inefficiency in Indonesian weaving industry. *Journal of Development Economics*, 9(1):43–64.

Podinovski, V. V. and Thanassoulis, E. (2007). Improving discrimination in Data Envelopment Analysis: Some practical suggestions. *Journal of Productivity Analysis*, 28(1-2):117–126.

Politis, D. N., Romano, J. P., and Wolf, M. (2001). On the asymptotic theory of subsampling. *Statistica Sinica*, 11(4):1105 –1124.

Pommerehne, W. W. (1990). Genügt bloßes Privatisieren? In Aufderheide, D., editor, *Deregulierung und Privatisierung*, Symposium Oeconomicum Münster, pages 27–63. Kohlhammer, Stuttgart.

Pucher, J. (1988). Urban public transport subsidies in Western Europe and North America. *Transportation Quarterly*, 42(3):377–402.

Racine, J. (2008). Nonparametric econometrics: A primer. *Foundations and Trends® in Econometrics*, 3(1):1–88.

Racine, J. and Li, Q. (2004). Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics*, 119(1):99–130.

Racine, J. S. (1997). Consistent significance testing for nonparametric regression. *Journal of Business and Economic Statistics*, 15(3):369–379.

Racine, J. S., Hart, J., and Li, Q. (2006). Testing the significance of categorical predictor variables in nonparametric regression models. *Econometric Reviews*, 25(4):523–544.

Riechmann, C. and Rodgarkia-Dara, A. (2006). Regulatorisches Benchmarking - Konzeption und praktische Interpretation. *Zeitschrift füer Energiewirtschaft*, 30(3):205–219.

Sakano, R. and Obeng, K. (1995). Re-examination of inefficiencies in urban transit systems: A stochastic frontier approach. *Logistics and Transportation Review*, 31(4):377–392.

Sakano, R., Obeng, K., and Azam, G. (1997). Subsidies and inefficiency: Stochastic frontier approach. *Contemporary Economic Policy*, 15(3):113–127.

Sampaio de Sousa, M. and Stošić, B. (2005). Technical efficiency of the Brazilian municipalities: Correcting nonparametric frontier measurements for outliers. *Journal of Productivity Analysis*, 24(3):157–181.

Schefczyk, M. (1993). Industrial benchmarking: A case study of performance analysis techniques. *International Journal of Production Economics*, 32(1):1–11.

Schmidt, P. and Lovell, K. C. A. (1979). Estimating technical and allocative inefficiency relative to stochastic production and cost frontiers. *Journal of Econometrics*, 9(3):343–366.

Schmidt, P. and Lovell, K. C. A. (1980). Estimating stochastic production and cost frontiers when technical and allocative inefficiency are correlated. *Journal of Econometrics*, 13(1):83–100.

Schmidt, P. and Sickles, R. C. (1984). Production frontiers and panel data. *Journal of Business & Economic Statistics*, 2(4):367–374.

Schubert, T. and Simar, L. (2011). Innovation and export activities in the German mechanical engineering sector: An application of testing restrictions in production analysis. *Journal of Productivity Analysis*, 36(1):55–69.

Seitz, H. (2008). Die Demographieabhängigkeit der Ausgaben und Einnahmen der öffentlichen Haushalte. Bertelsmann Stiftung, Gütersloh, website. `http://www.bertelsmann-stiftung.de/bst/de/media/xcms_bst_dms_23806_23829_2.pdf`.

Sharkey, W. W. (1982). *The theory of natural monopoly.* Cambridge University Press, Cambridge.

Shepard, R. W. (1970). *Theory of cost and production function.* Princeton University Press, Princeton.

Shleifer, A. (1985). A theroy of yardstick regulation. *RAND Journal of Economics*, 16(3):319–327.

Shuttleworth, G. (2003). Firm-specific productive efficiency: A response. *The Electricity Journal*, 16(3):42–50.

Shuttleworth, G. (2005). Benchmarking of electricity networks: Practical problems with its use for regulation. *Utilities Policy*, 13(4):310–317.

Sickles, R. C. and Streitwieser, M. L. (1992). Technical inefficiency and productivity decline in the U.S. interstate natural gas pipeline industry under the national gas policy act. *The Journal of Productivity Analysis*, 3(1-2):119–133.

Sickles, R. C. and Streitwieser, M. L. (1998). An analysis of technology, productivity, and regulatory distortion in the interstate natural gas transmission industry: 1977-1985. *Journal of Applied Econometrics*, 13(4):377–395.

Simar, L. (2003). Detecting outliers in frontier models: A simple approach. *Journal of Productivity Analysis*, 20(3):391–424.

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. (2000). Statistical inference in nonparametric frontier models: The state of the art. *Journal of Productivity Analysis*, 13(1):49–78.

Simar, L. and Wilson, P. W. (2001). Testing restrictions in nonparametric efficiency models. *Communications in Statistic: Simulation and Computation*, 30(1):159 – 184.

Simar, L. and Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, 136(1):31–64.

Simar, L. and Wilson, P. W. (2008). Statistical inference in nonparametric frontier models: Recent developments and perspectives. In Fried, H. O., Lovell, C. K., and Schmidt, S. S., editors, *The measurement of productive efficiency and productivity growth*, pages 421–521. Oxford University Press, Oxford.

Simar, L. and Wilson, P. W. (2011). Inference by the m out of n bootstrap in nonparametric frontier models. *Journal of Productivity Analysis*, 36(1):33–53.

Statistisches Bundesamt (2007a). Statistik lokal. Daten für die Gemeinden, Städte, Kreise Deutschland Ausgabe 2007, Statistisches Bundesamt, Wiesbaden.

Statistisches Bundesamt (2007b). *Statistisches Jahrbuch 2007*. Statistisches Bundesamt, Wiesbaden.

Statistisches Bundesamt (2008). Finanzen und Steuern, Jahresrechnungsergebnis der kommunalen Haushalte 2006. Fachserie 14 Reihe 3.3, Statistisches Bundesamt, Wiesbaden.

Statistisches Bundesamt (2009). Preise und Preisindizes für gewerbliche produkte (Erzeugerpreise) 2008. Fachserie 17 Reihe 2, Statistisches Bundesamt, Wiesbaden.

Statistisches Bundesamt (2010). Finanzen und Steuern, Jahresrechnungsergebnis der kommunalen Haushalte 2008. Fachserie 14 Reihe 3.3, Statistisches Bundesamt, Wiesbaden.

Statistisches Bundesamt (2011). Finanzen und Steuern, Jahresrechnungsergebnis der kommunalen Haushalte 2009. Fachserie 14 Reihe 3.3, Statistisches Bundesamt, Wiesbaden.

Stiglitz, J. E. (1999). More instruments and broader goals: Moving toward the post-Washington Consensus. *Revista de Economica Politica*, 19(1 (73)):94–120.

Stiglitz, J. E. (2000). *Economics of the public sector*. W. W. Norton, New York, 3rd edition.

Tanaka, T. (2006). Cost efficiency in Japanese local governments: the economic effect of information technology in Japanese local governments. ERSA conference papers ersa06p234, European Regional Science Association, Louvain-la-Neuve.

Thompson, R. G., Dharmapala, P. S., and Thrall, R. M. (1995). Linked-cone dea profit ratios and technical efficiency with application to illinois coal mines. *International Journal of Production Economics*, 39(1-2):99–115.

Tibout, C. M. (1956). A pure theory of local expenditures. *Journal of Political Economy*, 64(5):416–424.

Ueda, T. and Hoshiai, Y. (1997). Application of Principal Component Analysis for parsimonious summarization of DEA inputs and/ or outputs. *Journal of the Operations Research Society of Japan*, 40(4):466–478.

Umweltbundesamt (2003a). Umweltpolitische Handlungsempfehlungen für die Finanzierung des ÖPNV - Band A. Website. `http://www.umweltdaten.de/publikationen/fpdf-l/2322.pdf`, retrieved 25 November 2009.

Umweltbundesamt (2003b). Umweltpolitische Handlungsempfehlungen für die Finanzierung des ÖPNV - Band B. Website. `http://www.umweltdaten.de/publikationen/fpdf-l/2323.pdf`, retrieved 25 November 2009.

Van Klaveren, C. and De Witte, K. (2010). How are teachers teaching? A nonparametric approach. TIER Working Paper Series TIER WP 10/23, Top Institute for Evidence Based Education Research, Amsterdam.

Vanden Eeckaut, P., Tulkens, H., and Jamar, M.-A. (1993). Cost efficiency in Belgian municipalities. In Fried, H. O., Lovell, C. K., and Schmidt, S. S., editors, *The measurement of productive efficiency: Techniques and applications*, pages 300–334. Oxford University Press, New York, 1st edition.

Verband Deutscher Verkehrsunternehmen (2009a). Finanzierungsbedarf des ÖPNV bis 2025. Website. `http://www.staedtetag.de/imperia/md/content/pressedien/2009/9.pdf`, retrieved 28 October 2009.

Verband Deutscher Verkehrsunternehmen (2009b). *VDV Statistik 2008*. Verband Deutscher Verkehrsunternehmen, Cologne.

Vickrey, W. (1980). Optimal transit subsidy policy. *Transportation*, 9(4):389–409.

Villalonga, B. (2000). Privatization and efficiency: Differentiating ownership effects from political, organizational, and dynamic effects. *Journal of Economic Behavior & Organization*, 42(1):43–74.

Viton, P. A. (1981). A translog cost function for urban bus transit. *Journal of Industrial Economics*, 29(3):287–304.

Vogelsang, I. (2002). Incentive regulation and competition in public utility markets: A 20-year perspective. *Journal of Regulatory Economics*, 22(1):5–27.

Walter, M. (2011). Some determinants of cost efficiency in German public transport. *Journal of Transport Economics and Policy*, 45(1):1–20.

Wehling, H.-G. (2006). Kommunalpolitik. Information zur politischen Bildung Nr. 242/2006, Bundeszentrale für politische Bildung, Bonn.

Williamson, J. (1990). What Washington means by policy reform. Peterson Institute for International Economics, Washington D.C.

Willner, J. (2001). Ownership, efficiency, and political interference. *European Journal of Political Economy*, 17(4):723–748.

Wilson, P. (2008). FEAR 1.0: A software package for frontier efficiency analysis with R. *Socio-Economic Planning Sciences*, 42(4):247–254.

Wolfe, J. (1991). State power and ideology in Britain: Mrs. Thatcher's privatization programme. *Political Studies*, 39(2):237–252.

Worthington, A. (1999). Performance indicators and efficiency measurement in public libraries. *Australian Economic Review*, 32(1):31–42.

Worthington, A. and Dollery, B. (2000). Efficiency measurement in the local public sector: Econometric and mathematical programming frontier techniques. School of Economics and Finance Discussion Papers and Working Papers Series 078, School of Economics and Finance, Queensland University of Technology, Brisbane.

Worthington, A. C. (2000). Cost efficiency in australian local government: A comparative analysis of mathematical programming and econometric approaches. *Financial Accountability and Management*, 16(3):201–224.

Yarrow, G. (1986). Privatization in theory and practice. *Economic Policy*, 1(2):324–377.

Zelenyuk, V. and Simar, L. (2011). To smooth or not to smooth? Discrete variables in nonparametric regressions. CEPA Working Papers Series WP10/2011, School of Economics, University of Queensland, Brisbane.

Zschille, M. (2012). Consolidating the water industry: An analysis of the potential gains from horizontal integration in a conditional efficiency framework. CEPR Discussion Papers 8737, CEPR, London.