Essays in supply and demand in energy economics: A frontier analysis approach

Manuel Llorca Riego
Department of Economics
University of Oviedo

Oviedo, 2014

Essays in supply and demand in energy economics: A frontier analysis approach

Universidad de Oviedo

Manuel Llorca Riego

Department of Economics

University of Oviedo

Supervisors:

Luis Orea Sánchez and José F. Baños Pino

PhD Programme in Economics: *Economic Analysis Instruments*

Oviedo, 2014

indique que es más difícil gestionar aquellas empresas que operan en regiones con condiciones meteorológicas desfavorables.

En el segundo ensayo se propone el uso de un modelo de clases latentes para realizar segmentaciones de muestras de empresas energéticas antes de llevar a cabo evaluaciones comparativas de las mismas utilizando DEA, una técnica no paramétrica para medir la actuación de las empresas ampliamente utilizada en regulación energética. A través de una simulación, se muestra que este enfoque supera a otros procedimientos menos robustos y más arbitrarios de dividir muestras de empresas. Se presenta una aplicación práctica de este método para el sector de transporte eléctrico.

En este ensayo se destaca de nuevo la importancia de evaluar de forma adecuada y justa la actuación de las empresas en un marco de regulación por incentivos, pero en este caso se enfatiza la necesidad de controlar por la heterogeneidad y las diferencias tecnológicas inobservables, incluso cuando gran parte de esta heterogeneidad no es observada por el regulador. La propuesta de utilizar un enfoque de clases latentes para hacer frente a estas cuestiones se basa en las ventajas teóricas de este procedimiento para encontrar diferencias en el comportamiento de las empresas, ya que precisamente tiene en cuenta posibles diferencias en tecnología y condiciones ambientales para dividir la muestra. La simulación de datos llevada a cabo en la que se asumen diferentes grados de diferencias tecnológicas y tamaños de escala de las empresas, confirma la venta teórica que el enfoque de clases latentes tiene sobre otros métodos de segmentar la muestra como el ampliamente utilizado análisis clúster de k-medias. Comparado con otros métodos más simples y ad-hoc, se observa que el método propuesto es el mejor para identificar a qué grupo tecnológico pertenece realmente cada empresa, y los niveles de eficiencia obtenidos son los más próximos a los reales (es decir, aquellos generados en la simulación). Se muestra a su vez que la capacidad discriminatoria y el éxito en la asignación del modelo de clases latentes se incrementan cuando existen grandes diferencias en tecnología y escala, una cuestión que es crucial para cualquier procedimiento de segmentación.

El procedimiento es ilustrado con una aplicación práctica a la misma base de datos de transporte eléctrico analizada en el ensayo previo. Se encuentran dos grupos o tecnologías diferentes a través del enfoque de clases latentes, ya que a pesar de se observan aumentos continuados en la eficiencia promedio a medida que aumenta el número de clases, el mayor incremento en los niveles de eficiencia se produce al pasar de un modelo de una clase a un modelo con dos clases, modelo que resulta ser el elegido. Se obtienen resultados similares cuando se incluyen la meteorología y el crecimiento de la demanda como variables ambientales que pueden influir en la tecnología adoptada por las empresas.

En el tercer ensayo se explora el uso de modelos de fronteras estocásticas para estimar funciones de demandas de energía agregadas en el sector transporte de América Latina y el Caribe, donde este sector representa el 43% del consumo total de energía. El uso de estos modelos permite obtener medidas de los niveles de eficiencia energética en estos países que pueden ser consideradas alternativas a los tradicionales indicadores de intensidad energética utilizados comúnmente en comparaciones internacionales. Hasta donde sabemos, ésta es la primera vez que se aplica un enfoque de fronteras estocásticas a la estimación de una demanda de energía en el sector transporte. Debido a la posible heterogeneidad inobservable entre países, en este caso también se propone el uso de un enfoque de clases latentes, lo que permite comprobar la existencia de países con demandas diferenciadas que están asociadas a diferentes elasticidades precio y renta.

La base de datos utilizada en este capítulo consiste en una muestra de 24 países de América Latina y el Caribe para el período 1990-2010. El precio de la energía, que es junto a la renta, una de las variables más relevantes en un análisis de demanda, se obtiene a través de un índice multilateral y transitivo que permite realizar comparaciones robustas entre países a lo largo del tiempo. El consumo total de energía en el sector transporte viene dado por el consumo de varios componentes energéticos y por tanto requiere agregar los precios individuales de estos componentes para obtener un índice del precio de la energía. Sin embargo, las agencias estadísticas internacionales no proporcionan ningún tipo de índice de precios para el total de países de nuestra muestra y por tanto ha tenido que ser calculado para este trabajo.

En términos generales, el ranking de países que se puede elaborar a partir de los índices de eficiencia obtenidos cuando se estima una demanda frontera, está notablemente correlacionado con la clasificación que se deriva de un indicador habitual de intensidad energética. Sin embargo, la falta de correlación que se observa entre eficiencia e intensidad energética para algunos países a lo largo del tiempo, indica que las variaciones en los indicadores pueden estar asociadas a circunstancias distintas a cambios en la eficiencia energética. Por otro lado, a través del modelo de clases latentes en el análisis empírico se identifican tres demandas con elasticidades precio y renta claramente diferentes. Las probabilidades de pertenencia a cada clase dependen de la renta, el área y la población, y reflejan que los países con mayor renta per cápita y menor densidad de población tienden a tener demandas de menor elasticidad precio. La estimación de estas demandas permite la identificación de aquellos países con mayor eficiencia energética en cada clase, que precisamente coinciden con los que han desarrollado políticas para mejorar el transporte público en los últimos años.

Finalmente, en el último ensayo, se estiman modelos de demanda de energía frontera para el sector energético residencial estadounidense, pero en este caso se amplía el modelo básico para poder incluir el efecto rebote, que tiende a atenuar los ahorros esperados en el consumo de energía asociados a mejoras en eficiencia energética. Este concepto no ha sido analizado de esta forma hasta ahora, por lo que la mayor contribución de este trabajo es sencillamente conectar ambas literaturas. Numerosos estudios han presentado diferentes enfoques para medir los efectos rebote. Entre los diferentes métodos, el utilizado más comúnmente es obtener el efecto rebote a través de las elasticidades precio de las demandas de energía estimadas, aunque únicamente proporciona una medida indirecta del potencial efecto rebote bajo fuertes supuestos sobre el comportamiento de los consumidores.

En este ensayo, se presenta el modelo estándar de demanda de energía frontera y los supuestos implícitos que se asumen en el modelo. Para obtener información sobre el efecto rebote, se sugiere la extensión de este modelo básico incorporando un factor de ajuste que mitiga o intensifica el efecto de las mejoras en eficiencia energética sobre el consumo de energía. Dos especificaciones alternativas se proponen para esta función de efecto rebote, además de sugerirse una estrategia basada en las estimaciones del modelo ALS para tratar el problema que existe en la identificación de la constante en esta función. Las especificaciones de estas funciones de efecto rebote están relacionadas con la demanda de servicios energéticos.

En la aplicación empírica se utilizan datos de energía residencial para una muestra de 48 estados de EE.UU. para el período 1995-2011. El efecto rebote medio que se obtiene para la muestra es relativamente alto (entre 56% y 80%) comparado con los valores que se observan en la literatura, pero es posible que estos valores estén sobreestimados debido a la concavidad de la forma funcional asumida para las funciones de efecto rebote. A pesar de esto, nuestros modelos permiten una identificación robusta de aquellos países en los que políticas destinadas a promover la eficiencia energética serían más exitosas. Además, utilizando el modelo propuesto, las medidas de efecto rebote que se obtienen no se ven alteradas por el modo en que el efecto del paso del tiempo es incorporado en la demanda, lo que contrasta con el sesgo que surge cuando el efecto rebote es obtenido a través de las elasticidades precio de la demanda.

RESUMEN (en Inglés)

This thesis is formed by four essays that cover different topics in energy economics. In these essays, efficiency analyses that mainly involve the estimation of stochastic frontier models are applied to deal with energy-related issues in diverse sectors of the economy.

The first two essays are specifically focused on the efficiency analysis of electricity transmission firms. In the first essay, the effect of potential determinants of firms' efficiency, primarily weather conditions, is studied through the estimation of heteroscedastic models. In the second essay, a latent class model is proposed to deal with technological and environmental differences as a first step to carry out standard benchmarking. In the remaining essays, energy demand frontier models are estimated to obtain information about energy efficiency and the so-called rebound effect in energy consumption. In the third essay, a latent class model is applied again, but in this case with the aim of finding distinct energy demand functions across countries for the transport sector in Latin America and the Caribbean. The last essay merges two literatures due to the use of the stochastic frontier approach to directly measure the magnitude of the rebound effect, and applies this model to the US residential energy sector.

In the first essay, the economic characteristics of the technology and the managerial efficiency of 59 US electricity transmission firms for the period 2001-2009 are analysed. The main contribution of this work is to study the effect of weather on costs of the electricity transmission grid. The estimated set of heteroscedastic models allow us to conclude that weather adversely affects electricity transmission costs primarily through inefficient management and not through the technology. That is, the estimated costs of a fully efficient firm are not going to be increased because of adverse weather conditions; an issue that must be taken into account when the costs of these firms are analysed by regulators.

This essay allows us to study economic features of firms' technology such as the economies of scale and density. In particular, the empirical model that is estimated allows us to state that most electricity transmission networks exhibit natural monopoly characteristics. We have also found that, despite the US regulators' effort aiming to improve firms' performance, the average efficiency level in this industry has decreased over the period analysed, showing however an increasing divergence among firms. This essay analyses the potential effect of certain environmental variables such as weather conditions on firms' efficiency. In this sense, we have found that more adverse conditions generate higher levels of inefficiency which may indicate that it is more difficult to manage firms that operate in regions with unfavourable weather.

In the second essay, the use of a latent class model approach is advocated to segment samples of energy firms into homogeneous groups before carrying out a standard benchmarking analysis using DEA, the most widely technique applied in energy regulation to measure firms' performance. Through a simulation exercise, it is shown that combining the latent class model and DEA outperforms other, less robust and more arbitrary, procedures to split samples of firms. A practical application of this method to the US electricity transmission sector is presented.

A proper (and fair) measurement of firms' performance in an incentive regulation framework is highlighted again in this essay, but in this case it is emphasised that unobserved differences in technology or environmental conditions should be taken into account, even though most of that heterogeneity is not observed by the regulator. The proposal of using a latent class model to deal with this issue is based on the theoretical advantages of using a sample separating procedure that precisely takes into account potential differences in technology or environmental conditions to split the sample of firms. The performed simulation analysis that plays with different degrees of technological differences and firms scale, confirm the theoretical advantage of the latent class model over other sample separating methods such as the widely used k-means cluster analysis. Compared to other simpler and ad-hoc methods, the proposed

procedure is the best to identify the technological group each firm really belongs to, and the efficiency levels obtained using this procedure are the closest to real ones (i.e. those generated in the simulation). It is also shown that the discriminatory capacity and the assignment success of the latent class model increase when large differences in technologies and scale arise, a feature that is crucial for any sample separating method.

The procedure is illustrated with an empirical application to the same database of electricity transmission firms analysed in the previous essay. Two different groups or technologies are found using the latent class approach. Although consecutive increases in the efficiency scores are observed when we move from one class to two classes and so on, the largest change in efficiency scores arises when we move from a one-class model to a model with two classes, which is precisely the preferred model. Similar results are obtained when weather conditions and the demand growth are included as environmental factors that may influence the technologies adopted by the firms.

In the third essay, a stochastic frontier approach is used to estimate aggregate energy demand functions in the transport sector of Latin American and Caribbean countries where this sector represents the 43% of total energy consumption. The use of these models to measure energy efficiency allows us to deal with some of the disadvantages of the energy intensity indicators commonly used in international comparisons. As far as we are aware, this is the first application of the stochastic energy demand frontier approach to measure energy efficiency in the transportation sector. This model is nested in a latent class structure to control for the likely large unobserved heterogeneity among these countries, and test for the existence of groups of countries with different demands associated to distinct price and income elasticities.

The database used in this essay consists of a sample of 24 Latin American and Caribbean countries for the period 1990-2010. The energy price, which is, in addition to income, one of the most relevant variables in a demand analysis, is obtained using a transitive multilateral index that allows for proper comparisons across countries over time. Energy consumption in the transportation sector consists of several components and therefore, it is required to add each individual price to obtain an index of energy price. However, no index price is provided by any statistical agency for the total sample of countries and hence it has been calculated for this essay.

Generally speaking the ranking of efficiency scores that is obtained when a single demand is estimated is notably correlated with the ranking derived from the widely-used energy intensity indicators. The lack of correlation observed for some countries over time indicates, however, that variations in energy intensity indicators may be associated with circumstances other than changes in energy efficiency. On the other hand, three demands with quite different income and price elasticities are found using a latent class model. The class membership probabilities of this model depend on income, area and population; and reflect that those countries with higher per capita income and lower population density tend to have the lowest price elasticity. The estimation of this model allows identifying the most energy efficient countries in each class that, in fact, coincide with those that have developed policies to improve public transport in recent years.

Finally, in the last essay, energy demand frontier models are estimated for the US residential energy sector, but in this case we have extended the traditional model to measure the so-called rebound effect, that tends to attenuate the expected savings in energy consumption, associated to improvements in energy efficiency. So far this issue has not been analysed using the stochastic frontier approach and hence the major contribution of this essay is just linking both literatures. Numerous studies have used different approaches to measure the different types of rebound effects. The most popular method relies on the estimation of price elasticities of the demand for energy, but it only provides an indirect measure of the potential rebound effect under strong assumptions about consumer behaviour.

In this essay, the standard energy demand frontier model and the implicit assumptions of this model are presented, and it is suggested the estimation of a model that allows obtaining non-zero rebound effects through a correction factor that mitigates or intensifies the effect of efficiency improvements on energy consumption. Two alternative specifications are proposed for the rebound-effect function, and also a strategy to deal with the identification of the intercept of this function. The specifications of these rebound-effect functions are related to the demand for energy services.

The empirical application is based on residential energy data for a sample of 48 US states over the period 1995-2011. The average rebound effect obtained is somewhat high (between 56 and 80%) compared with the values obtained in the literature, but it may be overestimated due to the concavity of the functional form assumed for the rebound-effect function. Despite of this, our models allow a robust identification of states in which policies to promote energy efficiency would be more successful. Furthermore, using the proposed model, the rebound effect measures are not altered by the way in which the effect of time is incorporated in the demand, which contrasts with the bias that arises when they are obtained through their own price elasticity estimates.

SR. DIRECTOR DE DEPARTAMENTO DE ECONOMÍA
SR. PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO EN ECONOMÍA: INSTRUMENTOS DEL ANÁLISIS ECONÓMICO

A mis padres, a mi hermana, a mi abuela, a mi tía y a Raquel.

Acknowledgements

To start with, I would like to express my most sincere gratitude to my supervisors José F. Baños and Luis Orea. Their "energy" has fostered my progress with this thesis over the last four years. None of this would have been possible without their guidance, counsel and hard work. Thank you very much to both of you.

I would also like to acknowledge the support of other academics who have stimulated my interest in research: Antonio Álvarez, Plácido Rodríguez and Rosa Aza, among others; and the help and advice from the rest of the members of the Department of Economics at the University of Oviedo, particularly from those who are part of the Oviedo Efficiency Group. Likewise, I would also like to show my appreciation to the international researchers that have collaborated in some chapters of this dissertation; José Somoza from the University of Havana (Cuba), Massimo Filippini from the University of Lugano and the ETH Zürich (Switzerland), and especially to Michael G. Pollitt from the University of Cambridge (UK). The effect of the hospitality of Michael and the Energy Policy Research Group cannot be overstated in the way it aided me during my time in Cambridge, and consequently in the writing of this document.

Furthermore, special thanks to my girlfriend Raquel, who has been my biggest inspiration and source of bliss during the last year. I also want to recognise the support of my immediate family as one of the strongest drivers towards my success. I would like to thank my friends, all those moments of entertainment and all the interesting conversations that we have had. Among them I want to mention Miguel Hermo and Shingi Padya, whose comments and suggestions have always added value to my research.

I do not want to finish this section without expressing my gratefulness to the Gijón Rugby Club and to the people from Judo Gandoy. In both places I have found not only sport and entertainment to free my mind, but also two groups of really good friends. In the particular case of Judo Gandoy, I started to do judo in this gymnasium with 7 years old, and after 23 years I truly believe that the teachings of this discipline have largely contributed to the forging of my character. I must mention José Luis Relancio and Jose Antonio S. Gandoy as two of the most influential people in my life from whom I have learned many lessons since I was a child.

Finally I would like to acknowledge the financial support from the Government of the Principality of Asturias through their 'Severo Ochoa' PhD scholarships. The research presented here has only been possible by this funding.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Mankind has been interested in understanding what energy is and how its use can be exploited to its advantage since the dawn of time. The relevance of the energy for human beings is already highlighted by the Greek myth of Prometheus[1] who was a titan that stole the fire from the gods to give it back to men. This myth represents the importance of the discovery and mastery of fire as an essential starting point in the advancement of civilization through its use for cooking, lighting, warming, protecting against wild animals, making tools, etc. Moreover, in a spiritual sense, the fire is usually understood in a metaphorical way as a symbol of life, consciousness and intelligence; issues also related with human development.

Despite the great importance that energy has had throughout history, it was not until the eighteenth century when the use of energy became more essential in society due to the Industrial Revolution, and the beginning of the intensive use of coal as primary energy source. A crucial time in the development of energy use is the mid-nineteenth century, when there was a series of discoveries that revolutionised the world. To cite just two cases, in 1831, Faraday (among other scientists) discovered electromagnetism and also by this time in different parts of the world the modern oil extraction industry had begun. In the early twentieth century, the relevance of energy greatly increased after half a century of technological improvements and it became the main driver of industrial development and the facilitator of many human activities. As a consequence, during the twentieth century, electricity production increased until 15,000 TWh (terawatt-hours), world crude oil consumption reached approximately 4,000 Mtoe (million tonnes of oil equivalent) and world population growth expanded from about 1.6 billion people to 6 billion.[2] This dramatic growth depicts indeed that the great advances in technology of the twentieth century have been closely linked to an increasing dependence on energy.

Due to the concern that involves the use of energy resources for human activity, some energy-related topics have been analysed by economists in the past.[3] However, it is commonly accepted that the economics of energy was not developed as a specialised branch until the first oil shock in the 1970s (Edwards, 2003). The consequences of the high increase in oil prices highlighted the importance of energy in economic development of the countries and thereafter, researchers, academics and policy makers have shown great interest in energy studies.

[1] The word 'energy' precisely comes from ancient Greek word *energeia* (*ἐνέργεια*) which in English means being-at-work (Sachs, 1995).
[2] This information has been obtained from the World Bank, the US Energy Information Administration (EIA) and the US Census Bureau.
[3] One of the pioneering scholars on these issues was Jevons (1865) who already showed concern over the depletion of coal in his book "The Coal Question".

Following Bhattacharyya (2011), energy economics is focused, like any other branch of the economy, on the allocation of scarce resources in society. For this reason, it covers a wide range of microeconomic and macroeconomic issues in the demand and supply of energy. The topics that have been analysed within this field of research have been expanded over time to address new issues that have emerged in the energy industry. For example, the major concern of energy economics in the 70s was trying to understand the energy industry and particularly the oil industry, analyse the energy substitution and begin to study the alternative of renewable energy. The interest in environmental issues and economic development was increased in the following decade, while in the 90s, the issues that received most attention were those related with the liberalization and restructuring of energy markets, although concerns about climate change and other environmental issues still remained.

Finally, in recent years, attention has focused on high oil prices, energy scarcity, and the debate about state intervention and the role of the market in the management of energy supply. This discussion has basically attempted to find an answer to the concern of ensuring the supply of energy to the population in a carbon-constrained world. In summary, energy economics has provided tools for a better understanding of the role of energy in the society over time, trying to answer the most important questions about demand and supply of energy, given the political and economic situation of each moment.

Since energy economics encompasses a very wide range of topics, this thesis covers a small portion of this field of research. The core of the document is formed by four chapters in which efficiency analysis models are applied to the supply and demand for energy in different sectors of the economy such as electricity and transportation. It should be noted, however, that these chapters can be read separately and understood as independent essays with their respective specific motivations and conclusions. Regarding this issue, all of these essays have been submitted to international academic journals in an effort to make available, results obtained during the research process of this thesis. In fact, one of them (Chapter 2) has recently been published in an operational research journal.[4]

Chapters 2 and 3 are specifically focused on the efficiency analysis of electricity transmission companies. The electricity sector is undeniably one of the systemic sectors of the economy as it provides an essential input for the production system of the countries (and therefore affects the competitiveness of enterprises) and simultaneously, is a basic element for welfare and comfort in society. Over recent decades, this sector has been undergoing intense economic, technological and environmental changes, among others.

As mentioned before, since the 90s one of the objectives of the economics of energy has been to study the liberalization processes of the energy markets, as it has been in the case of the electricity sector. In most developed countries, this industry has been restructured with the aim of improving both service quality and firms' performance, ensuring that the consumers benefit from those gains (Jamasb and Pollitt, 2007). As a consequence of this, former state-owned utilities were privatised and electricity sectors were vertically separated into different segments: generation, transmission, distribution and retailing. The reforms that were carried out, led to the

[4] Llorca, M., Orea, L. and Pollitt, M.G. (2014), "Using the latent class approach to cluster firms in benchmarking: An application to the US electricity transmission industry", *Operations Research Perspectives*, 1(1), 6-17.

creation of some bodies to execute coordination functions that previously were internal to the companies, and in turn, also led to different treatments of the unbundled activities: generation and retailing have generally been opened to market competition, while transmission and distribution networks have been treated as natural monopolies that have to be regulated. Nowadays, regulated segments (i.e. transmission and distribution) still provide the infrastructure for the competitive segments and represent a significant share of the total price paid by final customers in many countries.

Different regulatory practices have been applied in electricity markets which are mostly based on benchmarking methods to determine the relative firms' cost efficiency or service quality through the comparison of each firm with those with best performance. Published papers have basically employed parametric, nonparametric and semi-parametric techniques in those analyses. A recent issue that is increasingly discussed in this literature is the need to control for the different environmental conditions under which each utility operates in order to obtain reliable (and fair) efficiency scores. This is not a minor issue in an incentive regulation setting as these conditions may have significant financial implications for the regulated companies. The concern about the inclusion of environmental variables in efficiency analysis has generated the development of many models for all of the abovementioned approaches (i.e., parametric, nonparametric and semi-parametric).

As Brunekreeft *et al.* (2005) and Joskow (2014) point out, the electricity transmission system is a critical segment of the electrical grid due to its influence over the whole electrical grid in features such as the efficient use and expansion of the network, and making investment decisions on locational choices of new generation and energy intensive users. Despite this, and while the distribution segment has received large amounts of attention, there is a lack of empirical studies that analyse firms' performance in the electricity transmission sector (Haney and Pollitt, 2013). Remarkable exceptions are Huettner and Landon (1978), Pollitt (1995), Dismukes *et al.* (1998), and von Geymueller (2009). Although the potential importance of weather conditions in electricity transportation (i.e. transmission and distribution) is highlighted in some engineering papers, it has only begun to be considered in economic research on firms' performance in electricity distribution quite recently (see Yu *et al.*, 2009). One of the most interesting issues with environmental conditions is to identify if companies are using them as an excuse for poor performance as Nillesen and Pollitt (2010) suggest for the US electricity distribution system.[5] Returning to the case of the limited empirical literature on electricity transmission, none of the previously mentioned papers controls for environmental conditions.

In an effort to fill this gap, the following two chapters of this thesis are focused on the analysis of the electricity transmission network. In Chapter 2, an empirical analysis of the economic characteristics of the technology and the inefficiency of the US electricity transmission firms is conducted through the estimation of several heteroscedastic models taken from the recent Stochastic Frontier Analysis (SFA) literature (for a detailed survey of this literature see Kumbhakar and Lovell, 2000). Unlike previous papers, the estimation of these models allows us to identify the determinants of firms' inefficiency in this industry and to control for weather conditions; one of the most decisive, uncontrollable factors in electricity transportation. In addition, the estimated models allow us to discuss whether the environmental factors

[5] More concretely, they suggest that companies which operate under unfavourable environmental conditions can become best-practice.

should be treated as determinants of firms' performance or as technological cost drivers, which may mean that companies are using environmental conditions as an excuse to avoid being penalized in regulatory processes due to their poor performance.

While Chapter 2 is mainly focused on observed differences across regulated electricity networks, Chapter 3 has to do with unobserved technological and environmental differences. In this chapter, the use of a Latent Class Model (LCM) approach is advocated to deal with this issue before carrying out a conventional efficiency analysis using DEA (Data Envelopment Analysis). As shown by Zhou *et al.* (2008), this nonparametric approach has become a widespread tool in energy and environmental studies, especially for benchmarking electric utilities. The characteristics that have made this approach appealing for regulators are that it imposes few assumptions on firms' technology features and it allows them to avoid the traditional convergence problems in the SFA literature. However, this approach does not easily address the effect of unobserved geographical and weather conditions on firms' performance.

It is common practice to assume that the whole set of benchmarked companies share the same technology in incentive regulation. Therefore, any difference in firms' performance is attributed to an inefficient use of the factors that are under the control of the companies. Possible differences among utilities associated with different technologies or environments are often overlooked or are not properly addressed, and hence the efficiency scores obtained from those analyses might be biased.

Our proposal in Chapter 3 is based on the fact that LCMs are designed to cluster firms by uncovering differences in technology parameters taking into account for clustering the data the same technological relationship that is going to be analysed later. This approach allows for a split of the electricity networks into a number of different classes where each class is associated with a specific technology or environment, before performing a traditional efficiency analysis of regulated electricity networks. A simulation analysis is carried out to examine whether the latent class approach outperforms other more arbitrary and less robust procedures of splitting a sample of observations. The procedure is illustrated with an application to the same US electricity transmission sample examined in the previous chapter, since these companies are expected to operate under different technologies, taking into account that they are located in different states and are subject to different environmental conditions.

In contrast to the previous supply-oriented essays, the remaining two chapters are focused on energy demand (for the case of transportation and residential sectors) and the efficiency in its use or consumption, another of the major concerns of energy economics due to the large dependence on non-renewable energy resources in human activity and the climatic change. The recent situation of economic crisis has led the national and supranational authorities to reassess the balance of priorities between the three objectives of energy policy: security of supply, competitiveness and environmental sustainability (Becker Zuazua, 2011).

Since the 1970s world oil crisis, previously mentioned as the origin of the energy economics, and the growing awareness of global warming in the late 1980s, reducing energy consumption and emissions has become a key energy policy objective for most governments across the globe. The promotion of energy efficiency as a means to that end, has converted the definition and measurement of this concept into an essential objective for the majority of the countries, prior to the design of economic and energy policies with the aim of reducing the use of energy.

To help achieving the ultimate goal of reducing energy consumption, various quantitative indicators that are related to the energy efficiency of each country have been developed and have been used in international comparisons. In that sense, there is no single definition universally accepted for the concept of energy efficiency. Ang (2006) indicates that the most common practice has been to link this idea with some thermodynamic, physical-based and monetary-based indicators that relate energy consumption to measurements of the economic activity or energy services derived from this consumption. Nonetheless, the value of energy intensity can vary significantly over time due to changes in the structure of GDP (Gross Domestic Product), which are difficult to assimilate into the concept of energy efficiency. In general, the most basic indicators do not allow cross-country comparisons and the calculation of potential energy savings. Regarding this issue, Filippini and Hunt (2011, 2012), suggest the use of an SFA approach to estimate energy demand frontier functions. The "frontier" nature of this approach allows the computation of alternative measures of energy efficiency that do not suffer from the same problems as the conventional indicators of energy intensity. The efficiency measures obtained are based on the comparison of the energy consumption of the countries with respect to the minimal energy consumption predicted by the frontier, which takes into account the optimizing behaviour of companies and individuals. The estimation of these models allows to control for characteristics such as economic or environmental factors that affect the sector and may bias the results obtained from standard energy intensity indicators. The frontier approach allows them to obtain a "pure" measure of the inefficient use of energy for each country.

The countries that aim to reduce their energy consumption and mitigate their greenhouse gas emissions should be especially concerned about the adoption of measures that improve the energy efficiency especially in those sectors in which are more energy intensive. In Chapter 4, the SFA model proposed by Filippini and Hunt (2011, 2012) is used to estimate energy demand functions in the transport sector in Latin America and the Caribbean. Transportation is the sector that involves the largest energy consumption in this region (43%), and the Economic Commission for Latin America and the Caribbean (ECLAC, 2010) indicates that the share of this sector respect to total energy consumption will even increase in the future. Despite this, there is a scarcity of empirical analysis on the transport sector in Latin America and the Caribbean which has been motivated by data unavailability and the absence of a formal link between institutions that are in charge of providing information on energy and transport. Given the amount of energy consumption in transport and the increase of energy prices that were experience by this region in recent years, it is thus necessary to conduct studies focused on the energy consumption of this sector that help raise awareness about the environmental sustainability issues that are mentioned in the "Millennium development goals" proposed by ECLAC (2005).

A Latent Class Stochastic Frontier Model (LCSFM) is proposed in Chapter 4 to capture unobserved demand heterogeneity due to the impossibility of including all relevant variables capturing the numerous and different features of the transport sector in each country. This approach allows us to test for the existence of groups of countries with clearly differentiated demands that are associated with distinct price and income elasticities. As the transport of both goods and passengers implies the consumption of different types of energy, an index that aggregates various energy prices is thus required for the analysis. However, international agencies do not provide specific indicators of aggregate energy prices in transport for the majority of the countries analysed. For this reason, the construction of a transitive multilateral index is proposed, which, in contrast to those frequently presented by the aforementioned agencies, facilitates international comparisons over time.

Finally, it should be mentioned that in practice, the achievement of savings in energy consumption through the promotion of energy efficiency do not depend only on the energy efficiency itself. Actual savings in energy consumption after an efficiency improvement might not coincide with the expected savings due to the so-called rebound effect, a phenomenon already highlighted by Jevons (1865) who observed that technological improvements that increased the efficiency in the use of coal in the 18th and 19th centuries, led to increases in its consumption in a broad range of industries. The rebound effect is a phenomenon that links both energy consumption and energy efficiency with energy services: as energy efficiency improvements make energy services cheaper, they may lead to increase the demand of those services. This response may thus offset the reduction in energy consumption that is predicted by engineering models.

Measuring the rebound effect is crucial in order to properly evaluate the effectiveness of energy policies that aim to promote energy efficiency improvements. There are many empirical studies that use econometric methods to estimate the rebound effect. In their review of the literature, Sorrell and Dimitropoulos (2008) have found a lack of consensus with regard to a consistent method to its measurement. In principle, it could be directly obtained from the elasticity of demand for energy services with respect to changes in energy efficiency. However, relatively few studies follow this approach because data on either energy services or energy efficiency is unavailable or is limited in terms of accuracy. As a consequence of this, the rebound effect is often measured indirectly through the estimate of different elasticities that are considered equivalent to the elasticity of the demand for energy with respect to changes in energy efficiency. The most commonly used is the own-price elasticity of the demand for energy, as is directly obtained from energy demand estimates.

Chapter 5 brings attention to the fact that the standard energy demand frontier model introduced by Filippini and Hunt (2011, 2012), that has been used in the previous chapter, is closely connected to the measurement of the rebound effect. In particular, it is shown that this model implicitly imposes a zero rebound effect, which contradicts most of the available empirical evidence on this issue and may explain why previous applications have not examined this issue. This restrictive assumption is relaxed through the modelling of a rebound-effect function that acts as a correction factor that mitigates or intensifies (i.e. adjusts) the effect of an efficiency improvement on energy consumption. The model is illustrated with an empirical application for the US residential energy sector where this issue can be particularly relevant since it accounts for 37% of the national electricity consumption, 17% of greenhouse gas emissions and 22% of primary energy consumption (International Risk Governance Council (IRGC), 2013).

Chapter 2

Efficiency and environmental factors in the US electricity transmission industry

2.1. Introduction

The electricity industry in most developed countries has been restructured over recent decades with the aim of reducing costs, improving service quality and encouraging electricity utilities to perform efficiently. As a result, former state-owned utilities were privatized and electricity sectors were vertically separated into generation, transmission, distribution and retailing, particularly in Europe (see Jamasb and Pollitt, 2005). Whereas some of these segments such as generation and retailing were opened to competition, other segments such as transmission and distribution are still regulated. In this sense, incentive-based regulation schemes have been implemented in several countries (e.g. UK, Norway) in order to encourage both transmission and distribution utilities to perform efficiently.

Joskow (2014) points out that for industries in which regulated segments provide the infrastructure platform upon which competitive segments rely, social welfare depends on firms' performance and reforms made in both regulated and competitive segments. Much of the research in the electricity industry has focused on competitive wholesale markets, although the regulated segments provide the infrastructure for the competitive segments and even though networks constitute a significant share of the final price paid by electricity consumers.[6] Even though electricity transmission is necessary for distribution and retailing, there is a lack of empirical studies that analyse both the economic characteristics of the technology and firms' inefficiency in the electricity transmission.

Statistical benchmarking methods have been largely used in the electricity industry to determine the relative efficiency of individual firms' costs compared to their peers (see Haney and Pollitt, 2009, 2013). Obtaining reliable (and fair) measures of firms' inefficiency requires controlling for the different environmental conditions under which each firm operates. This is especially acute in benchmarking because of the financial implications that this analysis can have on the firms and their effect over the whole network. One of the most interesting issues with environmental conditions is the question of whether firms are using them as an excuse for poor performance. In line with this, Nillesen and Pollitt (2010) find that firms which operate in unfavourable conditions can become best-practice for the case of US electricity distribution.

One of the most decisive uncontrollable factors in electricity transportation (i.e. in transmission and distribution) is the weather conditions of the area in which the companies operate. Billinton and Wenyuan (1991), and Billinton and Acharya (2005)

[6] Typically distribution and transmission charges combined compose around 25% of the residential bill (excluding taxes and environmental charges).

tried to explain changes in the probability of failure rate in the system using complex mathematical models. Generally speaking, they pointed out that most technical interruptions occur when weather is adverse and, in particular, extremely adverse. They also showed that assessing likely failure rates while ignoring weather tend to give too optimistic and erroneous predictions.

Regarding electricity transmission, Billinton and Wu (2001) pointed out that overhead transmission lines are exposed to a wide range of weather conditions and, that both failures rates and the probability of overlapping failures tend to increase sharply during periods of extremely adverse weather conditions. Rothstein and Halbig (2010) find that many atmospheric and hydrological parameters not only affect electricity generation and consumption, but also electricity transportation. Indeed, overhead lines are affected by atmospheric influences in several ways, such as lightning, wind, additional weight (e.g. ice or snow), low temperatures, humidity and moisture.

Despite the potential role of weather conditions in electricity transportation, only a few papers have analysed firms' performance in the electricity distribution sector controlling for environmental factors. In particular, Yu *et al.* (2009) showed using nine weather variables that severe weather conditions tend to increase service interruptions, and this in turn increases costs associated with replacing the damage equipment and restoring power. Jamasb *et al.* (2010, 2012) also concluded that the lack of inclusion of variables related to weather conditions might downward bias the estimated coefficients of other relevant variables, and, in particular, those associated with the marginal cost of quality improvements. Using weather and geographic composites, Growitsch *et al.* (2012) predicted up to 30% lower costs than average, for utilities that operate in areas with extremely good environmental conditions, and up to 39% higher costs than average, for utilities that operate in areas with extremely bad environmental conditions. On average, they predicted higher costs of about 5% as a result of hostile weather conditions.[7]

On the other hand, as far as we are aware there are only four published papers that separately study the performance of transmission firms, none of them include inefficiency determinants and only the most recent of them has controlled for environmental conditions. Using a sample of US firms, Pollitt (1995) analysed differences in efficiency between state-owned and private electricity transmission companies. He did not find significant differences between both types of firms using parametric and nonparametric specifications of the frontier model. Using also US data, Huettner and Landon (1978) and Dismukes *et al.* (1998) have examined the existence of returns to scale in the provision of electric transmission services. Huettner and Landon (1978) do not find increasing returns to scale, except for one category of sales expenses. In contrast, Dismukes *et al.* (1998) find significant economies of scale for all the NERC (North American Electric Reliability Corporation) reliability regions using data for the period 1986-1991. von Geymueller (2009) carried out a comparison of static and dynamic DEA models in electricity transmission using data of 50 US utilities for the period 2000-2006. The author finds that static models tend to overestimate firms' inefficiency because they do not take into account the existence of quasi-fixed inputs.

This chapter contributes to the literature analysing firms' performance in the electricity transmission industry with an empirical analysis of the US electricity transmission system for the period 2001-2009. The analysis of the economic

[7] In contrast, Nillesen and Pollitt (2010) do not find that US electricity distribution companies with unfavourable conditions are worse performers.

characteristics of the technology (such as economies of scale or economies of density) and the inefficiency of each US utility relies on the estimation of several specifications of heteroscedastic models taken from the recent SFA literature. Unlike previous papers, our SFA models allow us to identify the determinants of firms' inefficiency in this industry, and discuss whether the environmental factors should be treated as determinants of firms' performance or as technological cost drivers.[8] This is not a semantic point in an incentive-regulation framework as regulators should purge the data when environmental conditions are part of the technology, i.e. they are cost drivers independent from firms' performance, but not when they have an indirect effect through inefficiency. That is, conditional on a wide definition of the technology, firms cannot use unfavourable environmental conditions as an excuse to avoid being penalized due to their bad performance. To examine this issue we have applied a modified version of the 'zero inefficiency stochastic frontier model' recently introduced by Kumbhakar *et al.* (2013). To the best of our knowledge, this is the first time this model is used to capture differences in technology instead of differences in performance.

The estimated coefficients provide useful information about the firms' performance with both policy and managerial implications. We find using more recent data and larger firms than in previous papers that, given network infrastructure, most of the electricity transmission networks exhibit natural monopoly characteristics. Our results also indicate that more adverse conditions generate higher costs, mainly through higher levels of inefficiency. Furthermore, we find that investing in capital is a better strategy than incurring additional operating costs to deal with adverse weather conditions. On the other hand, we find that, as expected, firms' performance gets better when demand tends to be steady as firms cannot adjust their inputs without cost over time. The average efficiency at the beginning of the period is larger than in previous studies. But, using our preferred estimated model, the results indicate that efficiency has declined (and diverged) over time, suggesting that there is room for improvement in the performance of the US electricity transmission system. It should be mentioned that the use of US data to benchmark European and Australasian utilities is often suggested and has been undertaken by some regulators including the British regulator, Ofgem. Hence although the results obtained here relate to US transmission network, they are important for non-US regulators.

This chapter is organized as follows. Section 2.2 provides a brief review of the transmission and distribution literature and the most commonly used approaches to benchmark firm performance in incentive regulation schemes. Section 2.3 describes the theoretical cost function that we estimate as well as the empirical specifications of the estimated models. Section 2.4 presents the data and variables used in the empirical analysis. Section 2.5 reports the parameter estimates and the results obtained from those estimates. Section 2.6 presents the main conclusions.

2.2. Benchmarking in electricity transmission

The electricity sector is an industry with different and interrelated activities, which are affected by production and consumption decisions across the whole system.

[8] An additional contribution of the present chapter is that we control for weather characteristics by including a set of weather variables as determinants of firms' inefficiency that were gathered specifically for the present application. In addition, as our sample period is more recent than those analysed in previous papers we can see whether there has been an improvement in average efficiency in the US electricity transmission industry.

The US electricity system traditionally has been composed of large vertically integrated utilities. Nevertheless, in the last two decades several reforms have been implemented with the aim of disaggregating most utilities into differentiated segments. These reforms have led to different treatments of the separated activities: generation and retail are regarded as potentially competitive markets, while transmission and distribution networks are treated as natural monopolies that have to be regulated (see Joskow, 2014). As Jamasb and Pollitt (2007) point out, from an economic perspective, the aim of electricity unbundling is to provide utilities with incentives to improve their operating and investment efficiency and to ensure that consumers benefit from the gains. The main methods used to achieve these objectives are incentive regulation mechanisms, which include financial rewards and penalties for the firms linked with their performance.

Joskow (2014) notes that much of the research in this sector has focused on the competitive markets although the regulated segments provide the infrastructure for the competitive segments and represent an important amount of the total price paid by final consumers and have an important joint effect with competitive segments on social welfare. For these reasons, electricity transmission has played an important role in the success of liberalised power markets. Electricity reforms have led to the creation of some bodies to perform the coordination functions that formerly were internal to the firms. To deal with this issue and the stresses in transmission system after years of underinvestment, the Federal Energy Regulatory Commission (FERC) pursued the implementation of a Standard Market Design in the US and encouraged the so-called Regional Transmission Organizations (RTO) to facilitate efficient trade over wide areas and transmission investment. According to Greenfield and Kwoka (2011), the RTOs - such as PJM - provide transmission services but do not own transmission facilities and they are not responsible for the maintenance and repair, or fixed investment costs, of the transmission facilities over which they direct the flow of power. Their essential role is as an independent service provider that administers the terms and conditions of transmission services and maintains the short-term reliability of the network.

Despite the importance of RTOs in the overall performance of the electricity system, the transmission utilities and the structure of the network charges have a great effect on network use and its development. Following Brunekreeft *et al.* (2005, p.74-75), the setting of the charges at an appropriate level is a key issue because it affects "the locational choices of new generation (and of energy intensive users), as well as influencing the bidding behavior of generators, and the willingness of neighboring electricity markets to trade and cooperate". As a result, "ideally the structure of network charges should encourage: *i*) the efficient short-run use of the network (dispatch order and congestion management); *ii*) efficient investment in expanding the network; *iii*) efficient signals to guide investment decisions by generation and load (where and at what scale to locate and with what choice of technology-base-load, peaking, etc.); *iv*) fairness and political feasibility, and *v*) cost-recovery" (Brunekreeft *et al.*, 2005, p.75).

There are different regulatory practices across the world to set the total amount of network charges in the electricity market which are mostly based on benchmarking, i.e. on measuring firm's efficiency against the firms with best practice performance (see Haney and Pollitt, 2013). As regulators reward or punish firms according to their (in)efficiency level, the reliability of these scores is particularly crucial for regulatory credibility. Any efficiency estimate tries to measure the gap between actual cost (production) and the optimal point on the cost (production) frontier, which must be estimated from the available data. Published papers have basically employed parametric (e.g. SFA), nonparametric (e.g. DEA), and semi-parametric (e.g. StoNED, Stochastic Nonparametric Envelopment of Data)[9] techniques to estimate cost (production) frontiers. As all techniques have their advantages and disadvantages,[10] the selection of an appropriate estimation method is contentious and may influence the obtained results and the consequent regulatory policy implications (see, for instance, Coelli *et al.*, 2005).

Despite the relevance of transmission networks in the electric power industry it is very difficult to implement a statistical benchmarking for most of the countries due to the lack of domestic comparators (Haney and Pollitt, 2013). International benchmarking can be an alternative to deal with this issue, but the regulators face several problems. Joskow (2014, pp.54-55) notes that the layout of the transmission network depends on countless factors, such as "the distribution of generators and load, population density, geographic topography, the attributes and age of the legacy networks' components and various environmental constraints affecting siting of new lines, transformers and substations". Moreover, there is no standardization or homogeneity among countries about the voltage boundaries between transmission and distribution networks. For instance, in the UK the transmission network is formed by elements that run at 275 kV and above, while in other countries like the US or France transmission network is formed by elements that run above 60 kV, making an international comparison a challenging task. Regarding the inputs and outputs that should be taken into account in an empirical analysis on efficiency of transmission systems, Pollitt (1995) pointed out that it might be desirable to take every specific factor of the company into account due to the complexity of the network. Each transmission system is unique because of the different kinds of inputs that they use and the environment in which they operate.

By contrast, statistical benchmarking methods have been largely used in electricity distribution to determine the relative efficiency of individual firms' operating costs and service quality compared to their peers.[11] Some countries such Germany, Nordic countries and Switzerland have a large number of utilities. This provides a suitable basis for the use of advanced benchmarking techniques and without necessarily having recourse to international benchmarking. It is generally desirable for regulators to have a large number of utilities for comparison and efficiency benchmarking.

As mentioned above, obtaining reliable (and fair) measures of firms' inefficiency requires controlling for the different environmental conditions under which each utility operates. This is especially acute in benchmarking because of the financial implications that this analysis can have over the firms and their effect over the whole network. The concern about the inclusion of environmental variables (also called contextual variables or z-variables) has generated the development of several models either using parametric, nonparametric or semi-parametric techniques. Although in this chapter we do not pretend to provide a complete survey on the alternatives for including z-variables, we present in Figure 2.1 a brief summary of some models that can be applied following each approach.[12] Given the wide range of models that have been developed, here we only mention the methods most frequently applied.

---

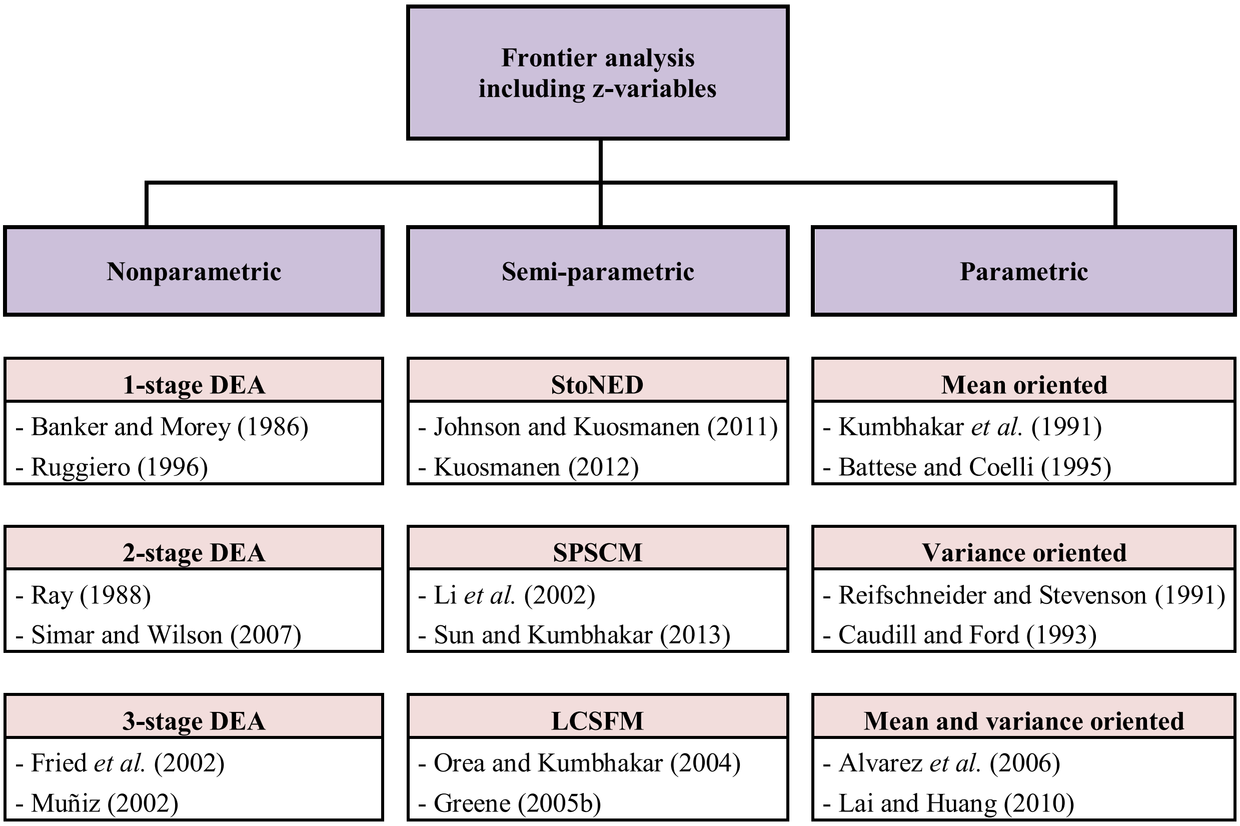
[9] These models are also labelled as semi-nonparametric.

[10] For instance, SFA has advantages over DEA when noise is a problem, and this can arise from measurement errors or other sources of statistical noise such as luck, weather, equipment failure or similar factors that are beyond firms' control.

[11] Jamasb and Pollitt (2001) show the most used approaches and provide a survey of benchmarking studies applied mainly in OECD countries. For a more current review of applied papers on electricity distribution see for instance Kuosmanen (2012).

[12] For a more detailed review of this topic in SFA and DEA, see Johnson and Kuosmanen (2011, 2012).

Figure 2.1. Approaches that allow including environmental variables in efficiency analysis (with key papers)



The inclusion of environmental variables in DEA has been done in one, two or even more stages. Ruggiero (1996) and other authors have highlighted that the one-stage model introduced in the seminal paper of Banker and Morey (1986) might lead to bias. To solve this problem, other models using several stages have been developed in the literature. Ray (1988) was the first who proposed a second stage where standard DEA efficiency scores were regressed on a set of contextual variables. This practice was widespread until Simar and Wilson (2007) demonstrated that this procedure is not consistent because the first-stage DEA efficiency estimates are serially correlated. Although the bootstrap procedure proposed by these authors to solve this problem in two stages became a widely used method in DEA to identify inefficiency determinants, three-stage models have also been developed (see, for instance, Fried *et al.* 2002; and Muñiz, 2002).

In the recently developed semi-parametric literature, we could mention three types of models. The first one is the extension of the StoNED method developed by Johnson and Kuosmanen (2011) where the z-variables are incorporated additively to the parametric part of the function which is estimated jointly with the nonparametric frontier. Kuosmanen (2012) has recently applied this approach for the case of the electricity distribution sector in Finland.[13] Alternatively, Li *et al.* (2002) introduced the Semiparametric Smooth Coefficient Model (SPSCM) where the regression coefficients are unknown functions which depend on a set of contextual variables. Sun and Kumbhakar (2013) extend this model by allowing the environmental variables to also

[13] This method has been adopted by the Finnish regulator since 2012.

enter through the inefficiency. Finally, the use of an LCSFM approach allows the identification of different technology parameters for different groups of firms that share environmental features. In an LCSFM the z-variables enter in non-linear form in the probabilities of belonging to the classes, and hence they can be viewed as a "discrete, semi-parametric approximation to the random parameters model" (Greene, 2005b, p.299). The use of the LCSFM in efficiency analysis was proposed by Orea and Kumbhakar (2004) and Greene (2005b).

The third approach included in Figure 2.1 involves several parametric models where the contextual variables are treated as inefficiency determinants.[14] They can be divided in three groups depending on how the z-variables are introduced in the model. As the inefficiency term in these models is defined as the truncation (over zero) of a normal distributed random variable, the contextual variables can be introduced in the model either through the mean as in Kumbhakar *et al.* (1991) and Battese and Coelli (1995), the variance as in Reifschneider and Stevenson (1991) or Caudill and Ford (1993), or simultaneously through the mean and variance, as in Alvarez *et al.* (2006) or Lai and Huang (2010). As this is the approach used in this chapter, more details about these models can be found in the next section.

2.3. Theoretical model and empirical specification

In this section we introduce the theoretical cost model that allows us to analyse the economic characteristics of the technology, such as economies of scale or economies of density, of US electricity transmission firms. In general terms, the cost function to be estimated can be written as:

$$\ln C = \ln C\left(y, n, p, d, t\right) \tag{2.1}$$

where C is a measure of total costs, y is a vector of outputs, n measures the network length, p stands for input prices, d is a set of regional dummies and t represents the time trend. As usual, if firms minimize cost, this function should be linearly homogeneous with respect input prices, and increasing in outputs.[15]

Economies of scale and density of electricity transmission firms can be computed once equation (2.1) is estimated. We associate economies of scale with *horizontal* system expansion, that is, increases in demand that require enlarging the current network to meet extra demand.[16] These economies can be then measured by the sum of cost elasticities with respect to the outputs, y, and the network length, n:

$$ES = \frac{\partial \ln C}{\partial \ln y} + \frac{\partial \ln C}{\partial \ln n} \tag{2.2}$$

On the other hand, we associate economies of density with *vertical* system expansion, i.e. expansion in transmitted electricity that do not require additional

[14] An interesting issue here is whether environmental variables should be included in the frontier as well (see later on the discussion in Section 2.4).

[15] Our cost variable is total expenditure (i.e. operating plus capital costs) due to the presence of possible trade-offs between operating and capital expenditures (Giannakis *et al.*, 2005). Regarding the set of output variables, we include the peak demand, transmission capacity and the energy delivered as cost drivers in electricity transmission (see Ofgem 2011, p.44-46).

[16] Note that here density is held constant because both output levels and network size is expanded simultaneously.

network. These economies can be measured by the sum of elasticity of cost with respect to the outputs, y:

$$ED = \frac{\partial \ln C}{\partial \ln y} \tag{2.3}$$

In this case, the cost elasticity of network is not taken into account, as we are considering an increase in output levels, given the actual length of the transmission network.

We next allow for deviations with respect to the above cost function. The stochastic frontier literature suggests that these deviations should not be entirely attributed to uncontrollable or unobservable factors (i.e. random noise) but also to (managerial) inefficiency. To capture both sources of deviations, Aigner, Lovell and Schmidt (1977) proposed using an econometric specification of the cost function (2.1) that includes two random terms. This model (ALS henceforth) can be presented as follows:

$$\ln C_{it} = \alpha + X_{it}'\beta + v_{it} + u_{it} \tag{2.4}$$

where i stands for firms and t for time, X_{it} is a vector of explanatory variables, α and β are parameters to be estimated, $v_{it} \sim N(0,\sigma_v^2)$ is the classical symmetric random noise, and u_{it} is a one-side error term which captures firms' inefficiency.

ALS assumed that this term follows a homoscedastic half-normal distribution, i.e. $u_{it} \sim N^+(0,\sigma_u^2)$. As the inefficiency term in ALS has constant variance, it does not allow the study of the determinants of firms' performance, which is the main issue examined in this chapter. It might also yield biased estimates of both frontier coefficients and firm-specific inefficiency scores (see Caudill and Ford, 1993). To deal with this issue, we propose estimating a heteroscedastic frontier model that allows incorporating z-variables in the model as efficiency determinants. As there are several options to achieve this aim using a parametric approach (see Figure 2.1) and the specific assumptions considered in these models might condition our results,[17] we explore several specifications of the model and carry out model selection tests to choose the "best" model. In that sense Coelli *et al.* (2005) suggest exploring alternative models to assess the adequacy and robustness of the results obtained when a parametric approach is applied.

The most general specification of u_{it} that we consider in this chapter is the General Exponential Model (GEM hereafter) introduced by Alvarez *et al.* (2006) that can be written as:[18]

---

[17] Similar problems might emerge when non- or semi-parametric approaches are used instead of a parametric approach. For instance, Martins-Filho and Yao (2013) point out that although the nonparametric approach considered by Kumbhakar *et al.* (2007) for estimating stochastic frontiers is quite general, the problem known as the curse of dimensionality could occur when the number of explanatory variables is large. This implies that one cannot be confident about the accuracy of the asymptotic approximation and the reliability of the efficiency estimates. Another example is the semi-parametric method known as StoNED presented by Kuosmanen (2012). This model allows introducing environmental variables in the model, but they can be interpreted either as factors that explain the inefficiency, or alternatively, as heterogeneity. Therefore this approach does not address whether environmental variables have direct or indirect effects. We discuss this issue in Section 2.4.

[18] Here we have adopted the notation used by Alvarez *et al.* (2006) and Lai and Huang (2010). Moreover, following Alvarez *et al.* (2006), we will use hereinafter the exponential functional form for the functions that incorporate environmental variables in all the estimated models.

$$u_{it} \sim N^{+}\left(\mu_{it}, \sigma_{uit}^{2}\right) \tag{2.5}$$

where

$$\mu_{it} = \exp\left(\delta_0 + z_{it}'\delta\right)$$

$$\sigma_{uit} = \exp\left(\gamma_0 + z_{it}'\gamma\right)$$

and δ_0, δ, γ_0 and γ are parameters to be estimated, and z_{it} is a vector of efficiency determinants. The two intercepts δ_0 and γ_0 in (2.5) allow us to get the homoscedastic frontier models. The environmental variables enter in this model both through the pre-truncated mean and the variance of the inefficiency term, and hence it allows for non-monotonic effects of the z-variables on firms' inefficiency (see Wang and Schmidt, 2002). Despite being a more comprehensive model than those usually presented in SFA, it is rarely estimated in the literature. For robustness grounds, we will also estimate more restricted models that are nested in the GEM and then some model selection tests will be performed for choosing the preferred specification.

The second estimated model is the proposed by Kumbhakar, Ghosh and McGuckin (1991), Huang and Liu (1994) and Battese and Coelli (1995) (hereafter KGMHLBC model). All of them consider a specification in which only the mean of the pre-truncated normal variable depends on environmental variables. In other words, it is assumed in this model that $\gamma=0$ in (2.5) and thus the variance of the pre-truncated normal variable is homoscedastic, i.e. $u_{it} \sim N^{+}(exp(\delta_0+z_{it}'\delta), \sigma_u^2)$, where for notational simplicity we have relabelled $exp(\gamma_0)$ as σ_u.

The last two models are similar to the one estimated by Reifschneider and Stevenson (1991), Caudill and Ford (1993) and Caudill, Ford and Gropper (1995) (henceforth RSCFG model). In these papers the environmental variables are treated as determinants of the variance of the pre-truncated normal variable. In other words, they assume that $\delta=0$ in (2.5) and thus $u_{it} \sim N^{+}(\mu, (exp(\gamma_0+z_{it}'\gamma))^2)$, where for notational ease $exp(\delta_0)$ has been relabelled as μ. If μ is allowed to be different from zero, we get the RSCFG-μ model introduced by Alvarez *et al.* (2006). This model nests the original RSCFG model in which μ=0 is imposed (i.e. δ_0=-∞ is assumed) and therefore it assumes that u_{it} follows a half-normal distribution, i.e. $u_{it} \sim N^{+}(0, (exp(\gamma_0+z_{it}'\gamma))^2)$. As a consequence of this assumption, the so-called *scaling property* is satisfied in this model in the sense that the inefficiency term can be written as a deterministic function of a set of efficiency covariates, i.e. $h(\cdot)$=exp($z_{it}'\gamma$), times a one-sided random variable that does not depend on any efficiency determinant, $u_{it}^{*} \sim N^{+}(0, \sigma_u^2)$.

The defining feature of models with the scaling property is that firms differ in their mean efficiencies, but not in the shape of the distribution of inefficiency. That is, the scaling property implies that changes in z_{it} affect the scale but not the shape of u_{it}. In this model u_{it}^{*} can be viewed as a measure of "basic" or "raw" inefficiency that does not depend on any observable determinant of firms' inefficiency. On the other hand, the scaling function $h(\cdot)$ can be interpreted as the portion of total estimated inefficiency that researchers are able to explain with the variables included in $h(\cdot)$. This function hence "adjusts" the underlying, and unexplained, inefficiency level upwards or downwards due to the influence of some potential inefficiency determinants. Although it has some features that make it attractive to some authors (see Wang and Schmidt, 2002), it is an empirical question whether or not the scaling property should be imposed, and not all commonly used models fulfil this property.[19]

To fully justify the choice of our preferred specification we will use the standard Likelihood Ratio (LR) test when comparing nested models (i.e. GEM vs. RSCFG-μ, GEM vs. KGMHLBC, RSCFG-μ vs. RSCFG, and RSCFG vs. ALS) and the Vuong (1989) test when they are non-nested (i.e. RSCFG-μ *vs.* KGMHLBC). It should be mentioned here that, although the standard RSCFG model is nested in the GEM model, they cannot be directly compared using standard LR tests because the GEM coefficients of the pre-truncated mean (i.e. δ) are not identified when μ=0 (as assumed in the RSCFG model). For the same reason, the ALS model cannot be compared against the KGMHLBC model using standard tests (i.e. δ is again not identified when μ=0). To test if μ=0, Alvarez *et al.* (2006) suggest carrying out a simple LR test using the RSCFG and RSCFG-μ models.

2.4. Data and sample

We use a panel data set of 59 US electricity transmission companies for the period 2001-2009. Most of these data were collected by various members of the EPRG at the University of Cambridge. That information was requested by Ofgem, in order to carry out an international benchmarking of electricity and gas utilities. Where the transmission operations are part of a larger utility - also involved in generation or distribution - shared costs are allocated on pro-rata basis. As can be seen in the data Appendix (Section 2.7), an allocation key based on the ratio between wages and salaries specific from transmission and the total labour expenses of the utility, were used for the assignment of shared costs to transmission. The main source of the electricity transmission data was the FERC form 1, an annual report of major electric utilities, and the variables collected included the quantity of assets, voltage levels by asset, maximum demand, load density, demand growth, maturity of service area, age/condition of network, network density and flow patterns.[20]

Although the choice of input and output variables is an important issue, there is no clear consensus about the variables that should be included to describe the performance of transmission and distribution companies. Jamasb and Pollitt (2001) show the wide range of variables that have been used in benchmarking analysis of electric utilities. They find that the most commonly used inputs in studies of electric

---

[19] Another model that also satisfies this property is the so-called *scaled Stevenson* (SS) model introduced by Alvarez *et al.* (2006). In this model, both the mean and the variance of the pre-truncated normal depend on the environmental variables but the coefficients of the environmental variables in the mean and variance of u in (2.5) are the same, i.e. δ=γ. We will not provide the parameter estimates of this model in Section 2.5 because it collapsed to the KGMHLBC.

[20] The original dataset was collected by the members of the EPRG and it includes information of electricity and gas utilities in the US from 1994 to 2009 and also contains information on non-US firms from other countries for a shorter period. Following Ofgem's (2011, p. 20) report, non-US transmission firms were not included in the analysis due to data limitations. Despite the initial proposal on international benchmarking in that report, so far, these data have not been used. In this chapter the sample was reduced to the last 9 years because labour costs in the electric power transmission industry are only available from 2001 to 2009. We have removed observations with missing and implausible values. We have also dropped a few isolated observations and maintained firms with (at least three) consecutive observations in order to minimize changes in our estimates when we change the specification of our model. It should be noted that this procedure does not give us a balanced panel, as we do not have the same number of observations per firm. Our final sample is an unbalanced panel data set of 402 observations without discontinuities across time.

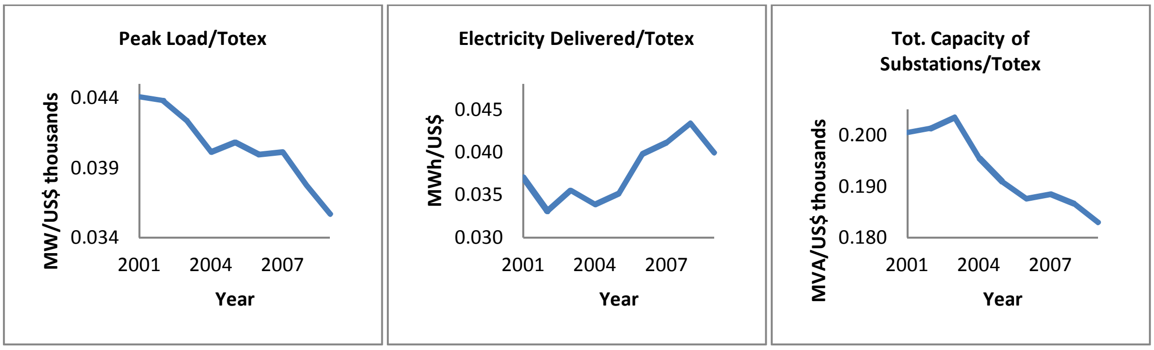
utilities are operating costs, number of employees, transformer capacity, and network length. Regarding the outputs, the most included variables are units of energy delivered, number of customers, and the size of service area.

As we have mentioned in Section 2.3, our cost variable is Totex. This variable is the sum of Opex, which includes operation and maintenance expenses incurred by the company over one year, and Capex, which is the sum of annual depreciation on capital assets and the annual return on the balance of capital. Both Opex and Capex (and hence also Totex) are measured in 2000 dollars.[21]

Following the basic economic theory of production and the literature on electricity networks, we use as explanatory variables of total cost: three types of outputs, a variable that measures the system size, labour and capital price, a set of regional dummies and a time trend. Our output variables are: *Peak Load* (PL), *Electricity Delivered* (DE) and *Total Capacity of Substations* (CS). While the first one is the maximum peak load of the year during 60 minutes and it might reflect transmission investment requirements given a fixed transmission capacity, the second one is the total annual energy delivered by the system which may imply an incremental effect in operating cost due to a greater use of electricity transmission assets. Due to a large amount of missing values in the data about voltage levels, we have introduced the CS as a proxy for the transmission capacity of the system. It is calculated as the sum of the total capacity of all substations in the transmission network.

In Figure 2.2 we show the evolution over time of the output variables divided by Totex, which can be interpreted as partial and observable productivity (efficiency) measures.[22] We can see in this figure a clear negative trend of the peak loads and the total capacity of substations given the total expenditure of each firm. In the case of electricity delivered, the temporal pattern of this variable is not so clear. These graphs give us a first idea about the negative evolution of the efficiency in our sample as the output level per dollar of cost, decreases, or in other words, the total unit cost per output, increases over time.

Figure 2.2. Annual evolution of outputs divided by Totex



[21] RTO costs are included in the total costs. For more information about the calculation of Totex and the rest of variables, see the Appendix.

[22] As we have an unbalanced panel of 59 firms, to depict this figure we have selected those firms that are observed during the whole sample period, i.e. 28 firms. This avoids comparing different sets of firms in different periods.

Network length (NL) is usually viewed as one of the most important cost drivers of an electricity network (Jamasb and Pollitt, 2001). To measure the network length we have used pole miles. This variable measures the total sum of all transmission lines in miles regardless of the number of power cables on each power line so it is essentially a measure of the geographic spread of each company. We thought about using circuit miles instead pole miles, but the problem of circuit miles is that this variable refers to the number of power cables on each line multiplied by the distance between two points, but it does not take into account the capacity of the cable so it is an unreliable measure of the physical infrastructure.

Regarding input prices, we include in the cost function a *Labour Price* variable (LPR) defined as the average annual wage for the electric power transmission and distribution industry by state. As in the case of Totex, this variable is also measured in 2000 dollars.[23] Regarding the *Capital Price* variable (KPR), we have finally used a producer price index for power transmission as a proxy for capital price.[24] The source of these two variables is the Quarterly Census of Employment and Wages from the Bureau of Labor Statistics.

Taking into account the importance of controlling for differences in business environment from the perspective of corporate structures after US market liberalization, we have also included seven regional dummies that represent the regional reliability councils of the NERC in which the transmission utilities of our sample are located: *SERC Reliability Corporation* (SERC), *Southwest Power Pool* (SPP), *Western Electricity Coordinating Council* (WECC), *Northeast Power Coordinating Council* (NPCC), *ReliabilityFirst Corporation* (RFC), *Midwest Reliability Organization* (MRO) and *Electric Reliability Council of Texas* (ERCOT). We expect that these regional dummy variables are capturing (jointly with the other variables included in the model) most of the unobserved differences between transmission companies' tasks, such as transportation of electricity, scheduling and dispatching of the plants, investment and maintenance of transmission assets, etc.[25]

Regardless the introduction of these variables, we think that there are three issues that should be mentioned related to the business environment as they might make a difference to transmission system efficiency in theory. The first is the presence or absence of incentive regulation in transmission. We have not included information about incentive regulation in our cost function as we do not have data on it. This is partly because each state is different and indeed each firm may have a different arrangement with its regulator. Identifying the arrangement for the transmission business as separate from the distribution business would be hard and a time series of the regimes would be needed for analysing this point. However we think this issue

[23] Unfortunately this information is not available at firm-level. Although it would be preferable to use firm-specific prices instead of state-level prices, as firm-level price data are not available in our application for both labour and capital, we have used the information that we found from statistical agencies. Clearly input prices do vary significantly across the US and it would be wrong not to adjust for them. In addition, we do not think there is much of a multi-state issue as the interesting thing is that transmission lines in the US are in fact mainly within one state.

[24] We have estimated our models using several indices and variables calculated with financial information of the companies. Their coefficients were not statistically significant or they even had unreasonable magnitudes from an economic point of view.

[25] Another option to deal with this issue is using a model with fixed effects (see Greene, 2005a, 2005b, and more recently Wang and Ho, 2010). However, this estimation strategy does not easily deal with rarely changing variables, i.e. variables with little within or temporal variation such as network length or energy delivered. For a discussion on this issue, see Greene *et al.* (2011).

should not affect the soundness of our results as it is not altogether clear what difference these things might actually make. Furthermore a detailed investigation of incentive regulation on efficiency is clearly out of the scope of this chapter.

The second issue is the introduction of nodal pricing into the RTO which might sharpen the pressure on transmission businesses to make lines available. However we have estimated a model including dummy variables that reflect the belonging to a certain RTO and this model is rejected in favour of our preferred model, which incorporates regional dummies for the NERC regions. Therefore our estimates suggest that once heterogeneity is controlled, the belonging to a certain RTO has a negligible impact on firms' efficiency. This may be because transmission systems have 99%+ availability, and hence the introduction of nodal prices may not have affected firms' performance. Furthermore RTOs do not 'regulate' total transmission revenue, so it is not clear why RTO membership should affect cost efficiency. This is because often it is overall revenue of a transmission business that is regulated and poor revenue performance due to low availability on one line may lead to increased charges elsewhere.

Lastly, the third issue is the degree of vertical integration. Vertical integration might be independently significant, simply because of cost allocation issues and fixed costs being spread. Our preferred model does not contain any vertical integration variables (in particular backward integration into generation and the forward integration into distribution), because they were only significant for the 3% of the observations in our sample.[26]

Regarding the stochastic part of our cost function, we use 9 variables that are expected to affect firms' performance and, hence, they are included as efficiency determinants. In particular, we include the following variables: another time trend, three weather variables (minimum temperature, wind and precipitation), the Capex/Opex ratio and two variables which measure the growth of the demand.

Our weather variables have been obtained from the surface daily weather information collected by the National Climatic Data Center for the 2001–2009 period. The files are available for around 3,000 weather stations located in the US and contain information about: mean, maximum and minimum temperatures, precipitation amount, wind speed, number of days with snow, hail, tornadoes, etc. Given the high correlation among several weather variables, we decided to include one variable for each one of these categories: *Temperature* (TMIN), *Precipitation* (PRCP) and *Wind* (WIND). The temperature variable is the annual minimum temperature in Fahrenheit degrees, wind speed is the average of the daily mean wind speeds in knots, and precipitation is the average of the daily precipitation in inches. These weather variables are measured at state-level, not at firm-level. In order to obtain a unique value of each variable per state and year, we have taken the average among the weather stations within a particular state except for the case of the temperature variable which is the minimum value measured by any of the above stations along the year. Then, each utility was associated with the weather of the state where its principal office is located.[27] We hereafter assume that more adverse conditions appear when wind speed and precipitation are high and minimum temperature is small. These weather variables have also been introduced in

[26] To examine this issue we have applied a modified version of the 'zero inefficiency stochastic frontier model' recently introduced by Kumbhakar *et al.* (2013).
[27] We recognise that this is a limitation especially when transmission companies may cover more than one state.

the cost function as determinants of the technology, i.e. of the frontier cost function, in some of our estimated models.

As utilities may adapt their operating and investment practices over time to prevent power interruptions and to reduce the effect of adverse weather conditions, we interact our weather variables with the mean of the ratio of Capex and Opex (COR) for each firm i over the T_i available observations for this firm. We expect a negative coefficient if investing in Capex is a better strategy rather than incurring additional operational and maintenance costs in dealing with adverse weather conditions.

Finally we have included two variables that measure the average *Growth in Demand* for each firm over time. We distinguish between positive growth (POSGR) and negative growth (NEGR). The coefficients of these two variables should not be statistically significant if there are no adjustment costs and all inputs can be adjusted (without cost) from one year to the next. However, as the electricity industry is highly intensive in capital with much of the assets becoming sunk cost upon investment, we expect significant coefficients for POSGR and NEGR. In particular, we expect a positive effect of POSGR on inefficiency indicating that utilities tend to anticipate future increases in the demand by investing in capital that is expected to be efficiently used in the future, but not in the present.[28] We expect a negative coefficient NEGR if there is a negative trend in demand and reducing quasi-fixed input levels is expensive due to the existence of adjustment costs.

The descriptive statistics of all monetary, physical and environmental variables used in the stochastic cost frontiers are shown in Table 2.1.

Table 2.1.Descriptive statistics

	Variable	*Units*	*Mean*	*Max.*	*Min.*	*Std. Dev.*
Totex	Cost	US$	145,111,000	667,127,000	20,713,600	120,627,000
Peak Load	Output	MW	6,208	23,111	380	5,539
Electricity Delivered	Output	MWh	6,279,730	74,584,700	56,730	8,872,920
Tot. Cap. of Subs.	Output	MVA	27,821	120,115	1,327	22,720
Network Length	Network	Miles	4,073	16,292	1,087	3,263
Annual Salary	Input Price	US$	62,144	94,005	34,024	10,531
Producer Price Index	Input Price	Index	179.21	222.40	155.00	21.35
SERC	Dummy	-	0.40	1	0	0.49
SPP	Dummy	-	0.22	1	0	0.41
WECC	Dummy	-	0.26	1	0	0.44
NPCC	Dummy	-	0.04	1	0	0.21
RFC	Dummy	-	0.25	1	0	0.43
MRO	Dummy	-	0.14	1	0	0.35
ERCOT	Dummy	-	0.04	1	0	0.21
Minimum Temp.	Weather	°F	-10.35	19.90	-59.80	16.57

Continued on next page

[28] As Jamasb and Pollitt (2007) note, achieving long-term efficiency improvements can involve short-term increases in Capex or Opex that may not generate immediate efficiency improvements. In fact, increases in short-term expenditure can deteriorate the firms' short-term relative performance. This might in turn discourage firms from efficiency-improving investments that have long-term gains.

Continued from previous page

	Variable	*Units*	*Mean*	*Max.*	*Min.*	*Std. Dev.*
Wind Speed	Weather	Knots	6.83	9.60	4.63	1.01
Precipitation	Weather	Inches	0.07	0.16	0.01	0.03
Capex/Opex	Other	Ratio	1.18	5.90	0.13	0.70
Growth in Demand	Other	%	0.03	244.11	-74.96	17.77

2.5. Empirical results

We estimate a Translog cost function that can be interpreted as a second-order approximation to the companies' underlying cost function.[29] All the variables are included in the model in logarithms, except the regional dummies and the time trend. Each explanatory variable is measured in deviations with respect to its mean, so the first-order coefficients can be interpreted as the cost elasticities evaluated at the sample mean. As usual, homogeneity of degree one in prices is imposed by normalizing cost and labour price with capital price. Thus, the estimated equation can be written as follows:

$$\begin{aligned}
\ln\left(\frac{C_{it}}{KPR_{it}}\right) &= \alpha + \sum\nolimits_{p=1}^{4} \beta_p \ln y_{pit} + \frac{1}{2}\sum\nolimits_{p=1}^{4}\sum\nolimits_{q=1}^{4} \beta_{pq} \ln y_{pit} \ln y_{qit} + \\
&\beta_L \ln\left(\frac{LPR_{it}}{KPR_{it}}\right) + \frac{1}{2}\beta_{LL}\left[\ln\left(\frac{LPR_{it}}{KPR_{it}}\right)\right]^2 + \sum\nolimits_{p=1}^{4} \beta_{pL} \ln y_{pit} \ln\left(\frac{LPR_{it}}{KPR_{it}}\right) + \\
&\sum\nolimits_{d=1}^{7} \beta_d NERC_d + \beta_t t + u_{it} + v_{it}
\end{aligned} \tag{2.6}$$

where for notational ease, the vector y stands for outputs and network length, i.e. y=(PL, DE, CS and NL).

As our results about firms' efficiency might depend on the empirical strategy followed to allow for inefficiency determinants, we first carry out several model selection tests to select the best specification supported by the data. Table 2.2 shows the LR tests for nested models, where the second model presented in each line is nested in the first model. Firstly we can see that the ALS model is rejected in favour of the RSCFG model due to the inclusion of environmental variables in the variance of the heteroscedastic inefficiency term. This latter model is in turn rejected in favour of the RSCFG-μ, indicating that the inefficiency term does not follow a half normal distribution. Table 2.2 also displays the Vuong test for the non-nested RSCFG-μ and KGMHLBC models. A positive value indicates that the first model is preferred to the second one. In this case we can see that the preferred model is the RSCFG-μ model. Lastly, the LR tests in Table 2.2 again indicate that the GEM clearly outperforms both RSCFG-μ and KGMHLBC models. Based on these comparisons, we will use the GEM model to examine in detail the estimated levels of cost efficiency.

In addition to the frontier parameters that are discussed later on, Table 2.3 displays the coefficients of the inefficiency term that have been estimated using the standard homoscedastic ALS model, and the heteroscedastic models presented before: RSCFG, RSCFG-μ, KGMHLBC and GEM. Although the environmental variables in the GEM model are included both in the mean and the variance of the inefficiency term,

[29] The more restricted Cobb-Douglas specification was always rejected in favour of the Translog specification.

their main effect is through the variance. Indeed, whereas most of the coefficients of the variables included as determinants of the variance of the inefficiency term are statistically significant, the estimated coefficients for the mean are not significant (except for the time trend which is negative and significant at a 90% confidence level). This is in line with our finding that the coefficients of these variables in the RSCFG and RSCFG-μ models are also significant, but not in the KGMHLBC model. The latter model clearly shows that the mean of the inefficiency is not able to capture the effect of the environmental variables on firms' inefficiency. Thus, we will focus our comments on the variance of the inefficiency term.

Table 2.2. Model selection tests

Comparison of nested models (LR test)	*Test value*	*D.o.f.*	*Preferred model*
RSCFG vs. ALS	74.052 ***	9	RSCFG
RSCFG-μ vs. RSCFG	37.137 ***	1	RSCFG-μ
GEM vs. RSCFG-μ	18.163 **	9	GEM
GEM vs. KGMHLBC	101.802 ***	9	GEM

Comparison of non-nested models (Vuong test)	*Test value*		*Preferred model*
RSCFG-μ vs. KGMHLBC	1.830 **		RSCFG-μ

Significance code: * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Regarding the inefficiency variance, our results indicate that weather is an important issue in this industry. [30] Wind speed and precipitation have a positive and significant coefficient indicating that more adverse conditions generate higher levels of inefficiency. The negative sign for the minimum temperature also suggests, although it is not significant, that lower minimum temperature slightly increases cost due to higher levels of firms' inefficiency. Obviously firms cannot adjust their performance by modifying their weather conditions as weather is an uncontrollable factor. Our results simply indicate that managing firms operating in regions with bad weather is more difficult than other firms enjoying better conditions. Therefore, policy measures aiming to improve managerial skills become more appropriate in firms operating in adverse conditions. In this sense, the introduction of the average ratio of Capex and Opex (COR) interacting with the weather variables allows us to catch an idea about the best strategy to deal with adverse weather conditions. The estimated coefficients have the opposite sign to those obtained for the isolated weather variables, indicating that, as expected, more capital-intensive utilities (e.g. with higher capital-to-opex ratios) are able to mitigate better the effect of unfavourable weather conditions. They are, *ceteris paribus*, more efficient than those utilities using a higher proportion of operating inputs.

[30] Note also that the coefficient of the time trend is positive, showing that the effect of time is different in both parts of the inefficiency.

Table 2.3. Parameter estimates of the Translog cost function

		ALS			*RSCFG*			*RSCFG-μ*			*KGMHLBC*			*GEM*		
	Parameters	***Est.***		***Est./s.e.***	***Est.***		***Est./s.e.***	***Est.***		***Est./s.e.***	***Est.***		***Est./s.e.***	***Est.***		***Est./s.e.***
Frontier	Intercept	13.208	***	169.471	13.391	***	231.108	12.573	***	83.435	13.394	***	65.985	13.294	***	202.786
	ln PL_{it}	0.508	***	8.148	0.523	***	11.160	0.452	***	11.161	0.616	***	11.249	0.545	***	12.233
	ln DE_{it}	0.057	***	2.808	0.060	***	3.527	0.024	**	2.002	0.040	*	1.853	0.061	***	3.945
	ln CS_{it}	0.138	**	2.034	0.164	***	2.896	0.349	***	6.917	0.056		0.884	0.140	**	2.517
	ln NL_{it}	0.145	***	3.918	0.135	***	4.296	0.016		0.583	0.064	*	1.776	0.145	***	4.841
	ln (LPR_{it}/KPR_{it})	0.582	***	3.722	0.528	***	4.399	0.448	***	4.398	0.422	**	2.502	0.497	***	4.113
	½ $(\ln PL_{it})^2$	-0.057		-0.288	-0.037		-0.223	-0.248	**	-1.999	0.271		1.308	0.044		0.292
	½ $(\ln DE_{it})^2$	0.038		1.415	0.040	**	2.198	0.032	**	2.021	0.017		0.639	0.044	**	2.338
	½ $(\ln CS_{it})^2$	0.167		0.526	0.109		0.422	0.077		0.372	0.304		0.825	0.119		0.536
	½ $(\ln NL_{it})^2$	0.270	**	2.077	0.247	***	2.962	0.380	***	6.213	0.196	*	1.677	0.253	***	3.278
	½ $(\ln (LPR_{it}/KPR_{it}))^2$	0.121		0.183	-0.139		-0.278	-0.319		-0.852	-0.159		-0.235	-0.012		-0.026
	ln PL_{it} · ln DE_{it}	-0.005		-0.085	-0.032		-0.764	0.009		0.273	0.037		0.619	-0.021		-0.540
	ln PL_{it} · ln CS_{it}	0.015		0.060	0.077		0.368	0.240		1.507	-0.346		-1.220	-0.002		-0.013
	ln PL_{it} · ln NL_{it}	0.061		0.482	0.182	**	2.098	0.182	***	2.837	0.002		0.019	0.155	*	1.800
	ln PL_{it} · ln (LPR_{it}/KPR_{it})	-0.152		-0.500	-0.085		-0.323	0.084		0.401	-0.277		-0.978	-0.160		-0.697
	ln DE_{it} · ln CS_{it}	-0.028		-0.455	0.013		0.258	-0.030		-0.827	0.004		0.077	-0.001		-0.020
	ln DE_{it} · ln NL_{it}	-0.042		-1.114	-0.082	***	-2.818	-0.046	*	-1.941	-0.052		-1.159	-0.093	***	-3.549
	ln DE_{it} · ln (LPR_{it}/KPR_{it})	0.100		1.239	0.084		1.279	0.086		1.396	0.162	**	2.035	0.091		1.394
	ln CSit · ln NLit	-0.050		-0.313	-0.165		-1.528	-0.313	***	-3.947	0.167		1.045	-0.091		-0.781
	ln CS_{it} · ln (LPR_{it}/KPR_{it})	0.056		0.165	0.051		0.182	-0.263		-1.127	0.013		0.038	-0.004		-0.013
	ln NL_{it} · ln (LPR_{it}/KPR_{it})	0.057		0.295	-0.212		-1.274	-0.013		-0.099	0.194		0.956	-0.168		-1.050
	SERC	-0.372	***	-5.889	-0.367	***	-7.050	-0.392	***	-9.280	-0.488	***	-7.050	-0.372	***	-7.291
	SPP	0.154	**	2.206	0.152	***	3.101	0.190	***	4.367	0.193	**	2.246	0.213	***	4.049
	WECC	-0.185	***	-2.584	-0.060		-1.071	-0.031		-0.554	-0.178	**	-2.021	0.013		0.202
	NPCC	0.130		0.923	0.106		0.870	0.251	***	3.014	0.163		1.103	0.153		1.299
	RFC	-0.127	*	-1.848	-0.161	***	-2.901	-0.151	***	-4.123	-0.068		-0.783	-0.081		-1.298
	MRO	0.051		0.585	0.060		0.778	0.040		0.680	0.046		0.446	0.145	*	1.909
	ERCOT	0.242	**	2.277	0.235	***	2.849	0.248	***	3.306	0.377	***	3.467	0.236	***	3.081
	t	0.000		-0.033	-0.029	***	-4.097	-0.014	**	-2.533	0.004		0.534	-0.025	***	-3.731

Continued on next page

Continued from previous page

		ALS			RSCFG			RSCFG-μ			KGMHLBC			GEM		
	Parameters	**Est.**		**Est./s.e.**	**Est.**		**Est./s.e.**	**Est.**		**Est./s.e.**	**Est.**		**Est./s.e.**	**Est.**		**Est./s.e.**
Noise term	ln (σ_v^2)	-1.946	***	-19.879	-1.825	***	-35.791	-2.185	***	-25.317	-1.558	***	-22.185	-1.918	***	-35.928
Inefficiency term (mean)	Intercept							-0.165		-0.966	-4.420		-0.946	-5.525	*	-1.732
	t										-0.338		-1.256	-0.310	*	-1.735
	$TMIN_{it}$										0.125		0.911	0.141		1.353
	$WIND_{it}$										-0.936		-1.080	-1.510		-1.357
	$PRCP_{it}$										7.659		0.425	0.725		0.045
	$TMIN_{it} \cdot COR_i$										-0.117		-0.626	-0.188		-1.086
	$WIND_{it} \cdot COR_i$										0.560		0.414	-1.405		-0.716
	$PRCP_{it} \cdot COR_i$										7.852		0.214	19.679		0.517
	$POSGR_i$										0.077		1.011	0.033		0.069
	$NEGR_i$										0.166		0.363	0.104		0.269
Inefficiency term (variance)	Intercept	-1.261	***	-12.217	-3.871	***	-7.650	-2.767	***	-12.197	-3.984		-0.312	-3.965	***	-8.248
	t				0.285	***	5.860	0.097	***	3.341				0.285	***	6.006
	$TMIN_{it}$				-0.015		-1.245	-0.001		-0.067				-0.013		-1.114
	$WIND_{it}$				0.286		1.618	0.340	**	2.546				0.502	***	2.733
	$PRCP_{it}$				27.643	***	4.001	15.809	***	3.680				29.302	***	4.251
	$TMIN_{it} \cdot COR_i$				0.062	**	2.362	0.037	*	1.868				0.078	***	2.955
	$WIND_{it} \cdot COR_i$				-0.323		-0.831	-0.678	**	-2.238				-0.139		-0.386
	$PRCP_{it} \cdot COR_i$				-23.963	**	-2.178	-38.825	***	-4.227				-30.327	***	-2.682
	$POSGR_i$				0.042	***	3.601	0.071	***	8.647				0.038	***	3.212
	$NEGR_i$				0.029		0.626	0.091	*	1.710				0.025		0.600
	Obs.	402			402			402			402			402		
	Log-likelihood	41.179			78.204			96.772			54.953			105.854		

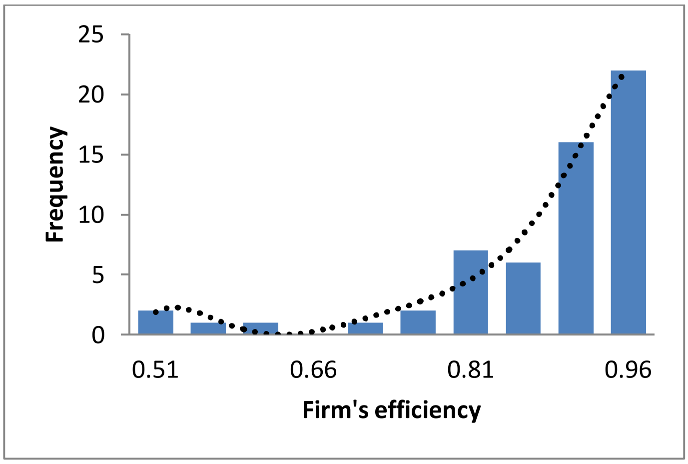
Significance code: * p<0.1, ** p<0.05, *** p<0.01

This result suggests therefore that investing in equipment is a better strategy than incurring additional operating costs in mitigating the effects of unfavourable weather conditions.

The last set of efficiency determinants has to do with growth in demand. We get a positive, and significant, coefficient for POSGR, indicating that utilities are more efficient when the demand is unchanging as they do not need to anticipate investments to meet future demand. However, the coefficient of NEGR is not significant in most of the models, indicating perhaps that reducing quasi-fixed inputs is not expensive for the companies or that maintaining the underused network is not very costly when there is negative trend in the demand growth.

In Figure 2.3 we depict the histogram of estimated levels of cost efficiency. The average efficiency in our sample is 88% using our preferred model. Pollitt (1995) using 1990 data found an average efficiency of 80% for the total of the companies in his sample and 88.3% for larger firms. The latter value is very similar to the one that we have found with our preferred model. This seems to indicate that the performance of the electricity transmission utilities has not experienced a significant improvement from one period to the next.

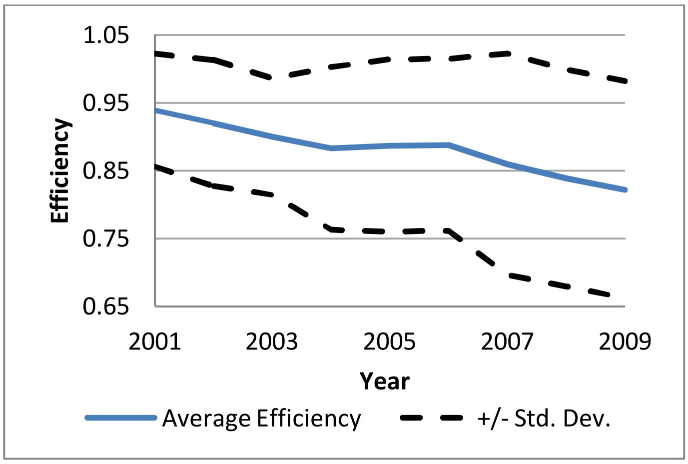
Figure 2.3. Histogram of efficiency scores for the firms using the GEM



We show in Figure 2.4 the temporal evolution of our efficiency scores using the GEM model.[31] The graph shows that the average level decreases over time, starting at 93.9% and finishing at 82.2%, and hence the negative sign of the coefficient for the time trend through the mean of u in our model seems not to offset its positive value through the variance. Our preferred model also indicates an increasing divergence in firms' performance over time. Overall, the estimated evolution in performance and the lack of convergence in firms' inefficiency scores seem to suggest that there is scope for improvements in the performance of the US electricity transmission system.

[31] Except for the RSCFG and the RSCFG-μ models which exhibit a similar evolution of the efficiency (not shown), the rest of the estimated models present clear differences with respect to the GEM, our preferred model. These differences might be taken as an anecdotal evidence of the biases that might appear in an empirical application when inefficiency determinants are not taken into account or are misplaced in the specification of the model.

Figure 2.4. Annual evolution of the efficiency in electric power transmission



We next focus our discussion on the estimated frontier parameters, also shown in Table 2.3. In general, all models perform quite well as most of the first-order coefficients have the expected sign and their magnitudes are quite reasonable from a theoretical point of view. Certainly, the coefficients of the three outputs and network length are always positive and mostly statistically different from zero when measuring the incremental costs associated with either higher maintenance and operational costs or the need for new capital. A similar statement can be made about the coefficients on input prices, which are also positive and statistically significant. The coefficients on many of the dummy variables for the NERC regions are also significant indicating that, regardless of the rest of firms' features, regional differences exist. The coefficient on the time trend is negative (nonetheless it is not significant in some of the models), which indicates that costs decrease over time, i.e. there is technical change.

As the selection of the output set is often quite contentious, we have carried out several LR tests to fully justifying the explanatory variables that were included in our cost frontier function. Table 2.4 shows several tests where our specification of the GEM model is compared to more restricted specifications where one of the outputs is excluded from the output set. As it can be seen all the output variables that were introduced in the model, and primarily peak load (PL) and network length (NL), are relevant cost drivers and should be taken into account in the analysis. We also test in this table the inclusion of dummy variables that reflect the belonging to a certain RTO or alternatively to a certain NERC region. In both cases the LR test values indicate that the model that is rejected is the one that does not include any regional dummy.[32]

Next, we will use our preferred model, GEM, to examine some characteristics of the estimated technology. Like in previous papers, the estimated elasticities allow us to measure economies of scale and density, but in this case using more recent data. Figure 2.5 depicts the elasticity of total cost with respect to peak load, delivered electricity, total capacity of substations and network length estimated for each observation, sorted in increased order at observation-level. Peak load seems to be the most important cost driver with an average elasticity equal to 0.54. This figure also allows us to examine the reliability of our estimated elasticities when we move away from the sample mean. Although the first derivative of our cost function just provides a first-order

[32] Moreover, a Vuong test not shown in Table 2.4 indicates that the model which includes RTO dummies is rejected in favour of our preferred model, which incorporates regional dummies for the NERC regions.

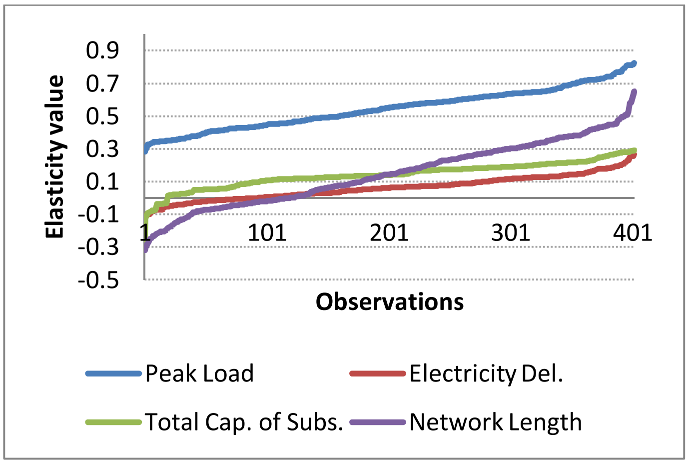
approximation to the underlying elasticity at the sample mean, most observation-specific elasticities are in a reasonable order of magnitude, except for the negative values on the left in three of the curves. In these cases, our estimates should be viewed with caution as they correspond to some observations which are far away from the sample mean.[33]

Table 2.4. Significance of variables in the frontier

	Variables	*Log LF*	*D.o.f.*	*LR Test*
	GEM	105.854	-	-
Output	PL	7.666	6	196.376 ***
Excluded	DE	78.116	6	55.476 ***
	CS	98.580	6	14.548 **
	NL	91.307	6	29.094 ***
	GEM (w/o Reg. Dum.)	54.151	-	-
Regional	NERC	105.854	7	103.406 ***
Dummies	RTO	79.782	6	51.261 ***

Significance code: * p<0.1, ** p<0.05, *** p<0.01

Figure 2.5. Elasticities of cost for outputs and network



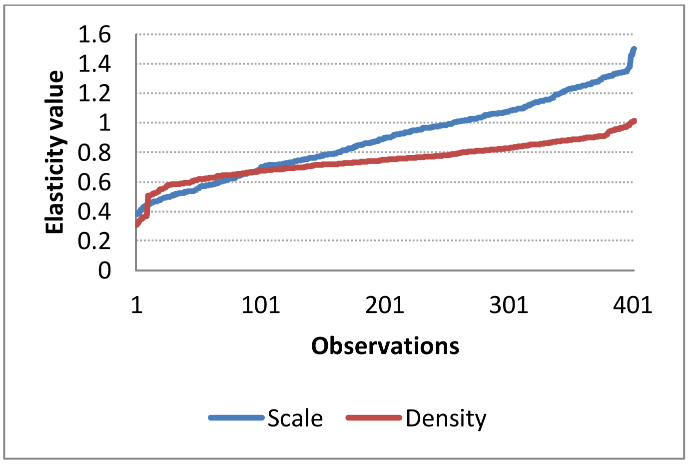
Adding the first-order coefficients of the three outputs we find that the elasticity of density evaluated at the sample mean is quite similar in all models, varying from 0.70 to 0.75. These values suggest the existence of important economies of density in the

[33] For most functional forms (e.g. the Translog function) there is a fundamental trade-off between flexibility and theoretical consistency. For instance, maintaining global monotonicity (e.g. positive elasticities and marginal costs) is impossible without losing second order flexibility. For example, Barnett *et al.* (1996) show that the monotonicity requirement is by no means automatically satisfied for most functional forms, and that violations are frequent. However to show the robustness of our estimates we have tested the monotonicity conditions using the well-known Wald test. We only find statistically negative values for 0.25% of the observations in the case of electricity delivered, 4.98% for the network length and zero for the other outputs. The small number of negative elasticities found gives us confidence about the fulfilment of the monotonicity conditions on outputs and hence about the suitable properties of the estimated cost function.

electricity transmission industry. That is, given network infrastructure, electricity transmission networks exhibit natural monopoly characteristics.

To analyse economies of scale, which involve expansions in both output and network, we need to add the cost elasticity of the network length to the elasticity of density. The elasticity of scale evaluated at the sample mean in the GEM model is 0.89. Figure 2.6 compares both elasticities. More than a half of the firms in our sample exhibit increasing returns to scale. These results suggest that electricity transmission networks still exhibit natural monopoly characteristics when network is expanded to meet the extra demand. Using data for 1990, Pollitt (1995) finds different degrees of economies of scale depending on firms' size for the US transmission utilities. In particular, he finds that decreasing returns to scale are more common in small utilities while increasing returns to scale are more common in medium and large companies. This seems to be consistent with the results obtained here, as in our sample we mainly have large firms. Dismukes *et al.* (1998) also show that all the NERC reliability regions in the US exhibit significant economies of scale for the transmission companies, while Huettner and Landon (1978) find that of six expenses categories, only sales expenses exhibits increasing returns to scale over the whole of the observed output range.

Figure 2.6. Elasticities of scale and density



Finally, as we mentioned in the introduction section, one of the most interesting issues in benchmarking is the question of whether firms are using environmental conditions as excuse for bad performance and whether their costs should be or not adjusted accordingly by the regulator. As pointed out by Nillesen and Pollitt (2010) for the case of US electricity distribution, firms' cost should be adjusted if there is a direct (frontier) effect of weather on cost, but not when the effect is indirect (i.e. through a larger inefficiency). In our case, the use of a LCSFM structure allows us to carry out an observation-by-observation analysis of this important issue, which may be missed in a cost model that only includes the weather variables in one part of the model.

To achieve this aim, we propose estimating a modified version of the so-called zero inefficiency stochastic frontier model introduced by Kumbhakar *et al.* (2013) to examine differences in performance (i.e. inefficiency). Here we have adapted this framework to capture differences in technology. Our LCSFM allows estimating two different cost frontiers: with and without weather variables. Like in the zero-inefficiency model, the other parameters of our model are assumed to be the same in both groups (classes).[34] If most firms belong to the class with no weather variables as determinants of the cost frontier, we then can conclude that our original cost frontier is already capturing their direct effect on firms' costs. The assigning of the firms to a particular group is performed by the model using class-membership probabilities, without any prior assumption by the researcher about the classification of the firms. Those firms with a large probability of belonging to the group where the cost frontier include weather variables can then use the weather conditions to justify (at least partially) their larger costs.

Table 2.5 shows the proportion of observations that are located in either the group with or without weather variables as relevant cost frontier drivers. In both cases, we have used our preferred specification of the inefficiency term, i.e. the GEM. It should be first pointed out that the discriminatory capacity of the LCSFM to allocate firms in different classes is quite robust as the posterior class probabilities are very large. The numbers in this table reflect that only 1% of the observations would be assigned to the class that include weather variables in the frontier (class 2).[35] Similar percentage is also obtained when we add the interactions of the weather variables with the variable measuring firms' cost structure. These results suggest that only the costs of a small number of firms could be adjusted downwards due to the negative influence of the bad weather in order to treat them fairly.

Table 2.5. Modified LCSFM

Basic LCSFM (Weather)		
Sample allocation	***Class 1 (No weather variables)***	***Class 2 (With weather variables)***
Number of observations	397	5
Percentage of observations	98.76%	1.24%
Posterior class probability	99.55%	87.81%

Extended LCSFM (Weather + Weather · COR)		
Sample allocation	***Class 1 (No weather variables)***	***Class 2 (With weather variables)***
Number of observations	397	5
Percentage of observations	98.76%	1.24%
Posterior class probability	98.95%	90.97%

[34] In particular, while the cost frontier of one group is simply $\ln C = \ln C(X, \beta)$, the cost frontier of the second group includes weather (i.e. z) variables and can be written as $\ln C = \ln C(X, \beta) + z'\psi$. Moreover, the frontier parameters associated to non-weather variables (β), and the parameters describing the distribution of v and $u(z)$ are imposed to be the same in both classes. Therefore, the issue here is to identify the set of firms with ψ=0 or not. As the value of ψ is not available to the econometrician, class membership probabilities should be estimated simultaneously alongside the other parameters of the model. See next chapter for more details about these models.

[35] These five observations come from 4 different firms and all of them, except one, show large posterior probabilities (higher than 90%) of belonging to class 2.

We have not been able to find a direct (or frontier) cost effect associated with (or "attributable to") different weather conditions, which may appear somewhat counterintuitive. However, this result would be expected if other explanatory variables, especially the regional dummy variables, were actually capturing the frontier effect of weather on firms' costs. We have checked this and have found (using a multinomial logit model) that our set of NERC regional dummies is jointly correlated with the set of weather variables. Moreover, some of the output variables (such as PL) and technological variables included in the cost frontier are also correlated with the weather variables. These correlation analyses indicate that any frontier effect of weather on firms' costs is already captured by the model, a result which indicates that the environmental factors are already taken into account in the design of networks, as pointed out by Jamasb *et al.* (2012). Therefore, it seems that advance planning has reduced the need to undertake corrective expenditure in response to outages caused by adverse weather conditions.

Overall, the above discussion suggests that the effect of weather on firms' costs estimated here is not actually capturing a direct (i.e. frontier) cost effect. Their indirect nature corroborates our main finding, i.e. most firms should not use unfavourable weather conditions as an excuse for their poor performance, and hence they should not demand that the regulators purge their cost data of weather effects.

2.6. Conclusions

The electricity industry in most developed countries has been restructured in recent decades with the aim of reducing costs, improving service quality and encouraging electricity utilities to perform efficiently. The remaining regulated segments (i.e. transmission and distribution) provide the infrastructure for the competitive segments and represent an important share of the total price paid by final customers. Despite the fact that electricity transmission is an essential part of the electricity supply sector there is a lack of empirical studies that analyse both economic characteristics of the technology and firms' inefficiency in electricity transmission.

To fill this gap in the literature we have analysed firms' performance in the US electricity transmission industry for the period 2001-2009. The analysis of the economic characteristics of the technology and inefficiency of US utilities relies on the estimation of several stochastic cost frontiers, which in turn are estimated using more recent data than in previous papers. The estimated coefficients provide useful information about firms' performance with both policy and managerial implications. For instance, we have found that, given network infrastructure, electricity transmission networks exhibit natural monopoly characteristics in most cases. This result provides support for the continuing regulation of electricity transmission. Moreover, our results indicate that efficiency in the US electricity transmission industry has declined (and diverged) over the period 2001-2009, suggesting that regulatory benchmarking techniques can identify room for improvement in performance of the US electricity transmission system.

Our stochastic frontier models also allow us to identify the determinants of firms' inefficiency in this industry. In particular, as determinants of firms' inefficiency, we have included several variables capturing weather conditions, companies' cost structure, and energy demand growth. The results indicate that more adverse conditions generate higher levels of inefficiency. As weather is an uncontrollable factor and firms cannot adjust their performance modifying weather conditions, our findings simply point out that it is more difficult to manage a firm operating in a region with bad weather, and hence policy measures aiming to improve managerial skills become more important in firms operating in adverse conditions.

We have also found that investing in capital is a better strategy to deal with adverse weather conditions rather than incurring additional operating costs. This might suggest a regulatory framework that favours capital investments to deal with unfavourable weather conditions. Finally we have found that, as expected, firms' performance gets better when demand tends to be steady as firms cannot adjust their inputs without cost over time. This result, combined with the previous finding on the importance of capital expenditure to deal with weather conditions, suggests that regulators should also take into account that achieving long-term efficiency improvements can involve short-term increases in both capital and operational costs and, hence, a deterioration in firms' short-term relative performance.

One of the most interesting issues in benchmarking is whether firms are actually using environmental conditions as an excuse for bad performance and whether their costs should be or not adjusted accordingly. Regulators should purge the data when environmental conditions are part of the technology, but not when they have an indirect effect through inefficiency. Unlike previous papers in electricity transmission, we examine this issue using a latent class (stochastic frontier) model. Our findings indicate that only the costs of a small number of firms should be adjusted downwards due to the negative influence of the bad weather. For the remainder firms, the environmental conditions mainly have an indirect effect on their costs by means of a higher inefficiency. Therefore, most firms of our sample might then be using unfavourable weather conditions as a simple excuse for their poor performance.

2.7. Appendix

Table A1. Variables and definitions from FERC FORM No. 1

Variable		*Definition*	*FERC pages*	*FERC account names/notes*
	AK	Allocation key (wages)	SWTR / (SWTT-SWAG)	
	SWTR		354-21b	Salaries and wages (transmission)
	SWTT		354-28b	Salaries and wages (total)
	SWAG		354-27b	Salaries and wages (admin. and general)
OPEX		Operational expenditure	100 * (TTE + AK * (TAGE - EPB - RCE - GAE)) / CPI	
	TTE		321-112b	Total transmission (op. and main.) expenses
	TAGE		323-197b	Total administrative and general expenses
	EPB		323-187b	Employee pensions and benefits
	RCE		323-189b	Regulatory commission expenses
	GAE		323-191b	General advertising expenses
CAPEX		Capital expenditure	100 * (DEP + IR * KBAL) / CPI	
	DEP	Depreciation	DETP + AK * (DEPGP + DEPCP)	
	DEPTP		336-7b	Depreciation (transmission plant)
	DEPGP		336-10b	Depreciation (general plant)
	DEPCP		336-11b	Depreciation (common plant)
	KBAL	Capital balance	OCK – ADEP	
	OCK	Original cost of capital	BTP + AK * BGP	
	BTP		207-58g	Balance end of year (total transmission plant)
	BGP		207-99g	Balance end of year (total general plant)
	ADEP	Accumulated depreciation	ADTTP + ADTRP + AK * ADTGP	
	ADTTP		219-25c	Accumulated depreciation total (transmission plant)
	ADTRP		219-27c	Accumulated depreciation total (regional plant)
	ADTGP		219-28c	Accumulated depreciation total (general plant)

Continued on next page

Continued from previous page

Variable	*Definition*	*FERC pages*	*FERC account names/notes*
TOTEX	Totex	OPEX + CAPEX	
PL	Peak Load	401b	(d) Peak load (MW)
DE	Electricity Delivered	401a-17	(b) MWh (total)
CS	Total Capacity of Substations	427	(f) Capacity of substation in service (MVA)
NL	Network Length	422	(f) + (g) Length of transmission lines (miles)
COR	Capex / Opex	CAPEX / OPEX (average over time for each firm)	
GROWTH	Growth in Demand	[(TE current year - TE previous year) / TE previous year] * 100	

Table A2. Variables from other sources

Variable	*Definition*	*Source*
LPR	Annual Salary	Data Quarterly Census of Employment and Wages (from the US Bureau of Labor Statistics)
KPR	Producer Price Index	US Bureau of Labor Statistics
NERC dummies	Regional dummy variables	North American Electric Reliability Corporation (NERC)
TMIN	Minimum Temperature	National Climatic Data Center (NCDC)
WIND	Average Wind Speed	National Climatic Data Center (NCDC)
PRCP	Average Precipitation	National Climatic Data Center (NCDC)
CPI	Consumer Price Index	International Labour Organisation - LABORSTA (Base Year = 2000)
IR	Interest rate (6%)	Nillesen and Pollitt (2010), p.63

Chapter 3

Using the latent class approach to cluster firms in benchmarking: An application to the US electricity transmission industry

3.1. Introduction

Electricity networks are often regulated by implementing incentive-based regulation schemes that use some types of benchmarking, i.e. a comparison of utilities' performance with best-practice references. As shown by Zhou *et al.* (2008), the nonparametric DEA has become a very popular tool in energy and environmental studies, especially for benchmarking electric utilities. Unlike the econometric SFA that requires the specification of a particular functional form for the cost or production functions to be estimated, DEA imposes fewer assumptions on the shape of firms' technology and it allows regulators to address traditional convergence problems and the well-known 'wrong skewness problem' in the SFA literature.

A key issue that is sometimes not taken into account by regulators (and researchers) is the heterogeneity or unobserved differences among firms, although utilities are usually quick to mention this issue to the regulators. This concern underlies the negotiations between regulators and utilities, where utilities wield uniqueness as a reason to avoid being compared with their peers. However, it is often assumed in this setting that the whole set of benchmarked firms share the same technology, and hence differences in behaviour are attributed to inefficient use of factors that are under the control of the companies. Possible differences among utilities associated with different technologies are either overlooked or are addressed using simple sample selection procedures, mostly based on factors that may affect performance such as geographic location or utilities' size. Therefore, the efficiency scores obtained from these analyses might be biased and some firms might be penalized (or rewarded) in excess if their underlying technology is less (more) productive than the technology used by other firms operating with more (less) advantageous conditions. This is particularly important in the case of incentive regulation and benchmarking of electricity networks where the results of efficiency analysis have important financial implications for the firms.

In this chapter we examine whether we should (a) split the sample arbitrarily on the basis of a single size variable, or (b) use a comprehensive statistical procedure to control for technological differences, before carrying out a traditional efficiency analysis of regulated electricity networks. We advocate using an LCM approach that allows us to split the electricity networks into a number of different classes, where each class is associated with a different technology. We advocate this approach for several reasons. First, LCM clusters firms by searching for differences in production or cost parameters, which is exactly what regulators are looking for. Second, our approach can be viewed as a "supervised" method for clustering data as it takes into account in the first stage the same (production or cost) relationship that is analysed later, often using nonparametric frontier techniques. Indeed, the literature on data dimension reduction uses this expression for those methods that not only use the information contained in the explanatory variables to be aggregated, but also the information of the dependent variable that will be predicted later on. And third, our approach is not more "technical" than other clustering methods as it can be implemented using standard software and using the same variables that will be used to get efficiency scores in a later stage. Having practicality in mind, we have proposed some simplifications such as the use of simple specifications for both the deterministic (e.g. Cobb-Douglas) and stochastic (e.g. normal distribution) parts of the model to facilitate its application. The use of the same variables in both the latent class stage and the second, DEA, stage also contributes to simplify the use of the proposed procedure.

The same idea is currently being developed by Agrell *et al.* (2013) in a very recent study where they use the LCM approach to control for technological differences in an application to Norwegian power distribution firms. Our research reinforces the approach from both a theoretical and an empirical point of view. In particular, we carry out a simulation analysis to examine whether the latent class approach outperforms other more arbitrary and less robust procedures for splitting a sample of observations - such as the k-means clustering algorithm or simply using the median of some relevant variables. The simulation exercises confirm our expectations and show that the proposed approach outperforms alternative sample selection procedures. We illustrate this procedure with an application to the US electricity transmission firms examined in the previous chapter. We find two statistically different groups of firms that should be compared or treated separately. In order to confirm the results from the simulation exercise, we compare the partition of the sample obtained through this method with those from alternative clustering procedures.

This chapter is organized as follows. Section 3.2 introduces the two-stage procedure that is proposed to control for unobservable differences in firms' technology (environment) in energy regulation. Section 3.3 introduces the simulation analysis performed and its main outcomes. Section 3.4 uses data from the US electricity transmission industry to compare the relative performance of our approach and alternative procedures. Section 3.5 concludes.

3.2. A two-stage procedure to address unobserved heterogeneity in utility regulation

As Haney and Pollitt (2009) pointed out in a recent survey, regulators have been using several statistical methods to determine the performance of energy utilities. Obtaining reliable measures of firms' performance requires dealing with controllable factors and monitoring for the different environmental conditions under which firms operate. However, both regulators' reports and academic studies do not usually deal with these technological differences. Statistical methods have recently been developed to address this issue. In most of these methods, heterogeneity is understood as an unobserved determinant of the production/cost frontier, while inefficiency is interpreted as the 'distance' to the frontier once heterogeneity has been taken into account.

Following Greene (2005a, 2005b) in the parametric (SFA) literature, we can basically distinguish two types of models that allow us to achieve our aim, namely the so-called True Fixed Effects (TFE) and True Random Effects (TRE) models introduced by this author, and the LCM[36], also known as finite mixture models, which have been broadly used in several fields of research (see Beard *et al.*, 1991; or Gropper *et al.*, 1999, for simple applications; and Battese *et al.,* 2004; or O'Donnell *et al.,* 2008, for more comprehensive applications that aim to examine technological gaps using a metafrontier approach). Both approaches have their own strengths and weaknesses. In the TFE/TRE models, unobserved heterogeneity is captured through a set of firm-specific intercepts that are simultaneously estimated with other parameters. Hence, this approach assumes that there are as many technologies as firms. However, as it imposes common slopes for all firms, all of them share the same marginal costs, economies of scale and other technological characteristics.

In contrast to the TFE/TRE models, the LCM approach allows the estimation of different parameters for firms belonging to different groups. This can be easily seen if the general specification of a cost function in this framework is expressed as follows:

$$\ln X_{it} = \alpha_j + \beta_j \ln Y_{it} + v_{it}\big|_j \tag{3.1}$$

where i stands for firms, t for time and $j = 1,\ldots, J$ for class. X_{it} is a measure of firms' cost, Y_{it} is a vector of explanatory variables, and the random term v_{it} follows a normal distribution with zero mean and variance σ_v^2. As both α_j and β_j, are j-specific parameters, the technological characteristics vary across classes.

Letting θ_j denote all parameters associated with class j, the conditional likelihood function of a firm i belonging to class j is $LF_{ij}(\theta_j)$. The unconditional likelihood for firm i is then obtained as the weighted sum of their j-class likelihood functions, where the weights are the probabilities of class membership, P_{ij}. That is:

$$LF_i(\theta,\delta) = \sum_{j=1}^{J} LF_{ij}(\theta_j) P_{ij}(\delta_j), \qquad 0 \le P_{ij}(\delta_j) \le 1, \qquad \sum_{j=1}^{J} P_{ij}(\delta_j) = 1 \tag{3.2}$$

where $\theta = (\theta_1,\ldots,\theta_J)$, $\delta = (\delta_1,\ldots,\delta_J)$ and the class probabilities are parameterized as a multinomial logit model:

$$P_{ij}(\delta_j) = \frac{\exp(\delta_j' q_i)}{\sum_{j=1}^{J} \exp(\delta_j' q_i)}, \qquad j = 1,\ldots,J, \qquad \delta_J = 0 \tag{3.3}$$

where q_i is either an intercept or a vector of individual-specific variables. Therefore, the overall likelihood function resulting from (3.2) and (3.3) is a continuous function of the vectors of parameters θ and δ, and can be written as:

$$\ln LF(\theta,\delta) = \sum_{i=1}^{N} \ln LF_i(\theta,\delta) = \sum_{i=1}^{N} \ln\left\{\sum_{j=1}^{J} LF_{ij}(\theta_j) P_{ij}(\delta_j)\right\} \tag{3.4}$$

Maximizing the above maximum likelihood gives asymptotically efficient estimates of all parameters. A necessary condition to identify the whole set of parameters is that the sample must be generated from at least two different technologies or two noise terms.

Several comments are in order. First, in this framework each firm belongs to one and only one class.[37] Therefore, the probabilities of class membership just reflect the

[36] More specifically the model proposed by this author (LCSFM) includes inefficiency in the error term. However, in this chapter we consider the use of a standard LCM since the efficiency scores are obtained in the second stage.
[37] This does not mean that a specific firm is going to be always in the same class. The clusters are created without taking into account the panel structure of the data, i.e. a particular firm can be in different clusters

uncertainty that researchers or regulators have about the true partition of the sample. The estimated parameters can be used to compute posterior class membership probabilities using the following expression:

$$P\left(j|i\right)=\frac{LF_{ij}\left(\hat{\theta}_j\right)P_{ij}\left(\hat{\delta}_j\right)}{\sum_{j=1}^{J}LF_{ij}\left(\hat{\theta}_j\right)P_{ij}\left(\hat{\delta}_j\right)} \tag{3.5}$$

These posterior probabilities of membership can then be used to allocate each firm to a particular class, e.g., each firm is allocated to the class with the higher posterior probability.

Second, only between-groups and not individual heterogeneity is controlled using a latent class model because all firms belonging to a particular group share the same technology. This situation is possible in energy economics if firms operating in areas with different environmental conditions must choose between a limited number of technical standards[38] to expand and maintain their networks. If firms have similar technologies, the estimated differences in technology (i.e. parameters) are likely to be capturing heterogeneity in operating environments.[39] Therefore, the differences in parameters between classes can be interpreted either as differences in technology or differences in environmental variables that might be unobserved.

Third, the number of classes J should be chosen in advance by the researcher or regulator. Selecting the number of classes is a key issue of the proposed approach, and is common to other clustering methods. Fortunately there are several statistical tests that are commonly used and accepted in the finite mixture models literature to choose the appropriate number of classes. For instance, the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) are frequently used in the LCM literature. These criteria involve minimizing an index that balances the lack of fit (too few classes) and overfitting (too many classes) as it includes a penalty that increases with the number of parameters. Models with lower AIC or BIC are generally preferred. The BIC considers a greater penalty for overfitting than AIC and, hence, BIC tends to favour more parsimonious models, which in turn help to estimate model coefficients with more precision (see Verbeek, 2008, p.61). Many authors (see for instance Koehler and Murphree, 1988) observed that the traditional AIC criterion and some of their variants tend to overestimate the correct number of classes. For those criteria that tend to overfit and favour more comprehensive models, it is very useful to examine a graph of the values of the computed statistic as the classes increase and look for the natural bend or break point where the curve flattens out. The number of data points till the "break" (i.e., including the point at which the break occurs) can be used as the number of classes to select. This method (labelled a "scree test") is often used in principal components or factor analyses to select the number of factors and it is described and pictured in every textbook discussion of factor analysis (see, for instance, Costello and Osborne, 2005, p.2-3).

Finally, it should be noted that the random term in (3.1) follows a symmetric distribution because it does not include a traditional one-sided inefficiency term. In

over time. It has been done in this way to give more flexibility to our model by allowing changes in firms' technology along the sample period. Moreover this type of approach usually yields similar results to those obtained if the belonging to a certain class is imposed for the whole sample period.

[38] These standards are either proposed by the International Electrotechnical Commission or the Institute of Electrical and Electronics Engineers.

[39] We are grateful to a reviewer for pointing this out.

other words, we advocate using a simple normal distribution in the first stage of our procedure and obtaining the efficiency scores later. There are three reasons for this. First, ignoring the asymmetric error term traditionally associated with inefficiency prevents the appearance of convergence problems in practice when estimating a latent class model, which by nature is highly non-linear. This facilitates replication of the procedure when researchers or regulators compare different specifications of the underlying technology. Second, this empirical strategy allows us to compute efficiency scores using more flexible representation of firms' technologies if nonparametric techniques such as DEA are employed. Finally, DEA is the method mainly used by regulators (see Haney and Pollitt, 2009).

The main advantage of using an LCM approach to cluster firms is that it allows us to control for environmental factors (i.e. contextual z-variables) that are not observable, difficult to measure accurately or even unknown in some cases. The LCM-DEA approach also allows the inclusion of z-variables to identify groups of comparable firms that share similar environmental or technological features (for a discussion on this topic in the DEA and SFA literature, see for instance Johnson and Kuosmanen, 2011, 2012). Thus it is more sophisticated than simply including z-variables without clustering. In this sense, our approach is consistent with the idea of benchmarking, which is based on the existence of comparable firms. However it extends this by avoiding the need for arbitrary clustering, which is often undertaken by researchers and regulators. Under arbitrary clustering larger samples are often split into sub-samples to be analysed separately on the basis of a single size metric (such as number of customers) or using subjective value judgement.

In a second stage DEA is separately applied for each class. DEA is a type of efficiency analysis which involves mathematical programming to construct a frontier of best performing companies.[40] Farrell (1957) was the first to propose this type of frontier analysis and since then there have been many authors who have developed and applied different models which have enlarged the literature in DEA methodology (see Coelli *et al*., 2005).

In this chapter, we will use an input-oriented DEA model as we assume that the output level cannot be modified by firms. This is a reasonable assumption for a network utility required to provide network capacity to service ultimate demand which is largely out of its control. Technical inefficiency can be then viewed as a proportional reduction in input usage or cost while maintaining the output levels constant. In our simulation exercise we impose constant returns to scale (CRS) as similar results are obtained if this assumption is relaxed. The optimization problem in this case can be represented as:

[40] Although DEA is a rather flexible method that does not impose implicitly the same 'parameters' on the whole sample of firms, we would like to point out, however, that obtaining different marginal products (elasticities) at different points of the sample does not mean in economics that we have estimated different technologies. From the engineering point of view, the term "technology" is often associated with a particular production process. Nevertheless in microeconomic theory, technology is more broadly defined as the set of processes technically feasible and available for firms in a moment of time, and movements along the frontier just represent different production processes within a certain technology. The differences in technology across firms (or over time) are captured in economic analysis by shifts of the production frontier. Thus, only differences in production processes are controlled when a single frontier is estimated either using DEA or SFA.

$$\begin{aligned} &\min_{\theta,\lambda} \theta, \\ &st \quad -y_i + Y\lambda \geq 0, \\ &\quad \theta x_i - X\lambda \geq 0, \\ &\quad \lambda \geq 0 \end{aligned} \tag{3.6}$$

where λ is a vector of constants and θ is a scalar calculated for each observation which represents the efficiency score for the i-th firm. y_i and x_i are the vectors of outputs and inputs for the i-th firm respectively, while Y and X are the output and input matrices for all I firms. This linear programming problem must be solved I times and gives an efficiency score θ equal or lower than one for each firm. In our empirical application we do not assume that all the firms exhibit constant returns to scale as electricity transmission firms are natural monopolies and increasing returns to scale were obtained in many applied studies.[41] A variable returns to scale (VRS) specification only requires adding the convexity constraint $I1'\lambda=1$ to the minimization problem in (3.6). $I1$ is a vector of ones, and multiplying by the vector of weights λ ensures that firms are only compared with firms of a similar size.

As pointed out by one reviewer, we are using a two-stage procedure that combines parametric and nonparametric techniques. Although it is unlikely that both techniques are fully compatible, we do not have to deal with the inconsistency problem that appears in the traditional two-stage DEA procedure. This problem arises in a different situation, when DEA is undertaken first and parametric analysis is then performed on the DEA results. Simar and Wilson (2007) have shown that applying a parametric regression in a second stage using the estimates obtained in a first stage through DEA is not consistent because firms' inefficiency is a relative measure and, hence, the nonparametric efficiency scores are serially correlated. As the order is reversed, this problem does not emerge in our case. Moreover, none of the variables used in our second stage are predicted or estimated variables.

Furthermore it should be mentioned that the DEA approach can be used as a clustering method. There is an evolving literature on this topic from Po *et al.* (2009), Krüger (2010) and Amin *et al.* (2011). However, Moazami Goudarzi and Jaber Ansari (2012) find that this approach, which is based on the piecewise production functions obtained from DEA models for clustering the data, faces several problems. Firstly, it may have alternative optimal solutions and hence the clusters produced are not unique. Secondly, they find that it is possible not to achieve any strictly positive multiplier weight for inputs and outputs in evaluating all firms. And finally, some of the obtained clusters may have overlapping units.

3.3. Simulation analysis

In this section we carry out a simulation exercise to examine whether a latent class approach is a good procedure to find groups of comparable companies within a sample when we aim to apply a benchmarking with DEA, commonly used in regulatory processes. It should be pointed out that the LCM is compared with other clustering methods as a point of comparison. However, the main objective in our simulation is to test the discriminatory power of the LCM under different scenarios when technological

[41] See for instance Huettner and Landon (1978), Pollitt (1995), Dismukes *et al.* (1998) and the previous chapter of this thesis.

and output differences arise, which as far as we know has not been performed before in the efficiency analysis literature.

The simulation exercise can be summarized as follows. Firms' costs are calculated using simulated data and following the normalized linear specification proposed by Bogetoft and Otto (2011) for the regulation of electrical Distribution System Operators in Germany. This functional form allows us to easily introduce heteroscedasticity in our data generation process. Following this specification, our cost function can be expressed as follows:

$$\frac{X_i}{Y_{1i}} = \beta_1 + \beta_2 \frac{Y_{2i}}{Y_{1i}} + u_i^+ + v_i \tag{3.7}$$

where X_i is our cost, while β_1 and β_2 stand for the marginal costs of the outputs Y_1 and Y_2 and define our technologies. Although we are imposing constant returns to scale in (3.7) to prevent size effects when comparing our sample separating methods, the use of variable returns to scale in the simulation produces the same partition of the sample and slightly larger efficiency scores.

In the papers in which simulations are carried out, the choice of the approach used in the Data Generation Process (DGP) is frequently quite contentious (see for instance Kuosmanen *et al.*, 2013). However, the way in which our DGP is defined here is not uncommon in efficiency analysis papers and can be found both in the SFA literature (see for instance Wang and Schmidt, 2002; or Kim and Schmidt, 2008) and in the DEA literature (see for instance Ruggiero, 1998; or Muñiz *et al.*, 2006).

Inefficiency levels are obtained assuming that the inefficiency term, u^+, is a positive half-normal distribution with zero mean and σ_u^2 variance. Random noise is simulated assuming that the noise term v follows a normal distribution with zero mean and σ_v^2 variance. We impose $\sigma = \sqrt{\sigma_u^2 + \sigma_v^2}$ equal to 1, which, given the specification that we have chosen, implies that the size of the random term in our function is relatively low, i.e. our levels of generated efficiency are quite high. We also fixed $\gamma = \sigma_u^2 / (\sigma_u^2 + \sigma_v^2)$ equal to 0.5, which implies that the weights of inefficiency and noise in the function are the same. Given the previous values, this implies that $\sigma_u = \sigma_v = 0.71$, and therefore is equivalent to generating a value of $\lambda = \sigma_u / \sigma_v$ equal to 1.[42]

We randomly generate 1,000 observations of two hypothetical outputs (Y_1, Y_2) using a uniform distribution between 0 and 1. We have chosen this distribution instead of the normal distribution because these variables cannot take negative values, and outputs in DEA must be positive. As the random noise term takes both positive and negative values, we impose on all technologies that $(\beta_1 + \beta_2) = 10$ to obtain positive costs. Technologies thus differ in relative marginal costs, i.e. the relative weight of each β. In particular, we have simulated three possible technologies:

- Technology A: $\beta_2 = \beta_1$, $(\beta_1 = 5,\ \beta_2 = 5)$
- Technology B: $\beta_2 = 2\beta_1$, $(\beta_1 = 10/3,\ \beta_2 = 20/3)$
- Technology C: $\beta_2 = 4\beta_1$, $(\beta_1 = 2,\ \beta_2 = 8)$

Both coefficients are the same in technology A, while marginal costs are increasingly different in the other two technologies, B and C. Although these

[42] Although the values of these parameters have been arbitrarily chosen, the results obtained from the simulation are consistent with respect to changes in them as long as we keep the underlying efficiency at 'normal' levels.

differences in parameters between classes are associated with different technologies, we have already mentioned before that they can be interpreted either as differences in technology or differences in environmental variables. Next, we will examine the robustness of our results by adding differences between outputs. In particular, we modify the original statistical distribution of the second output by doubling (Y_2~2·U(0,1)) and quadrupling (Y_2~4·U(0,1)) its range of values.

Taking into account that we always apply the technology A to the first 500 observations and then B or C to the following 500 observations, and that we have three output distributions, 6 possible scenarios are obtained. In Table 3.1, we show the scenarios and the percentage success in predicting the underlying class membership using different clustering methods. Percentages of success can be obtained through the identification of the groups, which is possible after comparing the real β-ratios with those obtained using group-specific OLS (Ordinary Least Squares) regressions. The estimated ratios that are also shown in Table 3.1 give an idea about how well each procedure is able to identify the underlying, but different, technologies.

The first empirical exercise has to do with the case in which DEA is applied using the real separation of our data. By construction, the percentage of success in this case is 100%. For this reason, this exercise is used as a benchmark to study the performance of four sample separation methods: the median of the cost, the k-means clustering algorithm considering the outputs, the k-means clustering algorithm including both outputs and cost, and the latent class model (that involves both output and cost information). Looking at the percentages of success and the β-ratios we can confirm that the LCM is the method that better allocates observations to specific technologies. It is also the best clustering method at identifying the relationship between technologies represented by the β-ratios. As we move to a different scenario where there are more uneven features among groups, we observe that there is a clear divergence in the behaviour of the procedures: whereas the LCM improves its percentage of prediction success,[43] the alternative procedures only slightly improve their performances.

We show in Table 3.2 the average efficiencies that are obtained after DEA is applied separately to each group of firms. The last column shows the sum of squared differences (SSD) with respect to the real separation case. The SSD is calculated as the total sum of squared differences of the predicted efficiency with respect to the value of the underlying efficiency of each observation. The smallest SSD value allows us to identify the best individual predictor procedure, i.e. the clustering method that better predicts the 'real' efficiencies. Leaving aside the real separation case where SSD is zero by construction, LCM provides by far the smallest SSD in all scenarios. As LCM is the procedure that gives the closest efficiency levels to the real separation case, it is the best at predicting individual efficiencies.

[43] The estimated probabilities for the most likely latent class also increase, so the LCM not only improves its prediction capacity but also the precision with which each observation is assigned.

Table 3.1. First stage simulation results: percentage success in identifying technologies

Simulation	*Procedure*	*% Success*	*Underlying technology*	
			Group 1 (β1 / β2)	*Group 2 (β1 / β2)*
A&B	Simulation	-	1.000	0.500
(OD 1)	Real separation	100.00	1.080	0.597
	Median (C)	49.60	0.890	0.756
	Cluster (Y_1, Y_2)	46.50	0.849	0.815
	Cluster (Y_1, Y_2, C)	49.30	0.894	0.763
	LCM	65.70	1.063	0.555
A&C	Simulation	-	1.000	0.250
(OD 1)	Real separation	100.00	1.080	0.331
	Median (C)	50.20	0.678	0.575
	Cluster (Y_1, Y_2)	46.50	0.646	0.642
	Cluster (Y_1, Y_2, C)	49.90	0.684	0.593
	LCM	79.20	1.162	0.337
A&B	Simulation	-	1.000	0.500
(OD 2)	Real separation	100.00	1.077	0.597
	Median (C)	55.00	0.834	0.799
	Cluster (Y_1, Y_2)	54.00	0.822	0.812
	Cluster (Y_1, Y_2, C)	55.10	0.822	0.802
	LCM	79.30	1.110	0.596
A&C	Simulation	-	1.000	0.250
(OD 2)	Real separation	100.00	1.077	0.331
	Median (C)	57.20	0.723	0.562
	Cluster (Y_1, Y_2)	54.00	0.656	0.609
	Cluster (Y_1, Y_2, C)	58.30	0.714	0.529
	LCM	87.90	1.099	0.337
A&B	Simulation	-	1.000	0.500
(OD 3)	Real separation	100.00	1.076	0.598
	Median (C)	57.40	0.868	0.765
	Cluster (Y_1, Y_2)	53.90	0.833	0.785
	Cluster (Y_1, Y_2, C)	57.80	0.863	0.754
	LCM	90.60	1.097	0.583
A&C	Simulation	-	1.000	0.250
(OD 3)	Real separation	100.00	1.076	0.331
	Median (C)	60.60	0.779	0.493
	Cluster (Y_1, Y_2)	53.90	0.674	0.576
	Cluster (Y_1, Y_2, C)	61.80	0.772	0.486
	LCM	94.70	1.102	0.328

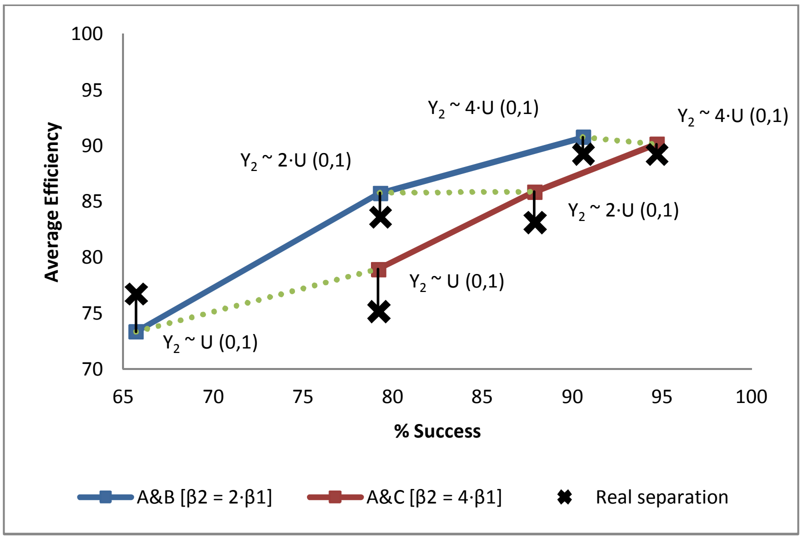
Table 3.2. Second stage (DEA) simulation results: predicted vs. underlying efficiency

Simulation	*Procedure*	*Av. Eff. (%)*	*Sum of Squared Differences*
A&B	Real separation	76.73	-
(OD 1)	No separation	67.15	135,252
	Median (C)	73.29	118,885
	Cluster (Y_1, Y_2)	70.38	108,590
	Cluster (Y_1, Y_2, C)	72.91	118,993
	LCM	73.35	40,268
A&C	Real separation	75.16	-
(OD 1)	No separation	54.70	533,029
	Median (C)	65.26	322,549
	Cluster (Y_1, Y_2)	61.65	379,683
	Cluster (Y_1, Y_2, C)	64.65	338,483
	LCM	78.93	139,993
A&B	Real separation	83.61	-
(OD 2)	No separation	73.28	151,038
	Median (C)	76.75	121,264
	Cluster (Y_1, Y_2)	75.93	119,387
	Cluster (Y_1, Y_2, C)	76.22	124,400
	LCM	85.75	51,232
A&C	Real separation	83.14	-
(OD 2)	No separation	63.22	507,703
	Median (C)	69.29	372,090
	Cluster (Y_1, Y_2)	68.14	386,773
	Cluster (Y_1, Y_2, C)	68.32	389,718
	LCM	85.87	45,663
A&B	Real separation	89.26	-
(OD 3)	No separation	78.75	180,309
	Median (C)	80.36	158,268
	Cluster (Y_1, Y_2)	79.99	160,646
	Cluster (Y_1, Y_2, C)	80.20	159,799
	LCM	90.76	29,598
A&C	Real separation	89.24	-
(OD 3)	No separation	70.49	511,481
	Median (C)	73.45	429,212
	Cluster (Y_1, Y_2)	72.86	446,596
	Cluster (Y_1, Y_2, C)	72.95	438,210
	LCM	90.15	16,511

When we move from a model with only one class to a model with two classes and unobserved heterogeneity is somehow taken into account (or 'removed') in a first stage, larger efficiency scores are obtained when carrying out a traditional DEA analysis. The computed efficiency improvements are partially caused by the fact that the number of peers necessarily decreases when a model with two classes is used,[44] regardless of the clustering method. However, our simulation exercise shows that these efficiency increases have also to do with the selection of a specific clustering method. In particular, Tables 3.1 and 3.2 (and Figure 3.1 below) indicate that the better the partition is, the larger the average efficiency scores are. Moreover, this result happens regardless of whether we carry out either traditional DEA or SFA (not shown) analyses in the second stage of our procedure.

Figure 3.1 shows the positive correlation that exists between efficiency and success in assigning observations to technologies using the LCM approach. This figure allows us to examine the discriminatory power of the model when there are either larger differences between technologies (illustrated as the shift from the blue to the red line) or between output data generation processes (illustrated as movements along the red and blue lines). As expected, the percentages of success are much larger when the two technologies differ notably in their characteristics. It is worth mentioning that this increase in percentages of success is especially important when there is no separating information on the output side, i.e. when both outputs are similarly generated. When outputs provide additional information to split the sample, both efficiency levels and percentages of success increase, regardless of whether the technologies are similar or diverse. On the other hand, Figure 3.1 also shows that as inequalities between groups rise, the average efficiency score obtained using LCM as a sample separation method even exceed the average efficiency score from the real separation case. This shows that an imperfect assignment of firms to groups can lead us to obtain higher levels of efficiency. In other words, making a good partition of the sample does not necessarily imply obtaining larger efficiencies.

Figure 3.1. Average efficiency and percentage of success for the LCM



[44] This does not necessarily happen when we move from 2 to 3 classes (and so on) because there is a reallocation of the observations into the different classes. Indeed, as a larger partition of the sample does not imply that one class is divided into two separable classes, some observations might have "new" peers and, hence, their (relative) efficiency might be less than before.

In summary, the above results clearly indicate that LCM deals with unobserved heterogeneity much better than the other clustering methods. We attribute this better performance to the fact that LCM splits the data taking into account the objective of the second stage, where a relationship between outputs and inputs (or costs) is estimated in order to compute inefficiency scores. In this sense, and borrowing the terminology used for dimension reduction, this approach can be interpreted as a 'supervised' method to split the data.

From a regulation point of view, the above results suggest that, given a number of classes, regulators could use this statistic (i.e. the mean efficiency) to compare the relative performance of several clustering methods in a real case in which they do not have information about the 'underlying partition' of the sample. Our proposed procedure thus can be labelled as a conservative approach. However, using a method that provides conservative efficiency estimates is common among regulators. For example, in Germany, the regulator assesses the performance of each firm using both DEA and SFA efficiency scores and chooses the larger of the two estimates (Agrell and Bogetoft, 2007). Here we provide an additional reason, based on simulation results, that justifies the use of a conservative approach when, and only when, clustering methods are used in benchmarking.

3.4. Application to the US electricity transmission industry

3.4.1. Data, sample and variables

We next illustrate the proposed procedure with an application to the US electricity transmission industry. As we have mentioned in Chapter 1, benchmarking of electricity transmission utilities is a challenging task due to the small number of transmission utilities that usually operate in the jurisdiction of a particular regulator. This likely explains why there are few empirical papers published on efficiency analysis of electricity transmission firms and moreover, none of these articles deal with unobserved heterogeneity or technological differences.[45]

The database used here is the same as in the previous chapter and contains 405 observations on 59 US electricity transmission firms for the period 2001-2009. Following the literature, we specify a standard cost function with four outputs where our cost variable is *Totex*, defined as before. The four outputs are: *Peak Load* (PL), *Electricity Delivered* (DE) and *Network length* (NL), which are again defined as in Section 2.4, and *Total Energy* (TE), which stands for the total energy of the system, including total net own generation, total purchases from others, net exchanges in the system (received-delivered), net transmission for others and transmission by others.[46] The four outputs considered (explanatory variables) and the cost variable (dependent variable) will be used later on in the DEA stage.

To analyse robustness, we extend the standard model by adding four time-invariant environmental variables to split the sample of transmission utilities. Three of

[45] On the contrary there is an extensive literature in electricity distribution (see for instance, Jamasb and Pollitt, 2003, for a European survey) and there are many articles that address the issue of heterogeneity including environmental factors in this sector (see for instance, Yu *et al.*, 2009, Nillesen and Pollitt, 2010, Jamasb *et al.*, 2012, or Growitsch *et al.*, 2012).

[46] In this chapter we incorporate an additional output, *Total Energy*, instead of the *Total Capacity of Substations*. The flexibility of the Translog models estimated in the previous chapter prevented the inclusion of this variable due to multicollinearity problems and it was substituted for a capacity variable.

these are the weather variables: TMIN, WIND and PRCP. The last environmental variable is the *Growth in Demand* (GDEM) for each firm over time. The descriptive statistics of the full set of variables are shown in Table 3.3.

Table 3.3. Descriptive statistics

	Variable	*Units*	*Mean*	*Max.*	*Min.*	*Std. Dev.*
Totex	Cost	US$	144,602,000	667,127,000	20,713,600	120,324,000
Peak Load	Output	MW	6,173	23,111	380	5,533
Electricity Delivered	Output	MWh	6,280,310	74,584,700	56,730	8,839,980
Total Energy	Output	MWh	34,557,900	116,415,000	2,339,000	26,752,600
Network Length	Output	Miles	4,064	16,292	1,087	3,253
Minimum Temp.	Weather	°F	-10.35	19.90	-59.80	16.51
Wind Speed	Weather	Knots	6.84	9.60	4.63	1.01
Precipitation	Weather	Inches	0.07	0.16	0.01	0.03
Growth in Demand	Other	%	0.03	244.11	-74.96	17.72

3.4.2. Empirical results

As above mentioned, we should initially use a simple specification of the cost function to split the sample in order to facilitate the replication of the procedure and to avoid convergence problems when more comprehensive models are estimated. In this sense, we use a Cobb-Douglas (or logarithm) specification of the cost function due to its widespread use and acceptance in previous empirical studies. Convergence problems prevented us from estimating the LCM for more than two classes with the linear specification that we used in our simulation exercise. However, these problems did not appear using the Cobb-Douglas functional form. As we do not know the true number of underlying technologies, this is an interesting advantage of the logarithm specification of the model. The coefficients for the Cobb-Douglas specification are shown in the Appendix (Section 3.6).

In Table 3.4 we show the descriptive statistics of the efficiency scores obtained using DEA as the number of classes is increased, and the number of observations (as a percentage) that improve their efficiency scores as we move from one class to two classes and so on. As expected, the average efficiency score for the so-called non-separation model, which can be considered as a model with one class, is 64.84%, much lower than the average efficiency obtained from the model with two classes, 77.03%. The most comprehensive model that is estimated is a LCM with 9 classes. Although the average efficiency score for this model goes up to 87.4%, the largest change in efficiencies occurs when we move from one class to two classes. We can also see that most observations have higher efficiency scores when more comprehensive models are estimated. This is compulsory for the 100% of the observations when we move from a model with one class to a model with two classes as the number of peers necessarily decreases in this case. This does not necessarily happen when we move from 2 to 3 classes (and so on) because there is a reallocation of the observations into the different classes and some observations might have "new" peers.[47]

[47] See footnote #44.

Table 3.4. Efficiencies with the LCM-DEA procedure

Number of classes	*Average efficiency*	*Δ*	*% Obs. improv.*
1	64.84	-	-
2	77.03	12.20	100.00
3	79.55	2.51	93.09
4	80.36	0.81	97.28
5	84.31	3.95	76.30
6	82.64	-1.66	56.79
7	86.71	4.06	69.14
8	87.51	0.80	61.23
9	87.41	-0.10	62.72

The choice of the number of classes is a key issue in any clustering method. The AIC and BIC model selection criteria and some of their variants are commonly used to choose the appropriate number of classes in the LCM literature. The general form of most information criteria can be written as follows:

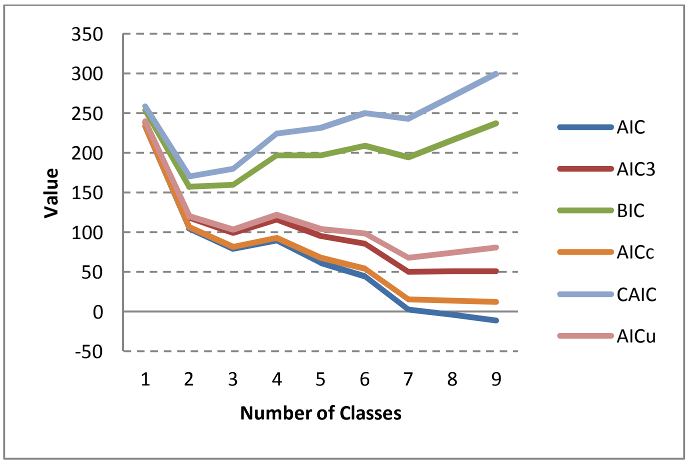
$$-2\ln LF + Penalty \tag{3.8}$$

where the first term is twice the negative logarithm of the maximum likelihood which decreases when the number of classes (complexity) increases. The penalty term penalizes too complex models, and increases with the number of parameters of the model. Thus, these criteria involve minimizing an index that balances the lack of fit (too few classes) and overfitting (too many classes). Models with lower values of (3.8) are generally preferred.

Several information criteria are shown in Figure 3.2 to illustrate robustness. The figure includes the traditional AIC and BIC criteria and some of their variants, the modified AIC criterion (AIC3), the corrected AIC (AICc), the so-called AICu, and the consistent AIC (CAIC) that can be considered either an AIC or BIC variant. For more details about these criteria and the associated penalty functions, see Fonseca and Cardoso (2007). All of them show a remarkable improvement in fitness-parsimony when we move from just one class to a model with two classes. While the traditional AIC criterion and some of their variants (AIC3, AICc, and AICu) show little improvements when more classes are added, the BIC and CAIC clearly deteriorate their performance with more than two classes.[48] Generally speaking, the abovementioned tests allow us to conclude that a reasonable and practical trade-off between good description of the data and complexity is provided by a model with two classes. We therefore choose this model as our preferred model.

[48] The same happens if we use a criterion (not shown in Figure 3.2) that penalizes poorly separated classes in LCMs with two or more classes, such as the so-called Complete Likelihood Classification (CLC) and the Integrated Classification Likelihood-BIC (ICL-BIC).

Figure 3.2. Choice of the number of classes



We also should expect notable differences in the composition of the two groups of firms if, instead of the LCM-based procedure, we use other procedures to split the sample. To give a sense of what difference this makes to the sample selection, we can pay attention to the percentage of observations for which this approach gives higher efficiencies than other methods. LCM provides not only the largest average efficiency score in the second stage, but also the majority of the observations obtain a higher value under this approach than from the others. More than 70% of observations are in an equal or better situation under LCM compared to the median of the network or cost and any cluster application. Thus, as discussed in the simulation section, the LCM approach provides the more favourable framework to benchmark firms. In addition, cluster-based separation procedures provide much more uneven sample partitions than the LCM approach. While in LCM there are 129 observations in a group and 276 in the other, under the two cluster applications the division is as follows: 49/356 when network is used as separating variable and 72/333 when also cost is included. This indicates the potential value to regulators of our LCM approach in reducing the need to rely on the small samples that can arise while using arbitrary approaches to sub-sample creation.

As a result of the above allocation, the alternative sample separation procedures provide different efficiency scores for each utility. The estimated efficiency levels are shown in Table 3.5. In accordance with the simulation results, the lowest levels are obtained not only when there is no separation of firms but also when we use cluster procedures using either network size or firms' cost as separating variables.[49] On the other hand, the largest efficiency scores are obtained when the LCM is used as a statistical tool to account for unobserved differences among firms. It is worth noting that most clustering procedures produce rather low efficiency scores for some observations. It should be noted, however, that this result has to do with application of DEA in the second stage. If we instead use an SFA approach, larger efficiencies would be obtained. This always happens because part of the measured inefficiency using DEA is now captured by the noise term of the model.

Using the median of cost as a sample-separating variable not only produces larger efficiency scores, but also a minimum efficiency (about 32%) that is much larger than in other clustering methods (including the LCM). This result is caused by the fact

[49] The sample partition is the same when we take into account all the outputs and cost, or network size and cost together.

that we have used the same variable to both split the sample and measure firms' inefficiency. If we use the median of cost to split the sample we are falsely minimizing the differences in costs within each group. For instance, some very inefficient small firms (with relatively high costs) might be allocated with large firms, and some very efficient large firms (with relatively low costs) might be allocated with small firms. The consequence of these movements is both a balance of the average efficiencies of both classes, and an increase of the minimum efficiency level as the small (large) firms allocated with large (small) firms will become more efficient because the lack of peers with similar output levels. Generally speaking, the above discussion highlights the fact that we should not split the sample using a variable that is also being used to measure firms' inefficiency.

Table 3.5. Efficiencies obtained with different clustering methods

Procedure	***Mean***	***Std. Dev.***	***Max.***	***Min.***
DEA (No separation)	64.84	21.87	100.00	9.15
LCM - DEA	77.03	19.22	100.00	9.39
Cluster (N) - DEA	65.52	22.45	100.00	9.15
Cluster (N, C) - DEA	67.05	21.23	100.00	11.77
Median of network - DEA	69.54	21.31	100.00	10.42
Median of cost - DEA	74.40	20.12	100.00	31.70

To give some intuition about the heterogeneity between classes that has been disentangled using the LCM procedure, we show in Table 3.6 the descriptive statistics of each one of the two groups that were found.[50] It can be seen that the average value of the cost and all the outputs is higher in class 1 than in class 2 so the largest companies are mainly located in the first class. However the standard deviations in class 1 for these variables are in general larger than in 2, indicating that there are more differences of size between firms in this class. Maybe this dissimilarity in the scale is because these firms operate in similar environments as it can be inferred from the smaller standard deviations of their environmental variables. The main difference on the average of these variables is observed for the temperature, indicating that in general firms of class 1 are located in colder regions, and the growth of the demand, which is mainly positive for firms in class 1 and negative for firms in class 2. These reasonable differences illustrate the nature of the heterogeneity that is controlled in our model. Clearly we could have arbitrarily allocated firms to two subsamples using a temperature threshold, however the LCM has allowed us to identify the number of subsamples to be analysed separately and then allocated firms between them in a statistically robust way.

[50] To confirm the point made in footnote #37, we have checked the number of observations that freely change between classes in our model for the sample period. It can be observed that about 90% of our firms' observations remain in the same cluster from one year to another and hence no erratic changes are observed between classes over time.

Table 3.6. Descriptive statistics of the classes found with LCM

CLASS 1 (129 observations)

Variable	*Mean*	*Max.*	*Min.*	*Std. Dev.*
TOTEX	224,218,000	667,127,000	20,713,600	166,005,000
PL	7,730	23,111	380	6,708
DE	6,376,580	39,484,700	82,304	7,948,270
TE	40,417,200	115,685,000	2,339,000	30,081,500
NL	5,078	16,292	1,088	4,733
TMIN	-13.04	19.00	-59.80	14.05
WIND	6.81	9.22	4.63	1.05
PRCP	0.07	0.15	0.02	0.03
GDEM	0.80	62.59	-40.35	11.68

CLASS 2 (276 observations)

Variable	*Mean*	*Max.*	*Min.*	*Std. Dev.*
TOTEX	107,391,000	309,969,000	25,559,800	63,868,300
PL	5,446	22,054	427	4,730
DE	6,235,310	74,584,700	56,730	9,240,720
TE	31,819,300	116,415,000	2,886,900	24,629,500
NL	3,591	10,451	1,087	2,099
TMIN	-9.09	19.90	-59.80	17.43
WIND	6.85	9.60	4.76	1.00
PRCP	0.07	0.16	0.01	0.03
GDEM	-0.33	244.11	-74.96	19.94

3.4.3. Robustness analyses

We next introduce some additional tables in which we show the results obtained from alternative approaches or specifications that help us to analyse the robustness of the proposed clustering procedure based on LCM.

We first show in Table 3.7 the results we get when we introduce three weather variables and demand growth as sample-separating variables in the first stage of our procedure. Table 3.7 shows that both LCMs give us larger efficiency scores than extended k-means procedures that include environmental variables (alone or with information about the cost function). Based on our simulation results, we could then conclude that LCM also outperforms other sample separating methods when information about firms' environmental conditions is available. On the other hand, the estimated coefficients of the sample-separating variables (see Appendix) are statistically significant, which implies that they have helped to better split the sample. Despite this, our sample partition does vaguely change when we try to control for environmental variables as the percentage of coincidence in allocating observations is quite high (88%). This means that the between-class differences in estimated parameters are already capturing heterogeneity in firms' operating environment. In other words, a simple latent class model is able to control for those differences without explicitly including environmental variables that regulators might find it very difficult or expensive to collect. To examine better this issue, we have carried out an auxiliary regression (not shown) where an environmental variable composite interacts with the rest of explanatory variables of the frontier function. As we cannot reject that these coefficients are statistically significant, we can conclude that the parameter differences identified in a LCM model are, at least partially, capturing differences in environmental conditions.

Table 3.7. Clustering methods including environmental variables

Procedure	Mean	Std. Dev.	Max.	Min.
LCM (W, D) - DEA	77.03	20.40	100.00	9.39
Cluster (W, D) - DEA	69.98	22.83	100.00	9.24
Cluster (W, D, N) - DEA	65.52	22.45	100.00	9.15
Cluster (W, D, N, C) - DEA	67.05	21.23	100.00	11.77

Regarding the specification of the functional form, Table 3.8 provides a brief comparison of both Cobb-Douglas and Translog results. Again, the parameter estimates are shown in the Appendix. The correlation of Cobb-Douglas and Translog efficiency scores is very high (about 93%) and the overlap between classes is also remarkable (almost 84%). Furthermore, the model selection analysis indicates that the best trade-off between fitness and complexity for the Translog specification is provided once more by a model with two classes. A Cobb-Douglas specification is still preferred on the grounds of simplicity and because, in our case, the way in which the technology is modelled is not very relevant.

Table 3.8. Cobb-Douglas vs. Translog using LCM-DEA

	Without including environmental variables			
	Number of obs.		***Av. eff.***	
	CD	**Translog**	**CD**	**Translog**
Class 1	129	102	66.83	63.61
Class 2	276	303	81.80	81.35
Both	405	405	77.03	76.88

	Including environmental variables			
	Number of obs.		***Av. eff.***	
	CD	**Translog**	**CD**	**Translog**
Class 1	174	117	69.09	62.78
Class 2	231	288	83.01	82.25
Both	405	405	77.03	76.63

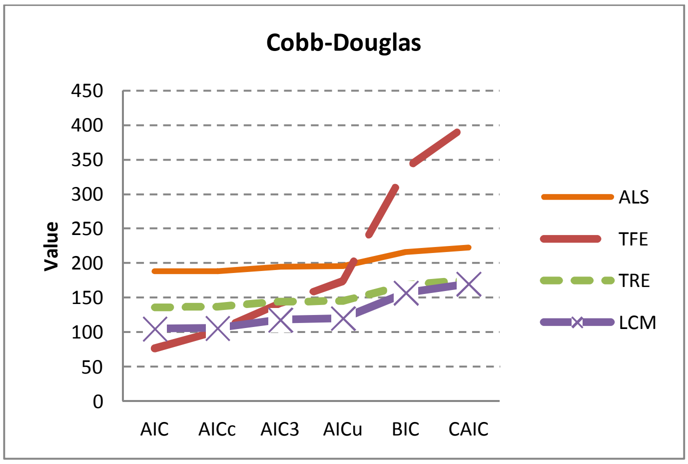
Although this chapter does not attempt to contribute to the current debate about the suitability of parametric and nonparametric approaches for purposes of benchmark regulation, we now try to compare the relative performance of our procedure that combines LCM and DEA and two fully parametric procedures based on LCM, using both Cobb-Douglas and Translog specifications for the cost function. Besides the LCM-DEA approach, Table 3.9 provides results for the LCM-ALS model in which the traditional ALS model is applied in the second stage as done in Agrell *et al.* (2013). The third model, LCSFM, is the one-stage Latent Class Stochastic Frontier Model introduced by Greene (2005b) that adds an inefficiency term to the LCM. Several interesting remarks are in order. First, when ALS is applied after the partition of the sample, the average efficiency for the whole sample is 100% because the estimated value of σ_u is equal to zero. This awkward result is often known as the 'wrong skewness problem' in the SFA literature, and might occur even when the model is correctly specified (Simar and Wilson, 2010). DEA likely became a very popular tool for benchmarking electric utilities because it allows regulators to address this issue. Second, LCSFM provides very similar partitions of the sample than the proposed procedure based on a non-frontier specification of the random term. For instance, the percentage of coincidences is about 98% when a Cobb-Douglas specification is used. This similarity is caused by the presence again of the 'wrong' skewness problem as the inefficiency term or σ_u (ignored in the proposed procedure) tends to vanish when a LCSFM is estimated. Therefore, it seems that a LCM model without a frontier specification and DEA is the best option to obtain proper efficiency levels in our application.

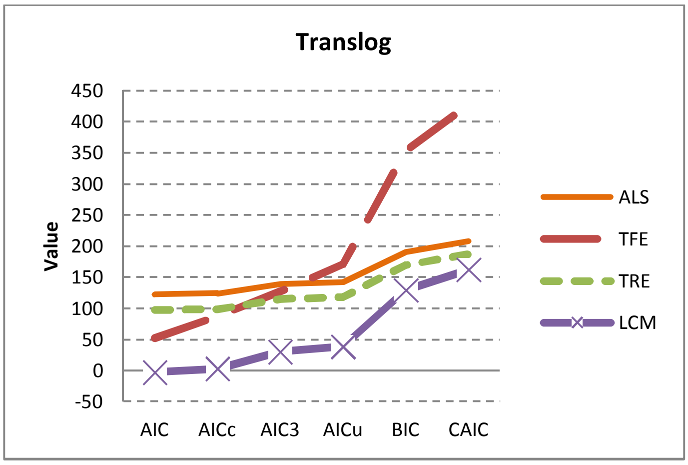
Table 3.9. 1 stage vs. 2 stages LCM clustering methods

		Class 1			*Class 2*		
Specif.	***Procedure***	***Number of obs.***	***Av. eff.***	σ_u	***Number of obs.***	***Av. eff.***	σ_u
CD	LCM-DEA	129	66.83	-	276	81.80	-
	LCM-ALS	129	100.00	0.00	276	100.00	0.00
	LCSFM	138	87.28	0.32	267	100.00	0.00
Translog	LCM-DEA	102	63.61	-	303	81.35	-
	LCM-ALS	102	100.00	0.00	303	100.00	0.00
	LCSFM	67	77.21	0.52	338	100.00	0.00

Finally, although the chapter is focused on clustering methods, we also try to compare the relative performance of our proposed procedure and two non-clustering methods broadly used in the literature to take into account unobserved heterogeneity: the TFE and TRE models introduced by Greene (2005a, 2005b). As shown in Figure 3.3, most of our earlier model selection criteria indicate that our empirical strategy based on estimating a LCM model provides a better fit than any of the stochastic frontier models introduced by Greene. This happens whether we use a Cobb-Douglas or Translog specification for the cost frontier. Our results seem to indicate that the underlying heterogeneity is better captured by a finite number of technologies rather than assuming that there are as many technologies as firms, but with the same marginal costs, economies of scale and other technological characteristics.

Figure 3.3. Clustering vs. Non-clustering model selection





3.5. Conclusions

Electricity networks are often regulated by implementing incentive-based regulation schemes based on a comparison of utilities' performance with best-practice references. A key issue that is sometimes not taken into account is the heterogeneity or unobserved differences among firms associated with different technologies or environmental conditions. As in Agrell *et al.* (2013), in this chapter we propose using a latent class approach as a statistical clustering method to split the sample into groups of more comparable firms before carrying out a traditional efficiency analysis using DEA, the most common frontier analysis technique used by regulators in utility benchmarking.

We advocate this approach for several reasons. First, latent class models are specifically designed to cluster firms by searching for differences in production or cost parameters, which is exactly what regulators are looking for. Second, our approach can be viewed as a "supervised" method for clustering data as it takes into account the same relationship that is analysed later, often using nonparametric frontier techniques. And third, our approach is not more "technical" than other clustering methods as it can be implemented using standard software. The use of the same variables in both the latent class stage and the DEA stage and the use of simple model specifications contribute to simplifying the proposed procedure. We have demonstrated through a simulation exercise that the latent class approach better allocates observations into different classes than alternative clustering procedures and better predicts the underlying efficiency of each observation. The discriminatory capacity and the assignment success of the proposed clustering method increase when large differences between technologies or output distributions arise. This, in turn, yields a convergence of estimated efficiency levels to the true underlying levels. Moreover, the better the partition is, the larger the average efficiency scores are, whether we carry out either parametric or nonparametric efficiency analyses in the second stage of our procedure. From a regulation point of view, this outcome indicates that, given a number of classes, regulators could use the average efficiency level to compare the relative performance of several clustering methods in a real case in which they do not have information about the 'underlying partition' of the sample. In this sense, our simulation exercise justifies the use of a method that provides conservative efficiency estimates in benchmarking when, and only when, clustering methods are used.

Finally, we illustrate the proposed method with an application to a sample of US electricity transmission firms for the period 2001-2009. Several model selection tests allow us to conclude that a reasonable and practical trade-off between good description of the data and complexity is provided by a latent class model with two classes. In this sense, we also find that the largest change in efficiency scores occurs when we move from a one-class model (without any partition of the sample) to a model with only two classes. In line with our earlier simulation results, the largest efficiency scores are obtained when the LCM is used as a statistical tool to account for unobserved differences among firms.

We have also found that a simple latent class model is able to control for heterogeneity in firms' operating environment without explicitly including environmental variables that regulators might find it very difficult or expensive to collect. Our results seem to indicate that the underlying heterogeneity is better captured by a finite number of technologies (identified by a clustering method) than by using non-clustering methods that, in contrast, assume that there are as many technologies as firms, but with the same technological characteristics.

3.6. Appendix

Table A3. Parameter estimates of the LCM using the US electricity transmission data

	Cobb-Douglas				*Translog*			
Variable	*Coeff.*	*t-ratio*	*Coeff.*	*t-ratio*	*Coeff.*	*t-ratio*	*Coeff.*	*t-ratio*
Class 1								
Intercept	18.686	326.6	18.668	383.3	18.586	185.5	18.634	209.0
ln PL_{it}	0.808	4.853	0.800	4.907	0.554	1.692	0.445	1.276
ln DE_{it}	0.044	1.900	0.042	1.457	0.040	0.492	0.087	1.170
ln TE_{it}	-0.261	-1.357	-0.237	-1.300	0.048	0.133	0.176	0.579
ln NL_{it}	0.184	2.038	0.182	2.085	0.047	0.212	0.004	0.022
½ (ln PL_{it})^2					0.347	0.148	-0.754	-0.727
½ (ln DE_{it})^2					0.049	0.783	0.098	1.170
½ (ln TE_{it})^2					1.086	0.344	-0.894	-0.769
½ (ln NL_{it})^2					0.362	0.513	0.183	0.294
ln PL_{it} · ln DE_{it}					0.521	2.002	0.327	1.295
ln PL_{it} · ln TE_{it}					-0.678	-0.258	0.849	0.850
ln PL_{it} · ln NLit					0.015	0.024	-0.412	-0.731
ln DE_{it} · ln TE_{it}					-0.567	-1.803	-0.485	-2.025
ln DE_{it} · ln NL_{it}					0.055	0.413	0.100	0.795
ln TE_{it} · ln NL_{it}					-0.142	-0.174	0.322	0.497
Sigma	0.380	22.982	0.381	22.078	0.356	15.218	0.332	15.719
Class 2								
Intercept	18.385	1664.0	18.390	1649.0	18.227	1186.4	18.201	1164.2
ln PL_{it}	0.144	3.109	0.166	3.881	0.273	6.622	0.423	8.491
ln DE_{it}	0.054	5.258	0.060	5.823	0.048	5.113	0.043	4.069
ln TE_{it}	0.415	7.817	0.401	7.785	0.295	5.794	0.106	1.754
ln NL_{it}	0.136	6.192	0.123	5.133	0.164	9.013	0.182	8.158
½ (ln PL_{it})^2					0.440	2.403	1.519	5.569
½ (ln DE_{it})^2					0.066	6.968	0.024	2.071
½ (ln TE_{it})^2					0.017	0.060	1.624	3.870
½ (ln NL_{it})^2					0.487	9.009	0.471	7.533
ln PL_{it} · ln DE_{it}					-0.142	-3.626	-0.126	-2.734
ln PL_{it} · ln TE_{it}					-0.177	-0.796	-1.457	-4.398
ln PL_{it} · ln NLit					0.182	3.035	0.324	4.064
ln DE_{it} · ln TE_{it}					0.090	1.789	0.093	1.649
ln DE_{it} · ln NL_{it}					-0.041	-2.606	0.006	0.308
ln TE_{it} · ln NL_{it}					-0.165	-2.127	-0.371	-3.318
Sigma	0.119	11.332	0.111	11.382	0.109	12.219	0.117	15.357
Class membership probabilities								
Intercept			-0.088	-0.416			-0.881	-3.524
$TMIN_i$			-0.065	-3.001			-0.074	-3.255
$WIND_i$			-0.373	-2.153			0.974	3.215
$PRCP_i$			11.910	1.535			31.827	3.268
$GDEM_i$			0.092	1.744			-0.086	-1.582
Prior class prob.	0.444	0.556	0.479	0.521	0.351	0.649	0.292	0.708
Log-likelihood	-39.666		-26.726		34.342		54.986	

Chapter 4

A latent class approach for estimating energy demands and efficiency in transport: An application to Latin America and the Caribbean

4.1. Introduction

Since the 1970s oil crisis, the measurement and control of energy efficiency has become an essential goal of the economic and energy policies of a large majority of countries, especially in those that import energy. This interest subsequently arose in the late 1980s as a result of the growing awareness of global warming. A key issue in the strategy of the countries that aim to reduce their energy consumption and mitigate their greenhouse gas emissions is the adoption of measures that improve the efficiency of energy use in all economic sectors and especially in those that are energy intensive, such as transport.

Figure 4.1 shows that transport is the sector with the highest energy consumption in Latin America and the Caribbean. In recent decades, this sector used, on average, 43% of the total energy consumption, followed by manufacturing at 37%, household consumption at 14% and the service sector at 6%. The ECLAC (2010) indicates that the transport of passengers and goods will increase in the future. Combined with the dissociated manner in which public policy on infrastructure and transport has been conducted, this will result in an increase in the future use of energy, which implies a significant amount of oil derivatives consumption in the near future. Nevertheless, little published information on the transport sector in Latin America and the Caribbean is available. It is thus necessary to conduct studies focused on the energy consumption of this sector that can help to mitigate the environmental sustainability issues that are mentioned in the "Millennium development goals" proposed by ECLAC (2005).

Per capita energy consumption in Latin America and the Caribbean is currently low in comparison with other world regions. However, since the 1990s, it has experienced significant growth, as shown in Figure 4.2. This low per capita consumption does not necessarily indicate high efficiency in the use of energy, as a significant part of the population of these countries lack the funds to have access to a private car. In this context, the rapid development of the region in the medium term might lead to unsustainable increases in the energy consumption of the transport sector and to the associated emissions of greenhouse gases. For example, between 1990 and 2007, the vehicle fleet that was used in Brazil, Mexico, Chile and Colombia increased by 53 million vehicles (the amount tripled), with 40% of this increase concentrated between 2003 and 2007. Therefore, it is crucial to elaborate orderly development strategies that favour public transport[51] and promote energy efficiency.

Figure 4.1. Final energy consumption by activity sector (average for Latin America and the Caribbean in the 1990-2010 period)

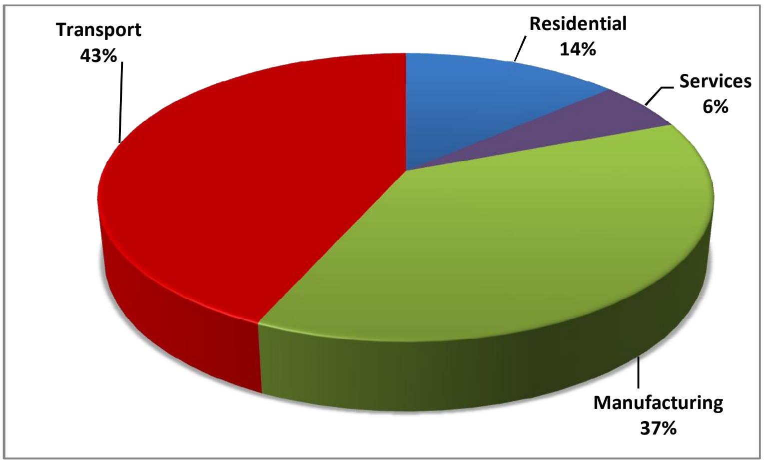


Figure 4.2. Energy consumption in toe per capita in transport (average for Latin America and the Caribbean in the 1990-2010 period)

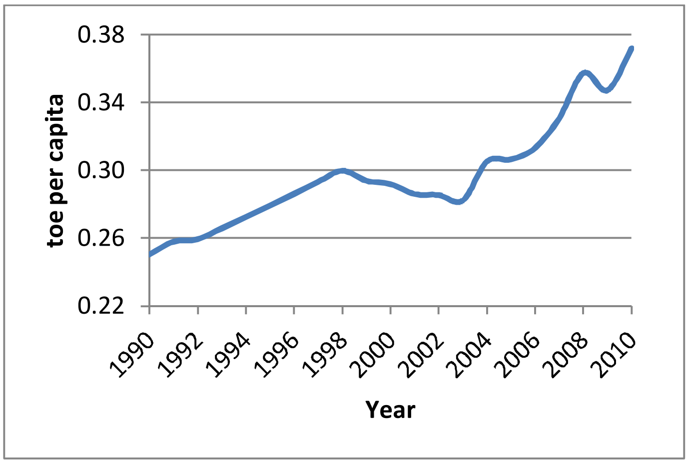
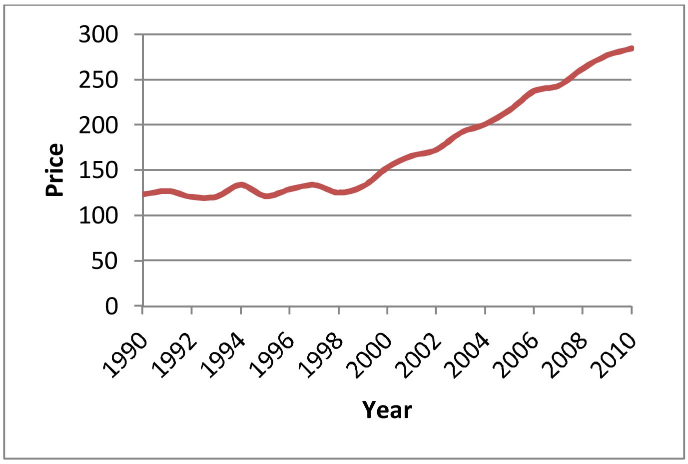


Figure 4.3 shows the change in energy price in the transport sector between 1990 and 2010. This price is a transitive multilateral index that the current authors have elaborated. It adds the weighted prices of the different types of energy that are consumed in transport (See Appendix in Section 4.6). The scenario of low energy prices in the 1990s contrasts with the inflationary process that was experienced in the first decade of the 21st century. This process also led many Latin American countries, especially those that were net importers of energy, to adopt programs to improve energy efficiency. These measures aim, on the one hand, to modernize public transport to

[51] The lack of public services can lead to inefficient individual consumption decisions. The deficit (in quantity, quality or both) in the public transport services, for example, incentivizes private transport, which may generate high costs for the user and cause contamination and congestion in the cities.

incentivize its use, renovate the vehicle fleets, introduce biofuels as alternatives to oil, promote the use of hybrid and electric vehicles, and promote the use of trains and subways in certain activities. On the other hand, the infrastructure network should be improved in tandem with logistical solutions to the provision of services, such as the adoption of intelligent measures that optimize transport routes and favour intermodality (ECLAC, 2010).

Figure 4.3. Price index for energy in the transport sector (average for Latin America and the Caribbean in the 1990-2010 period)



An additional issue that should be addressed is the review of subsidy policies on transport and on products derived from oil, with the aim of transmitting adequate signals to the economy and achieving improvements on energy efficiency. This goal has frequently clashed with the pressure that has been exerted by consumers from various countries who rally against increases in energy prices. In that sense, Latin American countries should introduce a fiscal, incentive and environmental regulations system similar to the ones that exist in other parts of the world, such as the European Union and the United States (Barros and Prieto-Rodríguez, 2008; Chavez-Baeza and Sheinbaum-Pardo, 2014).

To help achieve the goal of reducing energy consumption, various quantitative indicators that are related to the energy efficiency of each country have been developed and have been used in international comparisons. There is no single definition totally accepted for the concept of energy efficiency, both in terms of the economy as a whole or specifically the transport sector. However, Ang (2006) and Stead (2001) indicate that the most common practice has been to link this idea with some thermodynamic, physical-based and monetary-based indicators that relate energy consumption to measurements of the economic activity or energy services derived from this consumption. The most commonly used indicator is the ratio of energy consumed to GDP. This measure of energy intensity has the advantage of simplicity in its calculation and easy interpretation, thus leading to its continued use in international statistics of the International Energy Agency (IEA), the World Bank and the World Energy Council. Decreasing levels of this indicator represent, on average, a reduction in the energy that is required to generate a unit of national production. Therefore, energy intensity is simply the inverse of the energy productivity indicator. Nonetheless, the value of energy intensity can vary significantly over time due to the changes in the structure of GDP, which are difficult to assimilate into the concept of energy efficiency.[52] Furthermore, these types of measures are not "relative", i.e. do not allow cross-country comparisons for countries with better practices or the calculation of potential energy savings.

The main goal of this chapter is to adapt the methodological proposal of Filippini and Hunt (2011, 2012), for energy consumption in the transport sector of Latin America and the Caribbean. This adaptation is performed to obtain a relative measure of energy efficiency that overcomes the weaknesses of other indicators and can serve for international comparisons that are consistent throughout time. The measures of energy efficiency that are obtained are bounded and allow for the determination of potential energy savings given the characteristics of a country. Furthermore, the current study estimates various functions of frontier demand using a latent class approach. This type of approach takes into account the potential existing heterogeneity among the countries analysed, obtaining different demands that are associated with specific price and income elasticities for different country groups. To the best of our knowledge, this study is the first to apply this type of methodology for both the transport sector and the Latin American countries.[53]

This chapter is organized as follows. In Section 4.2, we define the general demand for energy in the transport sector by providing a brief review of the existing literature. Additionally, we propose the use of a Stochastic Frontier Analysis (SFA) approach and the application of a latent class model. In Section 4.3, we present the database and the econometric specification of our models. The results of the estimations are presented in Section 4.4 and finally, Section 4.5 ends the chapter with a summary and the presentation of conclusions.

4.2. Energy demand of the transport sector

Transport demand is derivative in nature, as the goal of moving goods and people is not to perform the journey but to reach a certain destination. In other words, demand is derived from the mobility of passengers and goods. This mobility, in turn, leads to energy or fuel demand, which is necessary for transport.

The previous research in the literature on the modelling of energy consumption for transport can be clustered into works that apply econometric techniques, those that use artificial intelligence approximations, those that use multi-criteria analysis and those that employ simulation methods (see Limanond *et al.*, 2011 or Suganthi and Samuel, 2012 for a review). The first group includes multiple linear regression models (Limanond *et al.*, 2011), partial least square regressions (Zhang *et al.*, 2009) and the analysis of time series and cointegration (Samimi, 2003; Galindo, 2005; Sa'ad, 2010; Hao, 2011). The second group includes studies of artificial neural networks (Dreher *et*

[52] The IEA (2014) recognises that the use of energy intensity as a proxy for energy efficiency can generate untrustworthy results. Despite significant interest in the measurement of energy efficiency, its calculation for the transport sector is a difficult task. This organization proposes indices of energy intensity for the sector that are calculated using various disaggregated indicators that are obtained from large quantities of information. Due to this requirement, it is impossible to calculate this measure for all Latin American and Caribbean countries.

[53] The scarcity of empirical analyses in this context has been conditioned by the availability of statistics. In fact, in many Latin American countries, there is no formal link between institutions that are in charge of providing information on energy and transport. Consequently, in this chapter, all variables that are relative to energy consumption are based on the author's own work on the data provided by the Latin American Energy Organization (OLADE in Spanish).

al., 1999; Murat and Ceylan, 2006; Limanond *et al.*, 2011) and harmony search algorithms (Haldenbilen and Ceylan, 2005; Ceylan *et al.*, 2008). Some studies have even combined the analysis of time series and fuzzy logic (Al-Ghandoor *et al.*, 2012). In the prediction of energy consumption for vehicles, the use of multi-criteria analysis should be noted, such as in the works of Lu *et al.* (2008) and Lu *et al.* (2009). Lastly, the most prominently used simulation model has been the Long-range Energy Alternatives Planning System (LEAP), which allows planning alternative scenarios for energy demand in the transport sector. The works that utilized this method include Bauer *et al.* (2003), Manzini (2006), Pradhan *et al.* (2006) and Islas *et al.* (2007).

Therefore, there is an extensive body of literature on the economics of transport that estimates various functions of energy consumption or the respective functions of fuel use for different types of vehicles. These studies have typically aimed at predictive purposes. The current study belongs to the line of econometric approximations of energy demand from the transport sector that calculates the price and income elasticities that are related to energy consumption (see, for example, Dahl, 1995). In their literature review, Graham and Glaister (2002) observe that, as a general rule, price elasticities that are obtained in the short term are commonly between -0.2 and -0.3 and that those obtained in the long term are between -0.6 and -0.8. For the case of the income elasticities, they find that are often greater than one (between 1.1 and 1.3) in the long term and between 0.35 and 0.55 in the short term. The papers that are included in their review generally analyse Organization for Economic Cooperation and Development (OECD) countries. Wohlgemuth (1997) presents elasticities for several countries that are not OECD members. In terms of Latin America and the Caribbean, the elasticities for Mexico[54] and Brazil are presented. In the long term, the income elasticities for Mexico are between 0.99 and 1.72 and the price elasticities are between -0.04 and -0.21. For the case of Brazil, the income elasticities are between 0.88 and 1.10 and the price elasticities are between -0.10 and -0.26.

In general, in the traditional transport literature, energy demand is understood as a standard demand function. As previously mentioned, in the proposal that is presented below, a stochastic frontier function, which is similar to the production/cost functions that are commonly estimated in efficiency and productivity studies, is considered.

4.2.1. A stochastic frontier approach for energy demand in transport

A generic function of energy demand, which positively depends on income and inversely depends on prices, can be presented in the logarithmic form as follows:

$$\ln q = \ln f(P, Y, X, \beta) + \varepsilon \tag{4.1}$$

where q represents the quantity of the demanded energy, P is the price of energy, Y represents income, X refers to other control variables, β are the parameters that are associated with the variables that are included in the model and can be directly interpreted as elasticities, and ε is the random error, which is commonly assumed to follow a normal distribution with a mean of zero and constant variance, σ_ε^2.

This assumption for the stochastic part of the function indicates that the researcher assumes that any deviation in energy demand that is predicted by the deterministic part of the model is a result of random shocks such as measurement errors

[54] Although in the paper of Wohlgemuth (1997) Mexico is included in the group of countries that are not members of the OECD, this country was already a member since May 18, 1994.

or uncertainty. Therefore, this model can be estimated using the usual OLS estimator, which allows for consistent and unbiased estimates of the model parameters under certain assumptions.

Although this approach has traditionally been used in empirical work, it does not provide direct information on one of the main issues of interest in the field of energy consumption in recent decades, i.e., energy efficiency. As stated in the previous section, there has been debate about the definition and measurement of this concept. In essence, this concept attempts to capture the relation between energy consumption and the production or service that is derived from this consumption. It should be measured in such a way that an improvement in the indicator implies a lower use of energy to produce a certain amount of output in a given economy.

However, in contrast to the research in the energy economics literature, the production economics field has developed various approaches that allow for the inclusion of efficiency in the activities of companies (or countries) within the random part of the model, without the need to add new variables or rely on other indicators. Based on the efficiency and productivity literature, Filippini and Hunt (2011, 2012) suggest the use of an SFA approach to estimate aggregate energy demand functions that are derived from a cost function in the provision of energy services. In this cost function, energy is an input and thus, following Shephard's lemma and deriving the function based on the price of energy, the demand function of this input can be obtained. The main goal of those authors is to obtain measures of energy efficiency that can be used as alternatives to the typical indicators of energy intensity. These efficiency measures are based on the comparison of the energy consumption of the countries with respect to the minimal energy consumption predicted by the frontier, which takes into account the optimizing behaviour of companies and individuals.

The basic model that is estimated by those authors is the standard ALS model, but they also estimate other models developed in the efficiency and productivity literature, such as the TRE presented by Greene (2004, 2005a, 2005b) or the formulation of Mundlak (1978) that was proposed for an estimator of random effects by Farsi *et al.* (2005). The ALS model can be presented for the case of energy demand as follows:

$$\ln q = \ln f(P, Y, X, \beta) + v + u \tag{4.2}$$

where the random term can be decomposed in v, which is a normal distribution that is analogous to that represented by ε in equation (4.1), and u, which is an asymmetric error that follows a half-normal positive distribution to capture the inefficiency of energy demand. As we have seen in the first chapter, in the SFA literature it is typically assumed that u is a negative half-normal (or truncated normal) if the function that is estimated is a production function with a maximum achievable production and positive if the estimated function is a cost function with an achievable minimum cost. In the case of a frontier demand, such as the proposed by Filippini and Hunt (2011, 2012), efficient energy demand represents a minimum feasible consumption. Thus, the approach that is used is the same as that for a cost function.

Based on the conditional mean of the inefficiency term proposed by Jondrow *et al.* (1982), the efficiency level for each observation can be obtained by applying the following expression:

$$EF_{it} = \frac{q_{it}^*}{q_{it}} = \exp\left(-\hat{u}_{it}\right) \quad (4.3)$$

where q_{it}^* represents the aggregate energy demand of the country i in period t on the frontier, i.e., the minimum level of energy necessary for this economy to produce its output level; q_{it} is the aggregate energy demand that is actually observed in this country; and EF_{it}, is thus a measure of efficiency that is bounded between zero and one. The difference between 1 and this measure of inefficiency shows the amount of energy consumption that could be reduced in this country (expressed as a decimal fraction) while maintaining the same level of transport services. Therefore, these are relative measures that, in contrast to energy intensity indicators, allow for direct comparisons between countries throughout time.

To explain the concept of stochastic frontier in a demand context, in Figure 4.4, we compare various approaches that could be used in the econometric estimation of energy demand functions. The blue line shows the energy demand function that is proposed in Equation (4.1) as estimated using OLS. With this approach, we obtain a function with a negative slope in relation to the prices that pass through the mean of the observed values. A basic frontier model that would allow the identification of countries' efficiency would simply assign the whole estimated error (i.e. $\hat{\varepsilon}$) that was obtained from applying OLS to the inefficiency. This simple approach does not allow the separation of inefficiency from noise because, by definition, a deviation from the minimum possible consumption that can be achieved is attributed to inefficiency. This type of frontier is typically known as deterministic frontier and can be obtained by moving the intercept of the OLS estimation until all observations are to the right of the estimated frontier. This form of frontier attainment is known as Corrected Ordinary Least Squares (COLS). In other words, it allows for the attainment of a function that envelops all observations. In the current case, it is represented by the blue dashed line. Although Filippini and Hunt (2011, 2012) do not represent it graphically, the demand that is estimated when an SFA approach is used, is a function such as that represented by the green line. The use of this type of methodology allows for certain observations to be to the left of the estimated frontier due to the existence of negative random shocks, although the majority of observations are to the right of the frontier due to the inefficiency effect.

Figure 4.5 represents the type of frontier that is estimated when using the SFA approach to obtain energy demand and how the Overall Random Error (ORE), i.e., the stochastic part of the model (v+u), can be decomposed into inefficiency and noise[55] in the various possible cases. As shown for observation 1, an observation lies only at the frontier that we have presented graphically when the inefficiency term compensates the negative value of the noise term (or both are equal to zero). By estimating a stochastic frontier demand, it is assumed that the majority of observations will be located to the right of the frontier. This can be due to an effect either of inefficiency or noise (if it is positive and inefficiency is equal to zero), as in observation 2, or to both of these effects together, as is the case in observation 3 (in which both are positive) and 4 (in which only one part of the inefficiency is compensated by the negative value of the noise

[55] The random component v includes events which cannot be controlled by transport companies or the individuals who use the private vehicle, such as those caused by the weather or natural disasters. Thus, by considering for Latin American countries the average amount of days per year when their transport infrastructure (roads, bridges, etc.) is cut off due to these causes, a deviation greater than (lower than) this value in a given year would produce a positive (negative) shock.

term). Nevertheless, as this is not a deterministic frontier, some observations can lie to the left of the estimated frontier, indicating that these countries use less energy than is predicted by the frontier for a specific price. Observation 5 is to the left of the estimated frontier because there is no inefficiency and the error term is negative. In observation 6, even with the existence of inefficiency, the negative noise term exceeds the value of u and thus, this observation is "super-efficient".

Figure 4.4. Approaches in the estimation of energy demand functions

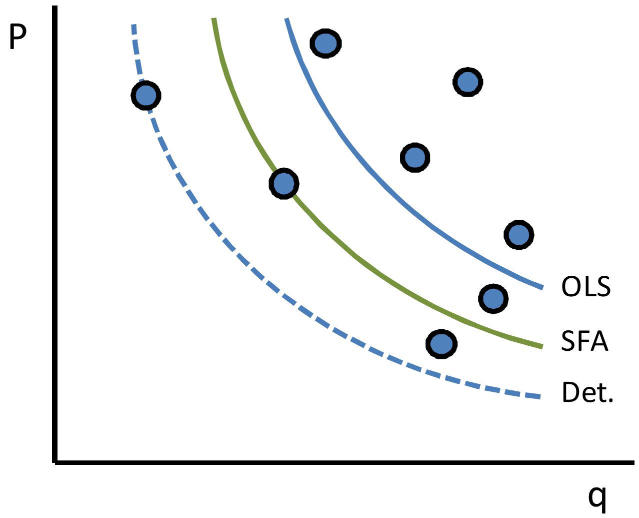
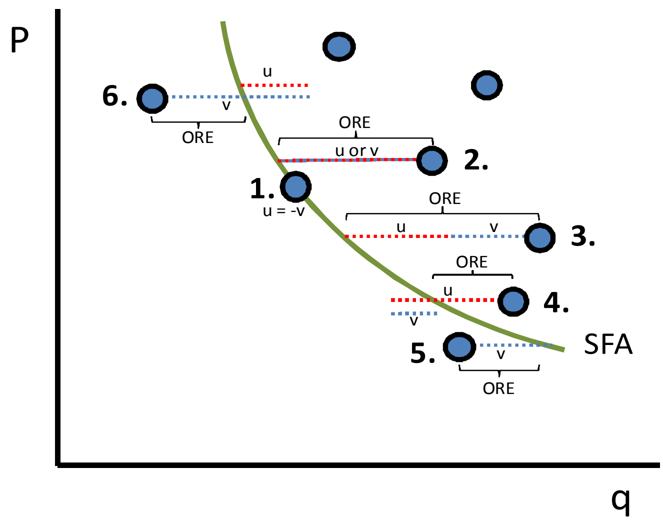


Figure 4.5. Decomposition of the random error term in a stochastic frontier demand



4.2.2. Treating unobserved heterogeneity with a latent class model

As we have seen in the first chapter, from the influential work of Aigner *et al.* (1977), a broad body of literature has been developed to attempt to precisely measure the efficiency of the studied individuals (firms, countries, etc.) with various methodological proposals that allow for solving specific problems that affect the obtained results. One of the main weaknesses of the basic model that is proposed in equation (4.2) is that despite the fact that its specification allows to control for random noise, the presence of unobserved heterogeneity between the studied individuals can bias the efficiency measures (see Greene, 2005a, 2005b).

In Chapter 2 we have mentioned that this heterogeneity is typically considered an unobserved determining factor of the estimated production or cost frontier, and inefficiency is interpreted as the distance to the frontier once heterogeneity has been taken into account. Multiple empirical strategies, each with specific advantages and drawbacks, have been developed to solve this problem. A first approach that can be applied, is the use of a specification that includes individual effects (fixed or random), as is the case for the TFE and TRE models proposed by Greene (2004, 2005a, 2005b). These models include a series of country-specific intercepts that are simultaneously estimated with remaining parameters of the model and allow the distinction between unobserved heterogeneity (which does not change over time) and inefficiency. In this approach, unobserved heterogeneity additionally enters the model as an individual-specific intercept and, therefore, is a neutral or parallel movement of the function that maintains the remaining common parameters for all individuals. In the case of energy demand, as estimated in the current chapter, this implies that specific characteristics of this demand, such as their price and income elasticities, are the same for all countries analysed. This assumption is difficult to justify for such a heterogeneous region as Latin America and the Caribbean. If there are different groups of countries in the sample with different demand characteristics, i.e., different parameters that are associated with the variables, we should estimate a model that allows us to take this feature into account.

An alternative approach to control for unobserved heterogeneity that seems to be adequate for the current context is the LCSFM, such as the proposed by Orea and Kumbhakar (2004) and Greene (2004, 2005b). This model allows for estimation of different parameters for countries that belong to distinct groups and share similar characteristics. The characteristics of the countries in each group differ and thus, given that the countries that belong to the same class share the same set of parameters, this approach controls for the existing heterogeneity between the groups. In other words, the latent class procedure allows us to control for heterogeneity in the slopes (the coefficients of the estimated variables), which is unobserved and associated with country groups. The estimation of a model of this type implies the existence of J groups of countries, which demonstrate differences between themselves in terms of their behaviour function:

$$\ln q_{it} = \ln f\left(P_{it}, Y_{it}, X_{it}, \beta_j\right) + v_{it}\big|_j + u_{it}\big|_j \tag{4.4}$$

where the subindex $j = 1,\ldots,J$ refers to class, β_j is the vector for the parameters that are estimated for class j, and the random term, as in prior models, is composed of $v_{it}|_j \sim N(0,\sigma_v^2|_j)$ and $u_{it}|_j \sim N^+(0,\sigma_u^2|_j)$, which are also specific for each class. The estimation of this model implies the maximization of the overall likelihood function from Equation (4.4), which is the sum of the likelihood functions at each point of the sample weighted by the probability of belonging to each class. This, in turn, is parameterized as a multinomial logit model. Additional variables can be included in the probabilities of class membership. If such variables are not included, the model uses the goodness of fit of each class to identify the distinct groups. The estimation procedure of latent class models and usual model selection criteria have been presented in the previous chapter.

After the model estimation, the posterior probabilities can be obtained to assign each country to a specific class and calculate the efficiency measures. One strategy to assign countries is assuming that the country belongs to a class to which it may belong with the highest probability. Therefore, only one of the demands is taken as the reference frontier to obtain the (in)efficiency measure for each country.[56] An alternative method, as Greene (2005b) proposes, is to take all classes into account when calculating country efficiency, i.e., adding the specific efficiencies of belonging to each of the classes weighted by the probability of belonging to them. However, here, as in Chapter 2, we use the first approach with the understanding that groups of countries actually have different demands.

4.3. Data and econometric specification

This section presents the data and the econometric specification of the models to be estimated that were presented above. Incomplete panel data are used, for the 1990-2010 period, from the following 24 countries in Latin America and the Caribbean: Argentina, Barbados, Bolivia, Brazil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Granada, Guatemala, Guyana, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, Dominican Republic, Suriname, Trinidad and Tobago, Uruguay and Venezuela.[57] The econometric specification of the basic model (ALS) is the following:

$$\begin{aligned} \ln q_{it} &= \beta_0 + \beta_Y \ln Y_{it} + \beta_{POP} \ln POP_{it} + \beta_P \ln P_{it} + \beta_{ST} ST_{it} + \\ &\quad \beta_{DEN} \ln DEN_{it} + \beta_t t + \frac{1}{2}\beta_{tt} t^2 + v_{it} + u_{it} \end{aligned} \tag{4.5}$$

where q, Y, P, v, u and β are defined as in the prior equations. Analogously to Filippini and Hunt (2011, 2012), we include other explanatory control variables such as *POP*, which represents the population; *ST*, which is the share of the transport sector in the economy; *DEN*, population density; and t, the time trend, which is also introduced squared.[58]

Table 4.1 shows the descriptive statistics of these variables. It should be mentioned that the dependent variable, q, represents the final energy consumption of the transport sector, expressed in thousands of toe. It is obtained by adding the total of the energy consumption in internal transport[59] for each country for both passengers and goods. The types of energy that are included in this aggregate are natural gas, liquid gas, electricity, gasoline (which includes biofuel), kerosene (jet fuel), diesel oil and fuel oil. Y, is the GDP of each country and is measured in millions of 2005 US dollars at Purchasing Power Parity (PPP). In international analyses, the use of this exchange rate

[56] In this chapter, we estimate the models assuming a panel data structure, i.e. the probabilities of belonging to each class are constant over time for each country. Therefore, each country is assigned to a single group throughout the sample period.

[57] The sample is composed of a total of 503 observations. The observation for Barbados in 2010 is not included because it is unavailable. Of the 27 country members of OLADE, Belize and Haiti are not included due to lack of information. Furthermore, Cuba is not included in the sample, as the inclusion of this country in the analysis does not allow for the convergence of estimates in some models because the estimated function does not fulfil the convexity property and, in other models, the obtained values for efficiency are practically zero. Due to these results, the observations for this country are considered to be outliers and, thus, we exclude them from the sample.

[58] However, in our model, we do not include meteorological variables because we analyse energy demand in the transport sector and such variables do not play a relevant role as in the modelling of total energy demand or the residential sector of a country. However, possible persistent meteorological differences would be controlled for in the latent class model, which precisely allows the treatment of unobserved heterogeneity.

[59] Internal transport includes domestic aviation, domestic shipping, roads and railways and excludes international maritime and air transport.

is indispensable for adequately comparing the GDP across countries. *POP* is the mean population for each country, as measured in thousands of inhabitants. *P* is an energy price index in the transport sector, calculated as the weighted sum of mean prices of the types of energy used in the sector. Because OLADE and other energy international agencies do not provide any price index for the total of the countries of Latin America and the Caribbean, we have calculated a transitive multilateral price index that allows for consistent comparisons between countries throughout the sample period (see Appendix in Section 4.6). *ST* is the ratio of Gross Value Added (GVA) in transport and the total GVA for each economy, and it is expressed in percentage. Lastly, *DEN* reflects the ratio between the population in thousands of inhabitants and the area of each country in km^2. This variable and per capita income (*Y/POP*) are also included in the LCSFM model within the class membership probabilities to help with the segmentation of the sample.[60] Concerning the data sources, the variables *q*, *P* and *POP* are derived from the Energy-Economic Information System of the OLADE. The variables *ST* and *DEN* are obtained from ECLAC. The variable *Y* is obtained from the data in the Penn World Table (PWT 7.1) presented by Heston *et al.* (2012).

Table 4.1. Descriptive statistics

Variable	*Units*	*Mean*	*Std. Dev.*	*Max.*	*Min.*
q	Thousands of toe	6,141	12,434	69,384	18
Y	Millions of US Dollars (2005)	164,968	339,168	1,800,000	713
POP	Thousands of inhabitants	20,517	38,114	195,498	91
P	Index	174.56	108.64	850.66	3.76
ST	%	4.02	1.59	12.74	1.07
DEN	Thousands of inhabitants / km^2	0.10	0.14	0.63	0.00

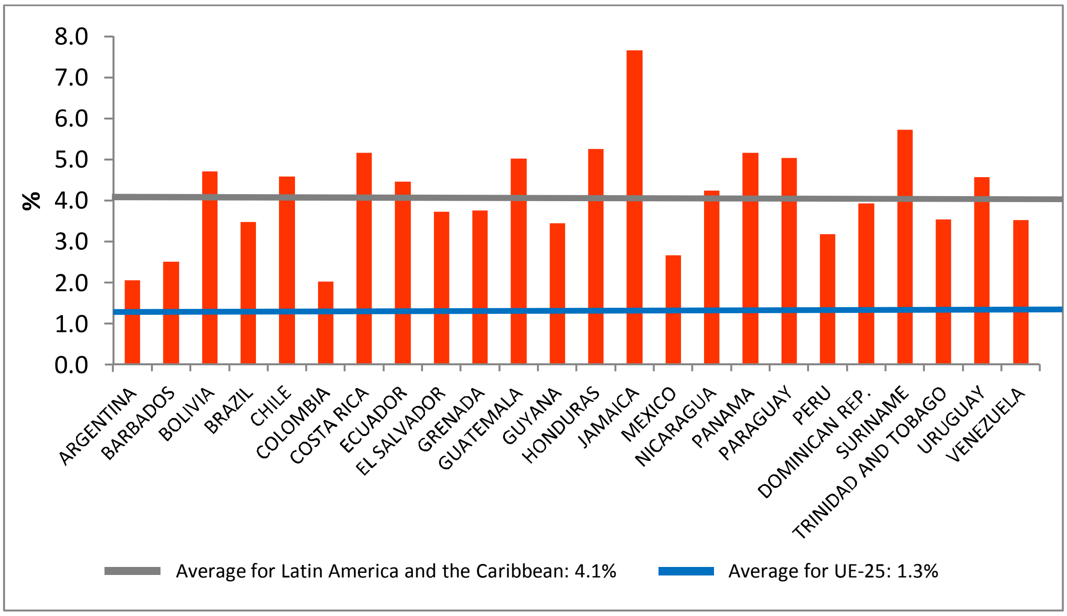
If we pay special attention to the quantity and price of the consumed energy (i.e., the most relevant variables in a demand analysis apart from income), significant differences between countries can be observed. Figure 4.6 shows that energy consumption in transport for Latin America and the Caribbean evidenced a significant dynamism during the period analysed, with an average annual growth of 4.1%, which is more than double that of the growth in the UE-25 (1.3%) for the same period. Nevertheless, the growth rates were quite different among countries, with Jamaica and Suriname displaying the highest growth and Argentina and Colombia presenting the lowest growth for the period of analysis.

In Figure 4.7, we represent the time evolution of the price index for the group of countries that evidenced the greatest differences in 2010. It should be noted that during the years analysed, Venezuela persistently maintained the lowest prices. Furthermore, it is noteworthy to mention the low cost of energy in Ecuador and Mexico. By contrast, the highest prices are found in Colombia, Bolivia, Brazil and Argentina. Furthermore, we observe that the price index that is used in this chapter does not require that its own value be equal for all countries in a base year, which is required when standard indices

[60] The lack of homogenous information or a sufficient timeframe on the transport infrastructure, stock of vehicles, distances travelled or goods and passenger traffic indicators, impedes the inclusion of these types of variables in the estimated demands.

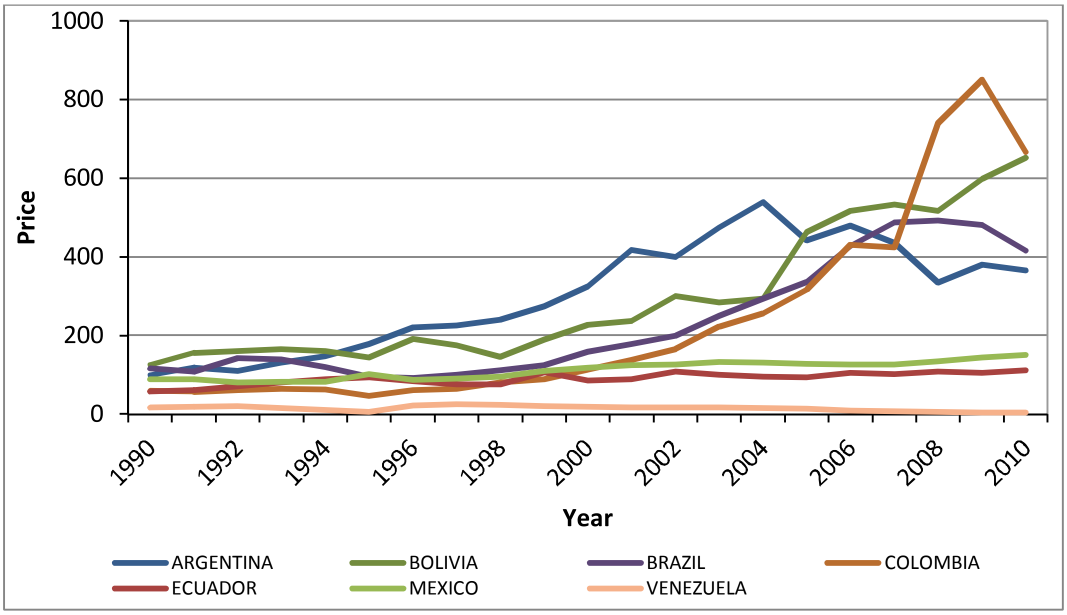
such as Laspeyres or Paasche are applied. As discussed in the Appendix, this facilitates a better fit of the estimated energy demand functions.

Figure 4.6. Average annual growth rate of energy consumption in transport for Latin America and the Caribbean, 1990-2010 (percentage)



Sources: ECLAC and EUROSTAT

Figure 4.7. Transitive multilateral price index of energy in the transport sector for Latin America and the Caribbean, 1990-2010



4.4. Estimates and results

Table 4.2 shows the results of the basic ALS model estimation. As previously mentioned, the model assumes the existence of a single demand and, therefore, does not allow for different elasticities for the various countries in the sample. All of the variables that are included in the models are statistically significant at 99% (except the time trend squared) and have the expected signs. The values of the income and price elasticities are 0.81 and -0.23, respectively. These elasticities are found within the value ranges that are obtained in the energy demand in transport papers, as discussed in Section 4.2. The coefficient of the population variable has a positive sign, which indicates (as expected) that a population increase leads to, *ceteris paribus*, an increase in the energy demand. A similar interpretation can be made for the share of the transport sector in the economy, which can be understood as a proxy for the degree of transport development. It can be expected that a more developed sector results in greater welfare for society, which is achieved through greater energy consumption. However, density presents a negative sign, indicating (as expected a priori) that the countries that are more densely populated have, *ceteris paribus*, lower transport energy demand due to the smaller average distances that companies and individuals travel. After controlling for the remaining variables in the estimation, the positive sign of the time trend shows that energy consumption increased throughout the sample period (as shown in Figure 4.2), which may indicate technical regress in the sector.[61] The mean value of efficiency is 87.4%. Nevertheless, great variability is found among the observations, with minimum and maximum values of 66.2% and 94.7% respectively.[62]

Table 4.3 shows the results of the LCSFM models for two and three classes[63], which include separating variables in the probabilities of class membership. If we analyse the prior probabilities of the two-class model, the separating variables (income per capita and density) are not statistically significant. Thus, this model is equivalent to a model that does not include separating variables. However, these variables are significant in the three-class model. The signs and values of the variables indicate that countries with higher income per capita and lower population density tend to be assigned to class 1 and, to a lesser degree, to class 2.

Figure 4.8 shows the different information criteria that are used as selection tests to choose the preferred model. As previously mentioned, all of these criteria are based on the maximum value of the likelihood function, which is obtained by estimating each model. These criteria only differ in that they penalize the increased number of parameters that are estimated for each model with different weights. The model with the best fit is that with the lowest criteria value. All of the presented criteria show a clear improvement in the fitness of the estimates when unobserved heterogeneity is addressed in the model through a latent class approach. Although a great improvement can be

[61] This model has alternatively been estimated by including a set of time dummies that capture the non-linear evolution of energy consumption over time. Nevertheless, we prefer the inclusion of a time trend and its square, as it allows the estimation of a latent class model without renouncing the inclusion of the time effect in the model.

[62] A reviewer's suggestion that the inefficiency in our model might include a behaviour that would be the consequence of low energy prices in certain countries led us to estimate a heteroscedastic model of the type proposed by Reifschneider and Stevenson (1991), Caudill and Ford (1993) and Caudill *et al.* (1995). The coefficients that are estimated with such a model for the variables in the frontier are practically identical to those obtained in the ALS model, and the price is not statistically significant in the inefficiency term. The values of the efficiencies that are obtained in this heteroscedastic model are similar to those obtained in the ALS model, with a 96% correlation between the two measures.

[63] Models with higher numbers of classes do not converge.

observed when moving from the ALS model to the LCSFM model with two classes (all criteria present a lower value), this heterogeneity is captured to an even greater extent by a three-class model. For the case of LCSFM models, the criteria values are also shown when these models are estimated without the inclusion of separating variables, although the estimated parameters are not presented in this chapter. In the model with two classes, no improvement is observed when separating variables are included. By contrast, in the model with three classes, these variables have a relevant influence. This three-class model is the one that fits best to the characteristics of our data and, thus, we consider it to be the preferred choice.

Table 4.3 shows that the majority of variables in these models are significant and have the expected signs, as in the ALS model. The preferred three-class model shows large differences in the coefficients between the classes for most of the variables. For example, the population variable coefficient varies between 0.247 and 0.742, the share of the transport sector in the economy is only significant for class 1, population density positively affects[64] energy consumption in group 2 and negatively affects it in group 1, and although all of the classes demonstrate a positive time trend, this growth in energy consumption is increasing in class 2 and decreasing in class 3 according to the sign of the square term in both cases.

Table 4.2. Standard frontier demand model

	ALS model		
Variable	***Coeff.***		***t-ratio***
Intercept	7.098	***	405.450
ln Y_{it}	0.810	***	39.720
ln POP_{it}	0.182	***	8.834
ln P_{it}	-0.229	***	-15.138
ST_{it}	0.047	***	7.103
ln DEN_{it}	-0.096	***	-12.031
t	0.013	***	6.960
½ t^2	-0.001		-1.537
σ	0.257	***	590.578
λ	0.886	***	7.411
σ_v	0.192		
σ_u	0.170		
Log-likelihood	52.689		

Significance code: * p<0.1, ** p<0.05, *** p<0.01

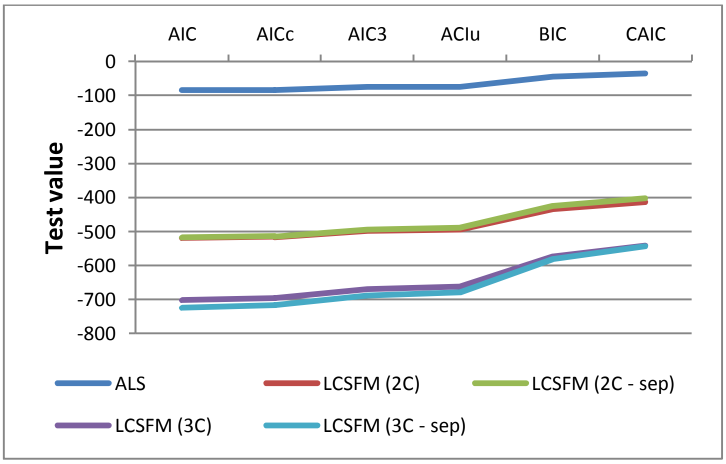
[64] As previously mentioned, a negative coefficient for *DEN* is expected because this variable mainly captures the effect of greater energy consumption as the territory of a country increases given its population. However, the coefficient of this variable is positive in class 2 of the latent class model. This result does not invalidate our intuition on this variable, as this ratio simply includes population divided by area and does not incorporate the degree of urbanization or whether the population is distributed homogeneously in the territory, a circumstance that may condition this result for this class.

Table 4.3. Frontier demands with latent class including separating variables

	LCSFM with two classes						*LCSFM with three classes*								
	Class 1			*Class 2*			*Class 1*			*Class 2*			*Class 3*		
Variable	***Coeff.***		***t-ratio***	***Coeff.***		***t-ratio***	***Coeff.***		***t-ratio***	***Coeff.***		***t-ratio***	***Coeff.***		***t-ratio***
Intercept	7.180	***	271.915	6.903	***	371.389	7.367	***	195.898	7.091	***	192.733	6.894	***	522.491
ln Y_{it}	0.784	***	24.069	0.637	***	35.489	0.566	***	22.754	0.179	***	3.687	0.649	***	35.921
ln POP_{it}	0.188	***	5.491	0.280	***	17.498	0.431	***	16.176	0.742	***	16.154	0.247	***	11.754
ln P_{it}	-0.188	***	-17.045	-0.175	***	-5.417	-0.161	***	-13.869	-0.288	***	-13.057	-0.407	***	-10.379
ST_{it}	0.094	***	8.745	0.037	***	7.048	0.044	***	4.966	0.002		0.307	-0.008		-0.811
ln DEN_{it}	-0.067	***	-5.137	-0.046	***	-6.349	0.007		0.714	0.125	***	6.868	-0.030	***	-4.016
t	0.006	***	2.774	0.016	***	8.642	0.009	***	4.985	0.042	***	16.986	0.030	***	12.527
½ t^2	0.000		0.426	-0.003	***	-5.835	0.000		-0.505	0.003	***	5.718	-0.001	***	-2.720
σ	0.246	***	12.009	0.171	***	11.022	0.166	***	7.336	0.112	***	5.391	0.135	***	12.703
λ	2.961	***	2.974	2.320	***	3.147	1.003	*	1.932	0.999		1.401	3.040	***	3.937
σ_v	0.079			0.068			0.117			0.079			0.042		
σ_u	0.233			0.157			0.118			0.079			0.128		
Class membership probabilities															
Intercept	0.236		0.507	-	-	-	1.036		1.237	0.736		0.870	-	-	-
ln $(Y/POP)_{it}$	0.639		0.674	-	-	-	4.293	**	2.250	3.082	*	1.770	-	-	-
ln DEN_{it}	-0.543		-1.397	-	-	-	-2.090	**	-2.485	-1.005		-1.387	-	-	-
Prior Prob.		0.559			0.441			0.477			0.354			0.169	
Log-likelihood			281.782								398.356				

Significance code: * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Figure 4.8. Model selection tests



The most relevant variables in demand analysis are income and price. As previously mentioned, the latent class model allows us to identify three classes with elasticities that clearly differ. The most inelastic demand for income can be found in class 2 (0.179), followed by class 1 (0.566) and finally, the most elastic class is 3 (0.649). The differences in price elasticities of the demand are also evident if we represent the data from the sample without performing any type of estimation and only use the partition of the sample that is generated by the preferred three-class model, as can be shown in Figure 4.9.[65] This figure shows that the demand of class 1 has the steepest slope and corresponds to the group with the lowest elasticity (-0.161) in the estimates. This group includes Argentina, Brazil, Chile, Ecuador, Guyana, Mexico, Paraguay, Suriname, Trinidad and Tobago and Venezuela. The demand of class 2 corresponds to the group of intermediate elasticity (-0.288) and is composed of Barbados, Bolivia, Colombia, Costa Rica, Jamaica and Panama. Finally, class 3, with the flattest slope in the graph, is the most elastic in the estimates (-0.407) and includes El Salvador, Granada, Guatemala, Honduras, Nicaragua, Peru, the Dominican Republic and Uruguay. This figure also represents the single demand that would be obtained if we did not take into account the heterogeneity between countries, thus obtaining a biased demand with an intermediate slope between class 2 and class 1, which would correspond to the price elasticity value obtained from the ALS model (-0.229).

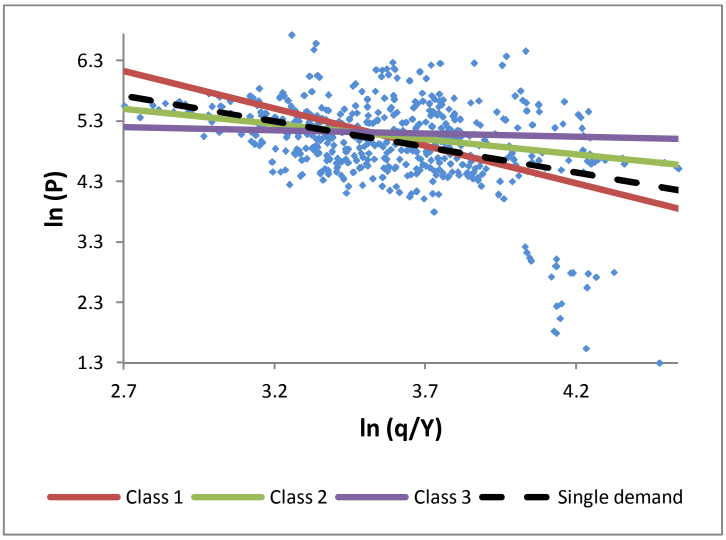
The mean efficiencies that are obtained in each class are around 95%, and the minimum value is consistently greater than 80%. These results indicate that the groups are more homogenous than when one single demand is estimated. These results reflect that more efficient countries can reduce their energy consumption up to 5% and that the less efficient countries have a margin of up to 20%.[66] The estimation of this latent class

[65] In this figure, we present price to energy divided by income. This consideration allows us to "relativize" the weight of income and isolate, to a certain degree, the price effect on demand, which is what we want to represent. On the other hand, logarithms in the units of both axes are calculated to reduce the measurement scale and facilitate the representation of the demand.

[66] These potential savings are however obtained without taking into account possible "rebound effects". This phenomenon basically captures the idea that part of the savings from increases in the efficiency level in the use of energy can be offset by increases in the demand for energy services derived from the marginal cost reduction of those energy services. This concept has recently received great attention in energy studies and especially in transportation (see for instance, Greene *et al.*, 1999b; Small and Van Dender, 2007; or Hymel *et al.*, 2010).

model allows us to identify the most efficient countries in each class (on average for the period considered). The remaining countries in each group, given their similar characteristics, should attempt to imitate these most efficient countries' energy policies. The two countries with the greatest energy efficiencies are Brazil and Mexico in class 1, Barbados and Colombia in class 2, and El Salvador and Guatemala in class 3.[67]

Figure 4.9. Linear demands obtained on the basis of observed values



As mentioned in the introductory section, energy indicators are typically used to measure energy efficiency in countries. The most commonly used indicator of energy intensity is the ratio of energy consumption to GDP in a country. Table 4.4 shows the value of this indicator for the transport sector of each country and presents a ranking of "energy intensity". The countries with a lower ratio of energy consumed in transport to GDP are identified according to this indicator as those that are the most energy efficient. The table also shows the mean efficiencies that are obtained for each country with a frontier demand such as the estimated demand.[68] The correlation coefficients of both measures for each country are in some cases, such as the Dominican Republic (-0.982) and Trinidad and Tobago (-0.986), quite high and negative. This result indicates, as expected, that energy efficiency improvements are associated with decreases in the energy intensity indicators. Although the correlation of these measures is high, on average, it is low in some countries (such as Brazil and Colombia). Furthermore, it is

[67] The reference countries for each of the demands seem to correspond to the countries that, according to ECLAC (2010), have adopted distinctive measures for the improvement of public transport in their cities. In this report, it is highlighted the Rapid Transit Bus (RTB) system implementation in Curitiba (Brazil). This system was started in 1972 as part of a general policy of urban planning. Other noted examples are the RTB TransMilenio, which has been developed since 2000 in Bogota (Colombia). The innovations of this system have made it the most solid RTB of the world and have led it to develop an extension plan of this system to seven additional cities. In Mexico City (Mexico), an RTB system has been implemented, named Metrobús, as a complement to the extensive subway system of the city. In Guatemala City (Guatemala), a trans-urban system was developed in 2009 with the aim of improving efficiency and reducing contamination indices of the transport sector in the city.

[68] For the comparisons with the rankings that are obtained based on energy intensity to make sense, the efficiency values that are shown in this table are obtained using the ALS model, as this is the only estimated model that assumes the existence of a single frontier.

positive in two countries (Chile and Venezuela), indicating that the evolution of energy intensity indicators is associated with circumstances other than energy efficiency. Using the Spearman's rank correlation coefficient, we observe that although the rankings that are obtained by alternatively applying the criterion of energy intensity and efficiency of an energy demand model can differ (for example, Barbados and Trinidad and Tobago fall by 10 places, and Panama moves from 15 to 3 when estimating a frontier model), on average these rankings are similar, with an approximately 70% correlation between them. In summary, these results seem to confirm that the efficiency measures that are derived from the estimation of energy demand frontier models are more appropriate than those that are provided by energy intensity indicators.

Table 4.4. Country ranking using energy intensity and energy efficiency

Country	***Indicator (Energy/GDP)***		***Frontier demand***		***Correlation (EI Vs Eff.)***
	EI	***Ranking***	***Eff.***	***Ranking***	
Argentina	0.037	14	0.845	19	-0.897
Barbados	0.019	1	0.885	11	-0.938
Bolivia	0.042	19	0.869	15	-0.883
Brazil	0.034	12	0.872	14	-0.241
Chile	0.044	20	0.844	20	0.161
Colombia	0.032	8	0.896	7	-0.061
Costa Rica	0.032	9	0.875	13	-0.720
Ecuador	0.055	22	0.828	22	-0.962
El Salvador	0.026	5	0.902	5	-0.931
Granada	0.029	7	0.877	12	-0.807
Guatemala	0.024	2	0.910	4	-0.952
Guyana	0.066	24	0.846	18	-0.956
Honduras	0.033	10	0.890	9	-0.925
Jamaica	0.038	16	0.813	24	-0.914
Mexico	0.040	18	0.861	17	-0.814
Nicaragua	0.040	17	0.888	10	-0.946
Panama	0.037	15	0.914	3	-0.893
Paraguay	0.054	21	0.815	23	-0.951
Peru	0.025	3	0.933	1	-0.763
Dominican Rep.	0.026	4	0.898	6	-0.982
Suriname	0.035	13	0.891	8	-0.906
Trinidad and Tobago	0.033	11	0.834	21	-0.986
Uruguay	0.028	6	0.924	2	-0.706
Venezuela	0.062	23	0.868	16	0.153
Spearman's rank correlation coefficient between both rankings					0.701

Note: *EI* stands for *Energy Intensity* and *Eff.* is the abbreviation of *Efficiency*

4.5. Conclusions

In this chapter, we estimate stochastic frontier demand functions to measure the level of energy efficiency of the transport sector in Latin America and the Caribbean by using panel data from 24 countries for the 1990-2010 period. The adopted approach constitutes a novel contribution to energy demand studies of the sector in this region, conferring great importance to the presented results. Due to the different types of energy that are used in the transport sector, it is necessary to employ an index that aggregates the set of energy prices for the estimation of these demands. International energy agencies do not provide a price index for all of the countries in the sample. Thus, we construct a transitive multilateral index, which allows for consistent comparisons of energy price among countries throughout time. The construction of this price index is a relevant issue often avoided in these studies.

The estimated models are a basic stochastic frontier and diverse latent class models that lead to obtaining differentiated demands. These models allow us to identify the reference countries in international comparisons of energy efficiency. The results indicate that the specification that best fits an energy demand is a model in which three classes are estimated using income per capita and population density as class-identifying variables. In this model, important differences in income and price elasticities are observed. Specifically, countries with higher income per capita and lower population density have a higher probability of having a more inelastic demand in terms of price. The estimation of the latent class model allows us to identify countries that have successfully implemented programs of improved public transport in some of their cities. The remaining countries of each class should follow the example of these countries and perform the extension or adaptation of the national transport sector policies implemented in the most efficient areas of the region, with the aim of improving energy efficiency and reducing the levels of urban contamination. Furthermore, general improvements in fuel efficiency and the transfer from private vehicle use to public transport ought to be additionally considered.

On the other hand, this chapter shows that the commonly used indicators of energy intensity cannot consistently be used as a reasonable reference for energy efficiency in the transport sector. Using efficiencies that are obtained through the frontier approach, we find that although the mean efficiency is relatively high, there is a margin for energy consumption savings and, thus, for a reduction of greenhouse gas emissions. Some measures that can be adapted for this purpose are as follows: correctly assign energy prices, plan the infrastructure and land use jointly to minimize distances, balance the modal distribution, establish fiscal incentives for the use of lower consumption engines, develop fuels with reduced levels of carbon and implement awareness programs that focus on the transformation of transport use toward rational and environmentally sustainable habits.

Finally, according to the "Jevons Paradox", it is possible that increases in energy efficiency do not involve a reduction in energy consumption and hence the energy savings predicted in the current model are not possible to reach. That situation, also called "back-fire", is a particular case of the phenomenon known as "rebound effect". This concept states that part of the savings from increases in the efficiency level in the use of energy can be offset by increases in the demand for energy services derived from the marginal cost reduction of those energy services. In other words, the increased efficiency in the use of a resource does not necessarily indicate a directly proportional decrease of total consumption. The rebound effect concept should be considered (as in next chapter) in future research that uses frontier approaches for the estimation of energy demands.

4.6. Appendix

Construction of the price index

The OLADE provides information on the prices and quantities consumed of the different types of energy that are used in the transport sector of Latin America and the Caribbean. The categories that appear in their database are as follows: natural gas, liquid gas, electricity, various types of gasoline, kerosene, diesel oil and fuel oil. However, this agency does not provide a general price of energy for these countries. Thus, to estimate aggregate energy demand in transport, it is necessary to obtain an indicator or index that accounts for the distinct components in the energy consumption of the sector. In general, a compound price index can be defined as follows:

$$PI_{0t} = \frac{\sum_{m=1}^{M} p_{mt} q_{mt}}{\sum_{m=1}^{M} p_{m0} q_{m0}} \tag{4.6}$$

where PI_{0t} measures the change in value of the total of the M energy components between the base period 0 and final period t. In this type of index, it is difficult to distinguish between the changes that only occur in prices and the change in consumed quantities. The two indices that are most commonly used in practice and calculated by international agencies for total energy consumption, such as those calculated by the IEA, are Laspeyres and Paasche. In the former, the quantities that are consumed in the base year (q_{m0}) are used as weights both in the numerator and in the denominator. Thus, this index isolates the change in prices without accounting for changes in consumption patterns. The second type of index uses energy quantities from the current period (q_{mt}) as weights, thus simultaneously including variations in prices and quantities. These two indices, therefore, represent two extreme cases and only coincide when relative prices do not experience any variation (i.e., p_{mt}/p_{m0} is constant).

However, there are alternatives that combine both approaches to address this issue, such as the Fisher and Törnqvist indices. Nevertheless, all of these indices present the same problem. Specifically, they allow for comparisons of a country with itself throughout time and comparisons between countries measured in price changes (if the same base year is imposed for all countries in the sample), but they do not allow for comparisons of price levels between countries throughout time.

Studies that use international data must employ an index that overcomes this difficulty. The solution to this problem involves obtaining transitive multilateral comparisons (as named in the literature on index numbers) between countries, as proposed by Elteto and Koves (1964) and Szulc (1964). This method, known as EKS, was used by Caves *et al.* (1982) to obtain transitive Törnqvist indices. The formula, in line with Coelli *et al.* (2005), is as follows:

$$\ln PI_{ij}^{CCD} = \frac{1}{2}\sum_{m=1}^{M}\left(\omega_{mj} + \bar{\omega}_m\right)\left(\ln p_{mj} - \overline{\ln p_m}\right) - \frac{1}{2}\sum_{m=1}^{M}\left(\omega_{mi} + \bar{\omega}_m\right)\left(\ln p_{mi} - \overline{\ln p_m}\right) \tag{4.7}$$

where ω_{mi} represents the importance held by component m in the energy expenditure of the transport sector of the country i and $\bar{\omega}_m$ is the arithmetic mean of these expenditure amounts. Furthermore, $\overline{\ln p_m}$ represents the average price of the energy component m for the set of countries.

The intuitive interpretation of equation (4.7) is that to compare the price indices of two countries, each of them is compared to the average country and then the differences from this mean are calculated. Logically, as opposed to other indices, when an observation is added or subtracted from the sample, all values should be recalculated due to changes in the mean of the sample.

It should be mentioned that in the current empirical application, the use of an approach such as the proposed by Caves *et al.* (1982) in the construction of the price index significantly improves the quality of fitting that is obtained when estimating the models. If a Paasche- or Laspeyres-type index is used rather than a transitive multilateral index, the logarithm of the likelihood function falls sharply and achieves negative values. The use of these simpler indices in practice implies the assumption that each country has a specific individual effect. In this case, we artificially introduce heterogeneity into the model. Thus, the model must be estimated by including individual effects, as in the TFE and TRE models.

Chapter 5

A new approach to measuring the rebound effect associated to energy efficiency improvements: An application to the US residential energy demand

5.1. Introduction

As we have commented before, reducing energy consumption and emissions is a key policy objective for most governments across the globe and the promotion of energy efficiency policies is seen as a key activity to achieving this goal. In practice, the achievement of savings in energy consumption depends on two issues. First, it is vital that policy makers be able to clearly measure the relative energy efficiency across states and over time. Second, the actual savings in energy consumption might not coincide with the expected savings due to the so-called rebound effect, a phenomenon associated with the consumption of energy and energy services. When the production of an energy service becomes more efficient, then the cost per unit of this service decreases. This cost reduction can produce an increase in the consumption of the energy service that might (at least partially) offset the expected savings in energy consumption derived from the energy efficiency improvements. Measuring the rebound effect is thus crucial in order to properly evaluate the effectiveness of any energy policy instrument that aims to promote energy efficiency improvements.

Regarding the first issue, Filippini and Hunt (2011, 2012) point out that defining and measuring energy efficiency and creating statistical measures as descriptors is a challenging task. They propose the use of an SFA approach to control for characteristics such as the structure of the economy that might bias the usual energy efficiency indicators. These authors illustrate their proposal by estimating an aggregate energy demand frontier model for the total energy consumption of a sample of OECD countries and for the residential energy consumption of the US states. The SFA approach allows them to obtain a "pure" measure of the inefficient use of energy (i.e. 'waste of energy') for each country or state.

Concerning the second issue, there is a large number of empirical studies that use econometric methods to estimate the rebound effect. In their review of the literature, Sorrell and Dimitropoulos (2008) have found a lack of consensus with regard to a consistent method to measure the rebound effect. In principle, it could be *directly* obtained from the elasticity of demand for energy services with respect to changes in energy efficiency. However, relatively few studies follow this approach because data on either energy services or energy efficiency are unavailable or are limited in terms of accuracy. As a consequence the rebound effect is often indirectly measured through the estimate of different elasticities that are considered measures of energy efficiency elasticities of the demand for energy, such as the own-price elasticity of the demand for energy.

The main contribution of this chapter is to link the energy demand frontier approach with the estimation of the rebound effect. We first bring attention to the fact that the frontier model introduced by Filippini and Hunt (2011, 2012) that has been applied in the previous chapter, also provides a direct measure of the rebound effect. However, we point out that a traditional specification of this model implicitly imposes a zero (or more accurately, constant) rebound effect, which contradicts most of the available empirical evidence. We next suggest estimating a more comprehensive model to relax the zero rebound effect assumption and examine the compliance with some of the restrictions used in previous studies focused on estimating the rebound effect using econometric techniques.

The chapter is organized as follows. The next section defines the rebound effect and provides a brief review of the empirical literature on measuring it using econometric models. Both standard and extended energy demand frontier models and the econometric specification of our model are introduced in Section 5.3. The data and results of the estimates are presented in Section 5.4 with a summary and conclusions in the final section.

5.2. Measuring the rebound effect: a short review of the empirical literature

The rebound effect is a phenomenon associated with energy consumption. This concept has to do with the idea that an increase in the level of efficiency in the use of energy decreases the marginal cost of supplying a certain energy service and hence may lead to an increase in the consumption of that service. This consumer reaction might therefore partially or totally offset the predicted reduction in energy consumption attributed to energy efficiency improvements using engineering models. Measuring the rebound effect is thus crucial in order to properly evaluate the effectiveness of any energy policy instrument that aims to promote energy efficiency improvements. This issue is particularly relevant for the US residential sector since it accounts for 37% of the national electricity consumption, 17% of greenhouse gas emissions and 22% of primary energy consumption (International Risk Governance Council (IRGC), 2013).

The definition of the rebound effect encompasses different mechanisms that may reduce potential energy savings derived from the improvements in energy efficiency. Frequently, three types of rebound effect are distinguished in the specialized literature. The first one is the *direct* rebound effect, which measures the increase in the use of the product or service that has experienced the efficiency gain. For instance, a homeowner may employ a portion of the energy savings from using an efficient heater to use the heater for longer periods during the winter to warm the house. The second type is the so-called *indirect* rebound effect and measures the reallocation of energy savings to spending on other goods and services that also require energy. For instance, the savings derived from the use of energy-efficient appliances at home can be spent on travel holidays which may lead to an increase in energy consumption and greenhouse gas emissions. The third type is the *economy-wide* rebound effect and captures the structural changes in the economy due to the variation of prices of goods and services as a consequence of energy efficiency improvements. These changes may produce a new equilibrium in the consumption of goods and services (including energy) in the economy.

There is an extensive literature on the concept and measurement of the rebound effect and several approaches have been applied with the aim of quantifying this phenomenon. For instance, in their report for the UK Energy Research Centre, Sorrell and Dimitropoulos (2007) find a wide range of methods that have been applied to measure the direct rebound effect. They identify at least four empirical approaches - single equation models, structural models, discrete/continuous models, and household production models - and several estimation techniques including ordinary least squares, instrumental variables or maximum likelihood. In addition, several empirical strategies have also been used to indirectly measure this rebound effect. An outline of these approaches can be found in Table 5.1. This table shows three theoretical relationships between two elasticities. The left-hand side elasticity is the energy efficiency elasticity of the demand for energy, which is used to calculate the clearest and most direct measure of the rebound effect (see Saunders, 2000, and Section 5.3 below). The lack of accurate data on energy services or energy efficiency typically precludes a direct measurement of the rebound effect based on this elasticity, so that its estimation is usually carried out using the right-hand side of the equations in Table 5.1.

Table 5.1. Approaches for measuring the direct rebound effect

Approach 1	$\varepsilon_E(q) = \varepsilon_E(S) - 1$
Approach 2	$\varepsilon_E(q) = -\varepsilon_P(S) - 1$
Approach 3	$\varepsilon_E(q) = -\varepsilon_{P_q}(q) - 1$

Notes: Letters in parentheses stand for elasticity numerators and subscripts for elasticity denominators. E: energy efficiency; q: Energy; S: Useful work; P_S: Energy cost of useful work; P_q: Energy price.

The first empirical approach relies on estimating the energy efficiency elasticity of the demand for energy services or useful work that is often available in personal transportation studies. For this reason, this engineering-based approach is generally used to measure the direct rebound effect associated with travelling by private cars (see for instance Greene *et al.,* 1999b; or Small and Van Dender, 2005). More studies follow the second empirical strategy, based on an estimate of the energy cost elasticity of the demand for useful work. This approach has been advocated by Khazzoom (1980), Greene *et al.* (1999a), Berkhout *et al.* (2000) and Binswanger (2001) and, unlike the first approach, it provides a way to estimate the magnitude of the rebound effect even when the available data provides little or no variation in energy efficiency. However, the validity of this approach relies on the assumption that consumers respond in the same way to decreases in energy prices as they do to improvements in energy efficiency (and vice versa). As Sorrell and Dimitropoulos (2008) pointed out, this assumption is likely to be flawed in many cases. These two approaches require accurate measures of the demand for useful work. This restriction has biased research studies towards personal transportation and household heating, where data about energy services can be easily calculated, e.g., vehicle kilometres in the case of transportation.

It is also possible to estimate the direct rebound effect from the own-price elasticity of the demand for energy, i.e., the third approach. While obtaining measures of useful work can be difficult, data on energy demand is more commonly available. The main advantage of the third approach over previous approaches is that data on either useful work or energy efficiency is not required. This explains why the approach based on the own-price elasticity of the demand for energy is the most popular empirical strategy to measure the rebound effect in other energy commodities or sectors (see for instance Zein-Elabdin, 1997; Berkhout *et al.*, 2000; Roy, 2000 and Bentzen, 2004).

However, Sorrell and Dimitropoulos (2008) pointed out that this empirical strategy might also yield biased estimates for the rebound effect if energy efficiency is not explicitly controlled for.[69] In this chapter, we propose another approach based on the estimation of an energy demand frontier function. In this framework, the rebound effect is directly estimated from the elasticity of the demand for energy with respect to changes in the level of energy efficiency.

There is a huge variety of estimated rebound effects in the literature not only because different methodological/empirical approaches have been used but also because they have been used to analyse the rebound effect for different energy commodities, sectors, countries or different levels of data aggregation. Since this chapter is focused on residential energy demand, we pay attention mainly to the results of papers on household energy demand. Sorrell and Dimitropoulos (2007) find that for household heating the rebound effect usually ranges from 10% to 58% in the short-term and from 1.4% to 60% in the long-term.[70] Household energy demand is dominated by the use of fuel and electricity for heating space. Focusing specifically on papers in which the price elasticity of total household electricity demand is estimated, the estimated values suggest an upper bound for the short-term rebound effect in the range of 20% to 35% and between 4% and 225% for the long-term rebound effect. Regarding other household energy services, the reviewed studies suggest a rebound effect up to about 26% for space cooling. Other studies produce rather different results. For instance, Guertin *et al.* (2003) estimate long-term rebound effects for both water heating and appliances/lighting and obtain values between 32% and 49%.

A recent survey can be found in a report on energy efficiency carried out by the IRGC (2013). This survey is based on the reviews of Greening *et al.* (2000), Sorrell (2007) and Jenkins *et al.* (2011) and summarises the large variety of results obtained from papers that measure rebound effects in the residential sector. This report shows that while for residential lighting there is a narrow range of results of the rebound effect from 5% to 12%, in the rest of energy services there is a wider range of values: for space heating the range goes from 2% to 60%, for space cooling from 0% to 50%, for water heating from less than 10% to 40%, and for other consumer energy services from 0% to 49%. As it can be seen, this more updated survey shows very similar values to the report previously mentioned.

However it should be noted that in this chapter we estimate a demand function aggregated at state-level for the US residential energy. Therefore our estimated rebound effect captures an overall effect composed of the sum of direct and indirect effects and hence the ideal lower and upper bounds for our estimates are not entirely clear. The literature has identified large positive as well as negative values for the indirect rebound effect, as found in Thomas and Azevedo (2013) for the household case. There are some papers that exhibit large direct rebound effects, such as Mizobuchi (2008) where a rebound effect of about 27% is found for Japanese households although the effect increases to 115% when capital costs are ignored in the analysis. Indirect rebound effects are usually larger than direct rebound effects and it is less ‘uncommon’ to find indirect rebound effects larger than 100%. Some examples can be found in Lenzen and Dey (2002) with an indirect rebound effect of 123% for Australia, Alfredsson (2004)

[69] In particular, this approach relies on the assumption that energy efficiency is unaffected by changes in energy prices.

[70] These rebound effects indicate percentage (expressed in relation to the predicted energy saving) by which the actual energy consumption is larger than the predicted energy consumption after an efficiency improvement. The measuring of the rebound effect is explained in detail in the next section.

with an indirect rebound effect up to 300% in Sweden or Brännlund *et al.* (2007) with an indirect rebound effect between 107-115% in CO_2 emissions in a simulation of an efficiency improvement in heating and transport sectors. In some cases this rebound measures can reach extremely large values, as in Druckman *et al.* (2010) who found indirect rebound effects up to 515% for the case of the UK.

5.3. Measuring rebound effects using energy demand frontier models

In this section, firstly we summarize the aggregate energy demand frontier model proposed by Filippini and Hunt (2012) to measure the level of "underlying energy efficiency" in the US residential sector. Subsequently, we link this model to the literature on the rebound effect and we introduce a more comprehensive model that allows estimating 'non-zero' rebound effects using an SFA approach. Once the econometric specification of the model is presented, we finally discuss new econometric issues that appear when the more general SFA approach is used to estimate rebound effects.

5.3.1. The standard energy demand frontier model

This approach treats energy as a production factor used in combination with other inputs to produce energy services, and attempts to measure inefficiency in the use of input energy as (positive) deviations from an energy demand frontier function that can be estimated for the whole economy or for a given sector. In general terms, the aggregate energy consumption can be written as follows:

$$q = F\left(Y, P, X, E, \beta\right)e^{v} \tag{5.1}$$

where q is the aggregate energy consumption, Y is the real income, P is the real energy price, β are parameters to be estimated, and X is a set of control variables such as population, average household size, heating degree days, cooling degree days, the share of detached houses, or time dummy variables. While v is the conventional noise term, E is the level of energy efficiency of a particular state. Since the energy efficiency level is not observed by the researcher, Filippini and Hunt (2012) made use of two assumptions in order to estimate equation (5.1). Firstly, they implicitly assumed that the energy demand function is *separable* in the sense that $F\left(Y, P, X, E, \beta\right)$ in (5.1) is decomposed into a function that does not depend on energy efficiency and an energy-efficiency function, that is:

$$F = f\left(Y, P, X, \beta\right)h\left(E\right) \tag{5.2}$$

where $h(E)$ is in turn assumed to be equal to $1/E$. The second assumption is that the unobserved energy efficiency term is bounded (i.e. $0 \leq E \leq 1$). These two assumptions allow using the stochastic frontier approach as the model to be estimated can now be written in logs as:

$$\ln q = \ln f\left(Y, P, X, \beta\right) + v + u \tag{5.3}$$

where u=-ln $E \geq 0$. The error term in (5.3) thereby comprises two independent parts. The first part, v, is the classical symmetric random noise, often assumed to be normally distributed, i.e. $v \sim N(0, \sigma_v^2)$. The second part, u, is a one-sided error term capturing the level of underlying energy inefficiency that can vary across states and over time.

Following Aigner *et al.* (1977) it is often assumed to follow a half-normal distribution, i.e. $u \sim N^+(0,\sigma_u^2)$. The identification of both random terms in this model relies on the asymmetric and one-sided distribution of u. If the inefficiency term could take both positive and negative values, it cannot be distinguishable from the noise term, v.

Equation (5.3) is the basic specification of the energy demand frontier that is estimated in Filippini and Hunt (2011, 2012) in order to get state-specific energy efficiency scores.[71] In the case of an aggregate residential energy demand function, $f(Y, P, X, \beta)$ reflects the demand of the residential sector of a state that has *and* uses fully efficient equipment and production processes. If a state is not on the frontier, the distance from the frontier measures the level of energy consumption above the minimum demand of reference, i.e. the level of energy inefficiency. Nevertheless, from an empirical perspective, the aggregate level of energy efficiency of US residential appliances is not observed directly, and therefore has to be estimated simultaneously with other parameters of the model. For this reason Filippini and Hunt (2011, 2012) use the expression 'underlying energy efficiency'.[72]

5.3.2. The (implicit) rebound effect in the standard energy demand frontier model

Although the basic concept of the rebound effect is not controversial, several mathematical definitions of this effect have been employed in the literature according to the availability of price and efficiency data.[73] Here we use the definition mentioned by Saunders (2000) which, in our opinion, provides one of the clearest and most direct measurements of the rebound effect. Following this author, the rebound effect is obtained as:

$$R = 1 + \varepsilon_E \tag{5.4}$$

where ε_E is the elasticity of energy demand with respect to changes in energy efficiency, i.e. $\varepsilon_E = \partial \ln q / \partial \ln E$. Table 5.2 shows the different rebound effects that we can find in a particular empirical application. The actual saving in energy consumption will only be equal to the predicted saving from engineering calculations when this elasticity is equal to minus one and hence there is no rebound effect ($R=0$). The rebound effect would be positive ($R>0$) if actual savings in energy consumption are less than expected, i.e. $-1 < \varepsilon_E$. The rebound effect could be larger than one ($R>1$) if improvements in energy efficiency increase energy consumption and hence the elasticity of energy demand with respect to changes in energy efficiency is positive, i.e. $\varepsilon_E > 0$. This somewhat counterintuitive outcome is termed 'backfire' in the literature (Saunders, 1992). In practice, negative rebound effects ($R<0$) can also be found for some observations if the improvements in energy efficiency produce larger decreases in energy use than predicted, i.e. $\varepsilon_E < -1$. Saunders (2008) labelled this - also rather counterintuitive - outcome as 'super-conservation'.[74]

---

[71] The estimation of (5.3) can be performed using either cross-sectional or panel data as in Filippini and Hunt (2011, 2012). They also propose to use a relatively simple log-log functional form.

[72] Filippini and Hunt (2011, 2012) advocate using panel data techniques to control for potential endogeneity problems caused by omitted variables or unobserved heterogeneity, an issue that is briefly discussed later on.

[73] See, for instance, Sorrell and Dimitropoulos (2008).

[74] For a more extended definition and some examples about this counterintuitive phenomenon see Saunders (2008).

Table 5.2. Possible values for the rebound effect and the energy efficiency elasticity

R > 1	Backfire	$\varepsilon_E > 0$
R = 1	Full rebound	$\varepsilon_E = 0$
0 < R < 1	Partial rebound	$-1 < \varepsilon_E < 0$
R = 0	Zero rebound	$\varepsilon_E = -1$
R < 0	Super-conservation	$\varepsilon_E < -1$

As the one-sided error term in (5.3) is measuring the level of underlying energy inefficiency, the elasticity of energy demand with respect to changes in energy efficiency is simply $\varepsilon_E = -\partial \ln q / \partial u$. Given the rebound effect definition provided by equation (5.4), we can then conclude that any energy demand frontier model that includes an inefficiency term as an explanatory variable implicitly provides a *direct* measure of the rebound effect. However, since ε_E in (5.3) is equal to –1, the standard SFA energy demand frontier model implicitly imposes a zero rebound effect, which contradicts most of the available empirical evidence surveyed in Section 5.2.

So far we have shown the implications of the standard SFA energy demand frontier model on the measurement of rebound effects. Next we will discuss the implications of the rebound effect story on both identification and measurement of the underlying energy efficiency. A key conclusion that one can get from the extensive literature focused on measuring the relationship between energy efficiency and energy demand is that the rebound effect tends to attenuate, exacerbate, or even reverse the effect of improvements in energy efficiency on energy consumption.[75] Therefore, the rebound effect issue can be introduced in an *energy demand* application of the SFA approach as a *correction factor* (1–*R*) that interacts with the energy inefficiency term (*u*) that is appended to the stochastic energy demand frontier. That is:

$$\ln q = \ln f\left(Y, P, X, \beta\right) + v + \left(1 - R\right)u \tag{5.5}$$

where again u=-ln $E \geq 0$. In this model, the effect on energy consumption is not necessarily proportional to the reduction in *u*; its effect is attenuated when the rebound effect is partial (i.e. when 0<*R*<1), exacerbated in case of super-conservation outcomes (i.e. when *R*<0), or reversed in case of extremely large rebound effects or backfire outcomes (i.e. when *R*>1).

Another interesting conclusion that can be inferred from the above equation is that any effort to improve energy efficiency of the current set of appliances (or their use) would not produce any change in energy consumption if consumers' reaction completely offset the potential energy savings, and hence the rebound effect is full. This implies that, in an energy demand setting, the underlying level of energy efficiency cannot be identified and estimated if *R*=1, since the energy demand model only have one (and symmetric) error term in this case. The other way around, this discussion

[75] It is not easy to find a similar phenomenon in production economics where SFA models have traditionally been applied. In that literature any improvement in firms' efficiency is assumed to have a proportional effect on firms' performance (outputs, cost, etc.). Just to conjecture an example, a sort of rebound effect might appear in public firms where employees' salary is not linked to their productivity. In this case, an employee who works efficiently could become "lazy" after a salary improvement since his earnings do not depend on his effort.

suggests that it only makes sense to estimate a stochastic energy demand frontier model when we believe that the rebound effect is not 100%.

5.3.3. A frontier energy demand model with non-zero rebound effects

Let us move to the estimation of a frontier energy demand model with non-zero rebound effects. To achieve this objective we should deal with several practical issues. The first one has to do with R in (5.5) that, like the energy inefficiency level, is not observed by the researcher because it is linked to the *demand for energy services*, *ES,* again a latent variable. To deal with this issue R can be approximated with a set of determinants of the demand for energy services, such as income and energy prices, i.e. z=(Y,P). This seems to be reasonable as most of the literature on the rebound effect associates the rebound effect with energy prices, and the theory often predicts that the rebound effect declines with income.[76]

If we replace the rebound effect variable R by a rebound-effect function, $R(\gamma' z)$, the model that can be estimated in practice is:

$$\ln q = \ln f(Y, P, X, \beta) + v + \left[1 - R(\gamma' z)\right] u \tag{5.6}$$

where γ are new parameters to be estimated. Several interesting remarks should be made regarding this specification. First, if the rebound-effect function does not depend on any covariate, our model simply collapses to the basic stochastic frontier demand model used in Filippini and Hunt (2011, 2012) that imposes zero (i.e. constant) rebound effects. In contrast, if $R(\gamma' z)$ varies across observations or states, the above equation allows us to get state-specific rebound effects that can be used for further analyses. Interesting enough, if z includes income and energy prices, the estimated γ can also be used to test whether both income and price elasticities of energy demand depend on energy efficiency.[77]

Second, unlike in production economics where a similar correction factor to our R function is often treated as part of firms' inefficiency, we point out in this chapter that $R(\gamma' z)$ is also -or mainly- capturing a rather different in nature phenomenon, i.e. the rebound effects associated to improvements in energy inefficiency.

Third, several specifications of $R(\gamma' z)$ can be used in a particular empirical application. Saunders (2008) recommends using extremely comprehensive (flexible) functional forms such as the Gallant and Fourier forms, which can depict the full range of rebound values. These forms are however intractable in our framework as they would

---

[76] Wang *et al.* (2012) point out, for instance, that the marginal utility of energy service consumption will decline as household income increases. Thus, energy efficiency improvements may not induce people to consume as much energy services as before. This means that the direct rebound effect might decline with the increase in household income. This is also confirmed in a limited number of studies, e.g. Small and Van Dender (2007) and Wang *et al.* (2012) that have found evidence of a negative relationship between the rebound effect and income.

[77] Indeed, both elasticities can be respectively written as:

$$\varepsilon_Y = \frac{\partial \ln q}{\partial \ln Y} = \frac{\partial \ln f(Y, P, X, \beta)}{\partial \ln Y} + \frac{\partial R(Y, P, \gamma)}{\partial \ln Y} \ln E$$

$$\varepsilon_P = \frac{\partial \ln q}{\partial \ln P} = \frac{\partial \ln f(Y, P, X, \beta)}{\partial \ln P} + \frac{\partial R(Y, P, \gamma)}{\partial \ln P} \ln E$$

interact with the stochastic part of the model and, hence, the maximum likelihood function would be highly non-linear in parameters. In this sense, as the choice of a particular function in this setting is limited by both methodological and practical issues, we propose exploring two simple rebound-effect functions:

$$R(\gamma' z) = \frac{e^{\gamma' z} - 1}{e^{\gamma' z}} \tag{5.7}$$

$$R(\gamma' z) = \frac{e^{\gamma' z}}{1 + e^{\gamma' z}} \tag{5.8}$$

Whereas the rebound-effect function in (5.7) can depict any value from full rebound to *super-conservation* (*SC*) outcomes (i.e. $R \leq 1$), the rebound-effect function in (5.8) precludes this somewhat counter-intuitive outcome as it only allows for *partial* (*PA*) rebounds-effects (i.e. $0 \leq R \leq 1$). In both cases, a positive (negative) value of γ indicates that the rebound effect increases (decreases) with z. It is worth noting that the SC and PA functions are respectively equal to 0 and 0.5 when $\gamma' z = 0$. This might occur when either all γ parameters are zero, or when R does not include a constant term and $z=0$.[78]

It should be noted that both specifications (5.7) and (5.8) of the rebound effect preclude the existence of backfire outcomes. This is not a coincidence as we must impose the restriction $R<1$ to our rebound-effect functions in order to distinguish inefficiency from noise. Otherwise, the second error term in equation (5.6) would no longer have a one-sided distribution and then we would not be able to take advantage of the asymmetric distribution of u to decompose the overall error term into two different stochastic components.

Other specifications have been examined in previous versions of this chapter, such as the simple cumulative density function of a standard normal variable, Φ, which like the PA function lies between zero and one, or the ratio $\Phi/(1-\Phi)$, which allows for super-conservation outcomes as does the SC model. The results of these models are not shown here as they are very similar to those obtained with the proposed models.

Finally, equations (5.6) with specification (5.7) or (5.8) for the rebound-effect function cannot be estimated if R includes a separated intercept and we assume that $\sigma_u = e^{\delta_0}$, as in the ALS model. This can be easily seen in the case of the SC function. In this case, we can rewrite (1–R) as $e^{-\gamma_0 - \gamma' z}$. A detail that is important here is that the estimated intercept of the rebound-effect function is *biased* because it also captures the parameter δ_0 that measures the standard deviation of the energy inefficiency term, u. That is, $\hat{\gamma}_0 = \gamma_0 - \delta_0$, and hence γ_0 and δ_0 cannot be estimated simultaneously.[79] A simple empirical strategy is proposed to deal with this issue. This strategy relies on the assumption that our energy inefficiency term follows the same distribution in both equations (5.3) and (5.6), so that the ALS estimate of γ_0 is used to adjust the estimated intercept of (1–R) accordingly.

[78] Most explanatory variables are centred at the sample mean to attenuate convergence problems when estimating the model using maximum likelihood techniques. Hence, $z=0$ for the representative observation.

[79] It is worth mentioning that this issue is not important in production economics as both options would yield exactly the same results when modelling *overall* firm inefficiency. However, if we estimate an energy demand frontier function, it matters as we would be either magnifying or diminishing the rebound effect.

5.4. Data and results

Our empirical application is based on a balanced US panel data set for a sample of 48 states over the period 1995 to 2011. That is, we have added four years to the data set used in Filippini and Hunt (2012). For the purposes of this chapter attention is restricted to the contiguous states (i.e. Alaska and Hawaii are excluded) except Rhode Island because of incomplete information: The District of Columbia is included and considered as a separate 'state'. The dataset is based on information taken from three sources. Residential energy consumption quantities and prices are provided by the Energy Information Administration (EIA). Population and real disposable personal income are from the Bureau of Economic Analysis of the US Census Bureau and the heating and cooling degree days are obtained from the National Climatic Data Center at NOAA. The number of housing units comes from the US Census Bureau and the share of detached houses for each state is based on the year 2000 census also obtained from the Census Bureau. Descriptive statistics of the key variables are presented in Table 5.3.

Table 5.3. Summary statistics of variables

Variable	*Description*	*Mean*	*Std. Dev.*	*Min.*	*Max.*
q	Energy consumption	229.60	209.42	19.02	932.92
Y	Real disposable personal income	92,620	105,635	6,072	654,780
P	Real price of energy	16.86	5.11	8.22	35.18
POP	Population	5,977	6,407	485	37,692
HDD	Heating degree days	5,134	2,007	555	10,745
CDD	Cooling degree days	1,147	805	128	3,870
AHS	Average household size	2.33	0.17	1.83	2.99
SDH	Share of detached houses	62.27	9.74	13.20	74.00

If we assume a Cobb-Douglas demand function, the econometric specification of the model can be written as:

$$\ln q_{it} = \left[\beta_0 + \beta_Y \ln Y_{it} + \beta_P \ln P_{it} + \beta_X \ln X_{it}\right] + \left(1 - R_{it}\right) u_{it} + v_{it} \tag{5.9}$$

where subscript i stands for state, subscript t is time, $v_{it} \sim N^+[0,\sigma_v]$, and $u_{it} \sim N^+[0,\sigma_u]$. Our dependent variable (q_{it}) is each state's aggregate residential energy consumption for each year in trillion BTUs. The income variable (Y_{it}) is each state's real disposable personal income for each year in million 1982 US$. The price variable ($P_{it}$) is each state's real energy price for each year in 1982 US$ per million BTUs. The set of control variables X_{it} includes Population (POP_{it}), the heating and cooling degree days (HDD_{it} and CDD_{it}), the average size of a household (AHS_{it}) obtained by dividing population by the number of housing units, and the share of detached houses for each state (SDH_i).

Regarding the rebound-effect function, it is modelled as a function of potential economic determinants of households' demand for energy services, such as household size, per capita income, and the price they must pay for energy. That is, $\gamma' z$ is specified as:

$$\gamma_0 + \gamma_Y \ln\left(Y/POP\right)_{it} + \gamma_P \ln P_{it} + \gamma_A \ln AHS_{it} \tag{5.10}$$

If we impose that the rebound-effect function does not depend on any covariate, we get the standard energy demand frontier model estimated in Filippini and Hunt (2011, 2012). Since the PA rebound-effect function prevents unlikely rebound effect outcomes, it is our preferred model. However the specification allowing for super-conservation outcomes, i.e. the SC model, is also estimated for robustness purposes. All models are estimated by maximum likelihood.

We show in Table 5.4 the estimation results of our preferred frontier energy demand models. The standard ALS model that imposes a zero rebound effects is also shown for comparison grounds. Simple LR tests indicate that both the PA and SC models outperform the ALS model. In general, both models perform quite well as most coefficients have the expected sign and almost all are statistically significant at the 5% level. This indicates that the results in terms of the estimated coefficients tend to be robust across the two different specifications of the rebound effect.

Table 5.4. Parameter estimates (models with time dummy variables)

		ALS			*PA*			*SC*		
	Parameters	***Est.***		***Std. E.***	***Est.***		***Std. E.***	***Est.***		***Std. E.***
Frontier										
	Intercept	5.012	***	0.022	5.043	***	0.018	5.042	***	0.018
	ln Y_{it}	0.364	***	0.037	0.238	***	0.046	0.236	***	0.046
	ln P_{it}	-0.101	***	0.025	-0.117	***	0.030	-0.114	***	0.030
	ln POP_{it}	0.670	***	0.038	0.797	***	0.047	0.799	***	0.047
	ln AHS_{it}	-1.117	***	0.053	-1.480	***	0.086	-1.469	***	0.088
	ln HDD_{it}	0.373	***	0.013	0.347	***	0.013	0.348	***	0.013
	ln CDD_{it}	0.084	***	0.007	0.080	***	0.008	0.080	***	0.008
	SDH_i	0.005	***	0.001	0.005	***	0.001	0.005	***	0.001
Noise term										
	ln (σ_v)	-2.633	***	0.120	-2.554	***	0.036	-2.555	***	0.037
Rebound-effect										
	Intercept				4.281	***	0.714	4.124	***	0.670
	ln $(Y/POP)_{it}$				-7.014	***	2.242	-6.148	***	2.034
	ln P_{it}				1.577	*	0.862	1.446	*	0.769
	ln AHS_{it}				-14.283	***	3.640	-12.187	***	3.165
Inefficiency term (homoscedastic)										
	ln (σ_u)	-2.530	***	0.258						
Log-likelihood		842.183			875.919			874.951		

Significance code: * p<0.1, ** p<0.05, *** p<0.01

Regarding the energy demand frontier, the estimated coefficients can be directly interpreted as elasticities as most of the variables are in logarithmic form. The estimated magnitudes of both price and income elasticities are quite reasonable from a theoretical point of view. The estimated frontier coefficients suggest that US residential energy demand is price-inelastic, with estimated elasticities of -0.10, -0.12 and -0.11 for the ALS, PA and SC models respectively. The results also suggest that US residential energy demand is income-inelastic, with an estimated elasticity of around 0.36 for the ALS model but only about 0.24 for the models allowing for non-zero rebound effects.

The positive coefficient on population obtained in all models suggests that energy consumption increases with population, given the total amount of disposable income in a particular state. For weather, the estimated cooling degree day elasticities for all three models are rather high, whereas the estimated heating degree day elasticities are much lower. The estimated coefficient of average household size suggests that as family size increases there is a tendency to use less energy, indicating that there are economies of scale with an estimated elasticity larger than unity in absolute terms. For the share of detached houses, the results suggest that there is only a marginal positive but significant influence on US residential energy demand.[80]

Table 5.5 provides descriptive statistics of the estimated energy efficiency for all US states. The ALS values are obtained directly using the Jondrow *et al.* (1982) formula. For the PA and SC models, the efficiency scores are computed dividing the estimated value of the overall one-sided term, i.e. $(1-R)u$, by (one minus) the estimated values of the rebound-effect function. We show three types of results in Table 5.6 in accordance with different adjustments of the intercept in the rebound-effect function. The first set of efficiency scores is obtained assuming that equation (5.10) has "no intercept" (i.e. $\gamma_0 = 0$) and hence u contains the whole estimated intercept. The second set of efficiency scores labelled "ALS-adjusted" follows the empirical strategy that uses the ALS estimate to adjust the estimated intercept of equation (5.10). The third efficiency scores are obtained following the opposite strategy to the first one, so in this case the rebound-effect function is "not adjusted" as it is assumed here that u does not contain an intercept.

Table 5.5. Energy efficiency scores using the PA and SC models (models with time dummy variables)

	Mean	*Std. Dev.*	*Min.*	*Max.*
ALS				
	0.939	0.025	0.831	0.977
PA				
No intercept ($\gamma_0 = 0$)	0.964	0.040	0.703	0.989
ALS-adjusted ($\gamma_0 = 1.833$)	0.911	0.041	0.651	0.958
Not adjusted ($\gamma_0 = 4.281$)	0.455	0.078	0.202	0.847
SC				
No intercept ($\gamma_0 = 0$)	0.987	0.003	0.974	0.997
ALS-adjusted ($\gamma_0 = 1.593$)	0.938	0.013	0.880	0.987
Not adjusted ($\gamma_0 = 4.124$)	0.455	0.079	0.200	0.848

Table 5.5 shows that the estimated *average* efficiency is between 45.5% and 98.7%. However, this wide range of results is due to the models that consider that the

[80] As in Filippini and Hunt (2012), the estimated coefficients of the time dummies (not shown) are significant in all models and although the overall trend in the coefficients is generally negative, they do not fall continually over the estimation period, reflecting the 'non-linear' impact of technical progress and other exogenous variables.

intercept may either be in the rebound-effect function or in the inefficiency term. If we focus on the ALS-adjusted results, the values obtained with the PA and the SC models are much more reasonable (91.1% and 93.8% respectively). Similar results were obtained by Filippini and Hunt (2012) using several specifications of the homoscedastic model. It is worth mentioning that the ALS model produces similar efficiency scores to those obtained when the intercept is properly adjusted. The efficiency scores clearly decrease when the intercepts of the PA and SC rebound-effect functions are not adjusted. By contrast, the largest efficiency scores are obtained when no intercept is considered in the rebound-effect function. These two cases define the lower and upper bound in the efficiency score estimates.

Regarding the rebound-effect function, recall from Table 5.4 that the coefficients of both income per capita and price are always statistically significant. The theory on rebound effects often predicts that they should decline with income, and the coefficient of this variable is negative in both models. This implies that the states with larger income levels have larger energy efficiency elasticities in absolute values, and therefore their rebound effects are lower. This seems to confirm the aforementioned hypothesis and is in line with the little available evidence on this issue in the empirical literature measuring rebound effects. On the other hand, the positive coefficient obtained for the price variable suggests that energy-inefficient states have more elastic energy demands. This result is expected in theory as energy-inefficient states tend to spend a larger share of their income on energy *ceteris paribus*, and hence the so-called income effect is more intense.

Our comprehensive frontier model of energy demand allows us to examine the compliance with some of the restrictions often assumed in previous studies devoted to estimating rebounds effects, but with different econometric techniques. For instance, most studies estimate the own-price elasticity of the demand for energy to get an indirect measure of the rebound effect. The validity of these papers hinges upon the assumption that consumers respond in the same way to decreases in energy prices as they do to improvements in energy efficiency. In particular, most of the empirical literature on rebound effects assumes that:

$$\varepsilon_E = -\varepsilon_P - 1 \tag{5.11}$$

We label this restriction as the *assumption of equivalence in responses*. Previous papers assume that equation (5.11) is fulfilled for all observations. As our model provides elasticities for both energy prices and energy efficiency, it allows us to examine (or even test) this issue in a very simple way. Thus, let us rewrite equation (5.11) as follows:

$$\varepsilon_P = a + b\varepsilon_E \quad , \quad a = b = -1 \tag{5.12}$$

Testing that $a = b = -1$ in an auxiliary regression allows us to examine the fulfilment of this assumption. In the Appendix (Section 5.6) we show that if we use a PA specification of the rebound-effect function, it is possible to directly test this assumption. In this sense, the Wald test carried out using the estimated parameters of our model suggests that energy and price elasticities are statistically different in our case. As a consequence, the absolute value of the elasticity of price in the frontier cannot be used for measuring the rebound effect as suggested by equation (5.11).

On the other hand, Sorrell and Dimitropoulos (2008) pointed out that the estimated price elasticities in previous studies might be biased if energy efficiency is not explicitly controlled for. The nature of this endogeneity problem is clear in our framework if the rebound-effect function depends on the energy price and efficiency is ignored because the overall error term in this case would include R and hence it would be correlated with the energy price in the frontier. In this sense, our extended frontier model clearly shows that it makes sense to follow Filippini and Hunt (2012) and estimate a standard energy demand model using the empirical strategy proposed by Mundlak (1978) to control for potential endogeneity problems.

We have also estimated our energy demand model including the Mundlak's adjustment but this adjustment does not affect our estimated rebound effects.[81] This is an expected result because our specification of the rebound-effect function already controls for potential endogeneity problems that would appear if we ignore that R is correlated with some of the energy demand drivers. The robustness of our results might also indicate that, given our specification of R, there are not significant traces of endogeneity associated to the inefficiency term, u, and hence there is no need to further extend our model to deal with this extra and cumbersome difficulty.[82]

Table 5.6 provides descriptive statistics for the overall US estimated rebound effects using the PA and SC models. It should be recalled that there are no values larger than unity in the PA model because its specification prevents backfire outcomes. In addition to the "ALS-adjusted" specification, for comparative purposes we show the estimated rebound effects that are obtained if the rebound-effect functions do not contain an intercept or if the estimated intercept is not adjusted (i.e. it completely belongs to the rebound-effect function). This table shows that the average rebound effect is 79% when our preferred PA model is used and the intercept is adjusted using the standard deviation of u of the ALS model. It decreases to 56% when the SC model is used.

Table 5.6. Rebound effects using the PA and SC models (models with time dummy variables)

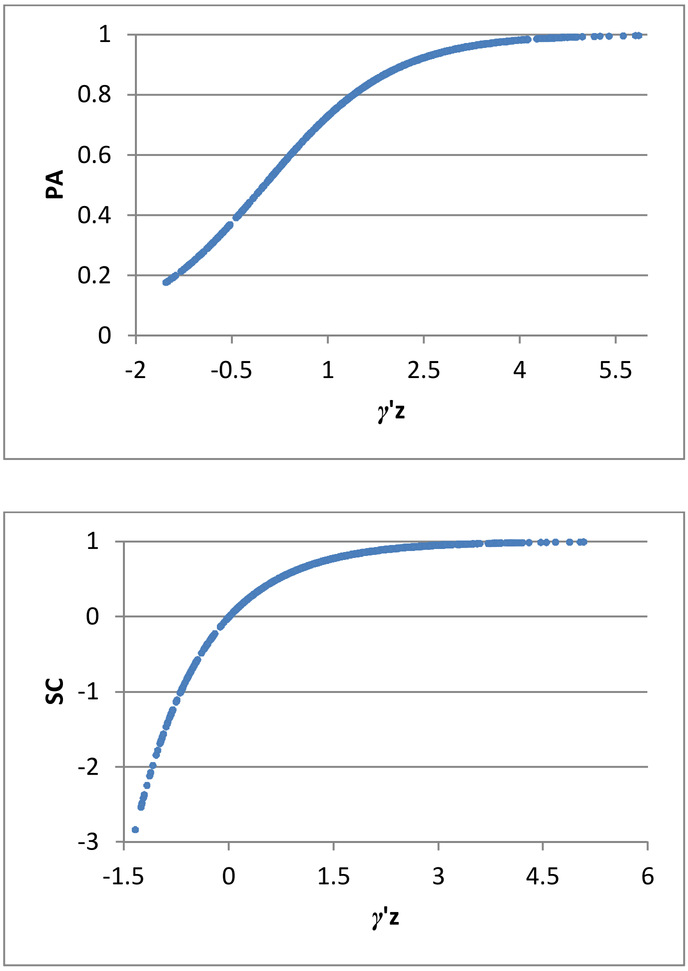
	Mean	*Std. Dev.*	*Min.*	*Max.*
PA				
No intercept ($\gamma_0 = 0$)	0.505	0.269	0.033	0.982
ALS-adjusted ($\gamma_0 = 1.833$)	0.791	0.208	0.177	0.997
Not adjusted ($\gamma_0 = 4.281$)	0.966	0.052	0.713	1.000
SC				
No intercept ($\gamma_0 = 0$)	-1.178	3.131	-17.866	0.969
ALS-adjusted ($\gamma_0 = 1.593$)	0.557	0.637	-2.835	0.994
Not adjusted ($\gamma_0 = 4.124$)	0.965	0.051	0.695	1.000

Generally speaking, our rebound effects tend to be larger than those obtained in the empirical literature using micro-data on the direct rebound effects of household energy demand (see our discussion in Section 5.2). Two different issues can partially

[81] Only a couple of the estimated coefficients lose statistically significance.
[82] This additional source of endogeneity could be addressed if we allow u to depend on a set of covariates (such as income and energy price). Actually, we have tried to estimate some versions of this model without success. This can be taken as evidence of the lack of additional endogeneity problems, but it also might be caused by the fact that the resulting likelihood function is much more complex (i.e. non-linear) than when u is homoscedastic.

explain this result. First, note that our estimated rebound effects involve more than one energy service, and hence they are not only capturing direct but also indirect effects. In addition, it should be pointed out that our results are even lower than those obtained in several papers - such as Lenzen and Dey (2002), Alfredsson (2004) or Mizobuchi (2008) - that also get large direct and indirect rebound effects, even reaching effects larger than 100%, i.e. backfire outcomes. A second reason has to do with the curvature of the estimated rebound-effects functions. In Figure 5.1 it is shown that the proposed rebound-effect functions are concave, at least when the value of $\gamma' z$ in (5.7) and (5.8) tends to be positive, as happens in our case due to the positive value of the intercept and the fact that all variables have been centred with respect to the sample mean. Thus, our rebound effect estimates are likely to be upwardly biased because the curvature imposed on our R functions "forces" the rebound effect to increase rapidly when we move away from the zero value. Research devoted to finding more flexible yet still simple rebound-effect functions that relax this curvature would be desirable in the near future.

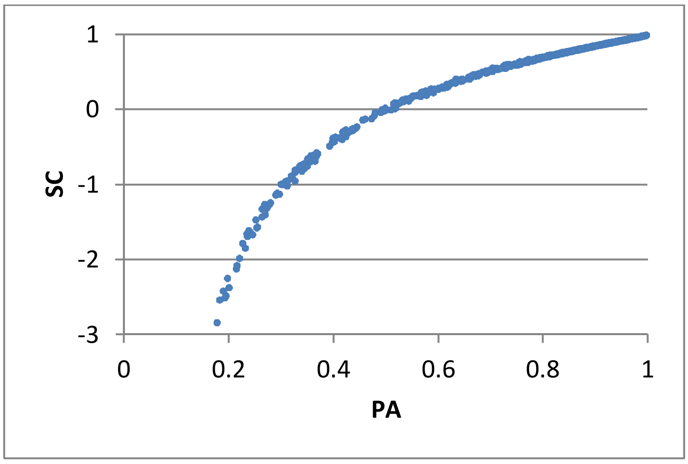
Figure 5.1. Curvature of the estimated rebound-effect functions



On the other hand, it is worth mentioning that the estimated rebound effects in Table 5.6 is about 97% in both models when it is assumed that the estimated intercept completely belongs to R. Hence, contrary to what happens in the efficiency estimates this procedure gives an upper bound for the rebound effect. These extremely large values probably suggest that R is upward biased, so that the estimated σ_u is also upward biased (i.e. the true σ_u is likely less than 1). Assuming, by contrast, that the rebound-effect function does not contain an intercept, the estimates produce an average rebound effect of about 50% for the PA rebound-effect function (and negative for the SC model). This outcome is, however, due to the fact that all variables have been centred with respect to the sample mean, and thus we are imposing that R=0.5 for the *average* state. These results thus point out the importance of adjusting the estimated intercepts when computing rebound effects using an SFA approach.

Regarding the issue of allowing or not for super-conservation outcomes, Figure 5.2 shows the relationship between the ALS-adjusted rebound effects obtained using our proposed models. This figure reveals that the rebound effects in which super-conservation outcomes are not restricted (SC model) are in practice monotonic transformations of the rebound effects obtained using models that only allow for partial rebound effects (PA model). In other words, allowing for super-conservation outcomes only has an effect on the magnitude of the rebound effects, but not on the relative values across observations. Overall, these results indicate that the ranking of rebound effects tend to be robust across different specifications of the R function.

Figure 5.2. Rebound effects with and without super-conservation outcomes (ALS-adjusted intercepts)



In Table 5.7 we show the parameter estimates of both the PA and SC models when they are estimated without time dummies for robustness analysis. These models are presented to check the sensitivity of the approach proposed to measure rebound effects, as these dummies are likely to be capturing - among other common temporal effects - technological improvements in the energy efficiency of households' equipment and appliances over time.

Again, both models perform quite well as most coefficients have the expected sign and almost all of them are statistically significant. Secondly, the income per capita and price variables of the rebound-effect function again have the expected signs and their coefficients are statistically significant. However, while the remaining coefficients are approximately in the same order of magnitude, the income and price elasticities in the frontier vary notably. This result is particularly striking and highlights the importance of a proper specification of technical progress (using a time trend or temporal dummies) in order to obtain unbiased estimates of the price and income elasticities. This may be a significant problem especially in those analyses aiming at estimating rebound effects through the own price elasticity. Moreover, in the rebound-effect function the coefficient of the price variable is positive and the coefficient of the income variable is negative, indicating that well-off states have lower rebound effects. In Table 5.8 we can see that both efficiency scores and rebound effects hardly change, indicating that the specification of technical progress in our model does not affect our results. As we have seen previously, the rebound-effect function without adjustment and the rebound effect without an intercept show the lower and upper bounds respectively for both the efficiency score estimates and the rebound effect estimates. Encouragingly, these results indicate that, overall, the estimated efficiencies and rebound effects tend to be robust to the different specifications of the technical progress in the frontier.

Table 5.7. Parameter estimates (models without time dummy variables)

		ALS			***PA***			***SC***		
	Parameters	***Est.***		***Std. E.***	***Est.***		***Std. E.***	***Est.***		***Std. E.***
Frontier										
	Intercept	4.937	***	0.009	4.992	***	0.008	4.990	***	0.008
	ln Y_{it}	0.259	***	0.033	0.114	***	0.042	0.113	***	0.041
	ln P_{it}	-0.207	***	0.017	-0.198	***	0.021	-0.196	***	0.021
	ln POP_{it}	0.776	***	0.035	0.921	***	0.043	0.923	***	0.043
	ln AHS_{it}	-1.113	***	0.058	-1.430	***	0.080	-1.422	***	0.081
	ln HDD_{it}	0.353	***	0.013	0.326	***	0.012	0.326	***	0.012
	ln CDD_{it}	0.079	***	0.007	0.070	***	0.007	0.069	***	0.007
	SDH_i	0.004	***	0.001	0.004	***	0.001	0.004	***	0.001
Noise term										
	ln (σ_v)	-2.738	***	0.108	-2.518	***	0.036	-2.520	***	0.037
Rebound-effect										
	Intercept				4.014	***	0.585	3.881	***	0.539
	ln $(Y/POP)_{it}$				-6.855	***	1.930	-5.979	***	1.720
	ln P_{it}				1.326	*	0.743	1.190	*	0.651
	ln AHS_{it}				-12.592	***	3.101	-10.719	***	2.696
Inefficiency term (homoscedastic)										
	ln (σ_u)	-2.239	***	0.117						
Log-likelihood		804.455			839.947			839.194		

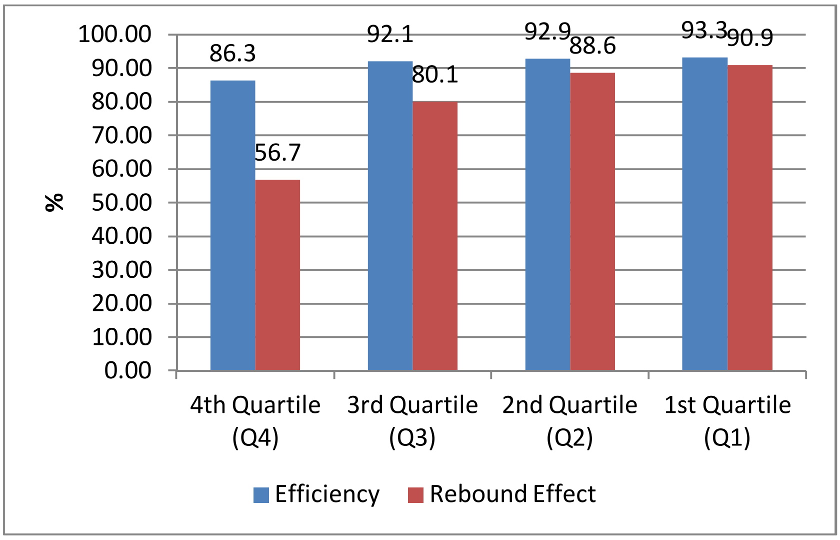
Significance code: * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Finally, our results might help policy makers to design more effective energy saving schemes. For instance, Figure 5.3 shows the overall relationship between energy efficiency and the rebound effect using our preferred model, the PA specification. If we sort the US states according to their average efficiency scores and then check their average rebound effects, we can get an idea about the correlation between these two measures. The average energy efficiency of the states in the fourth quartile is 86.3%. As usual in a frontier analysis framework, energy savings are *potentially* larger in those states with lower efficiency scores. Unlike standard SFA models, our models allow us to know whether the potential reductions in energy inefficiency are passed on entirely to final energy savings. As the states of the fourth quartile have also the lowest rebound effect (56.7%) we have more reasons to encourage energy efficiency improvements in these states. On the other hand, it is worth mentioning that although efficiency and rebound effects tend to increase as we move down the quartiles, the gap between both measures decreases and reaches a minimum difference in the first quartile where the most energy-efficient states (93.3%) are also those with the largest rebound effect (90.9%). This result indicates that as the efficiency of US states increases, households are less sensitive to changes in efficiency and they do not reduce their energy consumption as much as would be expected if we are swayed by what happens to the states with lower levels of efficiency.

Table 5.8. Energy efficiency scores and rebound effects using the preferred PA model (model without time dummy variables)

	Mean	*Std. Dev.*	*Min.*	*Max.*
Energy Efficiency Scores				
No intercept	0.957	0.044	0.671	0.985
ALS-adjusted	0.886	0.046	0.603	0.955
Not adjusted	0.456	0.083	0.150	0.861
Rebound effects				
No intercept	0.506	0.258	0.042	0.970
ALS-adjusted	0.805	0.190	0.224	0.995
Not adjusted	0.961	0.054	0.708	0.999

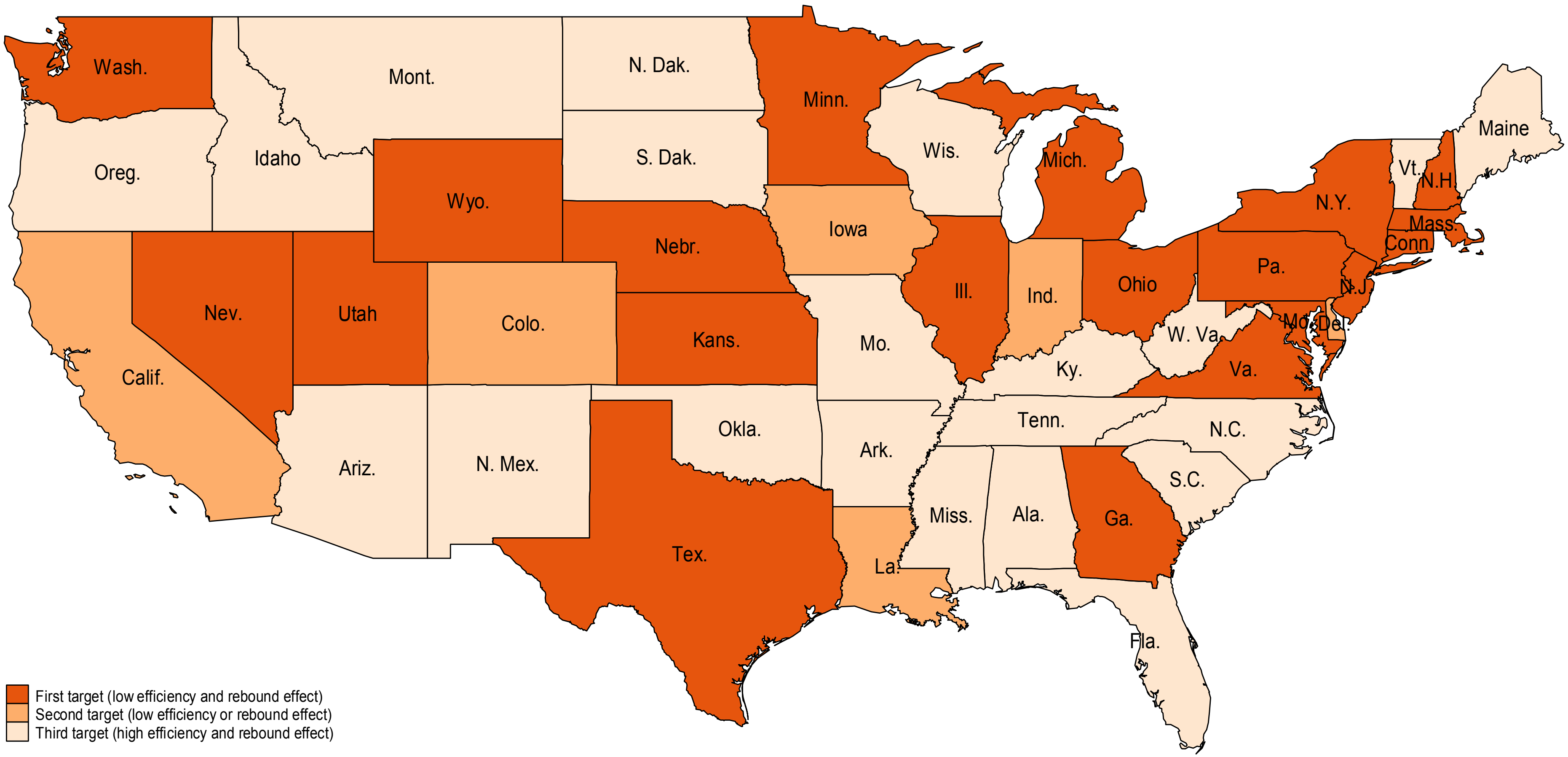
Figure 5.3. Average energy efficiency scores and rebound effects using the PA model



Focusing on the minimum rebound effects on Table 5.6, we can see that although the rebound effect is large on average, some US states have very small rebound effects compared to others. It can be seen in Figure 5.3 that there is a clear correlation between energy efficiency and rebound effects, but this does not mean that large energy efficiencies necessarily imply large rebound effects. Figure 5.4 reveals the heterogeneity that exists in our US sample. Those states with low energy efficiency (below the median) and a low rebound effect (also below the median) are highlighted in dark orange. These states are identified here as priority targets for energy policies, since improvements of energy efficiency in these states may yield large reductions in energy consumption (and probably greenhouse gas emissions).[83] On the other hand, those states marked in the lightest orange have large energy efficiencies as well as large rebound effects and therefore they should be labelled as the lower-priority targets. The intermediate orange highlights those states that have either low energy efficiency or a low rebound effect and hence cannot be identified as priority objectives. In summary, a sound policy would be not only focused on the most inefficient states but also on those with low rebound effects where the policy would have a greater overall effect over energy consumption.

[83] It should be stressed that if the average value is used instead the median to classify the states, just seven (Connecticut, Illinois, Maryland, Massachusetts, New Jersey, New York and Utah) would be below the average value of both efficiency and rebound effect, and hence only these would be primary targets.

Figure 5.4. Map of US states in which priority targets to reduce energy consumption are identified



5.5. Conclusions

This chapter highlights that the energy demand frontier model applied in the previous chapter and originally proposed by Filippini and Hunt (2011, 2012) to get country-specific energy efficiency scores, is closely linked with the so-called rebound effect, a phenomenon widely examined in the literature on energy economics. In particular, we have shown that the standard specification of the energy demand frontier model basically imposes a rebound effect equal to zero, something that clashes with the empirical evidence obtained in the literature on the rebound effect.

Based on the stochastic frontier approach, a new empirical strategy is proposed in this chapter to measure the rebound effect associated to energy efficiency improvements. Our more comprehensive energy demand frontier model avoids the 'zero' rebound-effect assumption through the estimation of a rebound-effect function that regulates the final effect of potential efficiency improvements on energy consumption. Two specifications for the rebound-effect function that preclude backfire outcomes are presented here. While the SC model allows for super-conservations outcomes, the PA model only allows for partial rebound effects. We however advocate using the latter model because it avoids obtaining too large (negative) rebound effects for some observations that are difficult to justify in economic terms.

We illustrate the approach proposed to measuring rebound effects with an empirical application of US residential energy demand data for 48 states over the period 1995-2011. The coefficients of the variables included in the models are highly significant, show the expected signs and have a quite reasonable magnitude regardless of the specification of the rebound-effect function used. Regarding the efficiency scores there is not much variation between estimated (PA and SC) models and they do not change much in response to the different options used to obtain the intercept of the rebound-effect function.

In relation to the rebound effects, values that are too large and too low are obtained if we ignore or do not adjust the estimated intercept of the rebound-effect function. Although the estimated rebound effects vary with the functional form, the position of each observation does not change as the SC rebound effects is a monotonic transformation of the rebound effects obtained with the PA model. This is an important result as the relative position of each state in terms of both energy efficiency and rebound effect rankings permits the identification of states where the enforcement of policies with the aim of promoting energy efficiency would be more effective. Compared to those analyses aiming at estimating rebound effects through the own price elasticity, our empirical approach suffers less from biases when technical progress is ignored.

To finish up, we would like to insist that this is the first attempt to use the stochastic frontier framework to measure rebound effects associated to energy efficiency improvements. In this sense, we have identified a few number of research areas that can be explored by other researchers in the next future in order to better estimate the rebound effects using a similar empirical strategy than the proposed here. This likely would imply the use of more sophisticated techniques than those proposed in this chapter.

For instance, a key issue is the identification of the true intercept of the rebound-effect function. We have proposed a simple empirical strategy to split the estimated intercept into its two components but other alternative approaches could be used to deal with this problem. A promising strategy could be treating the correction factor as an additional one-sided random term and, hence, estimating a model with *two* multiplicative one-sided random terms. Another issue has to do with the concavity problems of the proposed rebound-effect functions, which tend to overestimate the rebound effect. Although this is likely an issue related to our data set, future research will be likely focused on the use of alternative parametric specifications of the rebound function. In this sense, it should be also explored the potential use of semiparametric regression methods to relax the current concavity constraints. We also encourage specific research focused on the lack-of-backfire assumption used in our energy demand frontier model.

5.6. Appendix

Testing the assumption of equivalence in responses

Let us assume that the demand function is Cobb-Douglas and we use the SC rebound-effect function. In this case, the price elasticity of energy demand can be written as:

$$\varepsilon_P = \beta_P + \gamma_P \left[1 - R\right] \ln E \tag{5.13}$$

where β_P is the *frontier* price elasticity and γ_P is the coefficient of ln P in the SC rebound-effect function. As $\varepsilon_E = -\left(1 - R\right)$, equation (5.13) can be rewritten now as follows:

$$\varepsilon_P = \beta_P - \left(\gamma_P \ln E\right) \varepsilon_E \tag{5.14}$$

In summary, equations (5.12) and (5.14) jointly indicate that the equivalence of responses assumption will be satisfied in our model if we cannot reject the following null hypothesis:

$$H_0 : \hat{\beta}_P = \hat{\gamma}_P \ln E = -1 \tag{5.15}$$

Testing this hypothesis is difficult as energy efficiency varies across states and over time. An alternative way to test the equivalence of responses assumption is to test a sufficient (but weaker) condition for the fulfilment of the above hypothesis evaluated at the estimated mean of the energy inefficiency term:

$$H_0 : \hat{\beta}_P - \hat{\gamma}_P \hat{E}\left(u\right) = 0 \tag{5.16}$$

As we assume that u=-ln E follows a half-normal distribution, the expected mean in (5.16) is simply a function of σ_u and hence the sufficient condition in (5.16) can be finally expressed as follows:

$$H_0 : \hat{\beta}_P - \hat{\gamma}_P \sqrt{2/\pi}\hat{\sigma}_u = 0 \tag{5.17}$$

If instead we use the PA rebound-effect function, the sufficient condition in (5.16) becomes:

$$H_0 : \hat{\beta}_P - \hat{\gamma}_P \frac{e^{\hat{\gamma}'z}}{1 + e^{\hat{\gamma}'z}} \sqrt{2/\pi}\hat{\sigma}_u = 0 \tag{5.18}$$

Chapter 6

Conclusions and future research

This thesis is formed by four chaptered essays that cover different topics in energy economics. In these essays, efficiency analyses that mainly involve the estimation of stochastic frontier models are applied to deal with energy-related issues in diverse sectors of the economy.

The first two essays are specifically focused on the efficiency analysis of electricity transmission firms. In the first essay, the effect of potential determinants of firms' efficiency, primarily weather conditions, is studied through the estimation of heteroscedastic models. In the second essay, a latent class model is proposed to deal with technological and environmental differences as a first step to carry out standard benchmarking. In the remaining essays, energy demand frontier models are estimated to obtain information about energy efficiency, i.e. the potential reduction in energy required to provide products and services, and the so-called rebound effect in energy consumption. In the third essay, a latent class model is applied again, but in this case with the aim of finding distinct energy demand functions across countries for the transport sector in Latin America and the Caribbean. The last essay merges two literatures and proposes the use of the stochastic frontier approach to directly measure the magnitude of the rebound effect, and applies this model to the US residential energy sector.

In Chapter 2, the economic characteristics of the technology and the managerial efficiency of 59 US electricity transmission firms for the period 2001-2009 are analysed. The main contribution of this work is to study the effect of weather on costs of the electricity transmission grid. The estimated set of heteroscedastic models allow us to conclude that weather adversely affects electricity transmission costs primarily through inefficient management and not through the technology. That is, the estimated costs of a fully efficient firm are not going to be increased because of adverse weather conditions; an issue that must be taken into account when the costs of these firms are analysed by regulators.

This essay allows us to study economic features of firms' technology such as the economies of scale and density. In particular, the empirical model that is estimated allows us to state that most electricity transmission networks exhibit natural monopoly characteristics. We have also found that, despite the US regulator's effort aiming to improve firms' performance, the average efficiency level in this industry has decreased over the period analysed, showing however an increasing divergence among firms. This essay analyses the potential effect of certain variables such as weather conditions on firms' efficiency. In this sense, we have found that more adverse conditions generate higher levels of inefficiency which may indicate that it is more difficult to manage firms that operate in regions with unfavourable weather.

The estimation of a latent class model allows us to perform a case-by-case analysis to identify if weather conditions have a direct effect on firms' costs or if they indirectly affect those firms through inefficiency. The results suggest that only the costs of a small number of firms could be adjusted downwards by the regulators due to the influence of bad weather. Firms should therefore not use unfavourable weather conditions as an excuse to demand that the regulators purge their cost data.

There are natural extensions of the first chapter than could be carried out in the next future. Since weather can be understood as a 'single phenomenon' that is composed by a set of different events, the analysis could be extended by using some type of aggregates or composites that would include a much larger list of events that may affect firms' performance. Moreover, since the effect of weather variables on firms' cost can be complex and non linear, their inclusion may require abandoning the parametric framework and using a more flexible empirical approach. This will probably imply the application of semi-parametric models similar to the SPSCM introduced by Li *et al.* (2002) and extended by Sun and Kumbhakar (2013), or the StoNED method recently proposed by Johnson and Kuosmanen (2011) and Kuosmanen (2012).

In Chapter 3, the use of a latent class model approach is advocated to segment samples of energy firms into homogeneous groups before carrying out a standard benchmarking analysis using DEA, the most widely technique applied in energy regulation to measure firms' performance. Through a simulation exercise, it is shown that combining the latent class model and DEA outperforms other, less robust and more arbitrary, procedures to split samples of firms. A practical application of this method to the US electricity transmission sector is presented.

A proper (and fair) measurement of firms' performance in an incentive regulation framework is highlighted again in this chapter, but in this case it is emphasised that unobserved differences in technology or environmental conditions should be taken into account in energy regulation even though most of that heterogeneity is not observed by the regulator. The proposal of using a latent class model to deal with this issue is based on the theoretical advantages of using a sample separating procedure that precisely takes into account potential differences in technology or environmental conditions to split the sample of firms. The performed simulation analysis that plays with different degrees of technological differences and firms scale, confirms the theoretical advantage of the latent class model over other sample separating methods such as the widely used k-means cluster analysis. Compared to other simpler and ad-hoc methods, the proposed procedure is the best to identify the technological group each firm really belongs to, and the efficiency levels obtained using this procedure are the closest to real ones (i.e. those generated in the simulation). It is also shown that the discriminatory capacity and the assignment success of the latent class model increase when large differences in technologies and scale arise, a feature that is crucial for any sample separating method.

The procedure is illustrated with an empirical application to the same database of electricity transmission firms analysed in the previous chapter. Two different groups or technologies are found using the latent class approach. Although consecutive increases in the efficiency scores are observed when we move from one class to two classes and so on, the largest change in efficiency scores arises when we move from a one-class model to a model with two classes, which is precisely the preferred model. Similar results are obtained when weather variables and the demand growth are included as environmental factors that may influence the technologies adopted by the firms.

Future extensions of this chapter will involve the use of alternative DGPs in the simulation, i.e. exploring the use of different approaches for generating the data with the aim of knowing more about the performance of the latent class models. Following Kuosmanen *et al.* (2013), in addition to the specific characteristics of the regulatory models, the observed features of empirical data will be taken into account to adjust the DGP. In that sense, the simulated scenarios will be extended trying to incorporate additional technologies, correlations between outputs and the influence of environmental factors. The outcome of these simulations will probably be interesting not only from an energy industry point of view but also to the analysis of unobserved heterogeneity in other fields of research.

In Chapter 4, the stochastic frontier approach is used to estimate aggregate energy demand functions in the transport sector in a group of countries of Latin America and the Caribbean that represent the 43% of total energy consumption in this region. The use of these models to measure energy efficiency allows us to deal with some of the disadvantages of the energy intensity indicators commonly used in international comparisons. As far as we are aware, this is the first application of the stochastic energy demand frontier approach to measure energy efficiency in the transportation sector. This model is nested in a latent class structure to control for the likely large unobserved heterogeneity among these countries, and test for the existence of groups of countries with different demands associated to distinct price and income elasticities.

The database used in this essay consists of a sample of 24 Latin American and Caribbean countries for the period 1990-2010. The energy price, which is, in addition to income, one of the most relevant variables in a demand analysis, is obtained using a transitive multilateral index that allows for proper comparisons across countries over time. Energy consumption in the transportation sector consists of several components and therefore, it is required to add each individual price to obtain an index of energy price. However, no index price is provided by any statistical agency for the total sample of countries and hence it has been calculated for this essay.

Generally speaking the ranking of efficiency scores that is obtained when a single demand is estimated is strongly correlated (70%) with the ranking derived from the widely-used energy intensity indicators. The lack of correlation observed for some countries over time indicates, however, that variations in energy intensity indicators may be associated with circumstances other than changes in energy efficiency. On the other hand, three demands with quite different income and price elasticities are found using a latent class model. For instance, price elasticity goes from -0.16 for the most inelastic demand to -0.41 for the most elastic. However, when a single demand is estimated the price elasticity is the same for all countries and equal to -0.23. The class membership probabilities depend on income, area and population; and reflect that those countries with higher per capita income and lower population density tend to have the lowest price elasticity. The estimation of this model allows identifying the most energy efficient countries in each class that, in fact, coincide with those that have developed policies to improve public transport in recent years.

Regarding futures lines of research, the existence of possible asymmetric effects on the energy demand with respect to changes in energy prices and income will be analysed through the decomposition of price and income used in papers such as Gately and Huntington (2002), Adeyemi and Hunt (2007) or Adofo *et al.* (2013). Although the reversibility of prices has been previously analysed in the transport sector (see Dargay, 1992; Gately, 1992; and Dargay and Gately, 1997) this will be the first time that this approach is applied in an energy demand frontier model. Furthermore, potential energy and greenhouse gas emissions savings derived from the estimates presented here will be estimated. Another line of research is the analysis of rebound effects using a model analogous to the suggested in the following chapter. The rebound effect has been broadly studied in the transport sector, but there is a lack of empirical evidence for the case of Latin America and the Caribbean, which would surely make these results appealing for researchers and practitioners.

Finally, in Chapter 5, energy demand frontier models are estimated for the US residential energy sector, but in this case we have extended the traditional model to measure the so-called rebound effect that tends to attenuate the expected savings in energy consumption associated to improvements in energy efficiency. So far this issue has not been analysed using the stochastic frontier approach and hence the major contribution of this essay is just linking both literatures. Numerous studies have used different approaches to measure the different types of rebound effects. The most popular method relies on the estimation of price elasticities of the demand for energy, but it only provides an indirect measure of the potential rebound effect under strong assumptions about consumer behaviour.

In this chapter, the standard energy demand frontier model and the implicit assumptions of this model are presented, and it is suggested the estimation of a model that allows obtaining non-zero rebound effects through a *correction factor* that mitigates or intensifies the effect of efficiency improvements on energy consumption. Two alternative specifications are proposed for the rebound-effect function, and also a strategy to deal with the identification of the intercept of this function. The specifications of these rebound-effect functions are related to the demand for energy services.

The empirical application is based on a US panel data set for a sample of 48 states over the period 1995-2011. The average rebound effect obtained is somewhat high (between 56 and 80%) compared with the values obtained in the literature, but it may be overestimated due to the concavity of the functional form assumed for the rebound-effect function. Despite of this, our models allow a robust identification of states in which policies to promote energy efficiency would be more successful. Furthermore, using the proposed model, the rebound effect measures are not altered by the way in which the effect of time is incorporated in the demand, which contrasts with the bias that arises when they are obtained through their own price elasticity estimates.

Since this work is a first proposal of using a frontier approach to estimate rebound effects, future research on this topic must deal with different issues. Firstly, the identification of the true intercept of the rebound-effect function should be further analysed. A hopeful strategy is to consider the adjustment factor as an asymmetric random term and estimating the model including the product of two one-sided random terms, i.e. rebound effect multiplied by inefficiency. Secondly, flexible and tractable functional forms should be proposed in order to depict the full range of possible rebound effects in the results. The use of alternative parametric specifications of the rebound effect function could be deeply analysed. Continuing on from this, with the aim of relaxing the concavity restrictions on the rebound-effect functions proposed, the use of alternative approaches, such as semi-parametric regression methods, also could be explored.

Resumen y conclusiones en español

Esta tesis está formada por cuatro ensayos que tratan sobre diferentes temas encuadrados en la economía energética. En ellos se llevan a cabo análisis de eficiencia, generalmente a través de la estimación de fronteras estocásticas, con el objetivo de estudiar cuestiones relacionadas con la energía en distintos sectores de la economía.

Los dos primeros ensayos están centrados en el análisis de la eficiencia de las empresas de transmisión eléctrica estadounidenses. En el primero, se estudia la influencia de diversos factores, principalmente condiciones meteorológicas, que pueden afectar a la gestión de estas empresas. En el segundo, se aplica un enfoque de clases latentes para controlar por diferencias tecnológicas y/o ambientales antes de realizar una evaluación comparativa entre empresas. En los ensayos restantes, se estiman funciones de demanda de energía frontera para obtener información sobre la eficiencia energética de los países y el llamado efecto rebote en el consumo de energía. En el tercer ensayo se emplea de nuevo un enfoque de clases latentes, en este caso con el objetivo de identificar distintas funciones de demanda de energía en el sector transporte de América Latina y el Caribe. En el último ensayo se propone la estimación de modelos frontera que incorporan en su especificación funciones de efecto rebote, lo que permite medir directamente la magnitud de este fenómeno, mostrándose una aplicación práctica para el caso del sector de energía residencial estadounidense.

En el primer ensayo se analizan las características económicas de la tecnología y la eficiencia en la gestión de las empresas de transporte eléctrico en Estados Unidos. La principal contribución de este trabajo es estudiar el efecto de la meteorología en los costes de la red de transporte eléctrico. Los modelos heterocedásticos estimados permiten concluir que una meteorología adversa afecta a la red eléctrica a través de una gestión ineficiente de las empresas y no a través de la tecnología. En otras palabras, los costes de una empresa totalmente eficiente no se ven afectados por la meteorología, cuestión que debe ser tenida en cuenta cuando los costes de estas empresas son analizados por los reguladores.

En este ensayo se presenta una visión general del proceso de reestructuración en la industria eléctrica que ha sido común a lo largo de las últimas décadas en gran parte de países. Estas reformas han tenido como consecuencia la desintegración vertical de este sector en diferentes segmentos que han recibido distintos tratamientos por parte de los gobiernos. La aplicación de métodos de evaluación comparativa tratando de incorporar diferentes factores ambientales, se ha convertido en una cuestión primordial con el objetivo de obtener mediciones fiables de la actuación de las empresas reguladas. Sin embargo, esta cuestión ha sido desatendida en el análisis económico de las redes de transporte eléctrico debido a la escasez de datos empíricos. El establecimiento de tarifas adecuadas en la red, es destacado frecuentemente como una cuestión clave debido a su influencia sobre el funcionamiento de toda la red eléctrica y plantea la necesidad de realizar comparaciones apropiadas entre empresas. En el ensayo se presenta un resumen de los diversos modelos desarrollados dentro de los enfoques utilizados habitualmente en los análisis de eficiencia (paramétrico, no paramétrico y semi-paramétrico) que tienen en cuenta las condiciones ambientales.

El modelo empírico que se presenta es una frontera estocástica de costes que incluye una serie de outputs, el tamaño de la red, los precios de los inputs, un listado de *dummies* regionales y una tendencia temporal. La estimación de los parámetros de este modelo permite estudiar características económicas de las empresas como economías de densidad y de escala. De forma adicional al tradicional modelo ALS, se presentan varias especificaciones de modelos heterocedásticos surgidos recientemente en la literatura de fronteras estocásticas para incluir el efecto de determinantes de la eficiencia: RSCFG, RSCFG-μ, KGMHLBC y GEM. Además de las variables incluidas en la función de costes, se incorporan como determinantes de la eficiencia, variables meteorológicas (temperatura, precipitaciones y viento), su interacción con la estructura de costes de las empresas y dos variables que miden el crecimiento medio de la demanda para cada empresa en el período analizado.

La base de datos utilizada está compuesta por 59 empresas de transporte eléctrico estadounidenses para el período 2001-2009. Estos datos fueron obtenidos esencialmente de tres fuentes, el *FERC form 1*, el *US Bureau of Labor Statistics* y el *National Climatic Data Center* perteneciente al NOAA. A partir de las estimaciones, puede observarse que la mayor parte de las empresas analizadas exhiben características de monopolio natural y que la eficiencia en esta industria ha decrecido a lo largo del período analizado, apreciándose a su vez una creciente divergencia en la actuación de las empresas. Estos resultados sugieren que este segmento del sector eléctrico debería seguir siendo regulado y que existe un amplio margen para la intervención regulatoria con el fin de mejorar el funcionamiento de las empresas. En cuanto a los determinantes de la eficiencia, se observa que las condiciones meteorológicas adversas generan mayores niveles de ineficiencia, lo que quizá indique que es más difícil gestionar aquellas empresas que operan en regiones con condiciones meteorológicas desfavorables. Se observa también que invertir en capital es una mejor estrategia que incrementar los costes operativos para tratar de contrarrestar la influencia de la meteorología. Otro resultado destacable es que la actuación de las empresas es mejor cuando la demanda tiende a ser estable ya que las empresas no tienen que incurrir en costes de ajuste adicionales.

Por otra parte, se estima un modelo de clases latentes que permite realizar un análisis caso por caso para identificar si las condiciones meteorológicas tienen un efecto directo sobre el coste de las empresas o si por el contrario estas condiciones afectan indirectamente a las empresas a través de la ineficiencia. Los resultados sugieren que únicamente los costes de un pequeño número de empresas podrían ser ajustados a la baja por los reguladores debido a la influencia del mal tiempo. La gran mayoría de empresas no debería por tanto utilizar la meteorología como excusa para exigir que los reguladores purguen o ajusten sus datos sobre costes. Debido a que la meteorología puede ser individualmente entendida como un conjunto de fenómenos, en el futuro se espera extender el análisis utilizando algún tipo de agregado que permita incluir una mayor lista de eventos que puedan influir sobre la actuación de las empresas. Además, como el efecto de la meteorología es complejo y probablemente no lineal, su inclusión en el modelo quizá requiera abandonar el marco paramétrico y utilizar un enfoque empírico más flexible. Esto probablemente lleve a la aplicación de modelos semi-paramétricos similares al SPSCM presentado por Li *et al.* (2002) y extendido por Sun y

Kumbhakar (2013), o el método StoNED recientemente propuesto por Johnson y Kuosmanen (2011) y Kuosmanen (2012).

En el segundo ensayo se propone el uso de un modelo de clases latentes para realizar segmentaciones de muestras de empresas energéticas antes de llevar a cabo evaluaciones comparativas de las mismas utilizando DEA, una técnica no paramétrica para medir la actuación de las empresas ampliamente utilizada en regulación energética. A través de una simulación, se muestra que este enfoque supera a otros procedimientos menos robustos y más arbitrarios de dividir muestras de empresas. Se presenta una aplicación práctica de este método para el sector de transporte eléctrico.

La importancia de evaluar de forma justa la actuación de las empresas se menciona de nuevo en este ensayo, pero en este caso se enfatiza la necesidad de controlar por la heterogeneidad y las diferencias inobservables en la regulación energética. Si estas características no son tenidas en cuenta, posibles diferencias (vinculadas a distintos entornos o tecnologías) en la actuación de las empresas de servicios públicos, pueden ser erróneamente atribuidas a cuestiones que están bajo el control de las empresas y en consecuencia, los índices de eficiencia obtenidos de estos análisis pueden estar sesgados. En ese sentido, la propuesta de utilizar un enfoque de clases latentes para hacer frente a estas cuestiones se basa en las ventajas teóricas de este procedimiento para encontrar diferencias en el comportamiento de las empresas.

Este enfoque agrupa empresas mediante la búsqueda de diferencias en los parámetros de producción o de costes, tiene en cuenta la misma relación entre *inputs* y *outputs* que se analiza con posterioridad en la segunda etapa, y no es mucho más sofisticado que otros métodos de segmentación, por lo que puede ser implementado utilizando software estándar. Otra cuestión que otorga sencillez a la aplicación del procedimiento propuesto para la regulación energética, es que pueden utilizarse las mismas variables tanto en el análisis de eficiencia como en la división previa de la muestra. El proceso de estimación del modelo de clases latentes en la primera etapa y el cálculo de los índices de eficiencia utilizando un modelo DEA orientado al *input* en la segunda etapa, son definidos en el ensayo. Se presentan además varios criterios de información estadística que son generalmente utilizados para seleccionar el número apropiado de clases en los modelos de clases latentes.

Se lleva a cabo un análisis de simulación para identificar si el procedimiento propuesto es capaz de tratar la heterogeneidad inobservable de forma más adecuada que otros métodos de segmentar muestras y si predice mejor la eficiencia individual de cada empresa. Los resultados confirman la ventaja teórica del enfoque de clases latentes sobre otros métodos. En la simulación de datos, se asume la existencia de distintos grados de diferencias tecnológicas y tamaños de escala de las empresas. A través de la comparación del enfoque de clases latentes con otros métodos de dividir muestras, tales como los análisis clúster de k-medias o la mediana de ciertas variables, se observa que el método propuesto es el mejor para identificar a qué grupo tecnológico pertenece realmente cada empresa, y los niveles de eficiencia obtenidos son los más próximos a los generados en la simulación. Se muestra a su vez que la capacidad discriminatoria y el éxito en la asignación del modelo de clases latentes se incrementa cuando surgen diferencias en tecnología y escala. La correlación de eficiencias y éxito en la asignación parece sugerir que los reguladores podrían usar el nivel de eficiencia medio para comparar el rendimiento relativo de los diferentes métodos de segmentación en una aplicación real.

El procedimiento es ilustrado con una aplicación práctica a la misma base de datos de transporte eléctrico analizada en el ensayo previo. Se identifican hasta nueve tecnologías diferentes; sin embargo a través de los criterios estadísticos aplicados puede verse que el número de clases que debe ser seleccionado es dos. Los modelos de clases latentes permiten incluir variables separadoras que añaden información a la estimación y ayudan en la segmentación de la muestra. En este caso, las variables meteorológicas y el crecimiento de la demanda han sido incluidas como variables ambientales que pueden influir en la tecnología adoptada por las empresas. Las estimaciones adicionales incluidas para comprobar la robustez del método producen resultados similares a los obtenidos sin la inclusión de estas variables ambientales. Futuras extensiones de este trabajo implicarán el uso de procesos generadores de datos alternativos en la simulación, es decir, explorar diferentes mecanismos de generar datos con el objetivo de conocer más acerca del funcionamiento de los modelos de clases latentes. Siguiendo a Kuosmanen *et al.* (2013), además de los rasgos específicos de los modelos regulatorios, las características observadas de los datos empíricos serán tenidas en cuenta para calibrar el proceso de generación de datos. En ese sentido, los escenarios simulados serán extendidos tratando de incorporar tecnologías adicionales, correlaciones entre outputs y la influencia de factores ambientales. El resultado de esta investigación puede resultar interesante no sólo desde el punto de vista de la industria energética sino también para el análisis de la heterogeneidad inobservable en otros campos de investigación.

En el cuarto ensayo se explora el uso de modelos de fronteras estocásticas para estimar funciones de demanda energética agregadas en el sector transporte de América Latina y el Caribe. El uso de estos modelos permite obtener medidas de los niveles de eficiencia energética en estos países que pueden ser consideradas alternativas a los tradicionales indicadores de intensidad energética utilizados comúnmente en comparaciones internacionales. La aplicación de este enfoque es novedosa en el sector transporte y es aplicada al caso de un grupo de países para los cuales existe poca investigación al respecto, a pesar de que el peso del sector suponga un 43% del consumo de energía total en la región. Debido a la posible heterogeneidad inobservable existente entre países, en este caso también se propone el uso de un enfoque de clases latentes, lo que permite comprobar la existencia de países con demandas diferenciadas que están asociadas a distintas elasticidades precio y renta.

La preocupación por la medición y control de la eficiencia energética, especialmente en aquellos sectores más intensivos en el uso de energía, surge fundamentalmente a partir de la crisis mundial del petróleo de los años 70. En la sección introductoria del ensayo se resalta la importancia del sector transporte en América Latina y el Caribe a través de la presentación de datos del consumo energético y del precio de la energía en el sector durante las últimas décadas. Posteriormente se muestra una revisión de los diferentes enfoques que han sido utilizados para modelizar el consumo energético en transporte y se presenta la adaptación de un modelo de demanda de energía frontera al caso del sector transporte. Para tratar las diferencias entre países, se plantea el uso de un enfoque de clases latentes en este contexto, lo que permite obtener tantas demandas como grupos de países identificados.

La base de datos utilizada en este capítulo consiste en una muestra de 24 países de América Latina y el Caribe para el período 1990-2010 y la información ha sido obtenida de CEPAL, OLADE y la *Penn World Table* (PWT 7.1). Se estima una demanda energética en la que la variable dependiente es el consumo energético y las variables explicativas son la renta, la población, el precio de la energía, la participación del sector transporte en la economía, la densidad de población y una tendencia temporal (también incluida al cuadrado). El precio de la energía, que es junto a la renta, una de las variables más relevantes en un análisis de demanda, se obtiene a través de un índice multilateral y transitivo. El consumo total de energía en el sector transporte viene dado por el consumo de varios componentes energéticos (gas natural, gasolina, queroseno, etc.) y por tanto requiere agregar los precios individuales de estos componentes para obtener un índice del precio de la energía. Sin embargo, las agencias estadísticas internacionales no proporcionan ningún tipo de índice de precios para el total de países de nuestra muestra y por tanto ha tenido que ser calculado para este trabajo. El uso de un índice de precios multilateral y transitivo como el obtenido permite comparaciones apropiadas de los países a lo largo del tiempo.

En términos generales, el ranking de países que se puede elaborar a partir de los índices de eficiencia obtenidos cuando se estima una demanda frontera, está notablemente correlacionado (70%) con la clasificación que se deriva de un indicador de intensidad energética (consumo de energía en transporte dividido por PIB). Sin embargo, la falta de correlación que se observa entre eficiencia e intensidad energética para algunos países a lo largo del tiempo, indica que las variaciones en los indicadores pueden estar asociadas a otras circunstancias distintas a cambios en la eficiencia energética. Esto sugiere que las medidas de eficiencia que se derivan de la estimación de demandas frontera son más apropiadas que las proporcionadas por estos indicadores. Por otro lado, a través del modelo de clases latentes en el análisis empírico se identifican tres demandas con elasticidades precio y renta claramente diferentes. Por ejemplo, la elasticidad precio va desde -0,16 para la clase más inelástica hasta -0,41 para la clase más elástica. Si por el contrario se estima una única demanda sin tener en cuenta la heterogeneidad entre países, se obtiene una demanda sesgada con una elasticidad precio igual a -0.23. Las probabilidades de pertenencia a cada clase reflejan que los países con mayor renta *per cápita* y menor densidad de población tienden a tener demandas de menor elasticidad precio. La estimación de estas demandas permite la identificación de aquellos países con mayor eficiencia energética en cada clase, que precisamente coinciden con los que han desarrollado políticas para mejorar el transporte público en los últimos años.

En cuanto a futuras líneas de investigación, la existencia de posibles respuestas asimétricas en la demanda de energía ante cambios en el precio y la renta, será analizada a través de la descomposición de precio y renta utilizada por Gately y Huntington (2002), Adeyemi y Hunt (2007) o Adofo *et al.* (2013). Aunque la reversibilidad de precios ha sido previamente analizada en sector el transporte (ver Dargay, 1992; Gately, 1992; y Dargay y Gately, 1997) esta será la primera vez que este enfoque se aplique en un modelo de demanda de energía frontera. Además, se calcularán los potenciales ahorros energéticos derivados de las estimaciones presentadas en este trabajo y la consecuente reducción en emisiones de gases de efecto invernadero. Otra futura línea de investigación es el análisis del efecto rebote utilizando un modelo similar al que se presenta en el siguiente ensayo. El efecto rebote ha sido ampliamente estudiado en el sector transporte pero a través de enfoques convencionales. No obstante, no existe evidencia empírica para el caso de América Latina y el Caribe, lo que probablemente otorgue interés a los resultados que se obtengan.

Finalmente, en el último ensayo, se estiman modelos de demanda de energía frontera para el sector energético residencial estadounidense, pero en este caso se incorpora en la parte estocástica del modelo, una función asociada a la demanda de servicios energéticos que proporciona información sobre el efecto rebote. Este concepto no ha sido analizado de esta forma hasta ahora, por lo que se trata de una importante contribución en este particular campo de estudio. El efecto rebote es un fenómeno que refleja que tras una mejora en la eficiencia energética, el consumo energético no necesariamente disminuye de forma proporcional a la mejora en eficiencia experimentada. Esto se debe a que los consumidores perciben la mejora en eficiencia como un menor coste del servicio energético que demandan, lo que supone un incremento en la demanda de este servicio y por tanto el consumo energético no se reduce en la cuantía esperada.

Numerosos estudios han utilizado diversos enfoques para medir este concepto. En este ensayo se presenta una clasificación de los diferentes tipos de efecto rebote y de los enfoques alternativos que han sido aplicados para su cálculo. Entre los diferentes métodos, el más comúnmente usado es obtener el efecto rebote a través de las elasticidades precio de las demandas de energía estimadas. Se presenta un resumen de los resultados empíricos obtenidos para el caso del consumo de energía en los hogares, donde los valores estimados por lo general se sitúan entre el 0 y el 60%. Sin embargo, se observan casos en la literatura en los que importantes efectos rebote (incluso por encima del 100%) han sido obtenidos.

En este ensayo, se presenta el modelo estándar de demanda de energía frontera y los supuestos implícitos que se asumen en el modelo. Para obtener información sobre el efecto rebote, se sugiere la extensión de este modelo básico incorporando un factor de ajuste que mitiga o intensifica el efecto de las mejoras en eficiencia energética sobre el consumo de energía. Dos especificaciones alternativas se sugieren para este factor de ajuste (o función de efecto rebote), además de proponerse una estrategia basada en las estimaciones del modelo ALS para tratar el problema que surge en la identificación de la constante en esta función.

La aplicación empírica está basada en el consumo de energía residencial de una muestra de 48 estados de EE.UU. para el período 1995-2011. Las fuentes de datos son el EIA, el *US Census Bureau* y el *National Climatic Data Center* de NOAA. Las variables incluidas para explicar el consumo de energía residencial son la renta, el precio, los días de frío, los días de calor, el tamaño medio de los hogares, la proporción de viviendas unifamiliares y un listado de *dummies* anuales. Para la función de efecto rebote, la cual está basada en la demanda de servicios energéticos, las variables que se incluyen son la renta *per cápita*, el precio de la energía y el tamaño medio de los hogares. El efecto rebote medio que se obtiene para la muestra es relativamente alto (56% y 80% para las dos especificaciones presentadas) comparado con los valores que se observan en la literatura, pero es posible que estos valores estén sobreestimados debido a la concavidad de la forma funcional asumida para las funciones de efecto rebote.

Sin embargo, esto parece ser únicamente un problema de nivel, ya que a pesar de los diferentes efectos rebote que se obtienen a partir de ambas especificaciones, la clasificación ordenada de estados que se deriva es la misma. Esto es, nuestro modelo permite una identificación robusta de aquellos estados que tienen un alto o bajo efecto rebote comparado con el resto de estados analizados. Este resultado puede resultar interesante para el diseño de políticas energéticas, ya que permite la identificación de regiones en las que las políticas destinadas a promover aumentos en la eficiencia energética serían más exitosas. Además, a partir del modelo propuesto, las medidas de efecto rebote que se obtienen, no se ven alteradas por la forma en la que el paso del tiempo es incorporado en la demanda, lo que contrasta con el sesgo que se observa cuando se estima el efecto rebote a través de la elasticidad precio de la demanda.

Debido a que este trabajo es una primera propuesta de utilizar un enfoque frontera para estimar efectos rebote, futura investigación sobre este tema queda pendiente. En primer lugar, la identificación de la "verdadera" constante en la función de efecto rebote debería ser analizada de forma más exhaustiva. Una estrategia prometedora es considerar el factor de ajuste como un término aleatorio asimétrico y estimar el modelo incluyendo el producto de dos términos aleatorios de una sola cola, es decir, el efecto rebote multiplicado por la ineficiencia. En segundo lugar, deben proponerse formas funcionales flexibles y manejables que permitan poder representar toda la gama posible de efectos rebote en las estimaciones. El uso de especificaciones paramétricas alternativas en la función de efecto rebote debería ser analizado en profundidad. En este sentido, con el objetivo de relajar las restricciones de concavidad de las funciones que se han propuesto, también podría explorarse el uso de enfoques alternativos como el uso de métodos de regresión semi-paramétricos.

References

Adeyemi, O.I. and Hunt, L.C. (2007), "Modelling OECD industrial energy demand: Asymmetric price responses and energy-saving technical change", *Energy Economics*, 29(4), 693-709.

Adofo, Y.O., Evans, J. and Hunt, L.C. (2013), "How sensitive to time period sampling is the asymmetric price response specification in energy demand modelling?", *Energy Economics*, 40, 90-109.

Agrell, P.J. and Bogetoft, P. (2007), *Development of benchmarking models for German electricity and gas distribution*, Consultation report, Bundesnetzagentur, Bonn, Germany.

Agrell, P.J., Farsi, M., Filippini, M. and Koller, M. (2013), *Unobserved heterogeneous effects in the cost efficiency analysis of electricity distribution systems*, CER-ETH Economics working papers series 13/171, CER-ETH - Center of Economic Research (CER-ETH) at ETH Zurich.

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

Al-Ghandoor, A., Samhouri, M., Al-Hinti, I., Jaber, J. and Al-Rawashdeh, M. (2012), "Projection of future transport energy demand of Jordan using adaptive neuro-fuzzy technique", *Energy,* 38(1), 128-135.

Alfredsson, E.C. (2004), "'Green' consumption–no solution for climate change", *Energy*, 29, 513-524.

Alvarez, A., Amsler, C., Orea, L. and Schmidt, P. (2006), "Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics", *Journal of Productivity Analysis*, 25, 201-212.

Amin, G.R., Emrouznejad, A. and Rezaei, S. (2011), "Some clarifications on the DEA clustering approach", *European Journal of Operational Research*, 215, 498-501.

Ang, B.W. (2006), "Monitoring changes in economy-wide energy efficiency: From energy-GDP ratio to composite efficiency index", *Energy Policy*, 34, 574-582.

Banker, R.D. and Morey, R. (1986), "Efficiency analysis for exogenously fixed inputs and outputs", *Operations Research*, 34, 513-521.

Barnett W.A., Kirova, M., and Pasupathy, M. (1996), "Estimating policy invariant deep parameters in the financial sector, when risk and growth matter", *Journal of Money Credit and Banking*, 27, 1402-1430.

Barros, C.P. and Prieto-Rodríguez, J. (2008), "A revenue-neutral tax reform to increase demand for public transport services", *Transportation Research Part A: Policy and Practice*, 42, 659-672.

Battese, G.E. and Coelli, T.J. (1995), "A model for technical efficiency for panel data", *Empirical Economics*, 20, 325-332.

Battese, G., Rao, D.S.P., and O'Donnell, C.J. (2004), "A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies", *Journal of Productivity Analysis*, Springer, vol. 21(1), 91-103.

Bauer, M., Mar, E. and Elizalde, A. (2003), "Transport and energy demand in Mexico: The personal income shock", *Energy Policy*, 31(14), 1475-1480.

Beard, T.R., Caudill, S.B. and Gropper, D.M. (1991), "Finite mixture estimation of multiproduct cost functions", *The Review of Economics and Statistics*, 73(4), 654-64.

Becker Zuazua, F. (2011), "Prospectiva de la electricidad en España", *Estudios de Economía Aplicada*, 29(2), 415-431.

Bentzen, J. (2004), "Estimating the rebound effect in US manufacturing energy consumption", *Energy Economics*, 26(1), 123-34.

Berkhout, P.H.G., Muskens, J.C. and Velthuijsen, J.W. (2000), "Defining the rebound effect", *Energy Policy*, 28(6-7), 425-32.

Bhattacharyya, S. (2011), *Energy economics: Concepts, issues, markets and governance*, London: Springer.

Billinton, R. and Acharya, J. (2005), "Consideration of multi-state weather models in reliability evaluation of transmission and distribution systems", *Canadian Conference on Electrical and Computer Engineering,* 916-922.

Billinton, R. and Wenyuan, L. (1991), "Novel method incorporating weather composite system adequacy evaluation", *Transactions on Power Systems, IEEE*, 6, 3, 1154-1160.

Billinton, R. and Wu, C. (2001), "Predictive reliability assessment of distribution system including extreme adverse weather", *Canadian Conference on Electrical and Computer Engineering*, 2, 719-724.

Binswanger, M., (2001), "Technological progress and sustainable development: What about the rebound effect?", *Ecological Economics*, 36(1), 119-32.

Bogetoft, P. and Otto, L. (2011), *Benchmarking with DEA, SFA, and R*, Springer, New York.

Brännlund, R., Ghalwash, T. and Norström, J. (2007), "Increased energy efficiency and the rebound effect: Effects on consumption and emissions", *Energy Economics*, 29, 1-17.

Brunekreeft, G., Neuhoff, K. and Newbery, D. (2005), "Electricity transmission: An overview of the current debate", *Utilities Policy*, 13, 2, 73-93.

Caudill S.B, and Ford J.M. (1993), "Biases in frontier estimation due to heteroscedasticity", *Economic Letters*, 41, 17-20.

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, 105-11.

Caves, D.W., Christensen, L.R. and Diewert, W.E. (1982), "The economic theory of index numbers and the measurement of input, output, and productivity", *Econometrica*, 50(6), 1393-1414.

Ceylan, H., Ceylan, H., Haldenbilen, S. and Baskan, O. (2008), "Transport energy modeling with meta-heuristic harmony search algorithm, an application to Turkey", *Energy Policy*, 36(7), 2527-2535.

Chavez-Baeza, C. and Sheinbaum-Pardo, C. (2014), "Sustainable passenger road transport scenarios to reduce fuel consumption, air pollutants and GHG (greenhouse gas) emissions in the Mexico City Metropolitan Area", *Energy*, 66(1), 624-634.

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

Costello, A.B. and Osborne, J.W. (2005), "Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis", *Practical Assessment, Research & Evaluation*, 10(7), 1-9.

Dahl, C. (1995), "Demand for transportation fuels: A survey of demand elasticities and their components", *The Journal of Energy Literature*, 1, 3-27.

Dargay, J.M. (1992), "The irreversible effects of high oil prices: Empirical evidence for the demand for motor fuels in France, Germany and the U.K.", in D. Hawdon (ed.), *Energy Demand: Evidence and Expectations*, London: Surrey University Press, 165-182.

Dargay, J.M. and Gately, D. (1997), "The demand for transportation fuels: Imperfect price reversibility?", *Transportation Research Part B: Methodological*, 31 (1), 71–82.

Dismukes, D.E., Cope III, R.F. and Mesyanzhinov, D. (1998), "Capacity and economies of scale in electric power transmission", *Utilities Policy*, 7, 3, 155-162.

Dreher, M., Wietschel, M., Göbelt, M. and Rentz, O. (1999), "Energy price elasticities of energy service demand for passenger traffic in the Federal Republic of Germany", *Energy*, 24(2), 133-140.

Druckman, A., Chitnis, M., Sorrell, S. and Jackson, T. (2010), *An investigation into the rebound and backfire effects from abatement actions by UK households*, RESOLVE Working Paper 05-10, Guildford, University of Surrey.

ECLAC (2005), *The millennium development goals: A Latin American and Caribbean perspective*, United Nations Publications, Santiago de Chile.

ECLAC (2010), *Sustainable development in Latin America and the Caribbean: Trends, progress, and challenges in sustainable consumption and production, mining, transport, chemicals and waste management*, United Nations, Publications, Santiago de Chile.

Edwards, B.K. (2003), *The economics of hydroelectric power*, Edward Elgar, Cheltenham, UK.

Elteto, O. and Koves, P. (1964), "On a problem of index number computation relating to international comparison", *Statisztikai Szemle*, 42, 507-18.

Farrell, M.J. (1957), "The measurement of productive efficiency", *Journal of the Royal Statistical Society*. A, 120, 253-281.

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

Filippini, M. and Hunt, L.C. (2011), "Energy demand and energy efficiency in the OECD countries: A stochastic demand frontier approach", *The Energy Journal*, 32(2), 59-80.

Filippini, M. and Hunt, L.C. (2012), "US residential energy demand and energy efficiency: A stochastic demand frontier approach", *Energy Economics*, 34(5), 1484-1491.

Fonseca, J.R.S. and Cardoso, M.G.M.S. (2007), "Mixture-model cluster analysis using information theoretical criteria", *Intelligent Data Analysis*, 11(2), 155-173.

Fried, H.O., Lovell, C.A.K., Schmidt, S.S. and Yaisawarng, S. (2002), "Accounting for environmental effects and statistical noise in data envelopment analysis", *Journal of Productivity Analysis*, 17, 157-174.

Galindo, L.M. (2005), "Short-and long-run demand for energy in Mexico: A cointegration approach", *Energy Policy* 33(9), 1179-85.

Gately, D. (1992), "Imperfect price-reversibility of U.S. gasoline demand: Asymmetric responses to price increases and decreases", *The Energy Journal*, 13 (4), 179-207.

Gately, D. and Huntington, H.G. (2002), "The Asymmetric Effects of Changes in Price and Income on Energy and Oil Demand", *The Energy Journal*, 23(1), 19-55.

Giannakis, D., Jamasb, T. and Pollitt, M. (2005), "Benchmarking and incentive regulation of quality of service: An application to the UK electricity distribution networks", *Energy Policy,* 33(1), 2256-2271.

Graham, D. and Glaister, S. (2002), "The demand for automobile fuel: A survey of elasticities", *Journal of Transport Economics and Policy*, 36, 1-26.

Greene, W. (2004), "Distinguishing between heterogeneity and inefficiency: Stochastic frontier analysis of the World Health Organization's panel data on national health care systems", *Health Economics*, 13, 959-980.

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

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

Greene, D.L., Kahn, J.R. and Gibson, R.C. (1999a), *An econometric analysis of the elasticity of vehicle travel with respect to fuel cost per mile using the RTEC survey data,* Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Greene, D. L., Kahn, J.R. and Gibson, R.C. (1999b), "Fuel economy rebound effect for US household vehicles", *The Energy Journal*, 20(3), 1-31.

Greene, W., Orea, L. and Wall, A. (2011), *A one-stage random effect counterpart of the fixed-effect vector decomposition model with an application to UK electricity distribution utilities*, Efficiency Series Papers 01/2011, Oviedo Efficiency Group.

Greenfield D. and Kwoka, J. (2011), "The cost structure of regional transmission organizations", *The Energy Journal*, 32, 4, 159-181.

Greening, L.A., Greene, D.L. and Difglio, C. (2000), "Energy efficiency and consumption – the rebound effect – a survey", *Energy Policy*, 28, 6-7, 389-401.

Gropper, D.M., Caudill, S.B. and Beard, T.R. (1999), "Estimating multiproduct cost functions over time using a mixture of normals", *Journal of Productivity Analysis*, 11, 201-218.

Growitsch, C., Jamasb, T. and Wetzel, H. (2012), "Efficiency effects of observed and unobserved heterogeneity: Evidence from Norwegian electricity distribution networks", *Energy Economics*, 34, 2, 542-548.

Guertin, C., Kumbhakar, S., and Duraiappah, A. (2003), *Determining demand for energy services: Investigating income-driven behaviours*, International Institute for Sustainable Development.

Haldenbilen, S. and Ceylan, H. (2005), "Genetic algorithm approach to estimate transport energy demand in Turkey", *Energy Policy*, 33(1), 89-98.

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

Haney, A.B. and Pollitt, M.G. (2013), "International benchmarking of electricity transmission by regulators: A contrast between theory and practice?", *Energy Policy*, 62, 267-281.

Hao, H., Wang, H. and Yi, R. (2011), "Hybrid modeling of China's vehicle ownership and projection through 2050", *Energy*, 36(2), 1351-1361.

Heston, A., Summers, R. and Aten, B. (2012), *Penn World Table Version 7.1*, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

Huang, C.J. and Liu, J.T. (1994), "Estimation of non-neutral stochastic frontier production function", *Journal of Productivity Analysis*, 5(2), 171-180.

Huettner, D.A., and Landon, J.H. (1978), "Electric utilities: Scale economies and diseconomies", *Southern Economic Journal*, 44, 4, 883-912.

Hymel, K.M., Small, K.A. and Van Dender, K. (2010), "Induced demand and rebound effects in road transport", *Transportation Research Part B: Methodological*, 44(10), 1220-1241.

IRGC (2013), *The Rebound Effect: Implications of consumer behaviour for robust energy policies*, A review of the literature on the rebound effect in energy efficiency and report from expert workshops, Lausanne.

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

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

Jamasb, T. and Pollitt, M. (2005), “Electricity market reform in the European Union: Review of progress toward liberalization & integration”, *The Energy Journal*, 26(special I), 11-42.

Jamasb, T. and Pollitt, M. (2007), “Incentive regulation of electricity distribution networks: Lessons of experience from Britain”, *Energy Policy*, 35, 12, 6163-6187.

Jamasb, T., Orea, L. and Pollitt, M. (2010), *Weather factors and performance of network utilities: A methodology and application to electricity distribution*, Electricity Policy Research Group Working Paper 1020, University of Cambridge.

Jamasb, T., Orea, L. and Pollitt, M. (2012), “Estimating the marginal cost of quality improvements: The case of the UK electricity distribution companies”, *Energy Economics*, 34, 5, 1498-1506.

Jenkins, J., Nordhaus, T. and Shellenberger, M. (2011), *Energy Emergence. Rebound & backfire as emergent phenomena*, Oakland, The Breakthrough Institute.

Jevons, W.S. (1865), *The Coal Question: Can Britain survive?*, Republished by Macmillan, London, UK, 1906.

Johnson, A.L. and Kuosmanen, T. (2011), “One-stage estimation of the effects of operational conditions and practices on productive performance: Asymptotically normal and efficient, root-n consistent StoNEZD method”, *Journal of Productivity Analysis*, 36(2), 219-230.

Johnson, A.L. and Kuosmanen, T. (2012), “One-stage and two-stage DEA estimation of the effects of contextual variables”, *European Journal of Operational Research*, 220(2), 559-570.

Jondrow, J., Lovell, C.A.K., Materov, I.S. and Schmidt, P. (1982), “On the estimation of technical efficiency in the stochastic frontier production function model”, *Journal of Econometrics*, 19, 233-238.

Joskow, P.L. (2014), “Incentive regulation in theory and practice: Electricity distribution and transmission networks”, Chapter in NBER book *Economic regulation and its reform: What have we learned?*, 291-344, National Bureau of Economic Research, Inc.

Khazzoom, J. D., (1980), “Economic implications of mandated efficiency in standards for household appliances”, *The Energy Journal*, 1(4), 21-40.

Kim, M. and Schmidt, P. (2008), “Valid tests of whether technical inefficiency depends on firm characteristics”, *Journal of Econometrics*, 144(2), 409-427.

Koehler, A.B. and Murphree, B.S. (1988), “A comparison of the Akaike and Schwarz criteria for selecting model order”, *Applied Statistics*, 37, 187-195.

Kruger, J.J. (2010), “Comment on “A new clustering approach using data envelopment analysis””, *European Journal of Operational Research*, 206, 269-270.

Kumbhakar, S.C. and Lovell, C.A.K. (2000), *Stochastic frontier analysis*, Cambridge University Press, Cambridge.

Kumbhakar, S.C., Ghosh, S. and McGuckin, J.T. (1991), "A generalized production frontier approach for estimating determinants of inefficiency in US dairy farms", *Journal of Business & Economic Statistics*, 9, 279-286.

Kumbhakar, S.C., Parmeter, C.F. and Tsionas, E.G. (2013), "A zero inefficiency stochastic frontier model", *Journal of Econometrics*, 172(1), 66-76.

Kumbhakar, S.C., Park, B.U., Simar, L. and Tsionas, E.G. (2007), "Nonparametric stochastic frontiers: A local maximum likelihood approach", *Journal of Econometrics*, 137, 1-27.

Kuosmanen, T. (2012), "Stochastic semi-nonparametric frontier estimation of electricity distribution networks: Application of the StoNED method in the Finnish regulatory model", *Energy Economics*, 34, 2189-2199.

Kuosmanen, T., Saastamoinen, A. and Sipiläinen, T. (2013), "What is the best practice for benchmark regulation of electricity distribution? Comparison of DEA, SFA and StoNED Methods", *Energy Policy*, 61, 740-750.

IEA (2014), *Energy efficiency indicators: Essentials for policy making*, International, IEA Publishing, Paris

Islas, J., Manzini, F. and Masera, O. (2007), "A prospective study of bioenergy use in Mexico", *Energy*, 32(12), 2306-2320.

Lai, H. and Huang, C.J. (2010), "Likelihood ratio tests for model selection of stochastic frontier models", *Journal of Productivity Analysis*, 34, 3-13.

Lenzen, M. and Dey, C.J. (2002), "Economic, energy and greenhouse emissions impacts of some consumer choice, technology and government outlay options", *Energy Economics*, 24, 377-403.

Li, Q., Huang, C., Li, D. and Fu, T. (2002), "Semiparametric smooth coefficient models", *Journal of Business and Economic Statistics*, 20(3), 412-422.

Limanond, T., Jomnonkwao, S. and Srikaew, A. (2011), "Projection of future transport energy demand of Thailand", *Energy Policy*, 39, 2754-2763.

Lu, I.J, Lewis, C. and Lin, S.J. (2009), "The forecast of motor vehicle, energy demand and CO_2 emission from Taiwan's road transportation sector", *Energy Policy*, 37(8), 2956-2961.

Lu, I.J., Lin, S.J. and Lewis, C. (2008), "Grey relation analysis of motor vehicular energy consumption in Taiwan", *Energy Policy*, 36(7), 2556-2561.

Manzini, F. (2006), "Inserting renewable fuels and technologies for transport in Mexico City Metropolitan Area", *Hydrogen Energy*, 31, 327-335.

Martins-Filho, C. and Yao, F. (2013), "Semiparametric stochastic frontier estimation via profile likelihood", *Econometric Reviews*, forthcoming.

Mizobuchi, K. (2008), "An empirical study on the rebound effect considering capital costs", *Energy Economics*, 30, 2486-2516.

Moazami Goudarzi, M.R. and Jaber Ansari, M.R. (2012), "Clustering Decision Making Units (DMUs) using Full Dimensional Efficient Facets (FDEFs) of PPS with BCC technology", *Applied Mathematical Sciences*, 6(29), 1431-1452.

Mundlak, Y. (1978), "On the pooling of time series and cross section data", *Econometrica*, 64, 69-85.

Muñiz, M.A. (2002), "Separating managerial inefficiency and external conditions in data envelopment analysis", *European Journal of Operational Research*, 143, 625-643.

Muñiz, M.A., Paradi, J., Ruggiero, J. and Yang, Z. (2006), "Evaluating alternative DEA models used to control for non-discretionary inputs", *Computers and Operations Research*, 33, 5, 1173-1183.

Murat, Y.S. and Ceylan, H. (2006), "Use of artificial neural networks for transport energy demand modeling", *Energy Policy*, 34(17), 3165-3172.

Nillesen, P. and Pollitt, M. (2010), "Using regulatory benchmarking techniques to set company performance targets: The case of US electricity", *Competition and Regulation in Network Industries*, 11(1), 50-85.

O'Donnell, C.J., Rao, D.S.P. and Battese, G. (2008), "Metafrontier frameworks for the study of firm-level efficiencies and technology ratios", *Empirical Economics*, 34(2), 231-255.

Ofgem (2011), *Decision on strategy for the next transmission price control - RIIO-T1. Tools for cost assessment*, March 2011. Office of Gas and Electricity Markets, London.

Orea, L. and Kumbhakar, S. (2004), "Efficiency measurement using stochastic frontier latent class model", *Empirical Economics*, 29, 169-183.

Po, R-W., Guh, Y-Y. and Yang, M-S. (2009), "A new clustering approach using data envelopment analysis", *European Journal of Operational Research*, 199, 276-284.

Pollitt, M. (1995), *Ownership and performance in electric utilities*, Oxford University Press, Oxford.

Pradhan, S., Bahadur, B. and Bhusan, V. (2006), "Mitigation potential of greenhouse gas emission and implications on fuel consumption due to clean energy vehicles as public passenger transport in Kathmandu Valley of Nepal: A case study of trolley buses in ring road", *Energy,* 31(12), 1748-1760.

Ray, S.C. (1988), "Data envelopment analysis, nondiscretionary inputs and efficiency: An alternative interpretation", *Socio-Economic Planning Sciences*, 22(4), 167-176.

Reifschneider D. and Stevenson R. (1991), "Systematic departures from the frontier: A framework for the analysis of firm inefficiency", *International Economic Review*, 32, 715-723.

Rothstein, B. and Halbig, G. (2010), *Weather sensitivity of electricity supply and data services of the German Met Office*, Management of Weather and Climate Risk in the Energy Industry, A. Troccoli (ed.), 253-266.

Roy, J. (2000), "The rebound effect: Some empirical evidence from India", *Energy Policy*, 28(6-7), 433-38.

Ruggiero, J. (1996), "On the measurement of technical efficiency in the public sector", *European Journal of Operational Research*, 90, 553-565.

Ruggiero, J. (1998), "Non-discretionary inputs in data envelopment analysis", *European Journal of Operational Research*, 111, 461-9.

Sa'ad, S. (2010), "Improved technical efficiency and exogenous factors in transportation demand for energy: An application of structural time series analysis to South Korean data", *Energy*, 35(7), 2745-2751.

Sachs, J. (1995), *Aristotle's physics: A guided study*, New Brunswick, N.J.: Rutgers University Press.

Samini, R. (2003), "Road transport energy demand modelling. A cointegration approach", *Energy Economics*, 17(4), 329-339.

Saunders, H.D. (1992), "The Khazzoom–Brookes postulate and neoclassical growth", *The Energy Journal*, 13 (4), 130-148.

Saunders, H.D. (2000), "A view from the macro side: Rebound, backfire, and Khazzoom-Brookes", *Energy Policy*, 28, 439-449.

Saunders, H.D. (2008), "Fuel conserving (and using) production functions", *Energy Economics*, 30, 2184-2235.

Schmidt, P. and Sickles, R.C. (1984), "Production frontiers and panel data", *Journal of Business and Economic Statistics*, 2(4), 367-374.

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. (2010), "Inferences from cross-sectional, Stochastic Frontier Models", *Econometric Reviews*, Taylor and Francis Journals, vol. 29(1), 62-98.

Small, K.A. and Van Dender, K. (2005), *A study to evaluate the effect of reduced greenhouse gas emissions on vehicle miles travelled*, Prepared for the State of California Air Resources Board, the California Environment Protection Agency and the California Energy Commission, Final Report ARB Contract Number 02-336, Department of Economics, University of California, Irvine.

Small, K.A. and Van Dender, K., (2007), "Fuel efficiency and motor vehicle travel: The declining rebound effect", *The Energy Journal*, 28 (1), 25-51.

Sorrell, S., (2007), *The rebound effect: An assessment of the evidence for economy-wide energy savings from improved energy efficiency*, A report produced by the Sussex Energy Group for the Technology and Policy Assessment function of the UK Energy Research Centre, UK Energy Research Centre.

Sorrell, S. and Dimitropoulos, J. (2007), *UKERC review of evidence for the rebound effect. Technical report 2: Econometric studies*, Working Paper, UK Energy Research Centre.

Sorrell, S. and Dimitropoulos, J. (2008), "The rebound effect: Microeconomic definitions, limitations and extensions", *Ecological Economics*, 65(3), 636-649.

Stead, D. (2001), "Transport intensity in Europe – indicators and trends", *Transport Policy*, 8, 29-46.

Stevenson, R.E. (1980), "Likelihood functions for generalized stochastic frontier estimation", *Journal of Econometrics*, 13, 57-66.

Suganthi, L. and Samuel, A.A. (2012), "Energy models for demand forecasting–A review", *Renewable and Sustainable Energy Reviews*, 16(2), 1223-1240.

Sun, K. and Kumbhakar, S.C. (2013), “Semiparametric smooth-coefficient stochastic frontier model”, *Economic Letters*, 120, 305-309.

Szulc, B. (1964), “Indices for multiregional comparisons”, *Przeglad Statystyezny 3*, *Statistical Review 3*, 239-54.

Thomas, B. and Azevedo, I.L. (2013), “Estimating direct and indirect rebound effects for U.S. households with Input-Output analysis. Part 1: Theoretical framework”, *Ecological Economics*, 86, 199-210.

Verbeek, M. (2008), *A guide to modern econometrics*, John Wiley & Sons, Hoboken, NJ, 3rd edition.

von Geymueller, P. (2009), “Static versus dynamic DEA in electricity regulation: The case of US transmission system operators”, *Central European Journal of Operations Research*, 17, 397-413.

Vuong Q.H. (1989), “Likelihood ratio tests for model selection and non-nested hypotheses”, *Econometrica*, 57(2), 307-333.

Wang, H.J. and Ho, C.W. (2010), “Estimating fixed-effect panel stochastic frontier models by model transformation”, *Journal of Econometrics*, 157, 286-296.

Wang, H.J. and Schmidt, P. (2002), “One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels”, *Journal of Productivity Analysis*, 18, 129-144.

Wang, H., Zhou, P. and Zhou, D.Q., (2012), “An empirical study of direct rebound effect for passenger transport in urban China”, *Energy Economics*, 34(2), 452-460.

Wohlgemuth, N. (1997), “World transport energy demand modeling: Methodology and elasticities”, *Energy Policy*, 25 (14–15), 1109-1119.

Yu, W., Jamasb, T. and Pollitt, M. (2009), “Does weather explain cost and quality performance? An analysis of UK electricity distribution companies”, *Energy Policy*, 37(11), 4177-4188.

Zein-Elabdin, E.O. (1997), “Improved stoves in Sub-Saharan Africa: The case of the Sudan”, *Energy Economics*, 19(4), 465-75.

Zhang, M., Mu, H., Li, G. and Ning, Y. (2009), “Forecasting the transport energy demand based on PLSR method in China”, *Energy*, 34(9), 1396-1400.

Zhou, P., Ang, B.W. and Poh, K.L., (2008), “A survey of Data Envelopment Analysis in energy and environmental studies”, *European Journal of Operational Research*, 189, 1-18.