UNIVERSITY OF COPENHAGEN

Master Thesis

Christine Windfeld Hansen & Anna Plum

The Relationship between Animal Health and Milk Quality and the Efficiency of Danish Dairy Farms - A Stochastic Distance Function Approach

Supervisors: Jakob Vesterlund Olsen & Tomasz Gerard Czekaj

Submitted on: 5 March 2017

Name of department:	University of Copenhagen, Faculty of Science, Department of Food and Resource Economics
Authors:	Christine Windfeld Hansen (wzx552) & Anna Plum (pfx296)
Title and subtitle:	The Relationship between Animal Health and Milk Quality and the Efficiency of Danish Dairy Farms - A Stochastic Distance Function Approach
Supervisors:	Jakob Vesterlund Olsen & Tomasz Gerard Czekaj
ECTS:	30
Submitted on:	5 March 2017
Front page photo:	Anna Plum

Abstract

This thesis has investigated the relationship between animal health and milk quality indicators and technical efficiency on Danish dairy farms to examine potential sources of inefficiency. To estimate the effects, translog *input* and *output* distance functions were estimated as stochastic frontier models, using panel data for full time dairy farms in Denmark from 2011 to 2015.

On the 1 April 2015, the milk quota system within the European Union was abolished, which changed the market conditions for dairy producers substantially. Because of the regulation changes during the period covered in this thesis, the choice of orientation describing the production process became more important, and thus both an input and an output-oriented model were estimated and compared. We find that the Danish dairy farmers have been more input than output oriented during the period.

As expected, we find that higher frequencies of mastitis and reproductive disorders have a small and negative impact on the technical efficiency of the farm, whereas other types of disorders are without significance. The results indicate that Danish dairy producers are aware of the importance of preventing and treating diseases. As anticipated, we find that a higher milk quality is associated with a higher level of technical efficiency. We also find that the type of breed and milking system are rather important factors in relation to productivity. The results of the model are used to evaluate the potential for Danish dairy producers to improve technical efficiency and productivity in the future, by discussing the economic gains and potential disadvantages of investing in new technology.

Overall, full time dairy producers in Denmark are highly efficient and have increased their level of technical efficiency from 2011 to 2015, and thus drastic adjustments and improvements in relation to animal health and milk quality are not necessary. We argue that future improvements should be in relation to the technology used in production.

Keywords: technical efficiency, distance functions, stochastic frontier analysis, Danish dairy farms, animal health, milk quality

Acknowledgements

This thesis is the final part of our Master Degrees in Agricultural Economics at University of Copenhagen, Faculty of Science.

We would like to express our sincerest thanks to SEGES and especially to Jannik Toft Andersen, for providing the data and for always answering our questions regarding the dairy sector.

We would also like to express our deepest gratitude for the advice and support provided by our supervisors, Jakob Vesterlund Olsen and Tomasz Gerard Czekaj, from the Department of Food and Resource Economics at University of Copenhagen.

Further we would like to thank our friends and classmates at the thesis office for continued support and discussions throughout the process.

Copenhagen, 5 March 2017

Christine Windfeld H.

Christine Windfeld Hansen

Anna Plum

Table of contents

List of Figures

List of Tables

1. Introduction

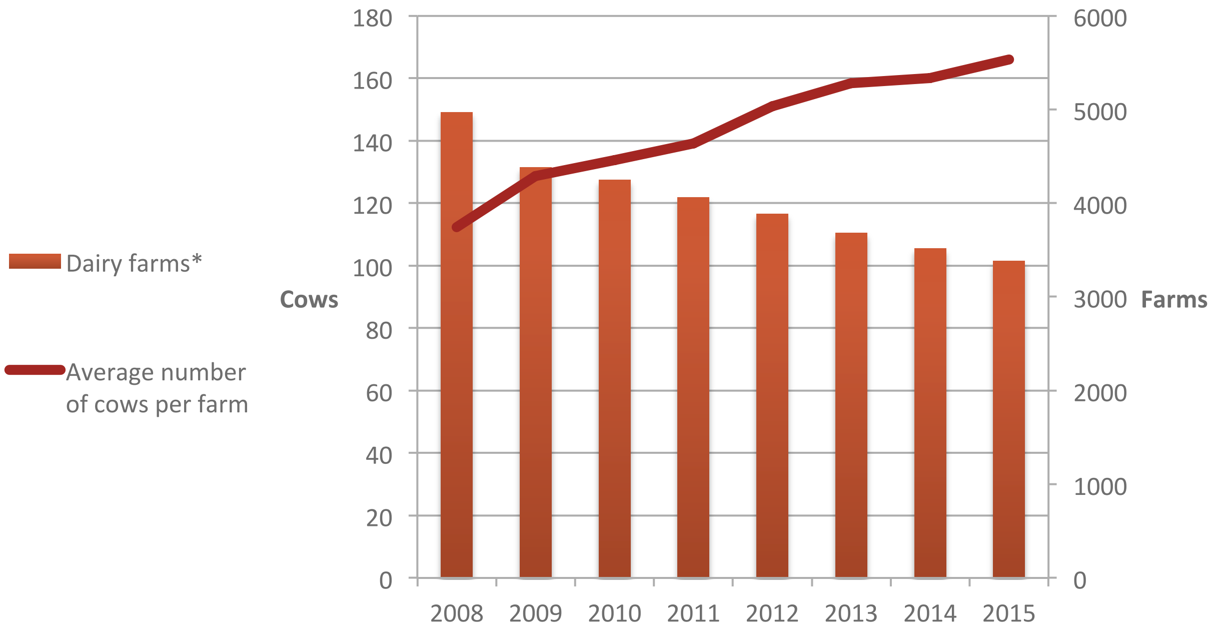
Farmers in Europe have operated in a dynamic environment for decades. The dynamic environment has occurred due to changes in demand for agricultural goods, overproduction, and new attitudes towards the agricultural sector. Adjustments and developments of the farm have been crucial to be able to survive as a competitive farmer.

1.1 The Danish dairy sector

The Danish dairy sector has developed throughout the last couple of decades despite the milk quota restrictions imposed by the European Union in 1984 (Landbrug og Fødevarer 2015). The Danish dairy farmers have experienced structural changes and made several investments, which has helped the sector to become one of the most efficient dairy sectors in the world. However, despite these improvements several dairy farmers are suffering economically and many have already gone out of business. Leaving the business could be either due to low efficiency or improper use of financial instruments. Common for all those who have left the business is that they failed in turning a profit. The farmers who are still at risk of shutting down are all subject to low equity or high debt (Pedersen et al. 2016).

Another development in the sector has been that the number of dairy farms in Denmark has decreased from nearly 5,000 farms in 2008 to approximately 3,400 farms in 2015. During the same time the average number of dairy cows per farm increased, as can be seen from Figure 1.1. The total number of dairy cows has also decreased since 2008, because the farmers until the 1 April 2015 were adjusting production according to the milk quota restrictions (Statistik 2015). While many farms have been closing, the remaining farms have gotten bigger. This suggests that the bigger farms are experiencing good results while the smaller farms cannot survive in the market (Pedersen et al. 2016).

Figure 1.1: The development in the farm size and the number of farms

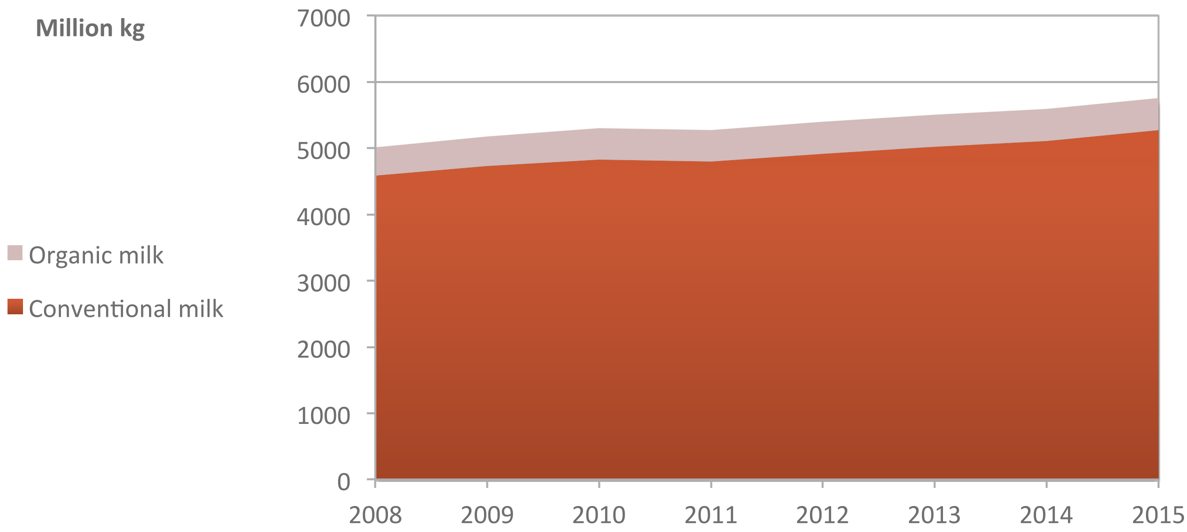


Note: *The figure contains all dairy cow farms in Denmark including part time and hobby farms.

Source: (Statistik 2016d) and own calculations.

However, despite this development, the milk production in Denmark has not decreased, as can be seen from Figure 1.2. The amount of organic milk produced has been stable at around 480 million kg a year from 2008 to 2015, while the amount of conventional milk produced has increased from less than 5,000 million kg per year to more than 5,200 million kg per year.

Figure 1.2: The development in the milk production

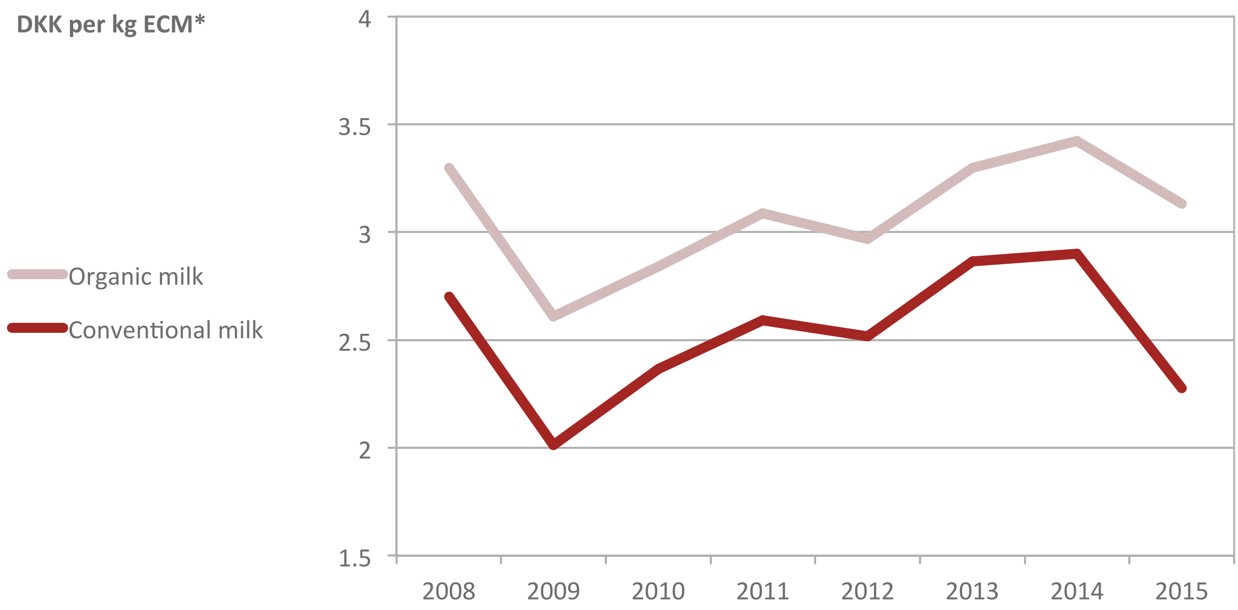


Source: (Statistik 2016a).

Even though there has been an increase in the amount of milk produced from 2008 and on, there has at the same time been a reduction in the number of dairy cows, which means that the yield per cow has increased. The farmers have managed to optimize their production, without increasing the number of dairy cows on the farm (Statistik 2016b).

In the years leading up to the removal of the milk quotas, the dairy producers have produced more milk than the previous years while there has been a decrease in the demand of dairy products from China. The increase in supply and decrease in demand has caused the price on milk to decrease, as can be seen from Figure 1.3 (Statistik 2016b; Vidø et al. 2016). This development was expected, because the incentives to limit the milk production were removed when the milk quotas were revoked on the 1 April 2015. This market change is one of the reasons why so many Danish dairy farmers have not been able to stay in production (Berthou 2016).

Figure 1.3: The development in the price on milk



Note: * Energy corrected milk (ECM)

Source: (Statistik 2016e).

1.2 Efficiency analysis

If a farmer knows how every part of his production works, he can more easily adjust to market changes, and hence always be efficient. For many years, the Danish dairy sector has overall performed poorly since not all farmers have managed to cover their total costs. However, despite the unstable market conditions some Danish dairy farmers have succeeded in generating a profit and therefore been successful through a period longer than just one season (Andersen et al. 2014).

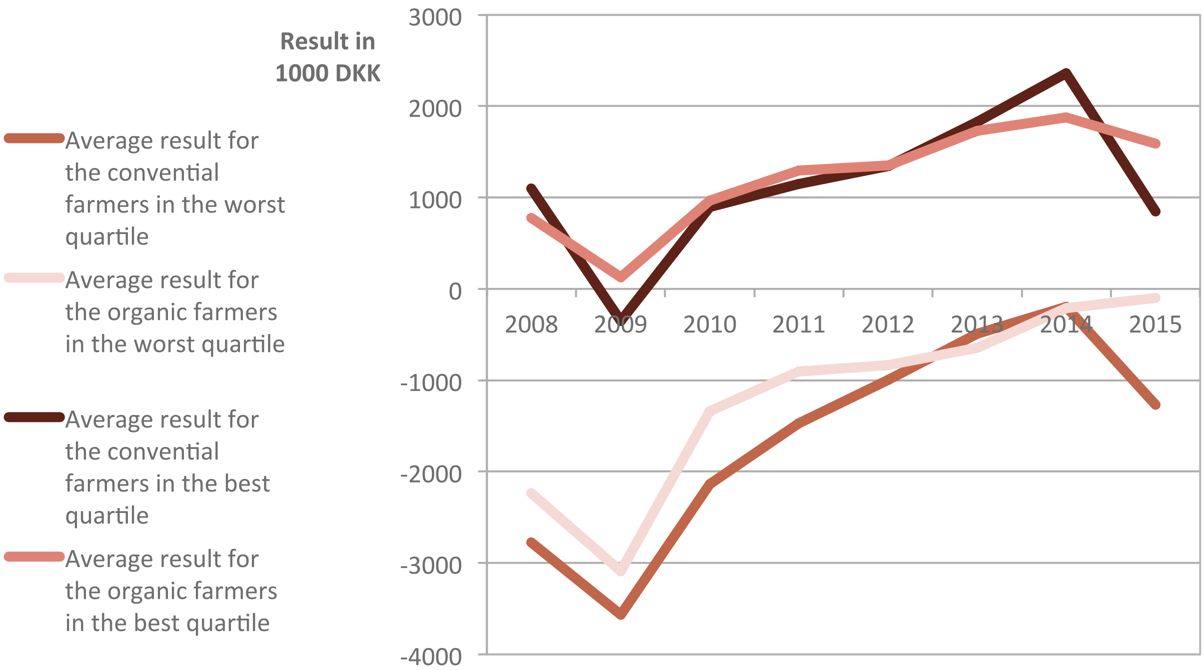
The literature concerning agricultural production has used efficiency analysis extensively. It is based on production theory, which studies the transformation of inputs into outputs. The method bundles partial productivities into aggregate performance measures. Thereby it compares the current performance level of a farm with the potential optimal performance level by determining a production frontier. The production frontier represents the best practice transformation of inputs into outputs. Farms that are placed below the frontier are not technically efficient, but have the possibility to change their input-output transformation and move closer to the frontier (van der Voort et al. 2014). It is possible to distinguish between producing the maximal feasible amount of outputs with a given amount of inputs (output-oriented technical efficiency) and producing with the minimal amount of input to obtain a given amount of output (input-oriented technical efficiency). In the case of an output oriented dairy farm, technical efficiency measures the ability to produce the maximum level of milk from the given set of input factors, whereas the input oriented dairy farmer will reduce inputs while maintaining the same level of milk output. A technical efficiency score is calculated for each individual farm by measuring the distance between the farms' production level and the best practice frontier (van der Voort et al. 2014).

1.3 Problem identification

The best third of all dairy farmers in Denmark has proven that it is possible to do well even during downturns in the market (SEGES P/S 2016). There are several ways in which the successful farmers manage to maintain economically good results during the different changes in the market. The success is dependent on an efficient production, where output is maximized or input is minimized (Andersen et al. 2014). The analysis by Pedersen et al. (2016) shows that the best dairy farmers, both conventional and organic, manage to keep total cost at around 80 per cent of the gross output whereas for the less successful ones the cost make up between 95-98 per cent of the gross output. The dairy farmers operate in a market close to perfect competition, and thus keeping cost at a minimum is crucial in order to survive.

From Figure 1.4 it is clear that there is a substantial difference in the results obtained by the best and the worst farmers in the sector. For all the years, except the year 2009, the average result for the best quartile with at least 100 dairy cows has been close to at least 1 million DKK for both the conventional and organic farmers. From 2009 to 2014 the best quartile experienced a positive development in the results. In 2015 both the conventional and organic farmers in the best quartile once again experienced a drop in the result alongside the conventional farmers in the worst quartile. Since 2009 the organic dairy farmers in the worst quartile have experienced a positive development. Overall the worst quartile has experienced negative results from 2008 to 2015.

Figure 1.4: The development and the differences in the results among dairy farmers for the best and worst quartile



Note: *The results are the average for those farms with 100 -199 dairy cows and those with more than 200.

Source: (Statistik 2016e) and own calculations.

The literature has focused mostly on why these differences occur in relation to equity and investments. Several reports show that the best farmers are good at investing and keeping debt to a minimum, as in Andersen et al. (2014), Kaiser et al. (2011), and Pedersen et al. (2016). In these reports, it is clearly shown that the best performing farms are in position of a manager who is aware of how resources can be allocated in an optimal way. However, these reports have focused mainly on the effects of capital and labour inputs on the farm results, but have left out the aspect of animal health and milk quality indicators, which could also help explain some of the differences in the results obtained among the dairy farmers in Denmark.

The milk quality is important, since it affects the price a farmer can receive for the milk. The milk quality is measured using different milk quality indicators and depending on the levels of these, the size of the supplement is determined. These supplements are of great importance to the industry and even though the Danish dairy producers have shown great improvements in relation to lowering the cell count, the Danish dairy farmers are still missing out on hundreds of millions DKK (Andersen et al. 2016). Producing organic milk has also proven to be of great importance for the economic performance of a farm, since organic farms have much higher input costs especially when it comes to feed (Andersen et al. 2016).

1.4 Research questions

The objective of this study is to analyse if there are any significant effects of the milk quality indicators and certain animal health indicators on technical efficiency among Danish dairy producers. By using an input and an output distance stochastic frontier approach it will be examined how both the milk quality and animal health indicators affect the technical efficiency of a farm. Further it will look at how these factors are related to technical efficiency along with other production characteristics such as the choice of breed and milking system.

The following research questions will be addressed:

1. Can the milk quality and level of animal health explain the differences in technical efficiency, which exist among Danish dairy producers?
2. To what extent can the production characteristics, such as the type of milking system and breed, describe the level of productivity obtained on the Danish dairy farms?
3. How can the specialised milk producers in Denmark improve technical efficiency in the future?

1.5 Scope and delimitation

The thesis aims to answer the research questions by analysing accounting data and cattle related data on full time specialised dairy farms in Denmark in the years 2011 to 2015. The dataset received from SEGES contained merged data from Økonomidatabasen (the accounting database) and Kvægdatabasen (the cattle database), sorted to enclose only farms with dairy production. A fulltime specialized dairy farm is in this thesis defined as one requiring at least 1665 norm hours on the farm per year, having two thirds or more of the gross margin stemming from dairy production, and having at least 49 cows in the herd. The final dataset is not considered to be representative of the whole sector, as the data was not collected using conditions of proper random sampling. Although the dataset is not representative, for the year 2015 it contains 1505 fulltime specialised dairy farms out of the total of 3,400 dairy farms in Denmark, corresponding to 44 percent.

The thesis will use panel data from two databases at SEGES and use econometric methods to analyse if and how much influence milk quality indicators and animal health indicators along with certain production characteristics have on the technical efficiency of the farm.
The study is limited to analysing primarily the effects of milk quality and health indicators and will not look in to the financial aspects of being a farmer such as the importance of investments, equity, profitability, and solvency. Farm level prices were not observed, and hence national level price indices were used to account for price changes in the analysed period, thereby assuming that all farmers have faced the same input and output prices.

1.6 Methodology

This thesis will use an econometric model that is based on economic theory of production economics. The model will be applied to a merged dataset of accounting and cattle data from the Danish Agricultural sector. All data manipulations and estimations will be conducted within the statistical software environment "RStudio". Basic statistics is presented through tables and graphs created in Excel.

1.7 Reader's Guide

The thesis is organized as follows: Chapter 1, 2 and 3 provide background knowledge of the Danish dairy sector and previous literature concerning the thesis objective. Chapter 4 and 5 introduce the theory and method to be used. Chapter 6 describes the datasets and decisions concerning the model. Chapter 7 presents the estimation procedure, while Chapter 8 presents the obtained results. Chapter 9 provides a discussion of the outcome of the investigation, whereas Chapter 10 concludes and Chapter 11 discusses possibilities for further investigations. In Chapter 12 all references are listed and Chapter 13 includes the appendices with extra tables, figures, and the data script.

2. Literature review

Extensive research has been made to investigate impacts on technical efficiency on farms. Often the social aspects such as the treatment of animals, has been left out in these analysis (Barnes et al. 2011). Only few studies are using real-world farm data. However, research focusing on animal welfare does exist and it often concentrates on animal welfare measures, using animal disease occurrences as measures (Lawson, Agger, et al. 2004; Barnes et al. 2011; Lawson, Bruun, et al. 2004). Often studies assessing the impact of production diseases on the economic performance have focused on evaluating the difference in milk output among diseased and non-diseased cows, the effect on reproductive performance or the effect of the treatment of the diseases (Lawson, Agger, et al. 2004). Animal welfare and health indicators are closely related, which is why we refer to literature regarding both aspects. However, it is important to emphasize that the expressions cannot be used interchangeably, since differences between the two terms exist.

The most commonly used methods in the literature regarding animal welfare and economic performance are the Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA) approaches. The DEA is a non-parametric approach, and therefore not as restrictive as the SFA. The DEA approach is more flexible because it does not require a functional form. Using the DEA, any possible noise is suppressed and any variation in the data is considered to be due to inefficiency. SFA is a parametric approach, and unlike the DEA, there should be made assumptions about the structure of the production possibility set and the data generation process. Even though the SFA requires a functional form, like a production function, which needs to fulfil certain restrictions, it allows us to assume that the deviations from the frontier can reflect inefficiencies and noise in the data (Bogetoft & Otto 2011).

Because diseases can affect both the quantity and the quality of the milk output and therefore also the economic performance of the farm, it can be argued that it is relevant to explain variations in management styles in relation to preventing and handling diseases. By looking at the whole farm, the farmer’s management abilities can be measured, by evaluating the level of technical efficiency (Lawson, Agger, et al. 2004).

This thesis will use the stochastic distance function approach to investigate changes in the Danish milk producers’ technical efficiency. This is a well-established method, which has been widely used to examine agricultural productivity. The stochastic frontier approach was presented by Meeusen & van Den Broeck (1977) and Aigner et al. (1977) independently, and distance functions were presented by Shephard (1970). The advantages from using the distance functions instead of a production function will be elaborated later.

Following these publications, many authors have used the approach to examine agricultural efficiency. An example of this is Newman & Matthews (2007), who used a stochastic output distance function to estimate the productivity growth of Irish agriculture for four farming systems. They find that variable costs and milk output have the largest elasticities for specialised milk producers. Brümmer et al. (2002) also used a stochastic output distance function to measure the productivity growth among German, Dutch, and Polish European dairy farms from 1991-1994, and find that efficiency has increased over the years, which is mainly due to technological improvements. They end up concluding that the results obtained can be used for further improvements in the dairy sector, since they have presented in more detail which areas of dairy production that can be adjusted.
Sipiläinen (2007) uses the distance function approach to estimate both an input and an output distance function for unbalanced panel data on Finnish farms specialising in milk production. The analysis finds that the orientation of the model is sensitive to explaining technical efficiency in relation to the size of the farm. Sipiläinen (2007) mainly focuses on the input distance function, given that Finland, by joining the European Union, had to adjust production according to the milk quota restriction. The results obtained show that there is a large elasticity of scale suggesting that Finnish farms can increase inefficiency by operating at a larger scale. Common for the studies is that the monotonicity assumption is violated for a small fraction of the elasticities, but it is fulfilled at the sample mean, and thus the estimated models hold.

Lawson, Bruun, et al. (2004) study the relationship between milk production efficiency and the incidence of reproductive disorders using a stochastic frontier production function approach to see whether the farmers reporting a higher number of incidences are less efficient. By using the stochastic frontier approach, they include the traditional behavioural assumption of farmers wanting to produce the maximal output from the available inputs. The results show that reproductive diseases do not have a negative effect on efficiency on Danish dairy farms. They argue that it is because managers compensate for the disorders by using early enrolment and replacement of cows (Lawson, Bruun, et al. 2004). Lawson, Agger, et al. (2004) investigate whether farms reporting more incidents of lameness and metabolic disorders are less technically efficient. Again, the stochastic frontier approach is used and it is found that milk producers reporting more treatments of the investigated diseases are the most efficient dairy farmers. It is found that the expected negative association between technical efficiency and the diseases is overshadowed by the productivity of some of the input variables.

Barnes et al. (2011) apply animal welfare as a discriminating technology within a technical efficiency framework. They investigate how lameness can be prevented by good management strategies by looking at 80 dairy farms across Great Britain, using observed data on inputs and outputs and collected lameness scores. Using the data envelopment analysis (DEA) approach they find that farms with low rates of lameness tend to have higher technical efficiencies than the farms with high rates of lameness. Furthermore, the study shows that low lameness rates are inefficient in terms of labour and stocking density but this factor is compensated by the higher milk yield obtained by these farms. Looking at the whole farm economy the results suggest giving more attention to the management of lameness, to increase the technical efficiency on the farm.

A theoretical framework for explaining the relationship between animal welfare and economic performance of livestock farms is proposed by Henningsen et al. (2016). They illustrate how each component in the production process is connected, and include animal welfare as being directly dependent on the production process. By incorporating animal welfare in the production process, they illustrate that animal welfare should not be ignored when optimizing. This theoretical framework will be discussed in the theory part of this thesis.

Overall it is clear that the well-being of the animals used in production affects the efficiency of the farm. Lawson, Agger, et al. (2004) find that the more efficient farms are those with the highest reporting of treatment. Barnes et al. (2011) also find that low rates of lameness are associated with greater efficiency. Common to all the studies is that they find that animal welfare and health is related to efficiency and hence the economic performance of the farm (Barnes et al. 2011). Further the studies conclude that management is very important to ensure a good economic result. The studies, do not discuss the implications of violating monotonicity, but are aware of the importance of fulfilling it at the mean. However, Henningsen & Henning (2009), who have invented a rather simple procedure to impose monotonicity, criticise authors for not imposing it, given the complications these violations might lead to .

The literature regarding animal welfare and health has not analysed the impact of various diseases and milk quality on efficiency over time, which justifies an attempt to estimate a model including these factors. Furthermore, according to our knowledge, there are no studies, which focus at the relationship between the milk quality and the technical efficiency of the farm. Thus, this thesis contributes to the literature by analysing the effects of milk quality on dairy farm efficiency.

3. Danish Dairy Production

3.1 Cow biology and production limitations

Milk production requires a lot of planning and cannot be adjusted from one day to the next. Before a heifer can start producing milk it must have its first calf. A calf is not ready for insemination before it is at least one year old. A gestation period lasts for about nine months and hence the heifer is approximately two years old, when it calves for the first time. Three months after the cow has calved it is ready for insemination again. In order to secure a good milk production after the cow has given birth again, the farmer stops milking the cow about two months before its due date. Shortly after the cow has given birth again, the milking is resumed. A typical Danish dairy cow only lives three to four years because its milk production is decreasing with age and hence it is more profitable to slaughter it at an early age and replace it with a younger cow (Landbrug og Fødevarer 2013).

Typically a cow is milked two or three times a day, depending on the feed and race. However, milking a cow three times a day requires more feed and labour input, and hence might not be optimal for all dairy farmers if the marginal costs associated with an increased production exceed the extra earnings (Martinussen & Sørensen 2015). In the short run certain production factors, such as the amount of land or the number of cows, are fixed for a dairy farmer. However, several factors related to the health of the animals can be adjusted in the short run. This includes milking routines and hygiene levels, which are dependent on the daily management. This thesis will investigate if those factors which can be quickly adjusted influence efficiency.

From this short overview it is clear that working with cattle is challenging. The production of milk has to be planned years in advance and cannot for the most part just be corrected on a daily basis. Working with live animals, the long run planning of production is even more difficult, because there can be more unforeseen challenges than with machines.

3.2 Health Indicators

The most common diseases in a herd are mastitis, hoof and limb disorders, and reproductive disorders. It is therefore these disorders that will be given focus in this thesis. The disorders will be used in general terms, as we do not have data concerning the different specifications of the diseases. The following knowledge of diseases and milk quality indicators was obtained during an interview with Veterinarian Peter Raundal from SEGES (Raundal 2016).

Mastitis occurs due to bacterial infections from environmental bacteria or cows being carrier of the bacteria. The clinical case lasts for 2-5 days and some end up being chronic, whereas others cause elevated cell count levels for a few weeks. If the occurrence is early in the lactation period, the cell counts can adjust to their normal stage rather fast, while it can take 2-3 weeks before the numbers are back to normal in other cases. Farms with larger herds often have lower treatment costs per cow, as they often have the knowledge to treat the animals, and therefore do not have the costs of a veterinarian. If the cow is given medication it is in detention for 8-10 days before the milk is acceptable to use again. The medication costs are around 150-250 DKK per cow for one treatment. While the cow is infected, it produces around 5-8 kg milk less a day.

Hoof and limb disorders are divided into skin related and horn related disorders, where the horn related can be more prolonged. It is very individual for the cow whether it ever gets a hoof disorder or if it has problems with it more often. The disorders can last for different periods of time depending on the individual cow and disease. The disease can last between 2 weeks and 7 months depending on when and how it is treated. The treatment takes approximately 2 weeks. Only a few are treated with antibiotics; it is only used related to infections and therefore the cow does not get any detention. Often the treatment is done with remedies as bandages, trimming of hoofs, and hoof shoes and therefore treatment costs vary, but the average cost per treatment is 100 DKK. Some farmers might choose to treat the animals themselves while others prefer assistance from a veterinarian. When a cow has a hoof disorder the milk yield is reduced by 300-800 kg milk for the whole lactation period.

Reproductive disorders are also very different from each other and include things as metritis, problems with the birth routes at calving, and lacking heat. Again, reproductive disorders demand very different treatments depending on the relevant case. As an example, does metritis create a loss of 2-3kg milk a day and it takes about a week for the treatment to work. If treated by a veterinarian the farmer has the cost of the vet visit, where the costs can vary, and the medication, which is around 150 DKK. If the cow has problems with lack of heat, it does not give direct production loss, but it is a management decision how many chances the cow is given.

3.3 Milk Quality Indicators

Cell counts, viable counts, and spores are all factors, which are important for the settlement of the milk prices given by the dairies. The cell count expresses the number of somatic cells in the milk, which is an expression of the health of the udder. Thus, high cell counts indicate that the cow is infected. The viable count is an expression of the number of bacteria in the milk, which is also related to the udder health and can help determine the level of hygiene at the farm. The third milk quality indicator, the spore count, is an indicator of the number of bacterial spores and is preferred to be low since a high spore count causes the milk to ferment (SEGES P/S n.d.). Hence, all three parameters express the level of bacteria in the milk and thus the quality.

In Denmark, the price, the farmer receives for his milk, depends on the average level of cells, viable counts, and spores in the milk he delivers. Since Arla is by far the biggest dairy in Denmark, receiving 90 percent of the milk produced in Denmark, the milk quality levels providing a supplement are assumed to be representative for the industry (Vidø et al. 2016). Generally, the farmers want to keep the numbers down to avoid the deductions described in column 5 of Table 3.1, which indicates that low milk quality indicators provide the best milk quality.

Table 3.1: Overview of Analysis, Limitations and Allowance & Deductions

Analysis	Analysis Frequency	Categories	Limitations	Allowance & Deductions, % of raw material value
Cell Counts, 1000 cells/ml	1 per delivery	1S	0-200	+ 2 %
		1E	201-300	+ 1 %
		1B	301-400	0 %
		2	401-500	- 4 %
		3	501-	- 10 %
Viable Counts, 1000 viable/ml	4-5 per month	1E	0-30	+ 1 %
		1B	31-50	0 %
		2	51-100	- 4 %
		3	101-	- 10 %
Spores, Spores/litre	2 per month	1E	0-400	+ 1 %
		1B	401-4000	0 %
		2	4001-	- 4 %

Source: (Arla Foods amba 2016)

Mastitis can, as mentioned, affect the cell count. In the case of mastitis, the cell count can increase to one million for one cow and it takes 4-5 weeks on average for the numbers to get back to normal. For some animals, the process is fast and for some it takes much longer. Some sorts of mastitis can also affect and increase the viable count.

The farmer can take certain precautions to ensure low and steady levels of cell counts, viable counts, and spores. Clean stables and clean animals can help prevent the spreading of bacteria, but most important is the handling of the animals during milking. The use of gloves is essential and thorough cleaning of the milking system is necessary to prevent bacteria from spreading.

If using a traditional milking system, there can be a risk with the large amount of handling of the animals. If one place is infected it can spread easily if the farm lacks high levels of cleaning. To minimize the risk of diseases spreading from one animal to the others, one should use milking gloves and make sure that there is a thorough cleaning of the teats.
With a robot milking system, the risks are different as there is less handling of animals, but instead bacteria can easily spread if the robot system is not maintained properly and cleaned daily. The farmer can through his daily management and technical decisions affect the cell count, the viable count, and the spore count by working with his milking procedures and hygiene.

On average the Danish milk producers had cell count values below 200,000 in 2015, which makes 2015 the best year for cell count numbers in history. Though there has been a tendency for the farmers to improve the cell count levels, there is still unused potential in this area, as the numbers varies between the farms, which indicates that some farms still have possibilities to improve the cell counts to obtain a higher supplement (Andersen et al. 2016).

Obtaining low milk quality indicators require attention to the hygiene in the stables, the milking routines, and the general health of the animals, which are all things that can increase the input use. Andersen et al. (2016) argue that attention should be made to lowering the cell counts, because the dairy sector was missing out on more than a 100 million DKK in supplements in 2015. They do not seem to reflect on the potential extra costs, which an improvement of the milk quality could require. If the extra costs of improving the hygiene in the stables or securing better milking routines exceed the gains of receiving a price supplement, then the farmer should refrain from doing it. The optimal level of the milk quality indicators can differ among the dairy producers.

It has been explained how dairy farming needs long term planning and that only some factors can be altered in the short run by the farmer. One area, in which the farmer can have an influence in the short run, is on how he manages diseases and thereby the health indicators of the herd. The most common diseases are mastitis, hoof and limb disorders and reproductive disorders and they can all be managed differently depending on how many resources the farmer spends on them. Some farmers choose to have the expenses of a veterinarian, while others foremost treat the animals themselves. Daily management can therefore also impact the levels of cell counts, viable counts, and the spore count. In his daily work the farmer is responsible of considering how many resources to be spend on managing health, for the farm to be as efficient as possible.

4. Theoretical Framework

This part of the study provides the theoretical and conceptual basis for the analysis. The section covers more general theory regarding production theory and the production function. We will also illustrate the role of management and the effect of animal health on the farmer's profit.

4.1 The Production Function

The production function is defined as the maximum amount of output which can be produced, given a certain technology with a certain amount of input (Rasmussen 2011).

One very basic assumption across all fields of economics is that there is a relationship between input and output in production. The basic production function is expressed as

$$Y(y, x) = 0$$

Here x represents a vector of non-negative inputs and y a vector of non-negative outputs, assuming the fact that production cannot take place without inputs and that output cannot be negative. Further restrictions are also imposed in the theory described by Chambers (1994). It is assumed that the function fulfils certain conditions, for instance: monotonicity, concavity, and essentiality.

The assumption of monotonicity in the production function is implemented because a rational decision maker would never increase the amount of input, if it would not lead to an increase in the output produced. This would only increase costs and hence not be optimal (Chambers 1994). If monotonicity is violated efficiency estimations become non-interpretable (Henningsen & Henning 2009).

Concavity is assumed because it ensures that the assumption of diminishing marginal returns is included (Chambers 1994). This condition states that when increasing the amount of input into a production, holding other inputs fixed, the additional returns generated will gradually diminish until they become negative. The marginal return expresses the extra output generated per input when the amount of input increases. If quasi-concavity is violated, the marginal rates of technical substitution are not decreasing and profit-maximizing behaviour is not reflected under microeconomic assumptions. However, multiple reasons exist why functions are not quasi-concave. It could be that inputs are not perfectly divisible, production activities are not almost independently applied, prices could be endogenous, and regulatory restrictions on inputs could exist (Henningsen & Henning 2009). Quasi-concavity requires the Hessian matrix of the second-order derivatives of output, with respect to the inputs, to be negative semi-definite. Henningsen & Henning (2009) suggest that one checks for quasi-concavity once the model is estimated; however, imposing it is not necessary.

Essentiality is the need for inputs to produce a product. Chambers (1994) looks at both weak essentiality and strict essentiality. Weak essentiality is when production of a positive output is impossible without at least one input, which is rather intuitive. Strict essentiality is when the production of a good cannot take place without a positive amount of a specific input. The input is then strictly essential to the production.

The Cobb-Douglas Production Function

There are several different production functions; however the Cobb-Douglas production function fulfils the restrictions that have been listed in the previous section, which is why we will explain the Cobb-Douglas production here (Chambers 1994). The choice of the functional form of the Cobb-Douglas production function can be discussed, but here the most general form of the function is presented:

$$f(x) = A\prod_{i=1}^{n} x_i^{\alpha_i}$$

Where $\alpha_i > 0$ and $i = 1,2, \dots, n$

The production function can be linearized using the natural logarithm:

$$\ln(y) = \alpha_0 + \sum_{i=1}^{N} \alpha_i \ln{(x_i)}$$

Where $\alpha_0 = \ln{(A)}$ (Henningsen 2014).

The Translog Production Function

A more flexible version of the Cobb-Douglas function is the translog production function. It was introduced by Berndt & Christensen (1973). It is a quadratic function and therefore an extension of the linear function. The translog function can thus be viewed as a combination of the Cobb-Douglas function and the quadratic function. The translog production function has the following specification:

$$\ln y = \alpha_0 + \sum_i \alpha_i \ln x_i + \frac{1}{2}\sum_i \sum_j \alpha_{ij} \ln x_i \ln x_j$$

Where $\alpha_{ij} = \alpha_{ji}$ (Henningsen 2014).

The translog function does not always fulfil the monotonicity assumption due to its quadratic nature. There will be a set of input quantities that provide a negative marginal product, but there are areas in the input space where all the conditions are satisfied. These areas might be large enough so that it is possible for the translog function to provide a relevant representation of the production possibilities (Berndt & Christensen 1973).

When deciding on which production function to use in the final estimation, a likelihood ratio test can be applied to compare the results obtained from the two production functions. This is possible because the Cobb-Douglas production function is "nested" in the translog production function (Henningsen 2014).

4.2 Input and Output Distance Functions

Distance functions allow us to describe a multi-input or multi-output production technology, without having to specify a behavioural objective such as cost minimisation or profit-maximisation, however specifying both may be an advantage (Coelli et al. 2005; Shephard 1970). Distance functions can be used to describe technology, by making it possible to measure efficiency and productivity, while accounting for multiple inputs and outputs without much aggregation. The concept of distance functions is closely related to production frontiers, as distance functions emphasize that the function is not only numbers. Procedures also map technologies and observations into real numbers (Bogetoft & Otto 2011). An input oriented distance function describes a production technology by looking at a minimal proportional contraction of the input vector, given an output vector. Contrary to this, an output distance function considers a maximal proportional expansion of the output vector, given an input vector (Coelli et al. 2005).

As we will account for more than one output in our model, we will use the multi-output generalization of a frontier production function: the Shepard distance function (Shephard 1970). In this thesis both the input and output distance approach will be estimated as the data set covers the period from 2011 to 2015, where a milk quota regulation until the 1 April 2015 restricted the farmers' output. We therefore assume that the milk quota restrictions called for an input oriented production in the period in the dataset from 2011 to 2014, while the cessation of the quota in 2015 would change the farmers' incentive towards an output oriented production. However, given that the farmers were well aware of the cessation of the quota years in advance, they might have planned to increase production. The Danish milk producers have to some extent exceeded their quotas, which could indicate that even with the quotas, the farmers have not kept output fixed every year (Wiese 2015).

The Input Distance Function

We start by considering the Shepard input distance function, where the output set, y, can be produced using the input vector, x. The input distance function includes, different from the output distance function, a scaling of the input vector. The Shephard input distance function is defined as (Bogetoft & Otto 2011):

$$D_I(x,y) = max\left\{D > 0 \mid \left(\frac{x}{D}, y\right) \in T\right\} = \frac{1}{E(x,y)}$$

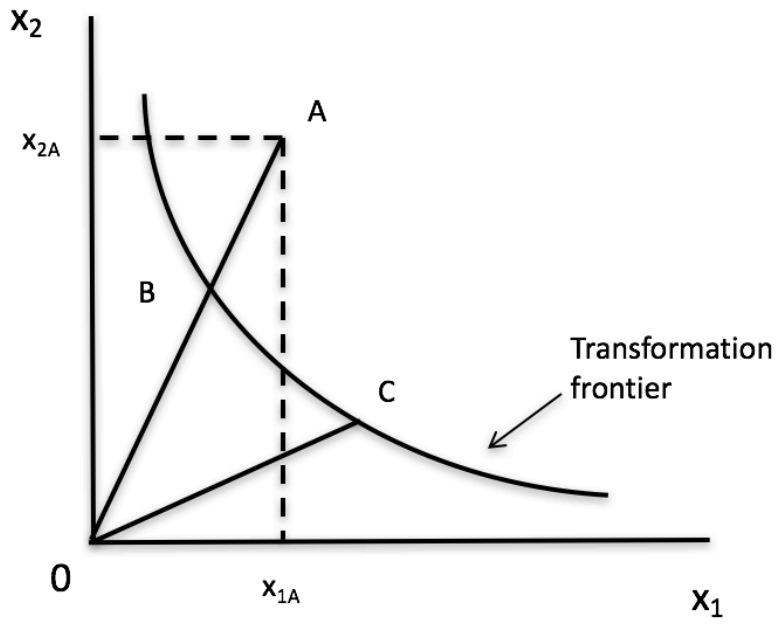
Where x is a vector of input quantities and y is a vector of output quantities, T is the technology set and $E(x,y)$ is the Farrell input efficiency (Bogetoft & Otto 2011).

The Farrell efficiency depends on the starting point (x,y) and the technology set T. The Shepard distance function is the inverse of the Farrell efficiency function (Bogetoft & Otto 2011).
In Coelli et al. (2005) it is stated that from the axioms on the technology set a few properties of $D_I(x,y)$ follow:

1. The input distance function is non-decreasing in x and non-increasing in y. This implies the assumption of monotonicity and therefore indicates that additional units of inputs can never decrease the level of output.
2. $D_I(x,y)$ is linearly homogeneous in x. The technology set is subject to constant returns to scale.
3. $D_I(x,y)$ is concave in x and quasi-concave in y.
4. If x belongs to the input set of y, then $D_I(x,y) \geq 1$.
5. The distance is equal to unity (i.e. $D_I(x,y) = 1$) if x belongs to the frontier of the input set (the isoquant of y). It follows from properties 4 and 5 that there is a potential radial expansion of the production up to the frontier of the production possibility set.

The input distance function can be illustrated using an example where two inputs, x_1 and x_2, are used to produce an output vector y. For a given output vector the production technology set is represented in Figure 4.1. The input set is the area bounded below the isoquant. For the point A, which defines the production point where firm A uses x_{1A} of input 1 and x_{2A} of input 2 to produce the output vector y, the value of the distance function is equal to the ratio $D_I = 0A/0B$.

Figure 4.1: Input distance function and input requirement set



Source: own drawings based on (Coelli et al. 2005)

The Output Distance Function

The Shepard output distance function can be defined as:

$$D_o(x, y) = min\left\{D > 0 \mid \left(x, \frac{y}{D}\right) \in T\right\} = \frac{1}{F(x, y)}$$

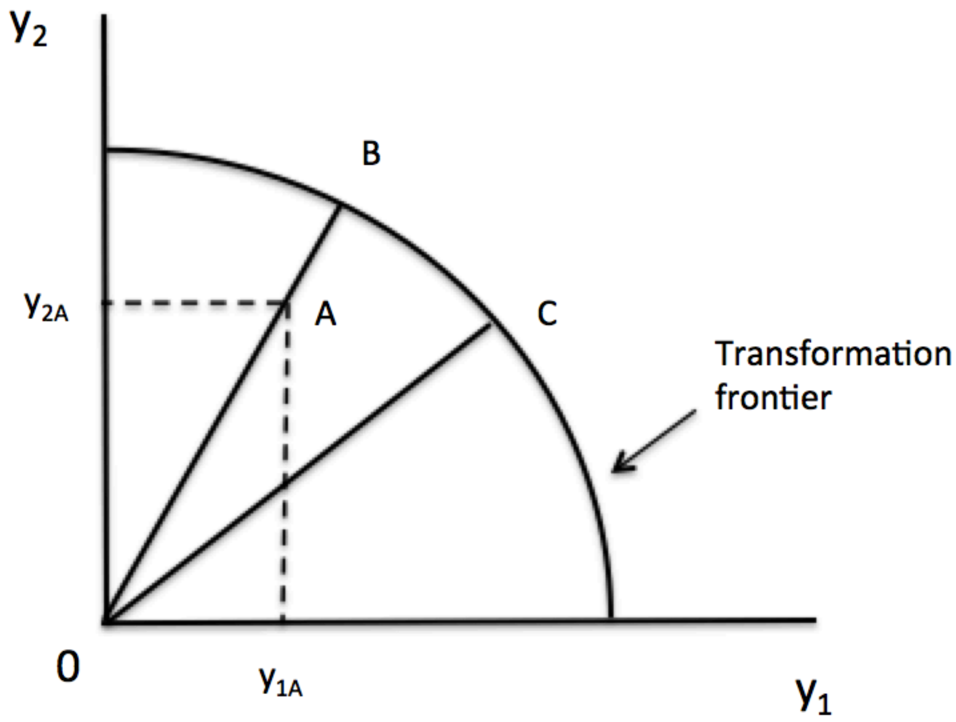
Where x is a vector of input quantities and y is a vector of output quantities. T is the technology set and $F(x, y)$ is the Farrell output efficiency (Bogetoft & Otto 2011).

Much like the properties for the input distance function, the properties for the output distance function can be listed (Coelli et al. 2005):

1. $D_o(x, 0) = 0$ for all non-negative x. This indicates that x cannot be non-negative for the technical efficiency to be different from zero.
2. $D_o(x, y)$ is non-decreasing in y and non-increasing in x. This implies the assumption of monotonicity and therefore indicates that additional units of inputs can never decrease the level of output.
3. $D_o(x, y)$ is linearly homogeneous in y, which follows from the distance function definition. The technology set is therefore subject to constant returns to scale.
4. $D_o(x, y)$ is quasi-convex in x and convex in y.
5. if y belongs to the production possibility set of x, then $D_o(x, y) \leq 1$
6. and distance is equal to unity (i.e. $D_o(x, y) = 1$) if y belongs to the frontier of the production possibility set. It then follows that properties 5 and 6 indicate the potential radial expansion of the production up to the frontier of the production possibility set.

The concept of an output distance function can be illustrated using an example where the outputs y_1 and y_2 are produced using the input vector x. In Figure 4.2 below, the production technology for a given input, x, is represented in a two-dimensional diagram. The technology possibility set, T, is the area bounded by the transformation frontier and the y_1 and y_2 axes. For the firm using input level x to produce the outputs defined by the point A, the value of the distance function is equal to the ratio $D_o = 0A/0B$. This distance measure is the reciprocal factor by which the production of all output quantities could be increased while remaining within the technical possibility set for the given input level. As the points B and C are on the production possibility surface they will have distance function values equal to one.

Figure 4.2: Output distance function and production possibility set



Source: own drawings based on (Coelli et al. 2005)

A few results can be stated for the connection of the input and output distance functions.
If y belongs to the production possibility set associated with input vector x, then x will belong to the input set associated with the output vector y.

If it is assumed that both inputs and outputs are weakly disposable it can be stated that:

$$D_I(x, y) \geq 1 \text{ if and only if } D_o(x, y) \leq 1$$

If the technology displays global constant returns to scale, it can be stated that:

$$D_I(x, y) = \frac{1}{D_o(x,y)}, \text{ for all } x \text{ and } y$$

Which indicates that, under constant returns to scale, the input distance function will be the reciprocal of the output distance function for any (x, y) (Coelli et al. 2005).

4.3 The influence of management on economic performance

In economic theory it is assumed that an individual always seeks to maximize profit. This means that the owner of a firm seeks to maximize output and minimize input while buying at a low cost and selling at a high price. Like any other theory there are exemptions. According to Varian (2010), business owners can have other objectives than to maximize profit. Typically, these owners will participate in the day-to-day operations to carry out their objectives for the firm. In agricultural firms it is common that there can be other objectives than just profit maximization due to the nature of the job. A farmer lives on his farm and hence his business is also his home. This means that he is more likely to have other motives for his business (Christensen et al. 1990). Another motive apart from profit maximization could be to have a healthy livestock, which means that the welfare of the animals on the farm might be more important than the costs of treating them.

There have already been conducted studies, which show the importance of coping with diseases and the general health of livestock to increase efficiency and the economic performance of a farm. Diseases can affect milk output in a number of ways, and it is therefore relevant to analyse how the farmers manage diseases (Lawson, Agger, et al. 2004). Mastitis, hoof problems, and the like are all management dependent, because the manager can prevent and treat such diseases by allocating resources and changing production inputs. When these disorders are present, it is a reflection of the farmer's prioritization of inputs and managerial abilities to choose the actions which can increase the health of the animals and thereby also their performance (Lawson, Agger, et al. 2004).

Henningsen et al. (2016) propose a theoretical framework, which can explain the relationship between animal welfare and the economic performance of livestock farms. Animal welfare is closely related to animal health, since the health of the animals contributes to their welfare. Further the milk quality is also affected by the health of the animals and welfare conditions, such as the cleanliness of stables. The levels of the milk quality indicators are sensitive to most diseases, and thus a high number of diseases will increase the cell counts, thereby lowering the milk quality. Management covers the managerial decisions such as the choice and quantity of input factors like the type of feed, veterinary products, work hours, and equipment.

Henningsen et al. (2016) consider how much a manager should invest in improving the welfare of his livestock. Increasing the welfare of his animals can result in better performing animals.

However, the relationship between economic performance and welfare is not linear and hence the welfare of the herd has a positive effect on economic performance until a point, where the resources spent on animal welfare is maximised. Beyond this point the extra resources spend on animal welfare do not generate more output, and hence would violate the assumption of monotonicity. It therefore becomes a question of daily management for the farmer to optimize his use of resources spent on animal welfare.

This thesis will, using a stochastic distance frontier, consider in which direction animal health and milk quality indicators, such as the number of reported diseases and values describing the udder health, influence the economic performance and technical efficiency of the farm.

5. Methodology

The idea behind benchmarking is to compare the relative performance of firms that use the same type of inputs to produce similar outputs. When benchmarking firms in the same industry, it is possible to learn more about how different production processes affect the technical efficiency (Bogetoft & Otto 2011). This chapter will add to the Theoretical Framework in section 4 by introducing how we will estimate an econometric model using the Maximum Likelihood Principle to estimate a distance function as a stochastic frontier model. The model and the underlying assumptions will be described in the following.

5.1 Stochastic Frontier Analysis

The Stochastic frontier analysis (SFA) is a parametric approach, meaning that there should be made assumptions about the functional form and the distribution of the composed error term. The SFA assumes that both the production possibility set and the data generation process are known in advance, which allows the assumption of a stochastic relationship between the inputs used and the outputs produced. Even though the assumption can be seen as a disadvantage, it allows us to assume that the deviations from the frontier can reflect inefficiencies and noise in the data (Aigner et al. 1977; Meeusen & van Den Broeck 1977).

The Parametric Approach

In the parametric approach, we use actual observations from different firms to estimate the production function and we then use the estimated function to measure the performance of the individual firms. We estimate a value for the unknown parameters β from the actual observations, (x^k, y^k), for $k = 1, \ldots, K$, where K is the number of individual observations. Maximum Likelihood Estimation (MLE) is most commonly used when estimating with a parametric approach, which means that we choose the values of $\hat{\beta}$ that makes the actual observations as likely as possible. The parametric approach suggests three options to explain why the actual observations deviate from the production function: 1) deviation can be viewed as noise as specified in an ordinary regression model, 2) deviation can be the result of inefficiency and 3) both explanations can be considered, which is the assumption behind a stochastic frontier approach (Bogetoft & Otto 2011).

The Ordinary Regression Model

When using a production function for empirical analysis, an ordinary regression technique (OLS) can be used to estimate the parameters of the average production function, and is often a good starting point for analysing data. The general specification is:

$$y^k = f(x^k; \beta) + v^k, \qquad v^k \sim iid\ N(0, \sigma^2), \qquad k = 1, \dots, K$$

All deviations from the frontier are interpreted as measurement noise such as measurement errors, omitted explanatory variables or unusual conditions, and is represented by the error term, v^k. The simplest way to estimate the regression is to assume that all deviations are symmetric around zero and follow a normal distribution. When using OLS to estimate the regression function the estimated function will lie in the middle of the observations, with observations above and below the estimated function, as the sum of the residuals is zero (Bogetoft & Otto 2011).

Deterministic Frontier Model

Instead of the OLS we can use a deterministic frontier model and assume that all the deviations are the result of inefficiency and use a model specified as:

$$y^k = f(x^k; \beta) - u^k, \qquad u^k \sim idd\ H, \qquad k = 1, \dots, K$$

Where H is a probability distribution of inefficiency.
It differs from the OLS, as this model assumes that there is no noise in the data and that a priori assumption is made regarding the functional form. This model can be an interesting starting point and is considered the other extreme together with OLS. Assuming no noise in the data and that the functional form is given a priori, it has the same drawbacks as the DEA without the flexible frontier specification (Bogetoft & Otto 2011).

Stochastic Frontier Model

Combining the ordinary regression model and the deterministic frontier model the stochastic frontier model (SFA) can be defined. SFA combines the efficiency term u with the error term v, and by doing this, the SFA model accounts for statistical noise, v, and technical efficiency, u (Aigner et al. 1977; Meeusen & van Den Broeck 1977). The SFA model is often estimated using a maximum likelihood estimation, which requires distributional assumptions of the error terms. The general model is presented here:

$$y^k = f(x^k;\beta) + v^k - u^k,$$
$$v^k \sim N(0,\sigma_v^2), \qquad u^k \sim N_+(\mu,\sigma_u^2), \qquad k = 1,\dots,K$$

Where $v^k \sim N(0,\sigma_v^2)$ is a random noise term that follows a normal distribution with zero mean and variance σ_v^2 and represents the possible measurement errors of the inputs and outputs. $u^k \sim N_+(\mu,\sigma_u^2)$ is an unobserved non-negative term, which accounts for technical inefficiency and follows a truncated normal distribution with location parameter μ and scale parameter σ_u^2. For $u^k = 0$ the firm is considered 100% efficient. If, however, $u^k \neq 0$, the firm is subject to inefficiency, leading to a left skewed distribution of the combined error term. N_+ indicate a half-normal positive distribution, which is a truncated normal distribution, where the point of truncation is 0 and the distribution is concentrated on the half-interval $[0,\infty[$ (Bogetoft & Otto 2011).

To make the estimation we need to know the density of the combined error term, ϵ

$$\epsilon = v - u$$

If we have a situation where v dominates u, meaning a situation where the variance of v, σ_v^2, is much larger than the scale parameter of u, σ_u^2, then the distribution of the combined error term, ϵ, will resemblance a normal distribution and will thereby look like the distribution of v. We could also imagine a situation where u dominates v. This would make the distribution of ϵ look like the distribution of u, which, as mentioned earlier, is a truncated normal distribution.

When estimating the stochastic frontier, the estimation algorithm re-parameterizes the variance parameter of σ_v^2 and the scale parameter of the inefficiency term σ_u^2, and instead it estimates the parameter γ (Battese & Corra 1977):

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

γ lies between 0 and 1 and gives an indication of the importance of the inefficiency term. When $\gamma = 0$, the inefficiency term, u, is irrelevant and the result should be equal to the OLS results.

When $\gamma = 1$, the noise term, v, is irrelevant and all deviations from the frontier are due to technical inefficiency.

Maximum Likelihood Estimation

The SFA method does not fulfil the minimal extrapolation principle. The principle states that the technology set should be the smallest set containing all data, but with the SFA method not all the data points will end up being below the estimated SFA line and the principle will not be fulfilled. This is the consequence for handling uncertainty in the model. Instead the model builds on an argument saying that data outside the derived technology set, are outside by pure chance and therefore it will not influence the way the technology set looks (Bogetoft & Otto 2011).

When estimating SFA models our interest is in the unknown parameters β and u. We want to decide on a value for the $\hat{\beta}$ that is as close to the true value of the unknown β as possible, i.e. parameter values are determined to make observations as likely as possible. The estimation software finds the parameter values by solving the maximization problem of the log-likelihood function. In large samples, maximum likelihood estimates are nearly unbiased, consistent, and efficient, thus they are close to the true value of the parameter and have variances close to the smallest possible variance. Furthermore, the estimated β, $\hat{\beta}$, is normally distributed (Bogetoft & Otto 2011).

The following log-likelihood function depends on the parameters to be estimated and the specific data (Bogetoft & Otto 2011):

$$l(\beta, \sigma^2, \lambda) = -\frac{1}{2} K \log\left(\frac{\pi}{2}\right) - \frac{1}{2} K \log \sigma^2 + \sum_{k=1}^{K} \log \Phi\left(\frac{\lambda\left(y^k - f(x^k;\beta)\right)}{\sqrt{\sigma^2}}\right)$$
$$-\frac{1}{2\sigma^2} \sum_{k=1}^{K} \left(y^k - f(x^k;\beta)\right)^2$$

The log-likelihood estimation is more convenient to work with than the likelihood estimation, since the natural logarithm makes the differentiation easier.

5.2 The Input Output Distance Function

There are several ways in which economic performance in a sector can be measured and evaluated. If information on prices is available along with a behavioural assumption, such as cost minimisation or profit maximisation, then performance measures can be estimated by including this information. The standard SFA models only allow for one output in the production functions, and thus they can be difficult to apply on real life data, where multiple outputs often exist. There are two solutions to this problem (Bogetoft & Otto 2011). One is to use the multiple input and multiple output cost frontier, however, given that individual prices are not available in our dataset, estimating a stochastic cost frontier is not possible (Coelli et al. 2005). The second is to measure efficiency by using distance functions. Distance functions can be used to estimate the characteristics of multiple-output production technologies when price information is not available or when one can, due to regulation, assume if firms minimize costs or maximize outputs. Input distance functions tend to be used instead of output distance functions when firms have more control over inputs than outputs (Bogetoft & Otto 2011). The choice to use the stochastic distance function approach was made because the distance function can be a good measure of inefficiency when we have multiple inputs and outputs and no prices.

A fully efficient farm will have an output distance function of $D_o(x,y) = 1$ or an input distance function $D_I(x,y) = 1$, which says that for an increasing D_o it makes a better performing (x,y) and for a decreasing $D_I(x,y)$ it makes a better performing (x,y). An inefficient farm has either $D_o(x,y) < 1$ or $D_I(x,y) > 1$. Distance functions can, in addition to describing the efficiency level, also describe the technology set, which can be defined as:

$$T = \{(x,y) \in \mathbb{R}_+^m \times \mathbb{R}_+^n | D_o(x,y) \leq 1\}$$
$$T = \{(x,y) \in \mathbb{R}_+^m \times \mathbb{R}_+^n | D_I(x,y) \geq 1\}$$

We can introduce the measure of inefficiency u into the output distance function, such that $u \leq 0$ and:

$$D_o(x,y) = e^u$$

From this it follows that $D_o(x,y) = e^u = 1$, when $u = 0$ and $D_o(x,y) = e^u < 1$, when $u < 0$.

And introduce $u \geq 0$ into the input distance function such that:

$$D_I(x,y) = e^u$$

It follows that $D_I(x,y) = e^u = 1$, when $u = 0$ and $D_I(x,y) = e^u > 1$, when $u > 0$.

Taking logs and due to the homogeneity condition, as stated in the properties, the distance functions can be turned into stochastic distance functions by also adding a random error term v (Bogetoft & Otto 2011):

$$\log y_M = -\log D_o\left(x, \frac{y_m}{y_M}\right) + v - u \qquad \text{and} \qquad -\log x_N = \log D_I\left(\frac{x_k}{x_N}, y\right) + v - u$$

As in the previous stochastic model we assume that v and u are independent and normally distributed, where u is only half-normal. Because this model has the same form as the models explained earlier we can use the same methods of estimation.

The Input Distance Function

The first step in econometric estimation of an input distance function is to choose a functional form. The ones that will be considered in this thesis are the two listed in the theoretical framework.

Assuming a Cobb-Douglas production function, the linearized general form of the Cobb-Douglas input distance function with multiple outputs can be written as:

$$lnD_I(x, y) = \alpha_0 + \sum_{m=1}^{M} \alpha_m lny_m + \sum_{k=1}^{N} \beta_k lnx_k$$

With $\alpha_0 = lnA$ (Coelli et al. 2003).

The Cobb-Douglas input distance function can be expressed as a stochastic frontier model, as linear homogeneity in inputs suggests that:

$$D_I(kx, y) = kD_I(x, y)$$

When exploiting that $\ln(D_I(x, y)) \geq 0$ by u (with $u \sim N_+(\mu, \sigma_u^2)$) and adding an error term v (with $v \sim N(0, \sigma_v^2)$) accounting for statistical noise, the following Cobb-Douglas input distance function can be estimated as a stochastic frontier model (Coelli et al. 2003):

$$-ln(x_N) = \alpha_0 + \sum_{m=1}^{M} \alpha_m lny_m + \sum_{k=1}^{N-1} \beta_k \, ln\left(\frac{x_k}{x_N}\right) + v - u$$

Since the Cobb-Douglas model is rather restrictive, the translog input distance should also be estimated and compared to the Cobb-Douglas model. Färe & Vardanyan (2016) present the general form of the translog input distance function as:

$$lnD_I(x,y) = \alpha_0 + \sum_{m=1}^{M} \alpha_m \, lny_m + \frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{M} \alpha_{mn} \, ln \, y_m lny_n + \sum_{k=1}^{N} \beta_k lnx_k + \frac{1}{2}\sum_{k=1}^{N}\sum_{j=1}^{N} \beta_{kj} lnx_k lnx_j + \sum_{k=1}^{N}\sum_{m=1}^{M} \zeta_{km} lnx_k lny_m$$

With $\alpha_{mn} = \alpha_{nm} \; \forall \, n,m$ and with $\beta_{kj} = \beta_{jk} \; \forall \, k,j$.

Like the Cobb-Douglas input distance function, the translog input distance function can, assuming linear homogeneity, be estimated as a stochastic frontier model.
Dividing all the inputs by one of the inputs imposes linear homogeneity in inputs. Defining $ln\big(D_I(x,y)\big) \geq 0$ as u (with $u \sim N_+(\mu, \sigma_u^2)$) and adding an error term v (with $v \sim N(0, \sigma_v^2)$), the stochastic translog input distance model can be written as (Sipiläinen 2007):

$$-lnx_N = \alpha_0 + \sum_{m=1}^{M} \alpha_m lny_m + \frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{M} \alpha_{mn} lny_m lny_n + \sum_{k=1}^{N-1} \beta_k ln\frac{x_k}{x_N} + \frac{1}{2}\sum_{k=1}^{N-1}\sum_{j=1}^{N-1} \beta_{kj} ln\frac{x_k}{x_N} ln\frac{x_j}{x_N} + \sum_{k=1}^{N-1}\sum_{m=1}^{M} \zeta_{km} ln\frac{x_k}{x_N} lny_m + v - u$$

The Output Distance Function

As the milk quota restriction was revoked in April 2015, most data for the last year in the dataset is from a time without an output restriction and hence we will also estimate a stochastic output distance function. Like with the input distance functions the output distance functions can be estimated as a stochastic frontier model by adding u (with $u \sim N_+(\mu, \sigma_u^2)$) and an error term v (with $v \sim N(0, \sigma_v^2)$) accounting for statistical noise. This results in the following specification of the Cobb-Douglas output distance function (Henningsen et al. 2016):

$$-lny_M = \alpha_0 + \sum_{m=1}^{M-1} \alpha_m ln\frac{y_m}{y_M} + \sum_{k=1}^{N} \beta_k lnx_k + v + u$$

Following the same approach as before the translog output distance function can be specified as (Olsen & Henningsen 2011):

$$-lny_M = \alpha_0 + \sum_{m=1}^{M-1} \alpha_m ln\frac{y_m}{y_M} + \frac{1}{2}\sum_{m=1}^{M-1}\sum_{n=1}^{M-1} \alpha_{mn}\, ln\frac{y_m}{y_M} ln\frac{y_n}{y_M}$$

$$+ \sum_{k=1}^{N} \beta_k lnx_k + \frac{1}{2}\sum_{k=1}^{N}\sum_{j=1}^{N} \beta_{kj} lnx_k lnx_j + \sum_{k=1}^{N}\sum_{m=1}^{M-1} \zeta_{km} lnx_k \frac{lny_m}{lny_M} + v + u$$

6. Description of Dataset

The data used in this thesis is provided by the Danish advisory firm SEGES. The dataset used in this thesis is based on the accounting data drawn from a large proportion of all commercial farms in Denmark. We were granted access to the data for all the farms in their database for which cattle related data was also available. The two datasets were merged by the specific observations “lbnr” and “Regnskabsaar” identifying each specific farm and year. The first database contains accounting data from 2011 to 2015 with 770 variables. The second database contains cattle related data also from 2011-2015 with 33 variables. In the combined unbalanced dataset, there are 3831 different farms and 14,323 observations.

The farms registered at the accounting database at SEGES have voluntarily chosen to provide SEGES with their financial data. Because it is voluntary and takes a bit of work for the farmers to be registered, there could be a selection bias among the participants. However, it is rather random who choose to be registered and maintain their registration, which is why we are not concerned with a selection bias in the unbalanced dataset. Unlike for the accounting database it is mandatory to register death and birth of cows to the cattle database. Both the health indicators and the milk quality indicators, which will be used in this thesis, are not mandatory and this can explain why the availability of these indicators can differ from farm to farm. Like with the accounting database, the information, which will be used from the cattle database in this thesis, could be subject to a selection bias.

The choice to use the cattle database from SEGES was made for us to be able to analyse the correlation between technical efficiency and milk quality and animal health indicators. However, the only information, which is mandatory for a farmer to report to the cattle database, is limited to the birth and death of an animal and hence all other factors are only available if the farmer wishes to report it. For the milk quality indicators, the farmers know the values of each of them, since the dairies measure the values whenever they collect the milk from the farmer, and hence reporting them only requires little time.

6.1 Variables

Given that the objective of this thesis is to analyse if milk quality and animal health indicators influence the efficiency of the farm, not all the available variables are relevant to this study. Only those factors, which are linked to the financial performance, milk quality, and the health of the livestock, will be included. A milk production equation is created where output is explained by several input variables, and further an inefficiency equation is created containing those factors which are related to (in)efficiency. The following variables are defined as to best describe the important input and output variables used in diary production. Further certain variables related to the technology set are defined. Given that we have access to the type of milking system used by the farmers, it is added to our model, because it can influence how quickly an infection can spread within a herd. However, the validity of the milking system variable can be questioned, since information regarding the type milking system is not always up to date in the database. The chosen variables for the analysis are listed below in Table 6.1.

Table 6.1: List of Variables

The milk production variables	
"grossmilk"	the gross output from milk production in DKK.
"grossother"	the gross output from all other products from production in DKK minus internally produced feed.
"feedexp"	the total expenditures in DKK for the farm to externally bought feed.
"vetmed"	the total expenditures in DKK for the farm to veterinaries, medicine, and vaccines.
"labour"	labour input as wages to hired labour and owner remuneration in DKK (listed "totalwages" in the data script).
"hours"	alternative labour input. Norm hours i.e. estimated total number of hours worked at the farm.
"land"	the total arable land owned and rented measured in hectares.
"materials"	the materials used in the production in DKK less feed expenditures, depreciation, maintenance, medicine, veterinary, wages, and land taxes.
"capital"	the capital input measured in DKK. Is contains the capital stock times an interest rate, the value of the cows, general maintenance, and depreciation.
Production characteristics (part of the technology set)	
"jersey"	a dummy for the breed Jersey.
"large"	a dummy for the large breed which contains RDM, SDM, mix of the two, and other types.
"AMS"	a dummy for the milking system AMS (Automated Milking System).
"fishbone"	a dummy for the milking system fishbone.
"othersys"	a dummy for other types of milking systems.
"organic"	a dummy for organic farms.
The inefficiency variables*	
"mastitis"	the number of mastitis disorders per cow per farm.
"hoofdis"	the number of hoof and limb disorders per cow per farm.
"reprodis"	the number of reproductive disorders per cow per farm.
"otherdis"	the number of other diseases per cow per farm.
"cell1"	a dummy for the farms which have an average cell count between 0-300 (1000 cells/ml)
"cell2"	a dummy for the farms which have an average cell count above 300 (1000 cells/ml)
"viable1"	a dummy for the farms which have an average viable count between 0-30 (1000 viable/ml)
"viable2"	a dummy for the farms which have an average viable count above 30 (1000 viable/ml)
"spore1"	a dummy for the farms which have an average spore count between 0-400 (spores/litre)
"spore2"	a dummy for the farms which have an average spore count above 400 (spores/litre)
"managerage"	the age of the manager of the farm.
"consultant"	a dummy for the use of a production consultant on the farm.

Note: * For the dummies: TRUE = 1 and FALSE = 0

Source: (SEGES 2013), (SEGES 2016), and own definitions and calculations

There are two outputs from production in our analysis: milk output and other outputs. Other outputs cover all other goods produced such as meat and cereal.

The feed expenditures are included separately as an input factor, given that we are analysing farms using livestock in production. It covers all costs to externally bought feed and hence does not include internally produced feed. As we do not include any feed expenditures to internally bought feed, other output is less income from feed sold. As most of the feed produced on specialised dairy farms is used on the farm, any income from feed sold is assumed not to be from outside buyers. As we do not account for internally bought feed on the input side, the income from this post is also taken out of the output. Those farms that do not buy feed from external producers, are more land, labour, and capital dependent, and thus will not appear more efficient.

The expenditures for veterinaries, medicine, and vaccines are included as a separate input factor. Because this thesis is concerned with the effect of animal welfare and technical efficiency it is relevant to look at an input factor directly linked to health.

There are two different measures of labour input, which can be used: “labour” and “hours”. The estimated norm hours used on the farm are listed in the accounting database, by using a standard assumption of how many hours of work are needed for certain amounts of livestock and land. It is based on production equipment and farm size, which is why it might not reflect the actual hours spent working on the farm. The use of the variable “hours” in the estimation, removes the variation in labour productivity between the farms. This indicates the fact that, using norm “hours” instead of “labour” would mean that no productivity difference in labour usage will be included in the model, which is a rather strong assumption.

The variable “labour” needs to be constructed using information from the dataset. Wages for hired labour are known in the dataset, so the problem is to determine the effort placed by the owner and his/her immediate family. The result before financing for personally owned businesses presented in the dataset, does not cover wages to the owner and his family. To compare results, it is necessary to construct an owner remuneration. If an owner remuneration is not assigned, then some farms can appear to be more efficient than others, due to the lower labour costs. The critical assumptions behind the construction of the labour input are that the presence of paid labour should lead to a bonus for the owner due to leadership. Income from an outside job most likely implies a lower effort on the farm by the owner.

The owner of a farm gets 300,000 DKK in owner remuneration. This is less any income he might have from an outside job and hence the wage represents his work on the farm. Further, the owner wage, which is constructed, assumes that for every hour spent working on the farm exceeding 1665 hours there will be a supplement to the owner, which increases with the number of employees as it would if the job as a leader was done in another company. The supplement is approximately 25,000 DKK per employee and the maximum supplement the owner can receive is 450,000 DKK. The maximum remuneration for the owner is therefore limited to 750,000 DKK.

If present, the spouse should be paid if the income from a job outside the farm is small or non-existing. There is believed to be a spouse, if there is family allowance, earnings from spouse or the family's manpower needs exceed 3,000 hours. If the spouse is working (part time) outside the farm, with an income below 300,000 DKK, then the spouse is getting paid for the work expectedly being done on the farm. The maximum remuneration for the spouse is 300,000 DKK. The composed owner remuneration was guided by Olsen (2016). The choice of labour input to use will be discussed later.

The input of land can either be self-owned or leased/rented. The two types are combined to create the variable containing the quantity of land used in the production measured in hectares. Combining the two types ensures that all land input is considered, whether it is used for production or leasing. This way the farmer will not be considered overly efficient if there is an income from land not used in production (SEGES 2013).

Materials cover all other intermediate inputs used in production. It is included because those farms buying feed do not use land, labour, and capital to produce the feed used in production and hence would appear to be more efficient. The variable covers variable costs such as seed, fertilizer, and pesticides.

The capital input or consumption includes the capital stock times an interest rate + depreciation + maintenance. From the dataset, these variables can easily be deducted. Capital stock is defined as the value of all agricultural assets minus value of land and housing. The capital stock is the value of the machinery, livestock, inventory, and agricultural buildings. Depreciation and maintenance is the sum of depreciation and maintenance on these assets plus investments that are fully depreciated in the specific year. The interest rate used in this thesis is an average bank interest rate faced by private owned businesses in from October 2013 to December 2015 based on numbers available at Denmark's Statistics (Statistik 2016c). The choice of capital valuation is based on the one used in Olsen & Henningsen (2011). The capital stock is the mean value of the capital stock in the beginning of the year and at the end of the year in order not to make farmers with investments either in the beginning or at the end of the year appear inefficient.

The variables chosen for the inefficiency equation can all be of importance for the gross output for the farm. As already mentioned several studies have found that animal health is correlated with efficiency and economic performance and hence we wish to incorporate animal health indicators and other confounding variables. Given that the price the farmer receives for his milk depends on the level of milk quality indicators, there could be an incentive not to reduce the values of these variables once a certain level is reached. To better explain the effects of low values with regard to the milk quality indicators two dummies for each type of milk quality indicator are created. The first dummy, e.g. "cell1", contains all those observations for which an allowance is obtained and the second dummy, e.g. "cell2", contains all those observations that do not obtain a higher price along with those who face a deduction. The levels and the corresponding allowances and deductions are shown in Table 3.1.

The production characteristics are part of the technology set used at the farm, and can help explain which type of technology improves productivity. Being organic can influence the choice of inputs and technology, however from Table A.1 in Appendix A it is clear that the production characteristics do not differ much among the organic and conventional farms used in the final dataset.

The age of the manager is included, since it is an indicator for management experience, and thereby management skills. The use of a production consultant is included to see if external knowledge can help the farm to become more efficient. Given that the amount a farm spends on a production consultant is most likely related to the size of the farm, it was decided to include the use of a production consultant as a dummy.

6.2 Imposing restrictions on the dataset

We wish to impose certain restrictions, which allow us to exclude those farms that are not classified as dairy units. The FADN (Farm Accountancy Data Network) defines a specialised dairy farm as one having at least 35 percent of the total output from milk production (European Commission - EU FADN 2014). In this thesis, a specialized dairy farm, is defined as a one having at least two thirds of standardized gross output from dairy production. This decision was made due to the fact that it is a general assumption in the literature analysing the production efficiency among full time dairy farmers (Sipiläinen 2007; Lawson, Bruun, et al. 2004).

Table 6.2 shows the individual restrictions for the variables and the number of observations violating each restriction. The first four rows contain restrictions related to being a full time dairy farmer. Dividing the restrictions into two groups allowed us to see how many violations were caused due to not being a full time dairy farmer and how many were due to negative or zero input values. It should be kept in mind that the same farm could possibly violate a restriction up to five times in this dataset (because we are working with panel data) until we impose the restraint of each farm having to be present in the dataset for at least three years (lbnr $\geq$ 3), meaning that each individual violation does not necessarily represent the number of farms which violate a restriction.

Table 6.2: Restrictions on the dataset

Restriction	Explanation	No. of violations
grossmilkshare > 0.66	At least 66 percent of the total gross output is from milk production.	3490
grossmilkshare ≤ 1	The gross output from milk cannot exceed the total gross output.	79
hours ≥ 1665	Alternative labour input. Norm hours i.e. estimated total number of hours worked at the farm. A full-time farm has to spend at least 1665 hours on the farm per year.	1287
yearcows > 49	A full time dairy farm cannot have less than 49 cows, corrected for incoming and outgoing cows during the year.	689
Total number of violations*	The total number of observations, which were excluded due to violations of the restrictions related to being a full-time dairy farmer.	4873
grossother > 0	The output from other goods produced should be positive.	81
feedexp > 0	The input of feed should be positive.	0
vetmed > 0	Expenditures for veterinaries and medicine should be positive.	5
totalwages > 0	Wages and owner remuneration combined should be positive.	0
land > 0	The input of land should be positive.	79
materials > 0	The material input should be positive.	2
capital > 0	The input of capital should be positive.	3
Total number of violations*	The total number of observations, which were excluded due to violations of the restrictions related to having non-negative inputs.	89
lbnr ≥ 3	Number of years with data per farm should be at least 3 years when all the above conditions are also fulfilled.	2021

Note: * The restrictions did not exclude any farms, only observations.

Source: Own calculations and definitions

The different restrictions limited the dataset to only have positive input and output quantities and only full time dairy farms with three or more observations during the five-year period from 2011 to 2015. The restrictions were imposed to remove non-meaningful observations and ensure strict essentiality. Further we had to exclude observations with no information regarding the cell, viable, and spore count as well as invalid values for the age of the manager. We end up having 8198 observations for 1810 different farms in the dataset distributed over the years, as can be seen from Table 6.3.

Table 6.3: Number of farms per year

Year	2011	2012	2013	2014	2015
No. of farms	1621	1615	1750	1707	1505

Source: Own calculations based on dataset

The distribution of the number of farms over the years can have been impacted by the price of milk, as we are using realized gross output. As seen in Figure 1.3 there was an increase in the price in 2013 and 2014 and a decrease in the price in 2015, which could partly explain the increase in number of farms in 2013 and 2014 and the decrease in number of farms in 2015. The price changes might have caused the gross output from milk to be less than two-thirds of the total output on some farms, in the years where the price on milk was low. How to handle price changes will be elaborated later.

Some of the restrictions imposed on the data do not need an elaboration. All the input factors should be positive to fulfil the theoretical assumptions made in section 4. Furthermore, negative values are not valid and hence would bias the obtained results, partly because the variables used in the estimations will be logged.

As already mentioned it is historically a standard assumption in the literature that a dairy farm should have at least two-thirds of the total gross output from milk, to be considered a full-time dairy farm. We use this assumption as we wish to focus on specialized dairy farms. 79 observations had a gross output from milk, which exceeded the total gross output. This should not be possible since the total gross output consists of the gross output from milk amongst other, and hence these farms were excluded. To ensure that only full-time dairy farms were included, the hours spent working on the farm should be at least 1665 norm hours a year. This restriction alone did not exclude those farms with only a few cows, which is why the farms included where restricted to those with at least 49 cows per year controlled for incoming and outgoing cows in the herd.

We have assumed that the gross output from other goods produced has to be positive. Since none of the farms in the dataset had a gross output from milk that was equal to the total gross output, this assumption is reasonable.

The input of medicine and veterinary services might not need to positive. However, it is rather unusual that a farmer does not have any expenses related to the health of his animals and given that we will use the natural logarithm to estimate our input and output distance functions, we need to exclude values of zero and less. The manager of the farm will always pay for the treatment of an animal if it is beneficial to the production, and hence those farms with higher expenditures to veterinarians and medicine are not necessarily less efficient than those with low expenditures. There are several ways in which the farm can choose to handle an infection and there can also be great variations in how long a cow is subject to an untreated infection (Raundal 2016). Because there are great differences among the treatment procedures and the need for outside assistance by a veterinarian, it cannot be assumed that the less efficient farms are also those with higher expenditures to veterinarians and medicine.

In the dataset, there were 439 observations for the age of the manager of the farm, which were listed as zero, and 12 observations, which were unrealistically high. Observations with a value smaller than 18 or greater than 90 were excluded to secure a true age distribution within the dataset. Since the estimation cannot be done with NA values, the choice of excluding unrealistic values had to be made. The value of zero is related to partnerships and limited liabilities companies, which will therefore not be included in the analysis.

One very important restriction is that there should be data for each farm for at least three years. Given that we wish to estimate an effect of the health indicators over time, we need several observations for each farm.

Overall, it is clear that the strictest restrictions are those requiring that the farms are full time dairy farms. It is unfortunate that these restrictions are excluding a rather large share of the farms in the original dataset. However, part-time dairy farms are not of interest because they might not be subject to the same problems and challenges or have the same resources as those who are producing milk full-time.

For all three milk quality indicators, the dataset had several missing observations, which could indicate that some farmers do not find it necessary to prioritize reporting these numbers to SEGES. As already mentioned, these observations were removed. Unlike with the milk quality indicators, the requirement of being a full time dairy farmer removed all observations listed as NA for all four disorders from the dataset. However, for several observations, there were values of zero reported cases of disorders. These observations were not removed, since they might not be false, however they could reflect that some farmers tend to report a value of zero instead of NA whenever they do not have the correct information available. This could affect how much influence a disorder has on efficiency in the estimation. Knowing that the variables we will use from the cattle database are reported voluntarily, a selection bias might exist.

From Table A.2 in Appendix A we see that for all three milk quality indicators, more observations in the dataset are in the first groups, indicating that overall the farms in the dataset have good values for the milk quality indicators, which is the general tendency among the Danish dairy farmers (Landbrug og Fødevarer 2016). Figure A.1 in Appendix A shows that the reported number of diseases per cow is rather low and Table A.3 also shows the percentage of zero reported cases of disorders. The potential lack of consistent data and selection bias will be discussed later.

6.3 Prices over time

Working with panel data we need to control for price changes over the years. To incorporate a price index, the price development of inputs and outputs are assumed to be identical for all farmers since input quantities are not known (except for land). Even though it is a common assumption within economics, it can be misleading since the differences in the economic performance among the farmers in the dairy sector could be due to different input prices over the years. However, it is a necessity to be able to handle inflation.

Following Rasmussen (2010), the Törnqvist index will be used to describe price changes over time. The Törnqvist index allows for more sector specific prices, which helps provide a more precise estimate of the price development experienced by the Danish dairy farmers compared to the general consumer price index. Since a few of the input variables and the output variable for other output contain several different elements, the problem is to determine the different types within the variables and their shares.

Other output contains grains and cattle related products other than milk but not internally produced fodder. By computing the share of each type of output for the whole sample we find the average share of each type of output out of the total. The shares have been used as allocation keys as to how much the price development of each good should influence the total price index for other outputs. The different output elements and the price index for each output is listed in

Table A.4 in Appendix A. The same approach was used to determine the different input types in the input variables “feedexp” and “materials”. For the inputs the different elements and the price index for each type of input is listed in Table A.5 in Appendix A.

We have chosen to use the price index for the agricultural sector available at Denmark’s Statistics (Statistik 2016f) for all outputs and inputs except for total wages. The choice was based on a wish to use the same database for all the goods, to ensure that that all prices have been subject to the same calculation methods. Due to the lack of a price index for wages, we used the nominal wages from 2010 to 2015 (Statistik 2017b) and converted each wage rate into 2010-level using a price converted from Denmark’s Statistics (Statistik 2017a) and calculated the development for each year in relation to 2010.

The prices were converted into 2010-values using appropriate price indices for variables expressed in monetary terms for all five years. Having deflated the prices, all values in our empirical analyses will be at the price level from 2010. The price index for agricultural goods at Denmark’s Statistics (Statistik 2016f) uses 2010 as the index year, which is the reason why all prices used in this thesis also will be at 2010-level. Having converted the nominal values into real values, differences in technical efficiency and performance cannot be explained by price differences and changes in the next chapters of this thesis.

7. Empirical analysis

Three steps will be used for the empirical analysis in this thesis. The first step is to test for non-random sample selection, the second is a correlation analysis, and the third is to use the stochastic frontier framework to perform the efficiency analysis.

7.1 Test for non-random sample selection

The first step is to check whether the final dataset is representative of the whole sector. Given that some observations were excluded due to not being meaningful, the test will not be based on the whole sample, but on the sample including only those observations with positive input quantities. Furthermore, we only wish to look at full time dairy producers, and therefore the restriction stating that at least two thirds of the total should be from milk production, needs to be included. We end up testing all 3831 farms, less the negative invalid observations, which were removed due to the restrictions listed in Table 6.2, against the final data sample that includes those 1810 farms for which data also was available for three or more years. We use a Welch's two-sample t-test to see if the mean values of the continuous variables such as recorded number of mastitis differ significantly between the two groups, and we compare the shares of the categorical variables (dummy variables) between the two groups.

Given the rather big sample size, excluding observations from the dataset with invalid values should not in itself create a selection bias. Further one could argue that invalid values would bias the dataset and hence the comparison would not be of much use.

7.2 Correlation Analysis

The second step is to investigate the relationship between the milk quality indicators, the animal health indicators, and the economic performance. This is done by using scatter plots and Pearson's correlation coefficient. As indicator of economic performance, we use the gross margin per cow, which is defined as the total revenues from milk for each farm less the costs associated with milk production, such as feed and veterinary costs. The milk quality indicators are the dummy variables for the cell, viable, and spore count in the milk and the animal health indicators are the number of reported cases of mastitis, hoof and limb disorders, reproductive diseases, and other types of disorders per cow. We also investigate the relationship between economic performance and other possible confounding factors such as whether the farm is organic and the age of the manager.

As a contribution to the correlation analysis an OLS regression is estimated to investigate the relationship between economic performance and the milk quality and health indicators. Contrary to the correlation analysis this method can control for confounding factors because it investigates the effects of multiple variables at one time. However, assuming both individual and time effects, we do not expect the OLS estimation to provide a true result. Specifications of the OLS estimation can be found in Equation B (Section 13.2.1) and Table B.2 in Appendix B.

7.3 Technical efficiency analysis

The third and last step is to use a stochastic frontier approach to examine how the milk quality indicators and the animal health indicators are related to technical efficiency. Since most of the data is collected during a time under the restriction of a milk quota, an input distance function might be a better fit to the data, given that output, for the most part, is restricted and hence could be assumed to be fixed. However, since the Danish dairy producers did tend to produce strategically above their quota and since the quota was abolished in 2015, which is the last year in our dataset, it can be relevant to estimate an output distance function to compare it with the input distance function, to see if the orientation of the model affects the efficiency.

In addition to the distance functions, an inefficiency equation explaining different input and output choices and results will be included in the estimations.

Input Distance Stochastic Frontier

The following Cobb-Douglas input distance stochastic frontier is estimated:

$$-ln(x_{Nit}) = \alpha_0 + \sum_{m=1}^{M} \alpha_m ln y_{mit} + \sum_{k=1}^{N-1} \beta_k \, ln\left(\frac{x_{kit}}{x_{Nit}}\right) + \sum_{r}^{R} \rho_r \; Regnskabsaar \qquad (1)$$

$$+ \varphi_1 milkingsystem2_{it} + \varphi_2 milkingsystem3_{it} + \varphi_3 jersey_{it} + \varphi_4 organic_{\,it} + \; v_{it} - u_{it}$$

Where we have $N = 6$ inputs of x and $M = 2$ outputs of y, t denotes the year, α_m is an estimate with respect to output quantities, β_k is an estimate with respect to input quantities, ρ_r is an estimate with respect to time, φ is an estimate with respect to production characteristics and are all coefficients to be estimated. Subscript i denotes the individual farm, $v_{it} \sim N(0, \sigma_v^2)$ is a random noise term which follows a normal distribution with zero mean and variance σ_v^2, and $u_{it} \sim N^+(\mu_{it}, \sigma_u^2)$ is an unobserved non-negative term which accounts for technical inefficiency and follows a truncated normal distribution with location parameter μ_{it} and scale parameter σ_u^2.

In many empirical cases, the output quantity does not only depend on the input quantities but also on some other variables, e.g. the manager's experience and in agricultural production also the soil quality and rainfall. If these factors influence the production process, they must be included in applied production analyses in order to avoid an omitted-variables bias (Battese & Coelli 1995). In our thesis, technical inefficiency can be related to milk quality indicators, animal health indicators, and other confounding factors. Following Battese and Coelli (1995) we use the following model specification of the location parameter of the technical inefficiency term:

$$\mu_{it} = \delta_0 + \delta_1 mastitis_{it} + \delta_2 hoofdis_{it} + \delta_3 reprodis_{it} + \delta_4 otherdis_{it} + \delta_5 cell1_{it} + \delta_6 viable1_{it} + \delta_7 spore1_{it} + \delta_8 managerage_{it} + \delta_9 consultant_{it} \tag{2}$$

Because of the nature of the data used in this thesis, and the fact the Cobb-Douglas input distance function is rather restrictive, the following time-dependent translog input distance function will also be estimated:

$$\begin{aligned} -lnx_{Nit} = {} & \alpha_0 + \sum_{m=1}^{M} \alpha_m lny_{mit} + \frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{M} \alpha_{mn} lny_{mit} lny_{nit} + \sum_{k=1}^{N-1} \beta_k ln\frac{x_{kit}}{x_{Nit}} \\ & + \frac{1}{2}\sum_{k=1}^{N-1}\sum_{j=1}^{N-1} \beta_{kj} ln\frac{x_{kit}}{x_{Nit}} ln\frac{x_{jit}}{x_{Nit}} + \sum_{k=1}^{N-1}\sum_{m=1}^{M} \zeta_{km} ln\frac{x_{kit}}{x_{Nit}} lny_{mit} \\ & + \sum_{r}^{R} \rho_r \; Regnskabsaar + \varphi_1 milkingsystem2_{it} + \varphi_2 milkingsystem3_{it} \\ & + \varphi_3 jersey_{it} + \varphi_4 organic_{\,it} + v_{it} - u_{it} \end{aligned} \tag{3}$$

Where we once again have $N = 6$ inputs of x and $M = 2$ outputs of y, t denotes the year, $\alpha, \beta, \zeta, \rho_r$, and φ are the coefficients to be estimated, and subscript i denotes the individual farm.

Following Battese and Coelli (1995), described in the methodology, the parameters of the model are estimated using the method of maximum likelihood. We jointly estimate the stochastic frontier input distance function and the inefficiency model using the maximum likelihood method, as this method results in consistent estimates of all the model parameters.

Output Distance Stochastic Frontier

The output distance function can be estimated using the same estimation methods as for the input distance function. The following Cobb-Douglas output distance stochastic frontier can be specified:

$$-lny_{Mit} = \alpha_0 + \sum_{m=1}^{M-1} \alpha_m ln\frac{y_{mit}}{y_{Mit}} + \sum_{k=1}^{N} \beta_k lnx_{kit} + \sum_{r}^{R} \rho_r \ Regnskabsaar \qquad (4)$$
$$+ \varphi_1 milkingsystem2_{it} + \varphi_2 milkingsystem3_{it} + \varphi_3 jersey_{it} + \varphi_4 organic_{\ it} + \ v_{it} - u_{it}$$

Following the same approach as before, the translog output distance stochastic frontier can be specified as:

$$-lny_{Mit} = \alpha_0 + \sum_{m=1}^{M-1} \alpha_m ln\frac{y_{mit}}{y_{Mit}} + \frac{1}{2}\sum_{m=1}^{M-1}\sum_{n=1}^{M-1} \alpha_{mn}\, ln\frac{y_{mit}}{y_{Mit}}\, ln\frac{y_{nit}}{y_{Mit}} \qquad (5)$$
$$+ \sum_{k=1}^{N} \beta_k lnx_{kit} + \frac{1}{2}\sum_{k=1}^{N}\sum_{j=1}^{N} \beta_{kj} lnx_{kit} lnx_{jit} + \sum_{k=1}^{N}\sum_{m=1}^{M-1} \zeta_{km} lnx_{kit} \frac{lny_{mit}}{lny_{Mit}}$$
$$+ \sum_{r}^{R} \rho_r \ Regnskabsaar + \varphi_1 milkingsystem2_{it} + \varphi_2 milkingsystem3_{it}$$
$$+ \varphi_3 jersey_{it} + \varphi_4 organic_{\ it} + \ v_{it} - u_{it}$$

8. Results

8.1 Test for non-random sample selection

Table 8.1 presents the results of the Welch's two sample t-tests for the continuous variables and the shares for the categorical variables. This is done to see the differences between the farms that are included in the analysis and the farms which are excluded from the original dataset, due to the implemented restrictions. The farms in the first column satisfy the restrictions related to being a full-time dairy farm with more than 49 dairy cows and only positive inputs, whereas the farms in the second column also satisfy the restriction of having data available for 3 or more years. Both datasets do not include invalid values for input factors. This is done in order not to avoid bias, as a comparison based on data including invalid values could lead to this. We test the null hypothesis that there is no difference between the two groups.

Table 8.1: Test for non-random sample selection

		Full time farms with positive input quantities	Full time farms with positive input quantities and data for 3 or more years	p-value
	Number of farms	3831	1810	
		Mean		
Reported diseases per cow	Mastitis	0.11	0.11	0.788
	Hoof and limb diseases	0.14	0.14	0.876
	Reproductive disorders	0.18	0.18	0.745
	Other disorders	0.03	0.03	0.898
	Manager age	49.0	49.0	0.886
		Percentage share		
Milk quality indicators	Cell 1	0.90	0.91	
	Viable 1	0.94	0.95	
	Spore 1	0.85	0.85	
Breed	Jersey	0.11	0.11	
Milking system	Fishbone	0.49	0.49	
	Other	0.27	0.26	
	Organic	0.17	0.18	
	Consultant	0.87	0.87	

Source: Result obtained from R

The tests show that removing approximately 53 percent of the farms from the dataset, due to the restrictions, does not bias the dataset in relation to the continuous variables. Further we see that the shares of the dummy categories presented in the first column, do not differ much from the shares presented in column two. The remaining farms seem representative for the whole sample of full-time dairy farms.

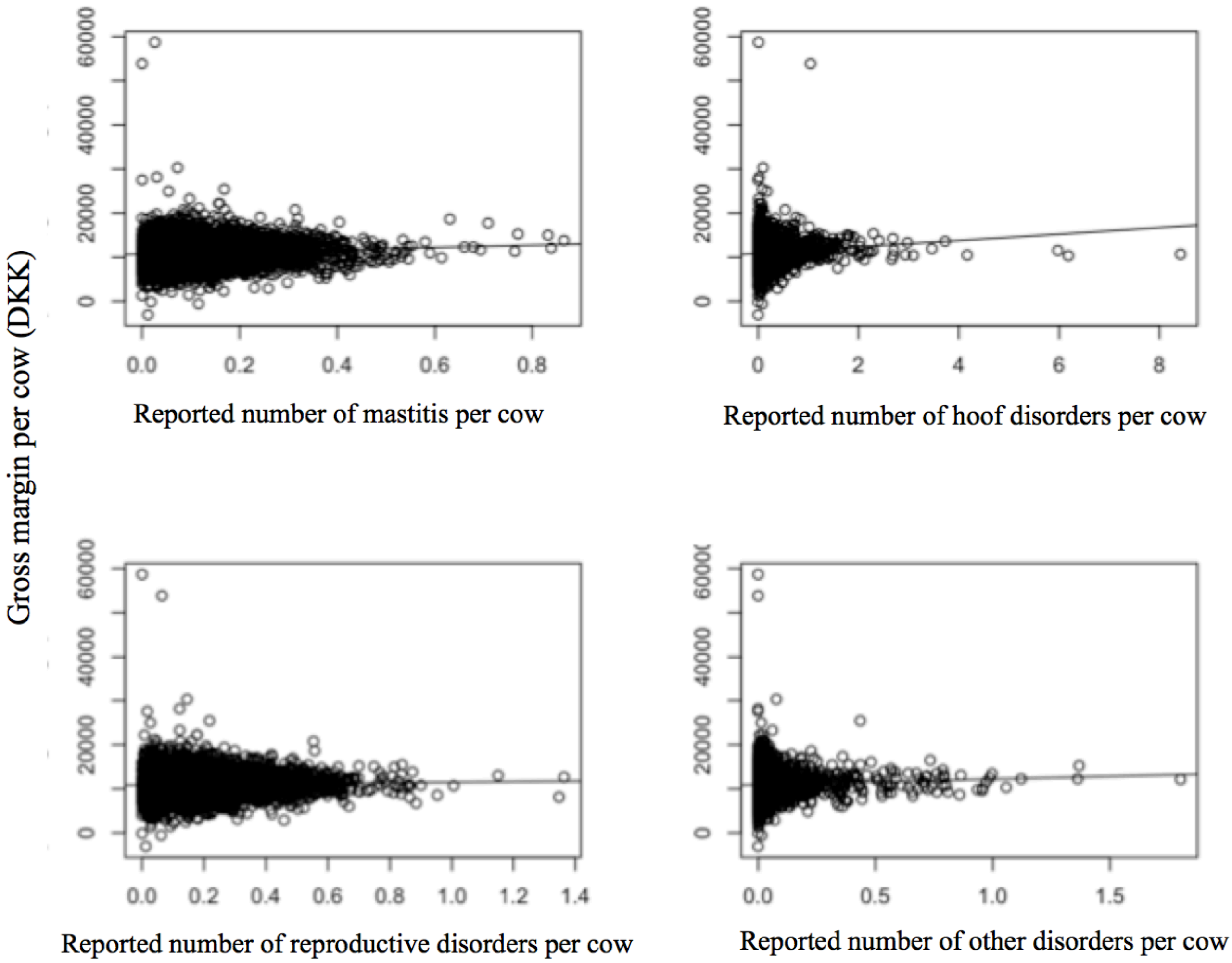
8.2 Correlation Analysis

Before estimating the input and output distance models, we simply test the correlations between the economic performance of the farm and the health and the milk quality indicators along with certain production characteristics. This is done to provide some background knowledge of the data before estimating our function. The gross margin (GM) is defined as the gross output from milk production less expenditures related to milk production, which are feed, veterinary, and medicine expenses. As we use the GM per cow in relation to milk output in the correlation and linear regression analysis, we need to account for the costs to internally bought feed, which has been left out until now. We assume that the costs of internally bought feed must be greater than zero, like for the other inputs. This assumption excludes 456 observations, however these observations are only left out in the correlation and linear regression analysis. Here the sole purpose is to determine to what extent health and milk quality indicators can explain the economic performance of the farm in relation to milk production. The expenditure to internally bought feed is deflated with the same price index as for the feed variable. For the efficiency analysis, internally bought feed is still not included.

The correlation analysis between the GM per cow and the number of reported diseases per cow is graphically presented in Figure 8.1 and the test-results can be read from Table B.1 in Appendix B. When the correlation coefficient is equal to 1, it corresponds to a perfect positive relationship between the two compared variables.

For all four types of diseases there is a weak but significant correlation between the number of reported diseases per cow and the GM per cow. The correlation coefficients obtained, show that there is no strong linear relationship between the two. However, the number of reported diseases per cow has a positive effect on the GM per cow, which is unlike what one would expect.

Figure 8.1: Gross margin per cow (DKK) and the number of reported diseases

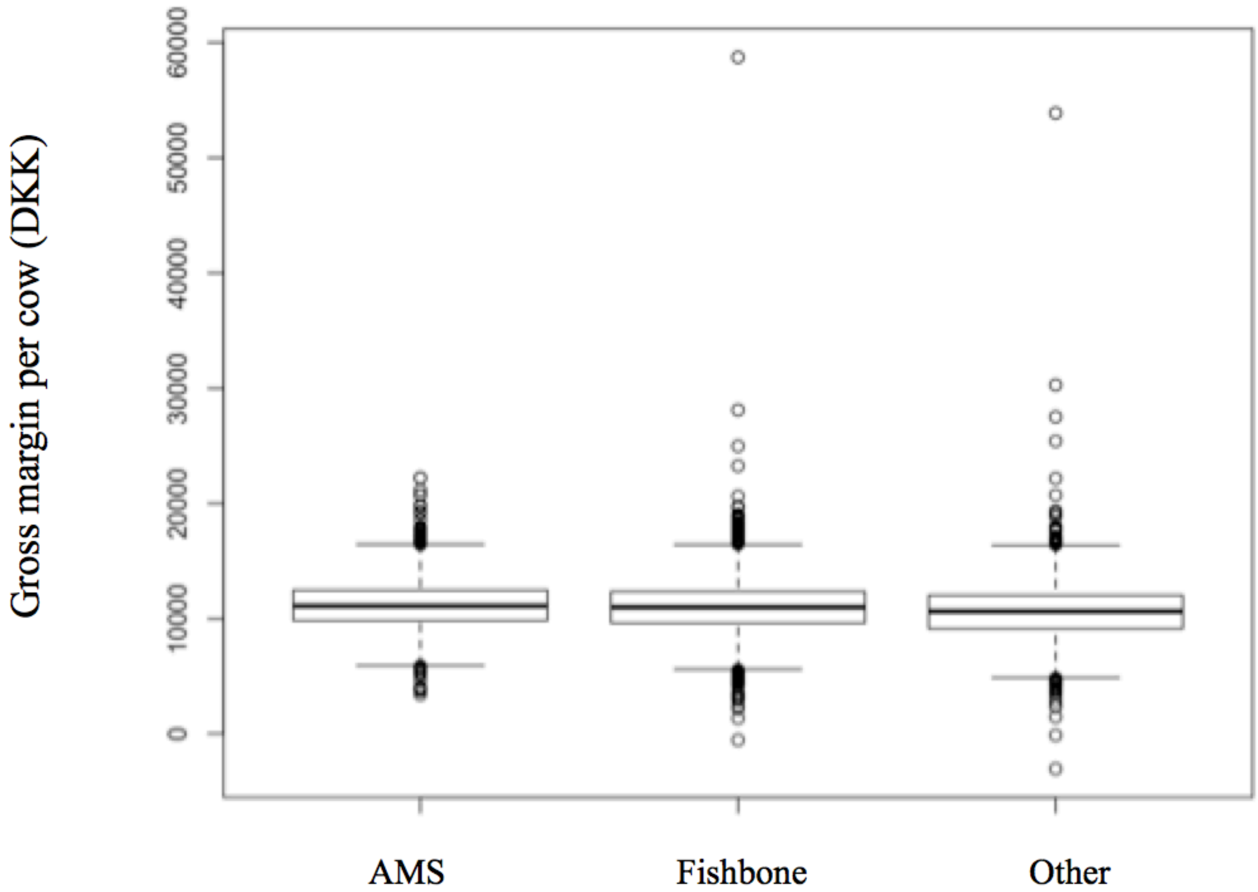


Source: Own plots from R.

We find that there is a rather small but significant and positive relationship between the dummies for low milk quality indicators and the GM per cow, as can be seen from Table B.1 in Appendix B. The positive correlation, is what one would expect, since a high milk quality leads to a higher price on the milk sold.

The correlations between the GM per cow and the different types of milking system do not show strong linear relationships. However, both fishbone and other systems are significantly different than the AMS. The fishbone system has a positive impact on the GM per cow, compared to the AMS, whereas it is the opposite for other types of systems. From the boxplot in Figure 8.2 it is clear to see that there are only very small differences in the effect on GM per cow for the three types of milking systems.

Figure 8.2: Gross margin per cow (DKK) and the type of milking systems



Source: Own plots from R.

Another factor that can influence the economic performance is the type of breed. Table B.1 in Appendix B shows the results of the correlation analysis between the GM per cow and the type of breed, which in this case again indicates a significant but positive small correlation in favour of the jersey breed. From Table B.1 in Appendix B we see that the GM per cow decreases when the manager gets older and that the presence of a production consultant increases the GM per cow. Both variables are significant. The only variable without significance is organic.

The overall picture of the correlation analysis gives an impression of very weak correlations between all the analysed variables and the gross margin per cow. The number of reported cases of diseases seem to have a positive and statistically significant effect on the GM, even though it does not make logical sense that an increasing number of diseases would increase the gross margin.

One explanation to this result is that the sector has been focusing more and more on how to increase the milk yield per cow while ignoring the health consequences of doing so. The focus on breeding better performing cows, has caused them to be less robust, and hence more vulnerable to catch and develop diseases (Beskyttelse n.d.). The better performing cows could therefore also be those with higher risk of getting more diseases. The milk quality dummies are all significant, and prove to have a positive effect on GM, which is what one would expect, since a higher milk quality yields a higher price. Both the type of milking system and breed seem to have significant impacts on the GM per cow, but still with very low correlation coefficients. Generally, we obtain very low (close to zero) correlation coefficients, indicating that the variables have very weak linear relationships with GM per cow and therefore one should be careful about concluding anything definite from these results.

Unlike the correlation analysis, we find that several of the estimated variables from the linear regression model have a significance level of either 5 or 1 percent. The results from the linear regression estimation can be found in Table B.2 in Appendix B, along with the interpretation of the estimated results.

8.3 Technical Efficiency Analysis

As presented in the theoretical framework, multiple parametric functional forms exist, and hence choosing the most suitable model for the data is arguably an empirical question. After creating the dataset and validating both the variables and the period in which the data is from, the decision with regards to functional form can be made.

Table 8.2: Descriptive Statistics

Variables	Measure	Mean	Std. dev
Outputs:			
grossmilk	DKK	4,540,745	2,996,938
grossother	DKK	929,628	696,328
Inputs:			
feedexp	DKK	1,480,601	1,104,571
vetmed	DKK	121,289	78,091
labour	DKK	817,819	559,407
hours	Hours	6,248	3,038
land	Hectares	165	90
materials	DKK	1,341,029	780,186
capital	DKK	1,514,013	938,433

Source: Own calculations.

The Input Distance Estimation

The first estimation is of the Cobb-Douglas input distance frontier function. This is done using the "frontier" package in R made by Coelli & Henningsen (2013). We use mean scaled quantities to be able to interpret the distance elasticities at the sample mean (Henningsen 2014). By dividing all the inputs by one of the inputs, linear homogeneity in inputs is imposed. In the estimation, all the inputs are divided by the material input. The estimation is done based on equation (1) and adding the inefficiency equation specified in equation (2). The results we obtain shows that the residuals from the Cobb-Douglas estimation are right skewed and that the estimation is very close to the OLS estimation, which indicates that no inefficiency is present. However, this result could also be due to a miss-specified model and hence, it is worth analysing if other models can handle the data differently. As the Cobb-Douglas functional form is restrictive and not suitable in this case, we additionally estimate a translog input distance function (TL IDF). The estimation of the TL IDF is done based on equation (3), adding the inefficiency equation, μ_{it}, specified in equation (2).

Several model specifications for the translog input distance function was estimated and tested prior to the selection of the final model. We first estimated our general translog model, which included all the inefficiency variables, specified in equation (2). We use a log-likelihood ratio test to compare the stochastic frontier (SFA) model with a standard ordinary least squares (OLS) model. The OLS model assumes that gamma, γ, is equal to zero. The null- hypothesis of the test states that there is no inefficiency, and therefore only statistical noise is present, and the OLS model is suitable. In this case, the log-likelihood ratio test provides a very small p-value (p-value: 0.000), which leads to a clear rejection of the null-hypothesis. This indicates that there is input oriented technical inefficiency and consequently leads to a favouring of the translog input distance model.

To see if more simple model specifications could be favoured, three restricted models were also estimated and compared to the general translog input distance frontier. The first restricted model limited the inefficiency term μ_{it} to only include the health indicators along with the age of the manager and the dummy for the production consultant; the second to only include the milk quality dummies along with the age of the manager and the dummy for the production consultant, and the third include the health indicators and the milk quality indicators. A log-likelihood ratio test clearly indicates that the general model is preferred to the first two restricted models; however, the fit of the third restricted model is not significantly worse than the fit of the general model (test statistic: 13.63; p-value: 0.001). The chosen model going forward thus contains seven inefficiency variables (the four health indicators and the three milk quality indicators) and all the variables linked to production. We found that both the age of the manager and the consultant dummy were insignificant to the models, and hence they are excluded. Summary statistics for all the estimated but rejected models can be found in Appendix C.

The summary statistics show the estimated γ parameter to be 0.56, which indicates that both statistical noise and technical inefficiency are important, but inefficiency is considered more important than noise. As σ_u^2 is not equal to the variance of the inefficiency term u_{it}, the estimated parameter cannot be interpreted as the proportion of the total variance that is due to inefficiency (Bogetoft & Otto 2011).

Following from this, one could test the null hypothesis of no inefficiency by simply testing whether γ is equal to (or does not significantly deviate from) zero. However, in this case a t-test of the null hypothesis $\gamma = 0$ is not valid, because γ is bound to the interval [0, 1] and hence, cannot follow a t-distribution. Another way of getting an indication of whether inefficiency is present is to check the skewness of the residuals. For the chosen model, a check for skewness returns a value of -0.20, and hence the residuals are left skewed, which is indicative of the presence of technical inefficiency among the dairy producers. All in all, the chosen TL IDF appears to be suitable.

The next decision is on which variable to use for labour input, as the suitability of both "labour", measured in wages and owner remuneration, and "hours" should be considered. The estimation parameters of the restricted model (4) can be compared to the same parameters of a similar model, where the only difference between the models is the use of "hours" instead of "labour" as the labour input. The differences between the two models are specified in Table 8.3.

Table 8.3: Comparison of labour and hours as inputs in the TL IDF model

	TL IDF labour	TL IDF hours
Mean efficiency	0.95	0.96
Mean input elasticity of labour input	0.08	0.30
Mean Elasticity of scale	1.05	1.10
No. of implausible estimates of elasticity of scale (Escale <0.5 | >2)	3	1
No. of observations with monotonicity (input)	7555	7647
Share of observations with monotonicity (input)	0.92	0.93
No. of observations with monotonicity (output)	8184	8180
Share of observations with monotonicity (output)	0.99	0.99
No. of quasi-concave observations	1034	7953
Share of observations with quasi-concavity	0.13	0.97

Source: Model result and calculations from R.

The mean efficiencies for the estimations are 0.95 and 0.96. Since they are both different than one, inefficiency is present. In a situation with no inefficiency it would imply that $u = 0$ and all observations would be on the frontier.

When looking at the mean elasticity of labour input we see a great difference. In the model using "labour", a one percent decrease in the labour input would, *ceteris paribus*, result in an efficiency gain of 0.08 percent, whereas it would be 0.30 percent in the model using "hours". This indicates that reducing the labour input in the model with "hours" would increase technical efficiency more than in the model with "labour".

For the input distance function, $D_I(x, y)$, the elasticity of scale is equal to the inverse of the negative sum of the distance elasticities with respect to the output quantities:

$$\epsilon = \left(-\sum_k \epsilon_k^I\right)^{-1}$$

When $\epsilon > 1$, the estimated elasticity of scale indicates increasing returns to scale (Henningsen 2014). For both models the estimated elasticity of scale is larger than one, but a difference is present between the elasticity of scale measures, as the model with "hours" has a substantially larger elasticity of scale at the mean. Both models contain at least one implausible estimate of the elasticity of scale. The model with the "labour" variable included has more than twice as many.

When the assumption of monotonicity is fulfilled, $D_I(x, y)$ is non-decreasing in x. This is the case if its first order derivatives with respect to input quantities are non-negative. This means that when all the elasticities related to the input quantities are positive in the input distance estimation, monotonicity in input quantities is satisfied. For both models this condition is satisfied at the mean.

It follows that, $D_I(x, y)$ is non-increasing in y if its first order derivatives with respect to output quantities are non-positive, when the assumption of monotonicity is fulfilled. When the distance elasticities of the estimated translog input distance function with respect to outputs are negative, it implies that the estimated first order derivatives are negative, and monotonicity in output quantities is fulfilled at the mean. For both the models we find that the output elasticities are negative, and that the assumption of monotonicity is fulfilled at the means and almost fulfilled for all observations. The total number of observations, which are monotone, is almost the same for the two models.

The elasticity for "materials" can be contained using the homogeneity restriction, where all input elasticities should sum to 1, which is what we find in both our estimated models.

We test whether the estimated translog input distance functions are concave in input quantities and quasi-concave in output quantities at the frontier, and find that both models violate the quasi-concave conditions for some of the observations; however, it is almost fulfilled for all observations for the model with "hours".

Conclusively, the model with "hours" appears to be a slightly better fit. This is not surprising since the estimated use of hours is based on the other inputs, and thus "hours" could be heavily correlated with the other inputs. The decision on what model to use going forward must also be based on which variables reflect the production decisions the best. Whether "labour" or "hours" account for the labour input better, is an empirical question as well as an arbitrary decision depending on beliefs of which variable represents the reality best. Using "hours" would indicate that no productivity difference in labour usage is included in the model, which is a rather problematic assumption to make. Considering this, the model with "labour" is selected for further analysis.

As already displayed in Table 8.3 the chosen TL IDF does not fulfil the monotonicity condition for all observations in the sample; however, the condition is satisfied at the sample mean for each variable, as can be seen from Table 8.4. From Table 8.4 the number of violations of the monotonicity assumption for each output and input elasticity are also listed. The two inputs, which violate the monotonicity assumption most frequently, are “veterinary and medicine” as well as “labour”. This is not rather surprising since the medicine consumption is not directly linked to production, which suggests that an increase in the medicine input will not necessarily directly generate more output, but only help the well-being of a sick cow.

The fact that “labour” violates monotonicity is more difficult to explain and could be due to the more complicated definition of the labour variable, which includes owner remuneration and wages. As a large share of the violated monotonicity can be explained by these two variables, it will not be imposed even though procedures to do so are available. Often monotonicity is not imposed as it can be rather complex to do so (Henningsen & Henning 2009). We did check for the implication of the monotonicity violations by removing the violated observations, but when looking at the means of the elasticities without violated observations, no severe differences were found. The mean elasticities for the observations fulfilling the monotonicity assumption can be found in Table B.3 in Appendix B.

The model violates the quasi-concavity conditions for most of the observations, which can be solved by imposing global restrictions, but this could also result in a less flexible functional form (Lawson, Agger, et al. 2004). This assumption is often violated when using SFA, and we will not impose the restriction in this case (Sauer. J, Frohberg. K, Hockmann 2006). As mentioned earlier Henningsen and Henning (2009) propose to check for quasi-concavity but not to impose it, as there in reality could exist several reasons as to why the estimation cannot fulfil this condition. The property of linear homogeneity in x is not fulfilled, as the quasi-concavity condition is violated and there is increasing return to scale. This can be due to the heterogeneity-causing factors, which were included in the inefficiency equation μ_{it} (Sipiläinen 2007).

The estimation results of the chosen translog input distance function can be found in Table B.4 in Appendix B. The results cannot be directly interpreted, and hence the elasticity for each variable is calculated. In Table 8.4 the mean elasticities for input and output are displayed along with the individual number of violations of the monotonicity condition.

Looking at our input variables we find that feed expenditure has the largest effect on technical efficiency. When feed expenditure decreases with one percent then, *ceteris paribus*, the distance to the frontier decreases and technical efficiency increases with 0.33 percent.

Table 8.4: Distance elasticities TL IDF

	Mean	St. Dev.	Violations of monotonicity
Milk output	-0.79	0.06	0
Other output	-0.17	0.05	14
Feed	0.33	0.07	1
Veterinary and medicine	0.03	0.02	186
Labour	0.08	0.04	380
Land	0.19	0.04	1
Capital	0.18	0.06	98
Materials	0.19	0.04	3
Mean Elasticity of Scale	1.05	0.12	

Source: Own calculations based on model result from R.

Like with the estimated coefficients for the inputs and outputs, the effects of the production variables should be calculated to find the exact effects. From the estimated results presented in Table B.4 we can obtain the standard deviations for the production variables and since they are all very small, we can trust the calculated marginal effects presented in Table 8.5 (Olsen & Henningsen 2011).

Table 8.5: Marginal Effects of production variables TL IDF

	Effect in percent	p-value
Regnskabsaar 2012	0.89	0.009
Regnskabsaar 2013	-5.09	0.000
Regnskabsaar 2014	-4.18	0.000
Regnskabsaar 2015	4.39	0.000
Milking system 2 (fishbone)	2.71	0.000
Milking system 3 (other)	1.91	0.000
Jersey	3.63	0.000
Organic	-0.27	0.336

Source: Own calculations based on model result from R.

From Table 8.5 we can see that most of the production variables are significant. The effect of both milking system 2 and 3 is positive and significant, meaning that those who use milking system 2 or 3, compared to milking system 1, lie on a lower frontier. Thus, both systems require less input than milking system 1 to produce the same amount of output. The results presented in Table 8.5 suggest that the AMS is not as productive as fishbone or other milking systems.

The production variable with the largest effect on productivity is "jersey", which indicates that having "large breed" requires more inputs in order to obtain the same level of output, compared to jersey and thus farms with jersey cows are, *ceteris paribus*, 3.63 percent more productive than farms with large breed.

It is worth noticing the year dummies for 2012 and 2015, since they show that the farms were better at transforming inputs to outputs in both years compared to 2011. In 2012 the farms were generally more productive and they experienced an increase in the farm size and milk yield, compared to 2011 (Andersen & Hansen 2012). The dummy for 2015 could be explained by the abolishment of the milk quotas, which made it possible to produce more milk than in previous years, hence caused production to expand.

The only variable without any significance is organic, meaning that organic productions are neither more nor less productive than conventional productions.

The last part of the estimation contains the effects of the inefficiency variables. Once again, the estimated coefficients cannot be directly interpreted. The calculated marginal effects of the inefficiency variables are presented in Table 8.6.

Table 8.6: Marginal effect of inefficiency variables TL IDF

	Mean effect in percent	Minimum	Maximum	p-value
Mastitis	-0.0157	-0.0345	-0.0019	0.015
Hoof and limb disorders	0.0007	0.0000	0.0015	0.706
Reproductive disorders	-0.0234	-0.0514	-0.0028	0.000
Other disorders	-0.0047	-0.0104	-0.0006	0.398
Cell 1	0.0182	0.0022	0.0399	0.000
Viable 1	0.0135	0.0016	0.0297	0.000
Spore 1	0.0124	0.0015	0.0273	0.000

Source: Own calculations based on model result from R.

From Table 8.6 we see that only two out of the four types of disorders are significant. Occurrences of hoof and limb disorders as well as other disorders cannot explain inefficiency. Experiencing more cases of mastitis and reproductive disorders will on the other hand decrease the inefficiency term u_{it}, which then increases the distance function and the distance to the frontier, thus making the farmer less efficient. This is what we expected as diseases tend to lower the quantity and quality of the milk output.

Overall, we see that all three milk quality indicators are highly significant but that they only affect inefficiency very little. All three variables are dummies representing those who receive a supplement to the standard price on milk. The marginal effect of all three variables is positive meaning that having low milk quality indicators will increase the inefficiency term u_{it}, which then decreases the distance to the frontier thus making the farmer more efficient. Not surprisingly we see that a higher milk quality has a positive effect on efficiency.

Dairy farmers operate in a market, which is close to perfect competition, meaning that they are price takers. Therefore, even small efficiency gains should be considered. We see that those who have a cell count below 300,000 have, *ceteris paribus*, a higher technical efficiency of 0.018 percent. As we see from Table 8.6 there is a small difference between the minimum values and the maximum values of the inefficiency variables, why we consider the estimated mean value to be close to the true mean value of the variables.

During the period, which is analysed in this thesis, we see generally high levels of efficiency. From 2011-2015 the Danish dairy farms experienced an increase in the technical efficiency as displayed in Table 8.7. This development might be explained by the fact that the farmers have experienced economic pressure and increased competition due to the abolishment of the quota. They have therefore worked to increase their technical efficiency to create performance growth (Philipp 2016).

Table 8.7: Technical Efficiency over time TL IDF

TE %	2011		2012		2013		2014		2015	
	N = 1621	%	N = 1615	%	N = 1750	%	N = 1707	%	N = 1505	%
0 - 80	7	0.43	6	0.37	2	0.11	3	0.18	2	0.13
80 - 85	31	0.91	19	1.18	17	0.97	12	0.70	8	0.53
85 - 90	99	6.11	97	6.01	83	0.74	78	4.57	67	4.45
90 - 95	609	37.57	628	38.89	669	38.23	672	39.37	569	37.81
95 - 100	875	53.98	865	53.56	979	55.94	942	55.18	859	57.08

Source: Own calculations based on model result from R.

The increased technical efficiency could also be due to a general development in the technology used on the farms, since the Danish dairy farms have gotten bigger and more specialised during the period. The increase in technical efficiency can also be a result of the farmers using the potential of productivity growth that exists under increasing returns to scale. From Table B.6 and Table B.7 in Appendix B we can see that the shares of the inputs used and the outputs produced have not changed, but the value of both the outputs and the inputs have increased. Since prices are all in 2010 values the development is due to an increase in the quantities used and produced, and hence an overall increase in the farm size.

The output distance estimation

Since the data used in this thesis is mainly from a period with a milk quota restriction, we have assumed that the farmers have been input minimizing. However, given that the quota did not regulate the market for the most part of 2015, it is worth investigating, if the output distance function is a better fit to the data. As already mentioned, the Danish milk producers have exceeded the quota from time to time, even with the restriction.

First, we estimate the Cobb-Douglas output distance frontier using the same approach as for the Cobb-Douglas input distance frontier. The estimation is done based on equation (4), using the inefficiency equation specified in (2), using the mean scaled quantities approach and gross milk as the explained variable. The obtained results prove to have left skewed residuals and being very close to the OLS estimation, which is an indicator of no inefficiency being present. However, as explained for the Cobb-Douglas input distance frontier, this result could also be due to a miss-specified model, which is why an translog output distance function (TL ODF) is estimated.

Like with the input distance estimation, several model specifications were estimated and tested against each other prior to selecting the final model to proceed with. The first estimation, the general model, is based on equation (5) adding equation (2) containing the inefficiency variables. Once again simpler model specifications were estimated in order to see if they could be favoured. Using a log-likelihood ratio test, the null hypothesis of no differences between the unrestricted model and a restricted model, including only the milk quality indicators, consultant and manager age as inefficiency variables, could not be rejected at a 10 percent level (p-value = 0.09286). Hence, we can proceed with the above-mentioned restricted model.

The chosen translog output distance model, including only milk quality indicators, manager age, and the dummy for consultant as inefficiency variables, has an estimated gamma equal to 0.69. This indicates that inefficiency is important in the model. Testing the frontier against an OLS, it can be rejected that there is no inefficiency in the model.

The mean efficiency is 94 percent, which indicates that the estimation does not lie on the frontier and inefficiency is present. The model fulfils the monotonicity assumption in inputs for 76 percent of the observations and for 99 percent of the observations in outputs. However, we find that the monotonicity assumption is satisfied at the mean for the calculated elasticities. From

Table 8.8, the number of violations of the monotonicity assumption for each output and input elasticity are listed. We find that the input variable "Veterinary and medicine" violates the monotonicity assumption most frequently. Which might be due to the fact that the use of medicine might not result in extra output. For the same reasons as described for the TL input distance frontier, we do not impose monotonicity.

The model violates the assumption of being quasi-convex in x and convex in y for all observations; however, as described earlier, this assumption is often violated when estimating a stochastic frontier (Sauer. J, Frohberg. K, Hockmann 2006).

As the mean elasticity of scale is larger than one, there is a small indication of increasing returns to scale, as was also found using the translog input distance function.

Table 8.8: Distance elasticities TL ODF

	Mean	St. Dev.	Violations of monotonicity
Milk output	0.85	0.05	0
Other output	0.18	0.05	9
Feed	-0.37	0.07	1
Veterinary and medicine	-0.01	0.02	1842
Labour	-0.09	0.03	84
Land	-0.18	0.04	1
Capital	-0.19	0.06	104
Materials	-0.17	0.04	4
Mean Elasticity of scale	1.01	0.06	

Source: Own calculations based on model result from R.

The mean elasticities for the translog output distance were also calculated based on the estimated coefficients and as in the input distance function, the feed expenditure is found to have the largest effect on the technical efficiency. Given that the dataset only contains specialised dairy farmers, it is not surprising that the biggest output elasticity is related to milk output. If milk output were to increase 1 percent then, *ceteris paribus*, the technical efficiency would increase 0.85 percent.

The effects of the production variables included in the technology set are estimated to obtain the most precise estimates as possible in relation to the technology set. The effects of the different production variables are presented in Table 8.9. The summary statistics of the TL output distance function can be found in Table B.5 in Appendix B, from which we can see that the standard errors for the coefficients are small, and therefore the estimates should be rather precise. The p-values in Table 8.9 show that the effects of the production variables are all significant, at least at a 5 percent level, except from the dummy for organic production.

Table 8.9: Marginal Effects of production variables TL ODF

	Effect in percent	p-value
Regnskabsaar 2012	-0.84	0.014
Regnskabsaar 2013	5.75	0.000
Regnskabsaar 2014	4.61	0.000
Regnskabsaar 2015	-4.49	0.000
Milkingsystem 2 (fishbone)	-2.41	0.000
Milkingsystem 3 (other)	-1.20	0.000
Jersey	-3.67	0.000
Organic	-0.21	0.461

Source: Own calculations based on model result from R.

From Table 8.9 we can see that the effects of both milking system 2 and 3 are negative and significant, meaning that those who use milking system 2 or 3 lie on a frontier above those who use milking system 1, and thus both systems can produce a larger amount of output using a fixed amount of inputs, compared to milking system 1. The production variable with the largest effect on productivity is “jersey”, which indicates that “large breed” is less productive than “jersey”. This could be due to differences in the protein and fat content in the milk produced by the two breeds. Like with the input distance function the years 2012 and 2015 seem to have been better for the productivity.

The calculated marginal effects of the inefficiency variables are presented in Table 8.10. The three milk quality indicators are all significant. All three milk quality dummies have a positive effect on efficiency compared to the omitted dummies, which indicate that having a high milk quality makes the farm more efficient. The age of the manager has a rather small negative but significant effect on efficiency, indicating that being older does not help the level of efficiency on the farm. The variable “Consultant” has no significant effect.

Table 8.10: Marginal effect of inefficiency variables TL ODF

Inefficiency variable	Mean effect in percent	Min	Max	p-value
Cell 1	0.0184	0.0015	0.0460	0.0000
Viable 1	0.0108	0.0009	0.0271	0.0000
Spore 1	0.0125	0.0010	0.0312	0.0000
Manager age	-0.0003	-0.0008	0.0000	0.0000
Consultant	0.0003	0.0000	0.0007	0.7305

Source: Own calculations based on model result from R.

When looking at the technical efficiency over time in Table 8.11, we find the same development as for the TL input distance estimation, where technical efficiency has increased over time.

Table 8.11: Technical Efficiency over time TL ODF

TE %	2011		2012		2013		2014		2015	
	N = 1621	%	N = 1615	%	N = 1750	%	N = 1707	%	N = 1505	%
0 - 80	23	1,42	15	0,93	6	0,34	7	0,41	6	0,40
80 - 85	33	2,04	37	2,29	31	1,77	23	1,35	16	1,06
85 - 90	110	6,79	103	6,38	102	5,83	86	5,04	75	4,98
90 - 95	558	34,42	556	34,43	612	34,97	618	36,20	536	35,61
95 - 100	897	55,34	904	55,98	999	57,09	973	57,00	872	57,94

Source: Own calculations based on model result from R.

Previous literature has found that the orientation of the model can influence how efficient different firm sizes appear, and thus estimating just the input oriented model can lead to a wrong view on how efficient the farms in the sample might be (Sipiläinen 2007). We find that both the smallest and the largest farms, when looking at the number of cows, all have an efficiency level close to the estimated average efficiency in both models, and hence deviations from the average efficiency cannot be explained by farm size.

From the results presented, there seems to be only small differences in the results obtained from the input and output distance functions, even though different model specifications were used. This can be due to the fact that the Danish dairy producers already are highly efficient. For both estimations, the technical efficiency has increased from 2011-2015. As shown in section 8.4.1, Table B.6, and B.7 the dairy farms in this dataset are very specialised and have increased production during the years from 2011 to 2015, indicating that the Danish dairy farms have gotten bigger and more specialised. From Figure B.2 in Appendix B the efficiency scores obtained from the input and output distance models are plotted. Although the farms in the dataset appear efficient in both models, it can from Figure B.2 be seen that a larger share of the observations is more efficient in the input distance model than in the output distance model. The output distance function violates the quasi-concave condition for all observations, whereas the input distance function has some quasi-concave observations. Due to this and the results presented in Figure B.1 the input distance function seems to be a better match to the data. The violation in the output distance model is, however, not surprising, given that the milk quota restriction has regulated the market, making it difficult to maximize output in the time period analysed.

8.4 Summary of Results

From the correlation analysis, it is found that there is a weak linear relationship between the health and milk quality indicators, and the gross margin per cow, while the regression analysis shows that having better milk quality indicators is positive for the gross margin per cow. The analyses show that the AMS and the jersey breed increase the GM per cow.

We see that the estimated TL input and output distance functions are good fits to the data, given that there are rather few implausible estimates for the elasticity of scale and that the monotonicity assumption, for the most part, is fulfilled. We find that the two estimated models, input and output, are much alike in distance elasticities and technical efficiency.

We find that the occurrence of certain types of diseases can explain variations in efficiency in the input distance function and that having good milk quality values have a small positive but significant effect on the technical efficiency. Unlike the animal health and milk quality indicators it is found that the production characteristics have a rather large effect on productivity.

Overall, the input distance frontier seems to be the better fit to the data, which is why the estimation results from this will be used in the further discussion.

9. Discussion

9.1 Choice of data and variables

We have decided only to focus on full-time dairy farms and also exclude those farms that were not in the dataset for three years, which means that our final data sample ends up consisting of specialised milk producers. It could affect our results, as this decision might have excluded the less specialised and hence the less technically efficient farmers from the data set. This assumption was necessary to create consistent data and to make sure that we could investigate a development over time. However, only allowing farms in the analysis, which have reported to SEGES for three or more years, could have excluded farms, which have gone bankrupt and not because they stopped reporting the data to SEGES. This indicates that we may have excluded the less efficient farms in favour of the more efficient. From 2011-2015 the development in the Danish dairy sector has tended towards fewer but larger and more specialised farms. The data used in this thesis proves to be representative of this, as we find that the farms in the dataset have increased production while at the same time increased technical efficiency. The analysis is therefore mainly based on larger specialised dairy farms.

The disease variables were included in the chosen input distance function. Not surprisingly, we find that experiencing more cases of mastitis and reproductive disorders decreases the technical efficiency on the farm. This is what we expected, as occurrences of diseases reduce the milk quality and quantity. However, we find that not all the defined disease variables are significant. This is surprising, but might have been the result of rather few occurrences of disorders or inconsistent data. As already mentioned, a rather large share of the disease variables has a value of zero. Since we cannot know if the farmers have reported all the cases with diseases, our data might not be representative of the true number of diseases actually experienced or treated. Lawson, Bruun, et al. (2004) find that reproductive disorders are without significance in relation to technical efficiency due to good management. This could also be an explanation of our findings of no significant correlation between all four types of diseases and technical efficiency. As mentioned, there are rather few reported cases of diseases per cow in our data set, which might reflect that more attention has been paid to securing fewer cases of various diseases during the period covered. Another reason why we only discover a small impact of diseases on efficiency could be the focus on breeding higher yielding cows. If the extra yield, the cow delivers, can cover the costs of the diseases it suffers from, there will be no effect of a disease on efficiency.

Changing the orientation of the model from being input oriented to being output oriented influenced the significance of certain variables. In the estimated output distance function the disease variables were not included, as they showed no significant explanatory power. We find that the input distance function is a better match to the data, as expected, since the dairy farmers have been restricted on output from 2011 to 2014. The output distance function is not the most suitable orientation for the data, which can explain the lack of effect of the disease variables. It is not an indication of output oriented dairy farms not being affected by more occurrences of different diseases, but rather that the orientation is not suitable to capture the effects of certain variables.

The marginal effects of the estimated milk quality variables show that the farmers with low milk quality indicators are more efficient. This indicates that it pays to have a high milk quality. Andersen et al. (2016) believe that the Danish dairy farmers will obtain better economic results if they have low milk quality indicators, as this will ensure them a higher price for their milk. We find that this is in line with our results, which show that having low cell, viable, and spore counts increase the technical efficiency, and thus the gains from obtaining low levels of the milk quality indicators are worth the costs of achieving them.

As mentioned, most of the farmers in our data set have good milk quality values and hence they receive a supplement of 1-2 percent in addition to the ordinary price on milk. From Landbrug og Fødevarer (2016) we see that more and more farmers have decreased their cell count from 2011 to 2015. In 2015, approximately 87 percent of the Danish farmers had a cell count below 300,000 meaning that it has become more common to receive a supplement than a deduction. We see the same development for the spore count, whereas there has been a small decrease in the share of farms with low viable count. This could be due to more focus on the matter through better management.

From our analysis, we see that the Danish dairy farmers have become more efficient from 2011 to 2015, which to some extent can be explained by the fact that more and more farmers have obtained lower values for the milk quality indicators. The farmers have managed to reduce the values of the milk quality indicators, e.g. through better management while at the same time increase efficiency. If the manager has succeeded in obtaining low milk quality values by changing routines and hygiene levels, the effort used to keep them low might not require a lot of extra resources allowing for efficiency to increase. The results seem to support the theory of optimal input use in relation to animal health.

As the Danish dairy farms have become bigger and more specialised the costs of diseases spreading and causing infections are more severe. As the milk quality is affected by the health of the cow, aiming for low milk quality indicators might be worth the extra costs, given that a good milk quality does increase efficiency. The results obtained suggest that there is an indirect effect of management, as knowledgeable management and decisions will affect the health of the animals and the milk quality indicators through daily routines. However, as it was found that the manager age had no effect on the technical efficiency on the farms, in the input distance function, and not a noticeable effect in the output distance function, it is difficult to say anything specific about the effects of a skilled and knowledgeable manager from our estimated models.

We find that farms using the AMS use more inputs in the production than those who use either the fishbone or other types of systems. AMS requires fewer hours of manual labour, but is at the same time more capital intensive, as it involves large investments to acquire and implement the system. On a daily basis, there are costs related to adjusting and maintaining the system. This could be why we find that the fishbone and other types of milking systems require less input to produce the same level of output. There are also several managerial challenges related to having an AMS, as bad hygiene can spread more easily amongst the cows in the herd since there is less supervision (Andersen et al. 2016). Because the AMS is a rather new system compared to the other milking systems included in the analysis, the lower productivity observed could be due to difficulties with adapting to the system. Another reason, why the AMS is unfavourable compared to other systems, is that the milk quality obtained from farms using the AMS is on average worse (Andersen et al. 2016). With the AMS the cows can be milked more times a day, than with other systems, however the extra yield do not seem to make up for the lower milk quality and thus lower price received for the milk.

As mentioned, the milk quality is affected by the general health of the cow, and it could therefore be interesting to investigate how the use of different milking systems affect the number of diseases and milk quality levels. However, this requires a more updated cattle database including more detailed and consistent information about the farmers.

Lawson, Agger, et al. (2004) find that farms with jersey cows are on average less technically efficient than farms with large breed. This is not in line with what we find in our results, where farms with jersey cows prove to be more productive than farms with large breed. The jersey breed is smaller in size than large breed, and the milk obtained from jersey cows has higher levels of protein and fat. Being smaller in size, jersey cows require less feed and combined with the higher protein and fat levels, they are more productive in our model. Adding to this, the higher protein and fat levels secure the farms with jersey cows a higher price per kg ECM, than received at the farms with large breed (Andersen et al. 2016). Unlike us, Andersen et al. (2016) find that large breed tend to make up for the lower energy in their milk by producing more, and thus farms with large breed do not seem to perform worse. The price difference is not directly accounted for in our model, as it is assumed that all farmers receive the same price, which might explain the large difference in productivity found between the breeds in the model. The result from the efficiency analysis is in line with the results found in the correlation and linear regression analysis. It could indicate that the farms with the jersey breed are good at milk production.

Organic dairy producers receive a higher price on milk due to an organic price supplement, which has not been considered in this thesis. This might be one of the explanations as to why we see no significant effect of being organic. Given that the price development for organic farmers has been close to that of the conventional farmers, as can be seen from Figure 1.3, and that the price supplement only caused large variations in the results between organic and conventional farmers in 2015, the assumption to use the same price index seems reasonable (Andersen et al. 2016). The price index used in this thesis is an aggregated price development based on both conventional and organic milk (Statistik 2016f).

The penalty (super tax) from exceeding the milk quota was not included in the analysis. If the farmers exceeded the national aggregated quota, the fee would be paid among those having exceeded their quota. The Danish farmers did exceed the quota in the period from 2011 to 2015, and therefore this cost could have been relevant to include in the analysis. The focus of this thesis has been on the effects of reported diseases, milk quality indicators, and production characteristics and not on the economic aspects of the milk quota restrictions. Given that we in this thesis are measuring how well the farmers can transform input quantities to output quantities and that we are assuming identical prices, including the fee was not considered relevant.

Furthermore, the risk of paying the fee is not new to the farmers, and we would therefore assume that they had taken it into consideration when planning their production. Given that the size of the fee might be dependent on farm size, including it in the estimation could be done using a dummy. This would find the effect of having to pay the fee independently of farm size.

As mentioned earlier, we see that for both models we have increasing returns to scale, indicating that there are gains from increasing the farm size. However, when looking at the relationship between technical efficiency and total output from the farms, it is clear that the level of technical efficiency is not dependent on the amount of output produced or the number of cows, and thereby not dependent on the size of the farm. This result might be due to the restrictions imposed on the dataset and the milk quota restriction.

The estimated models do not include a time trend, which means that a technological change over time is not considered. This could have been done, since previous studies have found that there exists a rather general trend. Given that we are only looking at a five year period and that caution should be taken when imposing a linear time trend, as it does not always sufficiently reflect the true development over time, it was not included (Karagiannis et al. 2004). Instead of a time trend, year dummies were included to be able to see the effect of the individual years.

In the years 2012 and 2015, the Danish dairy farmers were more productive compared to 2011. Given that monetary values have been deflated, price changes cannot explain the results obtained. We see that the year 2015, in which the milk quotas were removed, caused an increase in the productivity among the Danish dairy producers. This might be explained by the new market conditions, where output was no longer controlled by the milk quota system. The tough market conditions, which the farmers faced in the time after the milk quota was removed, might have affected the production processes as the farmers needed to "run a bit faster" in order to survive in the market. In 2012 the farmers were too quite productive. This might have been due to a higher milk yield (Andersen & Hansen 2012). Contrary to the years 2012 and 2015, the years 2013 and 2014 appear to have been less productive compared to 2011, in the estimated results. However, these two years actually produced good results for the Danish dairy farms, because of an increased price on milk. The difference in productivity obtained from the results might indicate a tendency of farmers focusing a little less on keeping productivity up in years with higher prices (Vidø et al. 2014). Our estimations have attempted to say something about dairy farmers' performance, without focusing on prices. The results indicate that more factors

than the price of milk are important for the performance of the dairy farm, as the year dummies prove to affect productivity in a different way than expected.
However, since most published reports regarding market conditions and performance for the dairy sector look at price changes, the estimated effects seem difficult to explain.

9.2 Choice of Method and Orientation

We have chosen to use the stochastic distance function approach, as this method is often used on data from the agricultural sector. It makes sense to be able to include more inputs and outputs in the estimation, when dealing with an agricultural production. Furthermore, the distance function approach does not have to include prices, which would have been difficult in this case. Given that prices were not available in the data set, using a cost function was not a possibility.

We assume that the farmers in the data set have faced identical prices. This assumption is unlikely to hold in real life, since one could expect larger farms to obtain price discounts due to large scale advantages. We have deflated all values using sector specific price indexes. To create a more precise price index, or a more precise idea of how efficient each farm is at transforming inputs to outputs, farm level prices or input quantities would have to be available. However, even so, creating price indexes using individual prices for all five years would have been too comprehensive given the time constraint.

When using a distance function, identical prices are implicitly assumed, and hence efficiency is based on a farm's ability to transform inputs into outputs, and not on the manager's ability to buy inputs at a low cost and sell outputs at a high cost. Excluding price differences means that we assume high efficiency to be associated with good economic performance. Keeping up a high efficiency score over time indicates that even during fluctuations in the economy, the farm still succeeds in performing well. The choice to use the distance function approach therefore has the implication of estimating technical efficiency and not economic efficiency or economic performance. The two concepts differ from each other as technical efficiency explains the ability to transform input into output, whereas economic efficiency concerns the capability to keep production costs as low as possible, while maximizing output production and selling output at the highest price possible. Therefore, the economic efficiency also depends a great deal on the prices related to the different factors of production. Because of this we have not been able to measure economic efficiency in this thesis, even though it might have provided a clearer picture of how the Danish milk producers are performing. Technical efficiency can be considered a prerequisite for economic efficiency, as the farmer needs to achieve technical efficiency to be able to achieve economic efficiency. One can consider perfect economic efficiency as the state where every input cost is minimized while the transformation of inputs into outputs is maximized.

The choice of orientation has proven not to affect the obtained elasticities evaluated at the sample mean. This could be due to the fact that the time period covered in this thesis is the years leading up to the removal of the milk quota restriction, so the farmers have been preparing to increase production as of 1 April 2015. However, when looking at the technical efficiency over time we do experience differences between the two orientations, confirming that the milk quota has called for more input oriented production. Another reason, why we might obtain similar results from the two estimations, could be that the farms in the data set are rather homogeneous given that they have chosen to specialize in milk production. We did find that efficiency did not depend on farm size measured in the total number of cows, and hence for the Danish dairy sector large differences in efficiency do not seem to exist among specialised milk producers. Given that we have excluded smaller farms from the data set, large variations in scale economics might not be reflected in the estimation.

This is not in line with other studies seeking to analyse efficiency among dairy farmers subject to the milk quotas imposed by EU. Sipiläinen (2007) finds that smaller farms tend to be more input oriented than larger farms and hence appear more efficient when using the input orientation. The choice of orientation can, as mentioned, have implications for the estimated efficiencies, which is why Orea et al. (2004) address the choice between an input oriented or output oriented model in a cost minimizing framework. They find that different orientations yield very different results, which emphasize the importance of choosing the correct model (Orea et al. 2004). However, given that the specialised Danish dairy producers are highly developed and efficient, the lack of effect of the orientation is not surprising.

The results of the two estimated models would have been expected to be the same if the models had exhibited constant returns to scale. We find that both models display increasing returns to scale, and thus the efficiency scores prove to be different for the models. This makes it difficult to conclude which model is better fit. Therefore, it might have been more relevant to consider a model mixing the orientations, which would involve the input and output distance functions in the context of a multi-input and multi-output production technology, as the production orientation as argued earlier could be shifting over our investigated period.

Orea et al. (2004) look at how one best chooses the orientation of the model specification by introducing efficiency as an input-oriented, an output-oriented, and a hyperbolic parameter. The hyperbolic parameter is a third way of measuring inefficiency, and unlike the two other methods, the hyperbolic method allows farmers to reduce inputs while they at the same time increase outputs. Knowing that different types of inefficiency might have affected data, testing several model specifications is advised.

Like Orea et al. (2004), Kumbhakar et al. (2007) look at how one can combine both the input and output orientation. They propose a latent class model (LCM) in which they mix the input and output distance approach, allowing for both input and output oriented firms to be efficient in the same estimation, by placing the firms in one of the two categories. The method differs from the one presented in Orea et al. (2004), because firms have to be placed into groups. It could therefore be suggested for further analysis on the subject to investigate using the approach of mixing the input and output orientations, as it can be assumed that both orientations have been used by the farmers over the period. This method might avoid the insecurity of deciding which model fits the data the best way (Kumbhakar et al. 2007). Both articles find that the input oriented model is a better fit. Since the milk quota regulation is no longer affecting the dairy sector, future research should be more aware of the orientation of the model and hence focus on new ways of analysing efficiency in European countries. The choice to estimate the input and output models separately was made due to limited time, and the fact that data used in this thesis is (mainly) from a time period where output was fixed given that the milk producers were subject to the milk quota restriction.

9.3 Violations of theoretical assumptions

Overall, we find that a sizeable share of observations violated the assumption of monotonicity and quasi-concavity. This problem could stem from the fact that we are looking at farmers, who can have other motives than to profit maximize, which is what the economic model assumes, and hence this condition could have been violated in the first place. In general, the literature regarding technical efficiency tends to accept the models violating the monotonicity assumption if the condition is satisfied at the mean. Imposing monotonicity is recommended by Henningsen & Henning (2009) using their rather simple three step procedure. However, given that the simple procedure does not apply to a multiple output framework and that time was limited, we have refrained from imposing it. Nevertheless, if future studies were to analyse efficiency among Danish dairy farmers, imposing monotonicity might improve the results.

Another problem when working with data is multicollinearity. When deciding on the variables to use, there is trade-off between showing more detailed information about input types thereby adding the risk of multicollinearity on one hand and aggregating the inputs and thus sacrificing the detailed information on the other hand (Brümmer et al. 2002). Having used six inputs and two outputs in the estimated models, problems of multicollinearity could exist. However, we have refrained from looking into any complications related to multicollinearity, since the literature does not seem to propose any solutions other than aggregating inputs.

Given that we are analysing technical efficiency over time, we should be aware of the potential endogeneity problems as well. Distance functions using normalized inputs and outputs could be subject to non-exogenous regressors. The mean-scaled model is less susceptible of endogeneity bias than the normal model, due to the fact that input and output variables often consist of several factors. Hence, Brümmer et al. (2002) argue that the problem is not likely to be more severe than in any production function type of study.

Overall it seems that the methods to handle the problems of monotonicity and endogeneity in the literature are of current debate. Given that most solutions to solve the existing problems are rather complicated and that time was limited, this study will not go into further details on how to deal with them.

9.4 Comparison of results with the literature

We find that the elasticity for milk output is approximately 0.80 percent for both of our models. Given that the sample only contains specialised full time dairy farmers whose output from milk is at least two thirds of the total output, this is not surprising. Common for the literature analysing efficiency among dairy farmers is that the elasticity for milk output is rather large. Sipiläinen (2007) also uses a multiple input distance function and finds that for the Finnish dairy farmers, the elasticity of milk is 0.54 whereas Newman & Matthews (2007) use an output distance function and find that the elasticity of milk is 0.77 for Irish dairy farmers. Overall, we see that the elasticity of milk output found in this study does not deviate from that found elsewhere.

The choice to include feed as a separate input variable was made because it is essential to animal production. Not surprisingly, we find that the elasticity for feed expenditure is by far the largest among all the input elasticities in our results with a value of 0.33 percent. Several of the studies regarding technical efficiency among milk producers have not included feed expenditure as a separate input variable; however, Lawson, Agger, et al. (2004) look at the effect of both concentrate feed and roughage feed for Danish dairy producers. They find that for concentrate the elasticity is 0.17 and for roughage 0.14. When combining the two we see that it is close to the result obtained from our estimations. As the type of feed can influence the milk yield, it would have been interesting to analyse the effects of different types of feed. However, different types of feed might violate the assumption of essentiality given that feed only in general is essential, but different, or more specialized types might not be. The relatively large elasticity of feed implies that the farmers have a larger potential of optimizing the production if adjusting the feed expenditure, than adjusting the other input variables. Furthermore, the use of feed can be considered a variable cost, and thus it is easier to adjust the quantity and kind of feed in the short run than other inputs like land or capital.

When looking at technical efficiency, the results presented in the literature are varying. Sipiläinen (2007) finds that the Finnish dairy farmers have experienced a decrease in the technical efficiency over time from 1989 to 2000. However, the development in the Finnish dairy sector is mainly because Finland joined the EU in 1995, and thus the dairy farmers became subject to the milk quota system. Lawson, Agger, et al. (2004) find that around 90 percent of the dairy farms analysed have a technical efficiency of 90 percent or more. Common for the studies mentioned here is that when the farmers are not 100 percent technical efficient there is increasing returns to scale at the mean as we also see from our estimations. This indicates that there is room for improvement and expansion. However, like Lawson, Agger, et al. (2004) we find that efficiency levels in general are very high and thus further improvement can mainly be obtained through technological change.

We find that our results in relation to production variables and elasticities are close to those presented in the literature regarding technical efficiency on dairy farms. We see that feed has a rather large impact on how efficiency can be improved compared to other inputs, which suggests for future studies to consider how different types of feed affect the milk yield, the milk quality, and the efficiency. Overall the results indicate that the farmers should consider new technologies like the AMS with caution, as the system requires more input and a skilled manager.

Furthermore, the milk producers should pay attention to the health of the animals, since they operate in a market close to perfect competition where even small improvements generate a gain.

9.5 Implications of results

The results obtained in this thesis have some interesting implications worth discussing. It appears that there are rather large productivity losses associated with the AMS compared to the other types of systems. This suggests that specialised milk producers, who consider shifting to the AMS, need to prepare carefully before investing in the system. Besides from requiring different management skills, the aspect of financing the new system can impact the choice to invest in the AMS. If the banks were to increase the interest rate for lending money, the economic gains, which one would expect, from replacing the old system with a new, might be reduced substantially. Non-scientific reports claim that the quality of the milk obtained from farms using the AMS is on average worse, and the farmers investing in the AMS should therefore be aware of the new challenges they can face. A productivity gain obtained from the AMS seems to be rather dependent on a skilled manager.

The higher price, which a good milk quality yields, has a very significant but small effect on the technical efficiency obtained by the farm. This is not surprising, since one should always try to maximize the value of the output, and our findings just support the documented tendency among the Danish dairy producers. However, if the price on milk changes, the incentive and the gain from achieving the supplement might change too. This would hardly affect efficiency, as the estimated effects are rather small. Nonetheless, the very significant effect of having good milk quality suggests that the Danish dairy farmers should keep aiming for low milk quality levels in the future.

The results also show that the type of breed can affect the productivity by more than 3 percent. Given that only around 11 percent of the farms in the data set use the jersey breed in their production, it seems that the sector as a whole could experience a rather large productivity gain, if more farmers where to replace the larger breeds with jersey cows. However the results obtained might not reflect the true effects of having jersey cows, and thus the true productivity gain might be smaller.

The very small effects of the diseases are also worth mentioning here. Based on the results obtained in this thesis, there seems to be no great effect of diseases on the technical efficiency among the Danish dairy farmers. This is rather surprising, but suggests that either the managers have had more focus on the matter or that the negative effect is overshadowed by the extra productivity. The findings can imply that the Danish dairy farmers are rather good at taking care of their herds and that securing the animals' health is considered to be of great importance, thus they do not need to improve much within this area. However, the results should be interpreted with caution since the small effect could be due to a lack of reported disorders and hence there might still be room for improvement in the handling of sick cows on the Danish dairy farms.

10. Conclusion

The objective of this study was to examine to what extent milk quality indicators and certain animal health indicators influence the technical efficiency of Danish dairy producers. The analysis was based on farm level data provided by SEGES from 2011 to 2015 and done using a stochastic distance frontier approach. After testing the data using a correlation analysis and an OLS estimation, different input and output distance functions were estimated as stochastic frontier models. The following investigations clarified that the input distance frontier seemed to be a better fit to the time period covered in the data.

We found that certain variables related to diseases in the dairy production did have a significant effect on technical efficiency. Having a high number of reported cases of mastitis and reproductive disorders in the herd lowers the technical efficiency on the farm. The effects of the milk quality indicators proved to be significant but small in relation to technical efficiency, and thus the farmers with a high milk quality has a higher technical efficiency.

Different production characteristics related to the technology set within dairy production were also examined. We found that the farms using the AMS are less productive, given that the AMS is known to create more managerial challenges than other more traditional milking systems. Furthermore, the results showed that there is a significant productivity loss from having large breed compared to jersey cows. This implies that the lower feed use and the higher energy content in the milk obtained from jersey cows can make up for the smaller yield.

Overall, the technical efficiency for the Danish dairy farmers has increased in the period from 2011 to 2015. This increase is compliant with the fact that specialized Danish dairy farms have been under economic pressure to increase efficiency by "running a bit faster" in the daily production routines. The results showed that there is still room for improvements in the future in relation to technical efficiency. Improvements can be done by changing the technology, securing fewer occurrences of disorders, a good milk quality, and by adjusting the use of feed. However, given that the Danish dairy farms are very efficient, and thereby not affected severely by challenges with efficiency, an increase in efficiency will require a change in the technology used, rather than more focus on the health of the animals and the milk quality.

11. Further Perspectives

Even though we consider the estimated results to be reasonable, the analysis could be developed further to obtain more knowledge on the topic of how milk quality indicators and different types of diseases influence a dairy farm's technical efficiency.

Additional analysis could include interaction terms between the production technology variables and the inefficiency variables. This might provide more knowledge on whether, and to what extent, different types of milking systems influence the milk quality or if the type of breed is more likely to suffer from various diseases. Further investigations could also examine in more depth whether the relatively low numbers of reported diseases are due to actual low levels of diseases in the Danish dairy herds, or if the numbers are more likely to reflect difficulties with correct reporting. Furthermore, it would be interesting to consider if the data, concerning milk quality indicators and reported diseases, is actually reported each month or if the year averages are more likely to only reflect a few months during the year. Another aspect could be to look into whether the milk quality and health indicators differ through different seasons.

As discussed earlier, it would also have made sense to estimate a model including prices if these had been present in the data. This would probably have helped to clarify whether the Danish dairy farmers are using their relatively high technical efficiency to ensure economic efficiency as well.

When prices are not available, further analysis would benefit from estimating a model taking into consideration the change in orientation that might have happened after the abolishment of the quota. A model allowing for both orientations to be present and for them to change over time might create a clearer picture of the farmers' optimizing behaviour related to the technical efficiency in the period.

Another aspect, which could benefit future analysis, is considering the possibility of imposing monotonicity and if potential problems with multicollinearity and endogeneity could be avoided in the estimated stochastic distance frontiers.

12. References

Aigner, D., Lovell, C.A.K. & Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6(1), pp.21–37.

Andersen, J.M. & Hansen, H.O., 2012. Landbrugets økonomi 2012. *University of Copenhagen*, pp.8–9.

Andersen, J.T. et al., 2016. *Produktionsøkonomi - Kvæg 2016*, SEGES P/S.

Andersen, J.T. et al., 2014. Sådan gør de bedste. *Videncentret for Landbrug P/S, Kvæg*.

Arla Foods amba, 2016. *Kvalitetsprogrammet Arlagården®*, Available at: https://www.arla.dk/globalassets/arla-dk/om-arla---oversigt/vores-ansvar/arlagarden/arlagarden-kvalitetsprogrammet-v.-5.0-januar-2016-dk.pdf [Accessed December 9, 2016].

Barnes, A.P. et al., 2011. The effect of lameness prevalence on technical efficiency at the dairy farm level: An adjusted data envelopment analysis approach. *American Dairy Science association*, 94, pp.5449–5457.

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

Battese, G.E. & Corra, G.S., 1977. Estimation of a production function model: with application to the pastoral zone of Eastern Australia. *Australian Journal of Agricultural Economics*, 21(3), pp.169–179.

Berndt, E.R. & Christensen, L.R., 1973. The translog function and the substitution of equipment, structures, and labor in U.S. manufacturing 1929-68. *Journal of Econometrics*, 1(1), pp.81–113.

Berthou, S., 2016. Økonomisk analyse. *Landbrug og Fødevarer*, (august), pp.4–5.

Beskyttelse, D., Målsætninger for malkekvæg. Available at: http://www.dyrenesbeskyttelse.dk/målsætninger-for-malkekvæg [Accessed February 10, 2017].

Bogetoft, P. & Otto, L., 2011. *Benchmarking with DEA, SFA, and R* 157th ed., Springer New York Dordrecht Heidelberg London.

Brümmer, B., Glauben, T. & Thijssen, G., 2002. Decomposition of productivity growth using distance functions: The case of dairy farms in three European countries. *American Journal of Agricultural Economics 84(3)*, 84(3), pp.628–644.

Chambers, R.G., 1994. *Applied production analysis - A dual approach.*, New York: Press Syndicate of the University of Cambridge.

Christensen, J. et al., 1990. *Managing Long-Term Developments of the Farm Firm*, Wissenschaftsverlag Vauk Kiel KG.

Coelli, T. & Henningsen, A., 2013. Stochastic Frontier Analysis. R package version 1.0. Available at: http://cran.r-project.org/package=frontier. [Accessed January 1, 2017].

Coelli, T., Singh, S. & Fleming, E., 2003. *AN INPUT DISTANCE FUNCTION APPROACH TO THE MEASUREMENT OF TECHNICAL AND ALLOCATIVE EFFICIENCY: WITH APPLICATION TO INDIAN DAIRY PROCESSING PLANTS*, Available at: https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=esam2003&paper_id=105.

Coelli, T.J. et al., 2005. *AN INTRODUCTION TO EFFICIENCY AND PRODUCTIVITY ANALYSIS* 2nd ed., Springer Science & Business Media.

European Commission - EU FADN, 2014. *EU dairy farms report 2013, based on FADN data*, Brussels, Belgium. Available at: http://ec.europa.eu/agriculture/rica/pdf/Dairy_Farms_report_2013_WEB.pdf.

Färe, R. & Vardanyan, M., 2016. A note on parameterizing input distance functions: does the choice of a functional form matter? *Journal of Productivity Analysis*, 45(2), pp.121–130.

Henningsen, A., 2014. Introduction to Econometric Production Analysis with R (Draft Version). Available at: https://files.itslearning.com/data/ku/103018/teaching/lecturenotes.pdf.

Henningsen, A. et al., 2016. The relationship between animal welfare and economic performance at farm level : A quantitative study of Danish pig producers.

Henningsen, A. & Henning, C.H.C.A., 2009. Imposing regional monotonicity on translog stochastic production frontiers with a simple three-step procedure. *Journal of Productivity Analysis*, 32(3), pp.217–229.

Kaiser, A.K., Kristensen, N.V. & Andersen, J.T., 2011. De bedste er bedre til at investere. *Kvæg*, pp.16–17.

Karagiannis, G., Midmore, P. & Tzouvelekas, V., 2004. PARAMETRIC DECOMPOSITION OF OUTPUT GROWTH USING A STOCHASTIC INPUT DISTANCE FUNCTION. *Amer. J. Agr. Econ.*, 86(4), pp.1044–1057.

Kumbhakar, S.C. et al., 2007. Do we estimate an input or an output distance function? An application of the mixture approach to European railways. *Journal of Productivity Analysis*, 27, pp.87–100.

Landbrug og Fødevarer, 2015. Europa står stærkest efter mælkekvoternes ophør. Available at: http://www.lf.dk/aktuelt/nyheder/2015/april/europa-staar-starkest-efter-maelkekvoternes-ophor#.V-0CemW7-t8 [Accessed September 29, 2016].

Landbrug og Fødevarer, 2013. Kvæg. Available at: http://www.lf.dk/viden-om/landbrugsproduktion/husdyr/kvag [Accessed October 6, 2016].

Landbrug og Fødevarer, 2016. *Statistik / Statistics 2015, Mejeri/dairy*,

Lawson, L.G., Agger, J.F., et al., 2004. Lameness, metabolic and digestive disorders, and technical efficiency in Danish dairy herds: A stochastic frontier production function approach. *Livestock Production Science*, 91(1–2), pp.157–172.

Lawson, L.G., Bruun, J., et al., 2004. Relationships of Efficiency to Reproductive Disorders in Danish Milk Production: A Stochastic Frontier Analysis. *Journal of Dairy Science*, 87(1), pp.212–224.

Martinussen, H. & Sørensen, M., 2015. *Produktionsøkonomi - KVÆG 2015*,

Meeusen, W. & van Den Broeck, J., 1977. Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. , 18(2), pp.435–444.

Newman, C. & Matthews, A., 2007. Evaluating the productivity performance of agricultural enterprises in Ireland using a multiple output distance function approach. *Journal of Agricultural Economics*, 58(1), pp.128–151.

Olsen, J. & Henningsen, A., 2011. *Investment utilisation, adjustment costs and technical efficiency in Danish pig farms*, Available at: https://core.ac.uk/download/pdf/6253952.pdf.

Olsen, J.V., 2016. Personal communication by email. *28.09.2016*.

Orea, L., Roibás, D. & Wall, A., 2004. Choosing the technical efficiency orientation to analyze firms' technology: a model selection test approach. *Journal of Productivity Analysis*, 22, pp.51–71.

Pedersen, H.B. et al., 2016. De bedste mælkeproducenter har overskud. *DST Analyse*, (november).

Philipp, P., 2016. Pressede mælkeproducenter er blevet mere effektive. *Landbrug og Fødevarer*.

Rasmussen, S., 2011. *Production Economics - The Basic Theory of Production Optimisation*, Springer-Verlag Berlin Heidelberg 2011.

Rasmussen, S., 2010. Scale efficiency in Danish agriculture: An input distance-function approach. *European Review of Agricultural Economics*, 37(3), pp.335–367.

Raundal, P., 2016. Personal interview with veterinarian from SEGES.

Sauer. J, Frohberg. K, Hockmann, H., 2006. Stochastic efficiency measurement: The curse of theoretical consistency. *Journal of Applied Economics*, IX(1), pp.139–165.

SEGES, 2013. Formler og konstanter i Økonomidatabasen, SEGES vedrørende 60-1.

SEGES P/S, Kvalitetsparametre. Available at: https://www.landbrugsinfo.dk/kvaeg/maelkekvalitet/kvalitetsparametre/sider/startside.aspx

[Accessed February 14, 2017].
SEGES P/S, 2016. *Prognose for kvægbrugets økonomiske resultater 2015-2017*,
Shephard, R.W., 1970. Theory of cost and production functions. In Princeton, N.J, pp. 64–78.
Sipiläinen, T., 2007. Sources of productivity growth on Finnish dairy farms—application of an input distance function. *Food Economics - Acta Agriculturae Scandinavica, Section C*, 4(2), pp.65–76.
Statistik, D., 2016a. ANI7. Available at: http://www.statistikbanken.dk/ANI71 [Accessed November 26, 2016].
Statistik, D., 2016b. Animalsk produktion 2. kvt. 2016. , (388). Available at: http://www.dst.dk/Site/Dst/Udgivelser/nyt/GetPdf.aspx?cid=20973 [Accessed September 29, 2016].
Statistik, D., 2016c. DNRUUPI. Available at: http://www.statistikbanken.dk/DNRUUPI [Accessed December 2, 2016].
Statistik, D., 2017a. Forbrugerprisindeks. Available at: http://www.dst.dk/da/Statistik/emner/forbrugerpriser/forbrugerprisindeks [Accessed January 12, 2017].
Statistik, D., 2016d. HDYR1. Available at: http://www.statistikbanken.dk/HDYR1 [Accessed December 5, 2016].
Statistik, D., 2016e. JORD2. Available at: http://www.statistikbanken.dk/JORD2 [Accessed December 2, 2016].
Statistik, D., 2016f. LPRIS25. Available at: http://www.statistikbanken.dk/LPRIS25 [Accessed December 8, 2016].
Statistik, D., 2017b. LPRIS35. Available at: http://www.statistikbanken.dk/LPRIS35 [Accessed January 11, 2017].
Statistik, D., 2015. *Nedgangen i antal malkekøer fortsaetter*, Available at: www.dst.dk/nyt/18808 [Accessed December 5, 2016].
Vidø, E. et al., 2014. Landbrugets økonomi 2014. *University of Copenhagen*, p.5.
Vidø, E. et al., 2016. Landbrugets økonomi 2016. *University of Copenhagen*, p.5.
van der Voort, M. et al., 2014. A stochastic frontier approach to study the relationship between gastrointestinal nematode infections and technical efficiency of dairy farms. *Journal of Dairy Science*, 97, pp.3498–3508.
Wiese, S., 2015. EU giver mælkeproducenter tre år til at betale superafgift. Available at: http://www.maskinbladet.dk/kvaegbrug/artikel/eu-giver-maelkeproducenter-tre-ar-betale-superafgift [Accessed January 30, 2017].

13. Appendices

13.1 Appendix A – Descriptions of the dataset

Table A.1: Production characteristics for organic and conventional farmers

	Organic	Conventional
Observations	1435	6763
Share of observations	0.18	0.82
Share of production characteristics		
Jersey	0.11	0.11
Large breed	0.89	0.89
AMS	0.24	0.25
Fishbone	0.51	0.49
Other systems	0.25	0.26

Source: Own calculations

Table A.2: Share of observations for the milk quality dummies

	Shares in the two groups	
	1: Price supplement	2: Deduction or no supplement
Cell	0.91	0.09
Viable	0.95	0.05
Spore	0.85	0.15

Source: Own calculations

Table A.3: Reported cases of the different disorders per cow

	Mean	Min	Max	Share with zero reported cases
Mastitis	0.11	0.00	0.87	2%
Hoof disorders	0.14	0.00	8.42	9%
Reproductive disorders	0.18	0.00	1.36	2%
Other disorders	0.03	0.00	1.80	23%

Source: Own calculations

Table A.4: The price index for the outputs

		Price index (2010=100)					
Output type	Output share	2010	2011	2012	2013	2014	2015
Milk output							
Total index for milk output	100%	100	107	105	118	121	97
Other output							
Cereal	25.69%	100	143	157	148	119	118
Seed	0.28%	100	108	115	115	115	114
Beets	0.21%	100	90	100	107	118	100
Potatoes	0.64%	100	104	99	121	150	154
Rape	0.99%	100	136	149	130	109	113
Peas	0.09%	100	111	156	118	105	103
Other crops	0.08%	100	115	126	131	116	108
Horticultural crops	0.05%	100	103	103	102	101	101
Energy crops	0.25%	100	154	164	172	180	189
Cattle	54.17%	100	118	134	135	123	127
Pigs	0.86%	100	109	122	123	113	101
Poultry	0.10%	100	118	124	132	120	114
Fur animals	0.14%	100	81	88	88	100	100
Sheep	0.05%	100	110	110	124	123	129
Other livestock	0.07%	100	118	134	135	123	127
Machine station	6.50%	100	101	105	107	108	110
Other agricultural income	9.35%	100	115	126	131	116	108
Total index for other output	100%	100	121.99	133.62	136.53	120.72	120.21

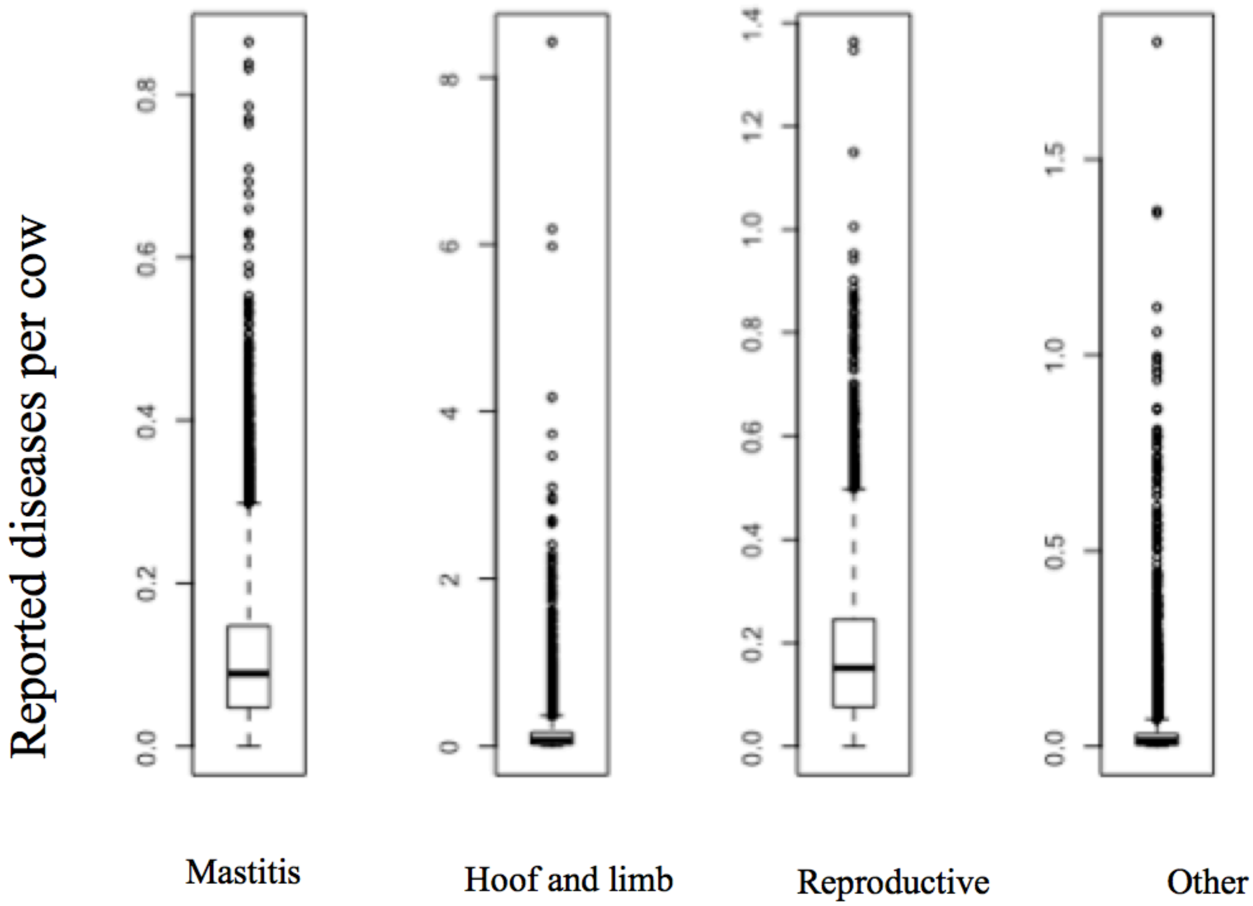
Source: (Statistik 2016f) and own calculations.

Table A.5: The price index for the inputs

		Price index (2010=100)					
Input variable	Share	2010	2011	2012	2013	2014	2015
Feed							
Cereal	14%	100	145	155	150	122	115
Readymix	73%	100	120	128	136	121	117
Other bought kinds	14%	100	125	140	153	152	155
Total index for feed expenditures	100%	100	125.4	134.74	134.74	126.69	123.21
Veterinary and medicine							
Total index for veterinary and medicine	100%	100	101	102	98	105	109
Wages							
Total index for wages	100%	100	0.99	0.97	0.98	0.98	0.99
Materials							
Sowing	9%	100	108	115	115	115	114
Fertilizer	8%	100	124	129	115	115	115
Crop protection	4%	100	92	94	126	191	194
Various, field	4%	100	107	111	119	137	150
Various, animal production	22%	100	101	103	106	108	100
Harvest of cereal and fodder	13%	100	102	105	105	103	102
Various machine	13%	100	101	105	107	108	110
Other costs	26%	100	111	117	121	119	118
Total index for materials	100%	100	106.1	110.3	112.9	116.3	115

Source: (Statistik 2016f), (Statistik 2017b),(Statistik 2017a), and own calculations.

Figure A.1: Boxplots of the different levels of the reported diseases per cow



Source: Own plots from R

13.2 Appendix B – Empirical Analysis and Estimation Results

Table B.6: Results of correlations analysis

	t-value	p-value	Correlation coefficient
The influence inefficiency indicators and production characteristics on GM per cow			
Mastitis	7.91	0.000	0.09
Hoof and limb disorders	7.14	0.000	0.08
Reproductive disorders	3.11	0.002	0.04
Other disorders	3.65	0.000	0.04
Cell 1	20.26	0.000	0.22
Viable 1	8.13	0.000	0.09
Spore 1	19.65	0.000	0.22
Jersey	9.87	0.000	0.11
Fishbone	2.26	0.024	0.03
Other systems	-6.58	0.000	-0.07
Organic	-1.12	0.263	-0.01
Manager age	-5.55	0.000	-0.06
Consultant	3.53	0.000	0.04

Source: Result from R.

13.2.1 Description of Equation B

We have used the same dependent variable as in the correlation analysis: the gross margin from milk per cow, which will be explained by the milk quality and animal health indicators. We also include the type of breed and milking system as these might have an influence on the productivity and thus the economic performance of the farms. This resulted in the following model specification of Equation B:

$$GM_{it} = \alpha_0 + \sum_{j=1}^{4} \alpha_{0+j} Regnskabsaar + \alpha_5 mastitis_{it} + \alpha_6 hoofdis_{it} + \alpha_7 reprodis_{it}$$

(B)

$$+\alpha_8 otherdis_{it} + \alpha_9 cell1_{it} + \alpha_{10} viable1_{it} + \alpha_{11} spore1_{it} + \alpha_{12} jersey_{it} + \alpha_{13} organic_{it}$$

$$+ \sum_{j=1}^{2} \alpha_{13+j} milkingsystem_{ijt} + \alpha_{16} consultant + \alpha_{17} managerage_{it} + \alpha_{18} managerage_{it}^2$$

$$+\varepsilon_i + a_{it}$$

Where α_j with $j = 0, \dots, 18$ are the parameters to be estimated, ε_i is a time independent noise term and a_{it} is a time dependent noise term which can take unobserved effects like management into account, “milkingsystem” represents the three dummy variables for the type of milking system used at the farm: “AMS”, “fishbone”, and “othersys” with the values of 1, 2, and 3 respectively. The dummy “jersey” is tested against the omitted dummy for the type of breed “large breed” and the dummies included for the milk quality indicators are tested against those who do not get an allowance in the price they receive for the milk they deliver. We also test if there is an effect of being organic compared to conventional and if the presence of a production consultant affects the GM. GM_{it} is the GM per cow, subscript i denotes the individual farm, and t denotes the year.

Table B.7: Results of the linear regression, Equation B

	Estimates	sd.err	p-value
(Intercept)	8424.87	664.47	0.000
Regnskabsaar2012	-55.01	84.50	0.515
Regnskabsaar2013	-278.99	83.36	0.001
Regnskabsaar2014	149.57	84.53	0.077
Regnskabsaar2015	971.91	87.71	0.000
mastitis	336.05	327.13	0.304
hoofdis	434.08	98.58	0.000
reprodis	338.72	221.88	0.127
otherdis	811.48	326.99	0.013
cell	1476.96	93.20	0.000
viable	633.98	116.08	0.000
spore	1190.03	75.89	0.000
jersey	1011.38	89.03	0.000
organic	-117.05	68.66	0.088
milkingsystem2	-248.59	64.22	0.000
milkingsystem3	-416.20	73.96	0.000
consultant	175.18	79.09	0.027
managerage	-9.58	26.40	0.717
I(managerage^2)	-0.11	0.27	0.675
R^2 (Adjusted R^2)	0.0.14363 (0.0.14328)		
F-statistics	71.962 (18,7723), p-value: 0.000		

Source: Model result from R.

For equation (B) we find that several of the estimated variables have a significance level of either 5 percent or 1 percent.

The effect of each year is significant. The years from 2012 and 2013 have negative coefficients, which indicate that the years have affected the gross margin per cow negatively compared to 2011, unlike 2014 and 2015 which were better years than 2011, when looking at the GM per cow. This can be explained by the fact that the Danish dairy farmers in general have gotten bigger and more specialised since 2011.

Looking at the health indicators, all but one is significant. All disorders have a positive effect on the GM per cow, which is different than what one might expect, but could be due to the fact that there has been more focus on breeding high yielding cows rather than robust cows, and hence the extra yield makes up for the additional occurrences of diseases (Beskyttelse n.d.). Among the milk quality indicators all three variables are highly significant. All three milk quality dummies have a positive effect on GM, which is what one would expect, since receiving a higher price on milk should increase the value of the gross output.

We also see that the type of breed matter. The dummy for jersey has a positive and significant effect on the GM per cow compared to large breed. Both the dummy for milking system 2 and 3 have a negative impact on the GM per cow compared to the AMS.

The age and squared age of the manager are not significant in the OLS estimation, which is a weak indication of that experience is not affecting the economic performance. From the F-statistic we see that all parameters together are significant on a 1 percent level. R^2 and adjusted R^2 are both found to be very low, which indicate that the data is fitted very poorly to the OLS model.

Overall the OLS estimates should be interpreted with caution. The linear regression model is always a good starting point for estimating a model, because it can be a supplement to the correlation analysis and provide an overview in which direction we might expect the variables to affect the explained variable, but in the case of panel data, it will most often provide biased and inconsistent estimates due to endogeneity.

Table B.3: Distance elasticities of TL IDF with zero violations of monotonicity

	Mean	St. Dev.	Violations of monotonicity
Milk output	-0.79	0.06	0.00
Other output	-0.17	0.05	0.00
Feed	0.33	0.06	0.00
Veterinary and medicine	0.03	0.01	0.00
Labour	0.09	0.03	0.00
Land	0.19	0.04	0.00
Capital	0.18	0.05	0.00
Materials	0.18	0.03	0.00
Mean Elasticity of Scale	1.06	0.11	

Source: Model result from R.

Table B.4: Estimated parameters of the TL IDF (TLidfres3)

Estimated Translog IDF with health and milk quality indicators

	Estimate	Std. Error	p-value
(Intercept)	0.04	0.01	0.000
log(grossmilkMS)	-0.77	0.00	0.000
log(grossotherMS)	-0.16	0.00	0.000
I(0.5 * log(grossmilkMS)^2)	-0.11	0.01	0.000
I(0.5 * log(grossotherMS)^2)	-0.07	0.00	0.000
I(log(grossmilkMS) * log(grossotherMS))	0.09	0.00	0.000
log(feedexpMS/materialsMS)	0.34	0.01	0.000
log(vetmedMS/materialsMS)	0.03	0.00	0.000
log(totalwagesMS/materialsMS)	0.09	0.00	0.000
log(landMS/materialsMS)	0.18	0.00	0.000
log(capitalMS/materialsMS)	0.17	0.00	0.000
I(0.5 * log(feedexpMS/materialsMS)^2)	0.24	0.02	0.000
I(0.5 * log(vetmedMS/materialsMS)^2)	0.03	0.01	0.000
I(0.5 * log(totalwagesMS/materialsMS)^2)	0.09	0.01	0.000
I(0.5 * log(landMS/materialsMS)^2)	0.02	0.00	0.000
I(0.5 * log(capitalMS/materialsMS)^2)	0.19	0.01	0.000
I(log(feedexpMS/materialsMS) * log(vetmedMS/materialsMS))	0.01	0.01	0.276
I(log(feedexpMS/materialsMS) * log(totalwagesMS/materialsMS))	-0.05	0.01	0.000
I(log(feedexpMS/materialsMS) * log(landMS/materialsMS))	-0.07	0.01	0.000
I(log(feedexpMS/materialsMS) * log(capitalMS/materialsMS))	-0.09	0.01	0.000
I(log(vetmedMS/materialsMS) * log(totalwagesMS/materialsMS))	0.01	0.01	0.261

I(log(vetmedMS/materialsMS) * log(landMS/materialsMS))	0.01	0.01	0.559
I(log(vetmedMS/materialsMS) * log(capitalMS/materialsMS))	-0.02	0.01	0.034
I(log(totalwagesMS/materialsMS) * log(landMS/materialsMS))	0.03	0.01	0.011
I(log(totalwagesMS/materialsMS) * log(capitalMS/materialsMS))	-0.03	0.01	0.001
I(log(landMS/materialsMS) * log(capitalMS/materialsMS))	-0.04	0.01	0.001
I(log(feedexpMS/materialsMS) * log(grossmilkMS))	-0.04	0.01	0.001
I(log(feedexpMS/materialsMS) * log(grossotherMS))	0.03	0.01	0.001
I(log(vetmedMS/materialsMS) * log(grossmilkMS))	-0.01	0.01	0.171
I(log(vetmedMS/materialsMS) * log(grossotherMS))	0.00	0.01	0.936
I(log(totalwagesMS/materialsMS) * log(grossmilkMS))	0.04	0.01	0.000
I(log(totalwagesMS/materialsMS) * log(grossotherMS))	-0.01	0.01	0.024
I(log(landMS/materialsMS) * log(grossmilkMS))	-0.04	0.01	0.000
I(log(landMS/materialsMS) * log(grossotherMS))	0.02	0.01	0.044
I(log(capitalMS/materialsMS) * log(grossmilkMS))	0.00	0.01	0.945
I(log(capitalMS/materialsMS) * log(grossotherMS))	-0.01	0.01	0.071
Regnskabsaar2012	0.01	0.00	0.009
Regnskabsaar2013	-0.05	0.00	0.000
Regnskabsaar2014	-0.04	0.00	0.000
Regnskabsaar2015	0.04	0.00	0.000
milkingsystem2	0.03	0.00	0.000
milkingsystem3	0.02	0.00	0.000
jersey	0.04	0.00	0.000
organic	0.00	0.00	0.336
Z_(Intercept)	0.14	0.02	0.000
Z_mastitis	0.09	0.04	0.015
Z_hoofdis	0.00	0.01	0.706
Z_reprodis	0.14	0.04	0.000
Z_otherdis	0.03	0.03	0.398
Z_cell1	-0.11	0.03	0.000
Z_viable1	-0.08	0.02	0.000
Z_spore1	-0.07	0.02	0.001
sigmaSq	0.02	0.00	0.000
gamma	0.56	0.09	0.000

panel data

number of cross-sections = 1810

number of time periods = 5

total number of observations = 8198

thus there are 852 observations not in the panel

mean efficiency of each year

2011	2012	2013	2014	2015
0.943	0.943	0.946	0.946	0.947

mean efficiency: 0.945

Source: Model result from R.

Table B.5: Estimated parameters of the TL ODF (TLodfres2)

Estimated Translog ODF with milk quality indicators, manager age, and consultant

	Estimate	Std. Error	p-value
(Intercept)	-0.03	0.01	0.000
log(grossotherMS/grossmilkMS)	0.16	0.00	0.000
I(0.5 * log(grossotherMS/grossmilkMS)^2)	0.07	0.00	0.000
log(feedexpMS)	-0.38	0.00	0.000
log(vetmedMS)	-0.02	0.00	0.000
log(totalwagesMS)	-0.10	0.00	0.000
log(landMS)	-0.17	0.01	0.000
log(capitalMS)	-0.19	0.00	0.000
log(materialsMS)	-0.17	0.01	0.000
I(0.5 * log(feedexpMS)^2)	-0.24	0.02	0.000
I(0.5 * log(vetmedMS)^2)	0.00	0.01	0.466
I(0.5 * log(totalwagesMS)^2)	-0.05	0.01	0.000
I(0.5 * log(landMS)^2)	-0.02	0.00	0.000
I(0.5 * log(capitalMS)^2)	-0.21	0.01	0.000
I(0.5 * log(materialsMS)^2)	-0.17	0.03	0.000
I(log(feedexpMS) * log(vetmedMS))	-0.01	0.01	0.214
I(log(feedexpMS) * log(totalwagesMS))	0.02	0.01	0.065
I(log(feedexpMS) * log(landMS))	0.05	0.01	0.000
I(log(feedexpMS) * log(capitalMS))	0.11	0.01	0.000
I(log(feedexpMS) * log(materialsMS))	0.06	0.02	0.002
I(log(vetmedMS) * log(totalwagesMS))	-0.01	0.01	0.120
I(log(vetmedMS) * log(landMS))	-0.01	0.01	0.220
I(log(vetmedMS) * log(capitalMS))	0.00	0.01	0.673
I(log(vetmedMS) * log(materialsMS))	0.04	0.01	0.005
I(log(totalwagesMS) * log(landMS))	-0.05	0.01	0.000
I(log(totalwagesMS) * log(capitalMS))	0.02	0.01	0.095
I(log(totalwagesMS) * log(materialsMS))	0.02	0.01	0.220
I(log(landMS) * log(capitalMS))	0.08	0.01	0.000
I(log(landMS) * log(materialsMS))	0.00	0.02	0.958
I(log(capitalMS) * log(materialsMS))	0.02	0.02	0.313
I(log(feedexpMS) * log(grossotherMS/grossmilkMS))	-0.03	0.01	0.000
I(log(vetmedMS) * log(grossotherMS/grossmilkMS))	0.00	0.01	0.560
I(log(totalwagesMS) * log(grossotherMS/grossmilkMS))	0.01	0.01	0.021
I(log(landMS) * log(grossotherMS/grossmilkMS))	-0.02	0.01	0.022
I(log(capitalMS) * log(grossotherMS/grossmilkMS))	0.00	0.01	0.741
I(log(materialsMS) * log(grossotherMS/grossmilkMS))	0.02	0.01	0.144

Regnskabsaar2012	-0.01	0.00	0.014
Regnskabsaar2013	0.06	0.00	0.000
Regnskabsaar2014	0.05	0.00	0.000
Regnskabsaar2015	-0.05	0.00	0.000
milkingsystem2	-0.02	0.00	0.000
milkingsystem3	-0.01	0.00	0.000
jersey	-0.04	0.00	0.000
organic	0.00	0.00	0.461
Z_(Intercept)	0.03	0.06	0.649
Z_cell1	-0.19	0.04	0.000
Z_viable1	-0.11	0.02	0.000
Z_spore1	-0.13	0.03	0.000
Z_managerage	0.00	0.00	0.000
Z_consultant	0.00	0.01	0.728
sigmaSq	0.02	0.00	0.000
gamma	0.69	0.05	0.000

panel data
number of cross-sections = 1810
number of time periods = 5
total number of observations = 8198
thus there are 852 observations not in the panel
mean efficiency of each year

2011	2012	2013	2014	2015
0.942	0.942	0.945	0.945	0.946

mean efficiency: 0.944

Source: Model result from R.

Table B.8: Average input and output shares from 2011-2015

Output	2011	2012	2013	2014	2015	Average share from 2011 to 2015
Milk	0.83	0.84	0.86	0.86	0.84	0.85
Other	0.17	0.16	0.14	0.14	0.16	0.15
Input						
Feed expenditure	0.22	0.23	0.25	0.24	0.24	0.24
Veterinary and medicine	0.03	0.03	0.03	0.02	0.02	0.03
Labour	0.17	0.17	0.17	0.18	0.18	0.17
Land	0.00	0.00	0.00	0.00	0.00	0.00
Capital	0.33	0.32	0.31	0.31	0.30	0.31
Materials	0.26	0.26	0.25	0.25	0.25	0.26

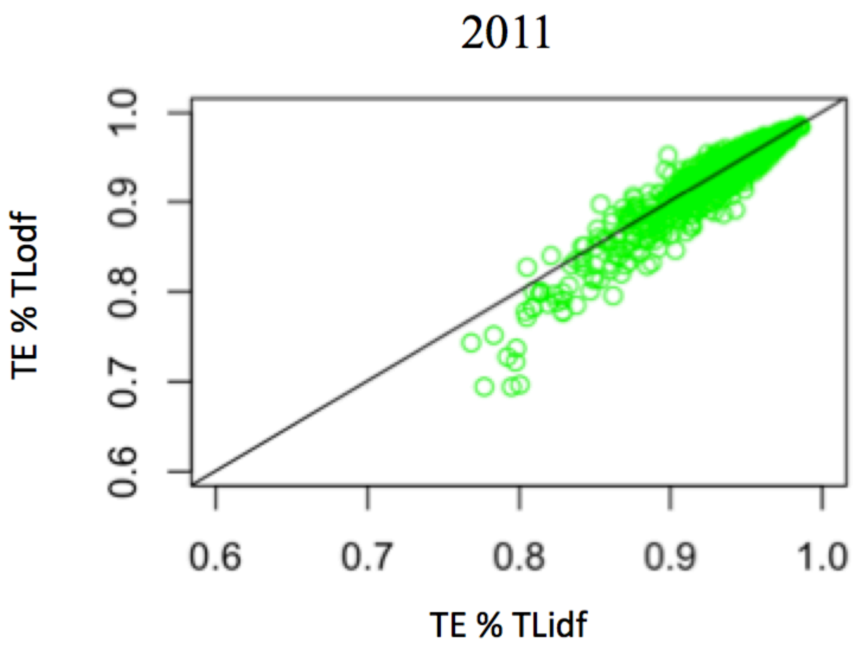
Source: Own calculations

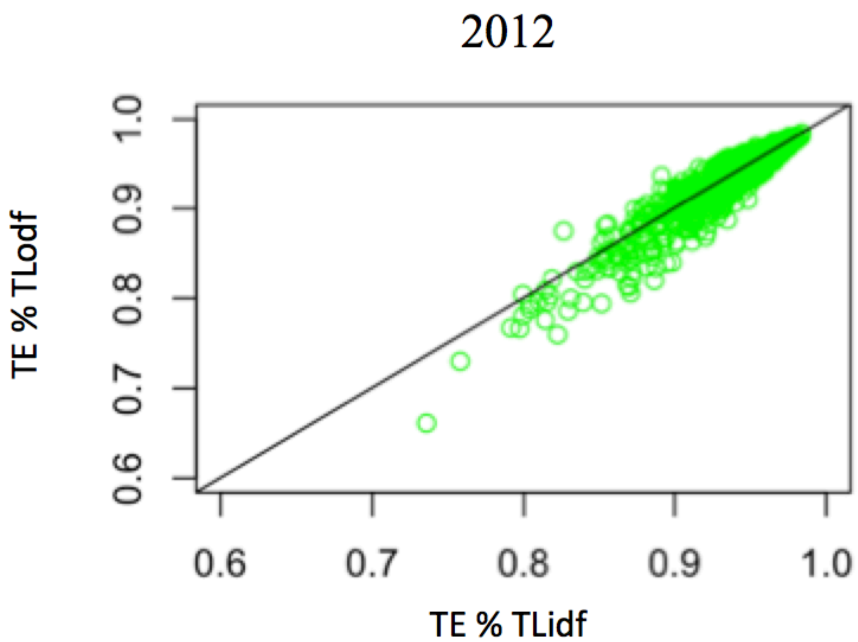
Table B.9: Average input and output costs from 2011-2015 (2010-values)

	2011	2012	2013	2014	2015
Milk	3757269.0	3944707.0	4150389.0	4242871.0	4579131.0
Other	764388.0	760526.8	633918.4	667035.8	872933.2
Feed expenditure	1023408.0	1087124.0	1259042.0	1238759.0	1269001.0
Veterinary and medicine	117526.4	115890.4	121031.8	119066.7	115840.6
Labour	752285.4	794242.1	818088.5	894183.4	917539.6
Land	157.2	160.4	166.4	171.4	174.5
Capital	1500271.0	1487390.0	1503115.0	1548202.0	1531278.0
Materials	1151687.0	1174796.0	1195350.0	1220485.0	1234701.0

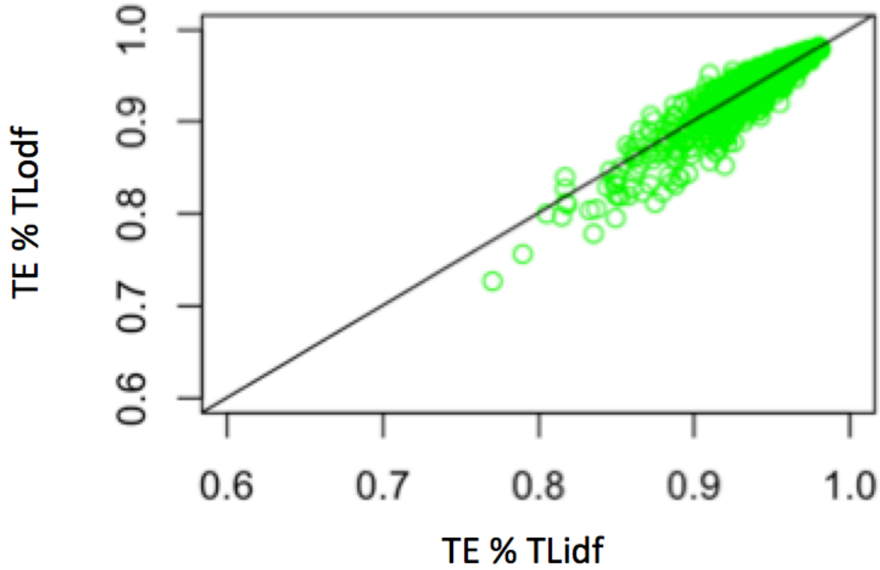
Source: Own calculations

Figure B.2: Comparison of technical efficiency between the TL IDF and the TL ODF from 2011-2015

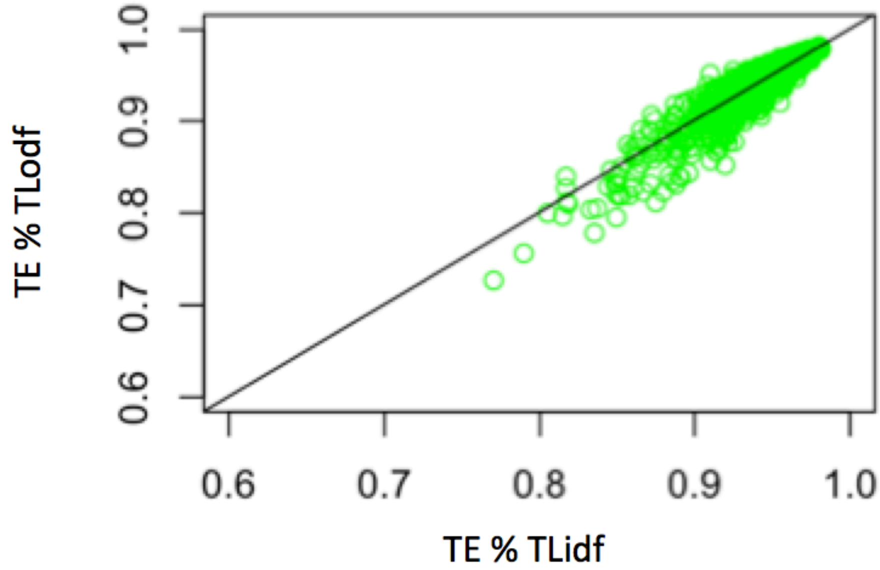




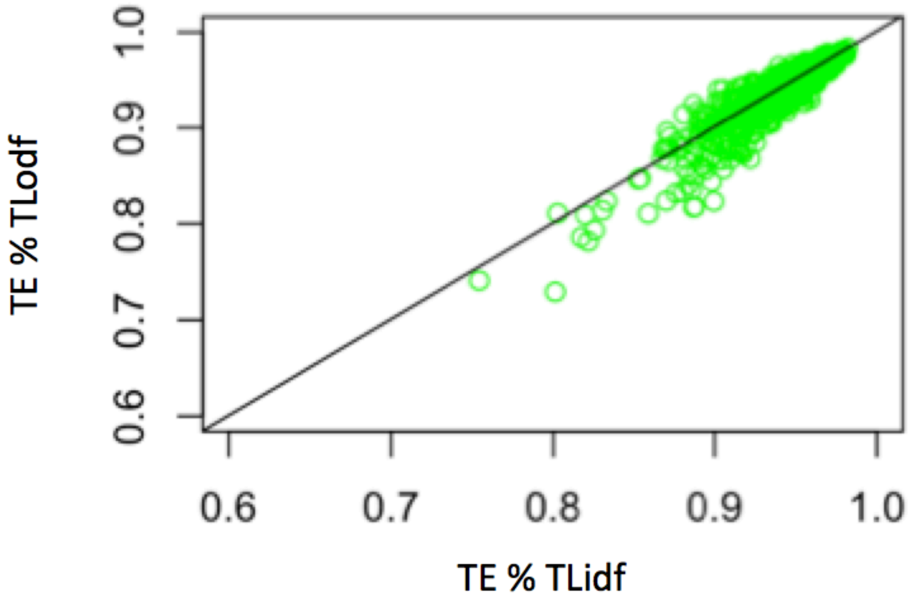
2013


TE % TLodf
1.0
0.9
0.8
0.7
0.6
0.6
0.7
0.8
0.9
1.0
TE % TLidf

2014


TE % TLodf
1.0
0.9
0.8
0.7
0.6
0.6
0.7
0.8
0.9
1.0
TE % TLidf

2015



Source: Own plots from R

13.3 Appendix C – Summary Statistics

Table C.10: Estimated parameters of the unrestricted Cobb-Douglas IDF (CDidfunres)

	Estimate	Std. Error	p-value
(Intercept)	0.04	0.00	0.000
log(grossmilkMS)	-0.81	0.00	0.000
log(grossotherMS)	-0.12	0.00	0.000
log(feedexpMS/materialsMS)	0.37	0.00	0.000
log(vetmedMS/materialsMS)	0.04	0.00	0.000
log(totalwagesMS/materialsMS)	0.07	0.00	0.000
log(landMS/materialsMS)	0.14	0.00	0.000
log(capitalMS/materialsMS)	0.16	0.00	0.000
Regnskabsaar2012	0.01	0.00	0.000
Regnskabsaar2013	-0.04	0.00	0.000
Regnskabsaar2014	-0.03	0.00	0.000
Regnskabsaar2015	0.05	0.00	0.000
milkingsystem2	0.02	0.00	0.000
milkingsystem3	0.02	0.00	0.000
jersey	0.03	0.00	0.000
organic	0.00	0.00	0.846
Z_(Intercept)	0.08	0.01	0.000
Z_mastitis	0.07	0.02	0.000
Z_hoofdis	0.00	0.00	0.984
Z_reprodis	0.09	0.02	0.000
Z_otherdis	0.02	0.02	0.461
Z_cell1	-0.04	0.00	0.000
Z_viable1	-0.03	0.01	0.000
Z_spore1	-0.02	0.00	0.000
Z_consultant	0.01	0.00	0.032
Z_managerage	0.00	0.00	0.047
sigmaSq	0.01	0.00	0.000
gamma	0.00	0.01	0.973

panel data
number of cross-sections = 1810
number of time periods = 5
total number of observations = 8198
thus there are 852 observations not in the panel
mean efficiency of each year

2011	2012	2013	2014	2015
0.954	0.950	0.953	0.951	0.952

mean efficiency: 0.952

Source: Model result from R.

Table C.11: Estimated parameters of the unrestricted Translog IDF (TLidfunres)

	Estimate	Std. Error	p-value
(Intercept)	0.04	0.01	0.000
log(grossmilkMS)	-0.77	0.00	0.000
log(grossotherMS)	-0.16	0.00	0.000
I(0.5 * log(grossmilkMS)^2)	-0.11	0.01	0.000
I(0.5 * log(grossotherMS)^2)	-0.07	0.00	0.000
I(log(grossmilkMS) * log(grossotherMS))	0.09	0.00	0.000
log(feedexpMS/materialsMS)	0.34	0.01	0.000
log(vetmedMS/materialsMS)	0.04	0.00	0.000
log(totalwagesMS/materialsMS)	0.09	0.00	0.000
log(landMS/materialsMS)	0.18	0.01	0.000
log(capitalMS/materialsMS)	0.17	0.00	0.000
I(0.5 * log(feedexpMS/materialsMS)^2)	0.24	0.02	0.000
I(0.5 * log(vetmedMS/materialsMS)^2)	0.03	0.01	0.000
I(0.5 * log(totalwagesMS/materialsMS)^2)	0.09	0.01	0.000
I(0.5 * log(landMS/materialsMS)^2)	0.01	0.00	0.000
I(0.5 * log(capitalMS/materialsMS)^2)	0.19	0.01	0.000
I(log(feedexpMS/materialsMS) * log(vetmedMS/materialsMS))	0.01	0.01	0.268
I(log(feedexpMS/materialsMS) * log(totalwagesMS/materialsMS))	-0.05	0.01	0.000
I(log(feedexpMS/materialsMS) * log(landMS/materialsMS))	-0.07	0.01	0.000
I(log(feedexpMS/materialsMS) * log(capitalMS/materialsMS))	-0.09	0.01	0.000
I(log(vetmedMS/materialsMS) * log(totalwagesMS/materialsMS))	0.01	0.01	0.263
I(log(vetmedMS/materialsMS) * log(landMS/materialsMS))	0.00	0.01	0.621
I(log(vetmedMS/materialsMS) * log(capitalMS/materialsMS))	-0.02	0.01	0.023
I(log(totalwagesMS/materialsMS) * log(landMS/materialsMS))	0.03	0.01	0.010
I(log(totalwagesMS/materialsMS) * log(capitalMS/materialsMS))	-0.03	0.01	0.001
I(log(landMS/materialsMS) * log(capitalMS/materialsMS))	-0.04	0.01	0.001
I(log(feedexpMS/materialsMS) * log(grossmilkMS))	-0.04	0.01	0.002
I(log(feedexpMS/materialsMS) * log(grossotherMS))	0.03	0.01	0.001
I(log(vetmedMS/materialsMS) * log(grossmilkMS))	-0.01	0.01	0.210
I(log(vetmedMS/materialsMS) * log(grossotherMS))	0.00	0.01	0.876
I(log(totalwagesMS/materialsMS) * log(grossmilkMS))	0.04	0.01	0.000
I(log(totalwagesMS/materialsMS) * log(grossotherMS))	-0.01	0.01	0.027
I(log(landMS/materialsMS) * log(grossmilkMS))	-0.04	0.01	0.000
I(log(landMS/materialsMS) * log(grossotherMS))	0.01	0.01	0.073
I(log(capitalMS/materialsMS) * log(grossmilkMS))	0.00	0.01	0.992
I(log(capitalMS/materialsMS) * log(grossotherMS))	-0.01	0.01	0.088

Regnskabsaar2012	0.01	0.00	0.002
Regnskabsaar2013	-0.05	0.00	0.000
Regnskabsaar2014	-0.04	0.00	0.000
Regnskabsaar2015	0.05	0.00	0.000
milkingsystem2	0.03	0.00	0.000
milkingsystem3	0.02	0.00	0.000
jersey	0.03	0.00	0.000
organic	0.00	0.00	0.264
Z_(Intercept)	0.09	0.03	0.005
Z_mastitis	0.07	0.03	0.029
Z_hoofdis	0.00	0.01	0.941
Z_reprodis	0.11	0.03	0.000
Z_otherdis	0.03	0.03	0.376
Z_cell1	-0.07	0.02	0.003
Z_viable1	-0.06	0.02	0.002
Z_spore1	-0.05	0.02	0.011
Z_managerage	0.00	0.00	0.022
Z_consultant	0.02	0.01	0.045
sigmaSq	0.01	0.00	0.000
gamma	0.42	0.12	0.001

panel data
number of cross-sections = 1810
number of time periods = 5
total number of observations = 8198
thus there are 852 observations not in the panel
mean efficiency of each year

2011	2012	2013	2014	2015
0.938	0.937	0.940	0.939	0.940

mean efficiency: 0.939

Source: Model result from R.

Table C.12: Estimated parameters of the restricted Translog IDF with indicators of diseases (TLidfres1)

	Estimate	Std. Error	p-value
(Intercept)	0.02	0.00	0.000
log(grossmilkMS)	-0.77	0.00	0.000
log(grossotherMS)	-0.16	0.00	0.000
I(0.5 * log(grossmilkMS)^2)	-0.11	0.01	0.000
I(0.5 * log(grossotherMS)^2)	-0.07	0.00	0.000
I(log(grossmilkMS) * log(grossotherMS))	0.09	0.00	0.000
log(feedexpMS/materialsMS)	0.34	0.00	0.000
log(vetmedMS/materialsMS)	0.03	0.00	0.000
log(totalwagesMS/materialsMS)	0.09	0.00	0.000
log(landMS/materialsMS)	0.18	0.00	0.000
log(capitalMS/materialsMS)	0.17	0.00	0.000
I(0.5 * log(feedexpMS/materialsMS)^2)	0.24	0.02	0.000
I(0.5 * log(vetmedMS/materialsMS)^2)	0.03	0.01	0.000
I(0.5 * log(totalwagesMS/materialsMS)^2)	0.09	0.01	0.000
I(0.5 * log(landMS/materialsMS)^2)	0.01	0.00	0.001
I(0.5 * log(capitalMS/materialsMS)^2)	0.20	0.01	0.000
I(log(feedexpMS/materialsMS) * log(vetmedMS/materialsMS))	0.02	0.01	0.119
I(log(feedexpMS/materialsMS) * log(totalwagesMS/materialsMS))	-0.05	0.01	0.000
I(log(feedexpMS/materialsMS) * log(landMS/materialsMS))	-0.06	0.01	0.000
I(log(feedexpMS/materialsMS) * log(capitalMS/materialsMS))	-0.09	0.01	0.000
I(log(vetmedMS/materialsMS) * log(totalwagesMS/materialsMS))	0.01	0.01	0.130
I(log(vetmedMS/materialsMS) * log(landMS/materialsMS))	0.00	0.01	0.661
I(log(vetmedMS/materialsMS) * log(capitalMS/materialsMS))	-0.02	0.01	0.013
I(log(totalwagesMS/materialsMS) * log(landMS/materialsMS))	0.03	0.01	0.008
I(log(totalwagesMS/materialsMS) * log(capitalMS/materialsMS))	-0.03	0.01	0.001
I(log(landMS/materialsMS) * log(capitalMS/materialsMS))	-0.04	0.01	0.000
I(log(feedexpMS/materialsMS) * log(grossmilkMS))	-0.04	0.01	0.001
I(log(feedexpMS/materialsMS) * log(grossotherMS))	0.03	0.01	0.000
I(log(vetmedMS/materialsMS) * log(grossmilkMS))	-0.01	0.01	0.084
I(log(vetmedMS/materialsMS) * log(grossotherMS))	0.00	0.01	0.739
I(log(totalwagesMS/materialsMS) * log(grossmilkMS))	0.04	0.01	0.000
I(log(totalwagesMS/materialsMS) * log(grossotherMS))	-0.01	0.01	0.019
I(log(landMS/materialsMS) * log(grossmilkMS))	-0.04	0.01	0.000
I(log(landMS/materialsMS) * log(grossotherMS))	0.02	0.01	0.025
I(log(capitalMS/materialsMS) * log(grossmilkMS))	0.00	0.01	0.787
I(log(capitalMS/materialsMS) * log(grossotherMS))	-0.02	0.01	0.040

Regnskabsaar2012	0.01	0.00	0.016
Regnskabsaar2013	-0.05	0.00	0.000
Regnskabsaar2014	-0.04	0.00	0.000
Regnskabsaar2015	0.05	0.00	0.000
milkingsystem2	0.03	0.00	0.000
milkingsystem3	0.02	0.00	0.000
jersey	0.04	0.00	0.000
organic	0.00	0.00	0.213
Z_(Intercept)	-3.67	0.60	0.000
Z_mastitis	0.27	0.21	0.212
Z_hoofdis	-0.26	0.04	0.000
Z_reprodis	1.40	0.29	0.000
Z_otherdis	-0.65	0.11	0.000
Z_managerage	0.01	0.00	0.000
Z_consultant	0.37	0.06	0.000
sigmaSq	0.13	0.02	0.000
gamma	0.95	0.01	0.000

panel data
number of cross-sections = 1810
number of time periods = 5
total number of observations = 8198
thus there are 852 observations not in the panel
mean efficiency of each year

2011	2012	2013	2014	2015
0.954	0.955	0.955	0.955	0.955

mean efficiency: 0.955

Source: Model result from R.

Table C.13: Estimated parameters of the restricted Translog IDF with milk quality indicators (TLidfres2)

	Estimate	Std. Error	p-value
(Intercept)	0.03	0.00	0.000
log(grossmilkMS)	-0.77	0.00	0.000
log(grossotherMS)	-0.16	0.00	0.000
I(0.5 * log(grossmilkMS)^2)	-0.11	0.01	0.000
I(0.5 * log(grossotherMS)^2)	-0.07	0.00	0.000
I(log(grossmilkMS) * log(grossotherMS))	0.09	0.00	0.000
log(feedexpMS/materialsMS)	0.34	0.00	0.000
log(vetmedMS/materialsMS)	0.03	0.00	0.000
log(totalwagesMS/materialsMS)	0.09	0.00	0.000
log(landMS/materialsMS)	0.19	0.00	0.000
log(capitalMS/materialsMS)	0.17	0.00	0.000
I(0.5 * log(feedexpMS/materialsMS)^2)	0.24	0.02	0.000
I(0.5 * log(vetmedMS/materialsMS)^2)	0.03	0.01	0.000
I(0.5 * log(totalwagesMS/materialsMS)^2)	0.09	0.01	0.000
I(0.5 * log(landMS/materialsMS)^2)	0.02	0.00	0.000
I(0.5 * log(capitalMS/materialsMS)^2)	0.19	0.01	0.000
I(log(feedexpMS/materialsMS) * log(vetmedMS/materialsMS))	0.01	0.01	0.214
I(log(feedexpMS/materialsMS) * log(totalwagesMS/materialsMS))	-0.05	0.01	0.000
I(log(feedexpMS/materialsMS) * log(landMS/materialsMS))	-0.07	0.01	0.000
I(log(feedexpMS/materialsMS) * log(capitalMS/materialsMS))	-0.09	0.01	0.000
I(log(vetmedMS/materialsMS) * log(totalwagesMS/materialsMS))	0.01	0.01	0.354
I(log(vetmedMS/materialsMS) * log(landMS/materialsMS))	0.01	0.01	0.386
I(log(vetmedMS/materialsMS) * log(capitalMS/materialsMS))	-0.02	0.01	0.027
I(log(totalwagesMS/materialsMS) * log(landMS/materialsMS))	0.03	0.01	0.013
I(log(totalwagesMS/materialsMS) * log(capitalMS/materialsMS))	-0.03	0.01	0.003
I(log(landMS/materialsMS) * log(capitalMS/materialsMS))	-0.04	0.01	0.001
I(log(feedexpMS/materialsMS) * log(grossmilkMS))	-0.04	0.01	0.001
I(log(feedexpMS/materialsMS) * log(grossotherMS))	0.03	0.01	0.000
I(log(vetmedMS/materialsMS) * log(grossmilkMS))	-0.01	0.01	0.298
I(log(vetmedMS/materialsMS) * log(grossotherMS))	0.00	0.01	0.845
I(log(totalwagesMS/materialsMS) * log(grossmilkMS))	0.04	0.01	0.000
I(log(totalwagesMS/materialsMS) * log(grossotherMS))	-0.01	0.01	0.021
I(log(landMS/materialsMS) * log(grossmilkMS))	-0.04	0.01	0.000
I(log(landMS/materialsMS) * log(grossotherMS))	0.02	0.01	0.041
I(log(capitalMS/materialsMS) * log(grossmilkMS))	0.00	0.01	0.957
I(log(capitalMS/materialsMS) * log(grossotherMS))	-0.01	0.01	0.060

Regnskabsaar2012	0.01	0.00	0.010
Regnskabsaar2013	-0.05	0.00	0.000
Regnskabsaar2014	-0.04	0.00	0.000
Regnskabsaar2015	0.04	0.00	0.000
milkingsystem2	0.03	0.00	0.000
milkingsystem3	0.02	0.00	0.000
jersey	0.04	0.00	0.000
organic	0.00	0.00	0.200
Z_(Intercept)	-0.02	0.09	0.808
Z_cell1	-0.21	0.04	0.000
Z_viable1	-0.15	0.04	0.000
Z_spore1	-0.15	0.04	0.000
Z_managerage	0.00	0.00	0.033
Z_consultant	0.05	0.02	0.011
sigmaSq	0.03	0.00	0.000
gamma	0.73	0.05	0.000

panel data
number of cross-sections = 1810
number of time periods = 5
total number of observations = 8198
thus there are 852 observations not in the panel
mean efficiency of each year

2011	2012	2013	2014	2015
0.951	0.951	0.954	0.954	0.955

mean efficiency: 0.953

Source: Model result from R.

Table C.14: Estimated parameters of the restricted Translog IDF with milk quality indicators. indicators of diseases and "hours" (TLidfhourshours)

	Estimate	Std. Error	p-value
(Intercept)	0.03	0.00	0.000
log(grossmilkMS)	-0.76	0.00	0.000
log(grossotherMS)	-0.13	0.00	0.000
I(0.5 * log(grossmilkMS)^2)	-0.09	0.01	0.000
I(0.5 * log(grossotherMS)^2)	-0.06	0.00	0.000
I(log(grossmilkMS) * log(grossotherMS))	0.08	0.00	0.000
log(feedexpMS/materialsMS)	0.30	0.01	0.000
log(vetmedMS/materialsMS)	0.02	0.00	0.000
log(hoursMS/materialsMS)	0.28	0.01	0.000
log(landMS/materialsMS)	0.08	0.01	0.000
log(capitalMS/materialsMS)	0.16	0.00	0.000
I(0.5 * log(feedexpMS/materialsMS)^2)	0.15	0.02	0.000
I(0.5 * log(vetmedMS/materialsMS)^2)	0.02	0.01	0.003
I(0.5 * log(hoursMS/materialsMS)^2)	0.62	0.03	0.000
I(0.5 * log(landMS/materialsMS)^2)	0.01	0.00	0.102
I(0.5 * log(capitalMS/materialsMS)^2)	0.16	0.01	0.000
I(log(feedexpMS/materialsMS) * log(vetmedMS/materialsMS))	0.03	0.01	0.009
I(log(feedexpMS/materialsMS) * log(hoursMS/materialsMS))	-0.15	0.02	0.000
I(log(feedexpMS/materialsMS) * log(landMS/materialsMS))	-0.03	0.01	0.050
I(log(feedexpMS/materialsMS) * log(capitalMS/materialsMS))	-0.04	0.01	0.009
I(log(vetmedMS/materialsMS) * log(hoursMS/materialsMS))	-0.04	0.02	0.007
I(log(vetmedMS/materialsMS) * log(landMS/materialsMS))	0.02	0.01	0.090
I(log(vetmedMS/materialsMS) * log(capitalMS/materialsMS))	-0.01	0.01	0.061
I(log(hoursMS/materialsMS) * log(landMS/materialsMS))	-0.06	0.02	0.002
I(log(hoursMS/materialsMS) * log(capitalMS/materialsMS))	-0.11	0.02	0.000
I(log(landMS/materialsMS) * log(capitalMS/materialsMS))	-0.01	0.01	0.358
I(log(feedexpMS/materialsMS) * log(grossmilkMS))	-0.01	0.01	0.279
I(log(feedexpMS/materialsMS) * log(grossotherMS))	0.02	0.01	0.014
I(log(vetmedMS/materialsMS) * log(grossmilkMS))	0.00	0.01	0.550
I(log(vetmedMS/materialsMS) * log(grossotherMS))	-0.01	0.01	0.282
I(log(hoursMS/materialsMS) * log(grossmilkMS))	0.05	0.02	0.002
I(log(hoursMS/materialsMS) * log(grossotherMS))	-0.01	0.01	0.487
I(log(landMS/materialsMS) * log(grossmilkMS))	-0.01	0.01	0.518
I(log(landMS/materialsMS) * log(grossotherMS))	0.01	0.01	0.137
I(log(capitalMS/materialsMS) * log(grossmilkMS))	-0.04	0.01	0.000
I(log(capitalMS/materialsMS) * log(grossotherMS))	-0.01	0.01	0.059
Regnskabsaar2012	-0.01	0.00	0.012
Regnskabsaar2013	-0.06	0.00	0.000

Regnskabsaar2014	-0.04	0.00	0.000
Regnskabsaar2015	0.03	0.00	0.000
milkingsystem2	0.01	0.00	0.014
milkingsystem3	-0.01	0.00	0.001
jersey	0.03	0.00	0.000
organic	-0.01	0.00	0.027
Z_(Intercept)	-0.13	0.06	0.024
Z_mastitis	0.20	0.09	0.025
Z_hoofdis	-0.70	0.20	0.000
Z_reprodis	0.39	0.19	0.039
Z_otherdis	-0.41	0.11	0.000
Z_cell1	-0.90	0.22	0.000
Z_viable1	-0.39	0.10	0.000
Z_spore1	-0.85	0.24	0.000
sigmaSq	0.09	0.02	0.000
gamma	0.93	0.02	0.000

panel data
number of cross-sections = 1810
number of time periods = 5
total number of observations = 8198
thus there are 852 observations not in the panel
mean efficiency of each year

2011	2012	2013	2014	2015
0.958	0.959	0.962	0.962	0.963

mean efficiency: 0.961

Source: Model result from R.

Table C.6: Estimated parameters of the unrestricted Cobb-Douglas ODF (CDodfunres)

	Estimate	Std. Error	p-value
(Intercept)	-0.05	0.01	0.000
log(grossotherMS/grossmilkMS)	0.13	0.00	0.000
log(feedexpMS)	-0.48	0.00	0.000
log(vetmedMS)	-0.04	0.00	0.000
log(totalwagesMS)	-0.08	0.00	0.000
log(landMS)	-0.20	0.00	0.000
log(capitalMS)	-0.18	0.00	0.000
Regnskabsaar2012	-0.01	0.00	0.017
Regnskabsaar2013	0.06	0.00	0.000
Regnskabsaar2014	0.05	0.00	0.000
Regnskabsaar2015	-0.04	0.00	0.000
milkingsystem2	-0.03	0.00	0.000
milkingsystem3	-0.02	0.00	0.000
jersey	-0.04	0.00	0.000
organic	-0.01	0.00	0.114
Z_(Intercept)	0.08	0.02	0.000
Z_mastitis	0.10	0.04	0.011
Z_hoofdis	-0.22	0.06	0.001
Z_reprodis	0.03	0.02	0.134
Z_otherdis	0.03	0.03	0.321
Z_cell1	-0.11	0.02	0.000
Z_viable1	-0.05	0.01	0.000
Z_spore1	-0.08	0.01	0.000
Z_consultant	-0.02	0.01	0.027
Z_managerage	0.00	0.00	0.000
sigmaSq	0.02	0.00	0.000
gamma	0.27	0.04	0.000

panel data
number of cross-sections = 1810
number of time periods = 5
total number of observations = 8198
thus there are 852 observations not in the panel
mean efficiency of each year

2011	2012	2013	2014	2015
0.951	0.951	0.954	0.953	0.955

mean efficiency: 0.953

Source: Model result from R.

Table C.7: Estimated parameters of the unrestricted Translog ODF (TLodfunres)

	Estimate	Std. Error	p-value
(Intercept)	-0.03	0.00	0.000
log(grossotherMS/grossmilkMS)	0.16	0.00	0.000
I(0.5 * log(grossotherMS/grossmilkMS)^2)	0.07	0.00	0.000
log(feedexpMS)	-0.38	0.00	0.000
log(vetmedMS)	-0.02	0.00	0.000
log(totalwagesMS)	-0.10	0.00	0.000
log(landMS)	-0.17	0.01	0.000
log(capitalMS)	-0.19	0.00	0.000
log(materialsMS)	-0.17	0.01	0.000
I(0.5 * log(feedexpMS)^2)	-0.24	0.02	0.000
I(0.5 * log(vetmedMS)^2)	-0.01	0.01	0.425
I(0.5 * log(totalwagesMS)^2)	-0.05	0.01	0.000
I(0.5 * log(landMS)^2)	-0.02	0.00	0.000
I(0.5 * log(capitalMS)^2)	-0.21	0.01	0.000
I(0.5 * log(materialsMS)^2)	-0.17	0.03	0.000
I(log(feedexpMS) * log(vetmedMS))	-0.01	0.01	0.210
I(log(feedexpMS) * log(totalwagesMS))	0.02	0.01	0.081
I(log(feedexpMS) * log(landMS))	0.05	0.01	0.000
I(log(feedexpMS) * log(capitalMS))	0.11	0.01	0.000
I(log(feedexpMS) * log(materialsMS))	0.07	0.02	0.001
I(log(vetmedMS) * log(totalwagesMS))	-0.01	0.01	0.129
I(log(vetmedMS) * log(landMS))	-0.01	0.01	0.207
I(log(vetmedMS) * log(capitalMS))	0.00	0.01	0.626
I(log(vetmedMS) * log(materialsMS))	0.04	0.01	0.005
I(log(totalwagesMS) * log(landMS))	-0.05	0.01	0.000
I(log(totalwagesMS) * log(capitalMS))	0.02	0.01	0.075
I(log(totalwagesMS) * log(materialsMS))	0.02	0.01	0.232
I(log(landMS) * log(capitalMS))	0.08	0.01	0.000
I(log(landMS) * log(materialsMS))	0.00	0.02	0.999
I(log(capitalMS) * log(materialsMS))	0.02	0.02	0.344
I(log(feedexpMS) * log(grossotherMS/grossmilkMS))	-0.03	0.01	0.000
I(log(vetmedMS) * log(grossotherMS/grossmilkMS))	0.00	0.01	0.581
I(log(totalwagesMS) * log(grossotherMS/grossmilkMS))	0.01	0.01	0.027
I(log(landMS) * log(grossotherMS/grossmilkMS))	-0.02	0.01	0.019
I(log(capitalMS) * log(grossotherMS/grossmilkMS))	0.00	0.01	0.779
I(log(materialsMS) * log(grossotherMS/grossmilkMS))	0.02	0.01	0.127
Regnskabsaar2012	-0.01	0.00	0.017
Regnskabsaar2013	0.06	0.00	0.000
Regnskabsaar2014	0.05	0.00	0.000
Regnskabsaar2015	-0.05	0.00	0.000
milkingsystem2	-0.02	0.00	0.000

milkingsystem3	-0.01	0.00	0.000
jersey	-0.04	0.00	0.000
organic	0.00	0.00	0.497
Z_(Intercept)	0.02	0.05	0.624
Z_mastitis	0.05	0.04	0.169
Z_hoofdis	-0.12	0.05	0.017
Z_reprodis	0.03	0.03	0.268
Z_otherdis	-0.03	0.03	0.468
Z_cell1	-0.19	0.03	0.000
Z_viable1	-0.11	0.02	0.000
Z_spore1	-0.13	0.03	0.000
Z_managerage	0.00	0.00	0.000
Z_consultant	0.00	0.01	0.772
sigmaSq	0.02	0.00	0.000
gamma	0.70	0.05	0.000

panel data
number of cross-sections = 1810
number of time periods = 5
total number of observations = 8198
thus there are 852 observations not in the panel
mean efficiency of each year

2011	2012	2013	2014	2015
0.942	0.943	0.946	0.946	0.947

mean efficiency: 0.945

Source: Model result from R.

Table C.8: Estimated parameters of the restricted Translog ODF with indicators of diseases (TLodfres1)

	Estimate	Std. Error	p-value
(Intercept)	-0.02	0.00	0.000
log(grossotherMS/grossmilkMS)	0.16	0.00	0.000
I(0.5 * log(grossotherMS/grossmilkMS)^2)	0.08	0.00	0.000
log(feedexpMS)	-0.38	0.01	0.000
log(vetmedMS)	-0.02	0.00	0.000
log(totalwagesMS)	-0.10	0.00	0.000
log(landMS)	-0.17	0.01	0.000
log(capitalMS)	-0.19	0.01	0.000
log(materialsMS)	-0.18	0.01	0.000
I(0.5 * log(feedexpMS)^2)	-0.23	0.02	0.000
I(0.5 * log(vetmedMS)^2)	0.00	0.01	0.803
I(0.5 * log(totalwagesMS)^2)	-0.04	0.01	0.000
I(0.5 * log(landMS)^2)	-0.02	0.00	0.000
I(0.5 * log(capitalMS)^2)	-0.21	0.01	0.000
I(0.5 * log(materialsMS)^2)	-0.17	0.03	0.000
I(log(feedexpMS) * log(vetmedMS))	-0.02	0.01	0.121
I(log(feedexpMS) * log(totalwagesMS))	0.02	0.01	0.116
I(log(feedexpMS) * log(landMS))	0.04	0.01	0.001
I(log(feedexpMS) * log(capitalMS))	0.11	0.01	0.000
I(log(feedexpMS) * log(materialsMS))	0.07	0.02	0.000
I(log(vetmedMS) * log(totalwagesMS))	-0.01	0.01	0.036
I(log(vetmedMS) * log(landMS))	-0.01	0.01	0.348
I(log(vetmedMS) * log(capitalMS))	0.01	0.01	0.466
I(log(vetmedMS) * log(materialsMS))	0.04	0.01	0.008
I(log(totalwagesMS) * log(landMS))	-0.05	0.01	0.000
I(log(totalwagesMS) * log(capitalMS))	0.02	0.01	0.034
I(log(totalwagesMS) * log(materialsMS))	0.02	0.01	0.135
I(log(landMS) * log(capitalMS))	0.08	0.01	0.000
I(log(landMS) * log(materialsMS))	0.00	0.02	0.776
I(log(capitalMS) * log(materialsMS))	0.01	0.02	0.430
I(log(feedexpMS) * log(grossotherMS/grossmilkMS))	-0.03	0.01	0.001
I(log(vetmedMS) * log(grossotherMS/grossmilkMS))	0.00	0.01	0.986
I(log(totalwagesMS) * log(grossotherMS/grossmilkMS))	0.01	0.01	0.028
I(log(landMS) * log(grossotherMS/grossmilkMS))	-0.02	0.01	0.011
I(log(capitalMS) * log(grossotherMS/grossmilkMS))	0.00	0.01	0.585
I(log(materialsMS) * log(grossotherMS/grossmilkMS))	0.02	0.01	0.073
Regnskabsaar2012	-0.01	0.00	0.047
Regnskabsaar2013	0.05	0.00	0.000
Regnskabsaar2014	0.04	0.00	0.000
Regnskabsaar2015	-0.05	0.00	0.000

milkingsystem2	-0.03	0.00	0.000
milkingsystem3	-0.01	0.00	0.000
jersey	-0.04	0.00	0.000
organic	0.00	0.00	0.496
Z_(Intercept)	-13.46	1.30	0.000
Z_mastitis	-3.35	0.46	0.000
Z_hoofdis	-8.43	0.98	0.000
Z_reprodis	0.46	0.16	0.004
Z_otherdis	-3.99	0.25	0.000
Z_managerage	0.08	0.00	0.000
Z_consultant	0.62	0.04	0.000
sigmaSq	0.52	0.07	0.000
gamma	0.98	0.00	0.000

panel data
number of cross-sections = 1810
number of time periods = 5
total number of observations = 8198
thus there are 852 observations not in the panel
mean efficiency of each year

2011	2012	2013	2014	2015
0.951	0.953	0.953	0.953	0.953

mean efficiency: 0.953

Source: Model result from R.

Table C.9: Estimated parameters of the restricted Translog ODF with indicators of milk quality and without manager age and consultant (TLodfres3)

	Estimate	Std. Error	p-value
(Intercept)	-0.03	0.00	0.000
log(grossotherMS/grossmilkMS)	0.16	0.00	0.000
I(0.5 * log(grossotherMS/grossmilkMS)^2)	0.07	0.00	0.000
log(feedexpMS)	-0.38	0.01	0.000
log(vetmedMS)	-0.02	0.00	0.000
log(totalwagesMS)	-0.10	0.00	0.000
log(landMS)	-0.17	0.01	0.000
log(capitalMS)	-0.19	0.00	0.000
log(materialsMS)	-0.17	0.01	0.000
I(0.5 * log(feedexpMS)^2)	-0.24	0.02	0.000
I(0.5 * log(vetmedMS)^2)	0.00	0.01	0.557
I(0.5 * log(totalwagesMS)^2)	-0.04	0.01	0.000
I(0.5 * log(landMS)^2)	-0.02	0.00	0.000
I(0.5 * log(capitalMS)^2)	-0.21	0.01	0.000
I(0.5 * log(materialsMS)^2)	-0.17	0.03	0.000
I(log(feedexpMS) * log(vetmedMS))	-0.01	0.01	0.241
I(log(feedexpMS) * log(totalwagesMS))	0.02	0.01	0.080
I(log(feedexpMS) * log(landMS))	0.05	0.01	0.000
I(log(feedexpMS) * log(capitalMS))	0.12	0.01	0.000
I(log(feedexpMS) * log(materialsMS))	0.06	0.02	0.002
I(log(vetmedMS) * log(totalwagesMS))	-0.01	0.01	0.085
I(log(vetmedMS) * log(landMS))	-0.01	0.01	0.228
I(log(vetmedMS) * log(capitalMS))	0.00	0.01	0.681
I(log(vetmedMS) * log(materialsMS))	0.04	0.01	0.005
I(log(totalwagesMS) * log(landMS))	-0.05	0.01	0.000
I(log(totalwagesMS) * log(capitalMS))	0.02	0.01	0.086
I(log(totalwagesMS) * log(materialsMS))	0.02	0.01	0.236
I(log(landMS) * log(capitalMS))	0.08	0.01	0.000
I(log(landMS) * log(materialsMS))	0.00	0.02	0.940
I(log(capitalMS) * log(materialsMS))	0.02	0.02	0.349
I(log(feedexpMS) * log(grossotherMS/grossmilkMS))	-0.03	0.01	0.000
I(log(vetmedMS) * log(grossotherMS/grossmilkMS))	0.00	0.01	0.617
I(log(totalwagesMS) * log(grossotherMS/grossmilkMS))	0.01	0.01	0.019
I(log(landMS) * log(grossotherMS/grossmilkMS))	-0.02	0.01	0.014
I(log(capitalMS) * log(grossotherMS/grossmilkMS))	0.00	0.01	0.741
I(log(materialsMS) * log(grossotherMS/grossmilkMS))	0.02	0.01	0.122
Regnskabsaar2012	-0.01	0.00	0.023
Regnskabsaar2013	0.06	0.00	0.000
Regnskabsaar2014	0.05	0.00	0.000
Regnskabsaar2015	-0.04	0.00	0.000

milkingsystem2	-0.03	0.00	0.000
milkingsystem3	-0.01	0.00	0.000
jersey	-0.04	0.00	0.000
organic	0.00	0.00	0.588
Z_(Intercept)	0.18	0.02	0.000
Z_cell1	-0.23	0.05	0.000
Z_viable1	-0.13	0.03	0.000
Z_spore1	-0.16	0.04	0.000
sigmaSq	0.03	0.01	0.000
gamma	0.74	0.06	0.000

panel data

number of cross-sections = 1810

number of time periods = 5

total number of observations = 8198

thus there are 852 observations not in the panel

mean efficiency of each year

2011	2012	2013	2014	2015
0.943	0.944	0.947	0.947	0.948

mean efficiency: 0.946

Source: Model result from R.

13.4 Appendix D – R-script

```
### loading yearly data ###
data1<-read.csv("aardata.csv",header=T,sep=";")
library(zoo)
library(dynlm)
library(Formula)
library(lmtest)
library(plm)
library(sandwich)
library(Hmisc)
library(boot)
library(pastecs)
dim(data1)
pdata<-plm.data(data1,index=c("lbnr","Regnskabsaar"))

### Defining variables ###
### OUTPUT ###

# Total gross output #
pdata$grosstotal<-with(pdata,bru_i_alt-X140)

# Gross output from milk #
pdata$grossmilk <-with(pdata,X170+X171+X172)

# Gross output from milk out of total gross output #
pdata$grossmilkshare <-(pdata$grossmilk)/(pdata$grosstotal)

# Gross output from other outputs than milk #
pdata$grossother<-as.numeric(with(pdata,grosstotal-grossmilk))

### INPUT FACTORS ###
# Feed expenditures minus internally produced/bought feed #
pdata$feedexp<-with(pdata,(X250+X255+X260)*-1)

# Internally produced/bought feed #
pdata$ownfeed<-(pdata$X265*-1)

# Vet, medicine, and vaccine expenditures #
pdata$vetmed<-pdata$X270*-1

# Hours #
pdata$hours<- as.numeric(pdata$norm_t_ejendom)
```

```
# Wages #
pdata$wage <- pdata$X360*-1

# Calculated manager wage #
# Labor costs
# Hired labor
pdata$hlabor <- ifelse(pdata$X360 < 0, pdata$X360, 0)
# Tillæg ved ansatte på 15 kr. pr. normtime - ca. tillæg på 25.000 kr. pr. ansat
pdata$til1 <- ifelse(pdata$norm_t_ejendom > 1665, (pdata$norm_t_ejendom - 1665) * 15, 0)
# Tillæg maksimalt 450.000 kr. dvs. maksimal driftslederløn på 750.000 kr.
pdata$til <- ifelse(pdata$til1 > 450000, 450000, pdata$til1)
# Ejeren får 300.000 kr.+tillæg for stor farm, dog trækkes anden lønindtægt fra
pdata$ownermanager <- ifelse(pdata$X7960 < 225000 &
        pdata$X7960 > 0, 300000 - pdata$X7960, 75000) + pdata$til
# ægtefællen får maksimalt 300.000 kr. Der antages, at være en ægtefælle,
# hvis der er børnefamilieydelse, lønindtægt fra ægtefælle
# eller familiens arbejdskraftbehov overstiger 3000 timer.
pdata$spouse <- ifelse((pdata$X7962 > 0 | pdata$X7964 > 0 | pdata$norm_t_ejerfam > 3000)
            & pdata$X7962 < 300000 & pdata$X7962 >= 0, 300000-pdata$X7962, 0)
pdata$spouse <- ifelse(pdata$spouse < 0, 0, pdata$spouse)
# Familiens aflønning består af driftleder + ægtefælle, børn forventes aflønnet
pdata$famlabor <- pdata$ownermanager + pdata$spouse

pdata$result <- pdata$res_f_fin-pdata$famlabor

# Total wages: other + owner #
pdata$totalwages <-with(pdata,wage+famlabor)

# Land #
pdata$land<-with(pdata,X3901+X3911)

# Materials used in the production minus feed expenditures,
# depretiation, maintenaince, wages, and land taxes #

pdata$materials <- with(pdata, (X230+X235+X240+X245+
                X275+X280+X310+X315+X320 +X325+X330+X375+X380)*-1)
# Agricultural assets #
pdata$assetsprimo<-
with(pdata,X8002+X8006+X8007+X8009+X1540+X8011+X1560+X8014+
            X8015+X8016+X8017+X8018+X8019+X8020+X8021+X8022+X8023)

pdata$assetsultimo<-
with(pdata,X1000+X1008+X1010+X1016+X1018+X1020+X1024+X1026+
          X1028+X1030+X1032+X1034+X1038+X1040+X1042+X1044+X1048+X1050)
```

```
pdata$assets<-(pdata$assetsprimo+pdata$assetsultimo)/2

#  Maintenance #
pdata$main <- with(pdata,(X335+X340+X345+X350+X355+X357)*-1)

# Depreciation #
pdata$dep <-pdata$afsk*-1

# Land value #
pdata$landvalue<-pdata$X8104

# Residence #
pdata$residence<-pdata$X8107

# Capital #
pdata$capital<-with(pdata,((assets-landvalue-residence)*0.05052+dep+main))

### INEFFICIENCY VARIABLES ###
# Manager age #
pdata$managerage <-as.numeric(pdata$X6422)

# Production consultant #
pdata$consultant<-ifelse((pdata$X7826*-1)>0,1,0)

### Diseases ###

# mastitis per cow #
pdata$mastitis<-pdata$k114

# hoof and limb diseases per cow #
pdata$hoofdis<-pdata$k115

# Reproductive disorders  per cow #
pdata$reprodis<-pdata$k116

# Other disorders  per cow #
pdata$otherdis<-pdata$k117


# Health indicators overall #
pdata$healthindicators<-with(pdata,mastitis+hoofdis+reprodis+otherdis)

#### Milk quality indicators ####
```

```
# Cell count #
pdata$cell<-with(pdata,k110)
summary(pdata$cell)
pdata$cell1<-ifelse(pdata$k110<=300,1,0)
pdata$cell2<-ifelse(pdata$k110>300,1,0)

# Viable count #
pdata$viable<-with(pdata,k111)
pdata$viable1<-ifelse(pdata$k111<=30,1,0)
pdata$viable2<-ifelse(pdata$k111>31,1,0)

# Spore #
pdata$spore<-with(pdata,k112)
pdata$spore1<-ifelse(pdata$k112<=400,1,0)
pdata$spore2<-ifelse(pdata$k112>401,1,0)

### PRODUCTION CHARACTERISTICS ###
#### Creating dummies ####
### Type of cow ###
table(pdata$X5100)
# Breed = Jersey when X5100 = 3 #
pdata$jersey<-ifelse(pdata$X5100 == 3,1,0)
# Breed = Large when X5100 != 3 #
pdata$large<-ifelse(pdata$X5100<3|pdata$X5100>3,1,0)

# Milking system #
table(pdata$X5101)
# Milking system = AMS when X5101 = 3 #
pdata$AMS<-ifelse(pdata$X5101 == 3,1,0)

# Milking system = Fishbone when X5101 = 4 #
pdata$fishbone<-ifelse(pdata$X5101 == 4,2,0)

# Milking system = other when not 3 or 4  #
pdata$othersys<-ifelse(pdata$X5101<3 | pdata$X5101>4,3,0)

# Milking system as categorial variable #
pdata$milkingsystem<-with(pdata,as.factor(AMS+fishbone+othersys))

# Organic #
pdata$organic<-ifelse(pdata$X6410==1,1,0)

# Herd size crontrolled for incoming and outgoing cows during the year #
pdata$yearcows<-pdata$k97
```

```
#### Only keeping defined variables ####
keeps<- c("lbnr","Regnskabsaar","grossmilk","grossmilkshare","grossother",
        "grosstotal","vetmed","hours","feedexp","ownfeed",
        "totalwages","land","materials","capital", "managerage","consultant",
        "mastitis","hoofdis","reprodis","otherdis","cell","viable","spore",
        "cell1","viable1","spore1", "cell2","viable2","spore2",
        "milkquality", "healthindicators",
        "jersey","large","AMS","fishbone","othersys","organic","yearcows",
        "milkingsystem")

pdata1 <-pdata[ , (names(pdata) %in% keeps)]

#### Summary statistics of variables and number of violations ####
summary(pdata1$grossmilkshare>=0.66)
summary(pdata1$grossmilkshare<=1)
summary(pdata1$yearcows>49)
summary(pdata1$hours>=1665)

### Excluding those who are not specialised in milk production ###
pdata11<-subset(pdata1,grossmilkshare>=0.66 & grossmilkshare<=1 & hours>=1665 &
              yearcows>49)

### Creating new dataset with zero inpu/output values ###
summary(pdata1$grossother>0)
summary(pdata11$vetmed>0)
summary(pdata11$feedexp>0)
summary(pdata11$totalwages>0)
summary(pdata11$land>0)
summary(pdata11$materials>0)
summary(pdata11$capital>0)
summary(pdata11$managerage>=18)
summary(pdata11$managerage<90)

pdata2<-subset(pdata11, grossother>0 & vetmed>0 & feedexp>0 & totalwages>0 & capital>0 &
             materials>0  & land>0 )

### Mangerage should lie between 18 and 90, otherwise NA  ###
summary(pdata11$managerage>=18)
summary(pdata11$managerage<90)
pdata2$managerage[pdata2$managerage<18 | pdata2$managerage>90] <- NA

### Redefining the data used for the estimations, such that NA values are not a problem ###
pdata20<-pdata2[!is.na(pdata2$cell1),]
```

```
pdata201<-pdata20[!is.na(pdata20$cell2),]
pdata21<-pdata201[!is.na(pdata201$viable1),]
pdata211<-pdata21[!is.na(pdata21$viable2),]
pdata22<-pdata211[!is.na(pdata211$spore1),]
pdata221<-pdata211[!is.na(pdata211$spore2),]
pdata23<-pdata221[!is.na(pdata221$managerage),]

### Excluding those with data for less than 3 years ###
tbl <- table(pdata23$lbnr)
pdata33 <- droplevels(pdata23[pdata23$lbnr %in% names(tbl)[tbl >= 3],,drop=FALSE])

### STEP 1 ###
# Testing the data for the farms removed against

#### Health indicators t-test ####
# mastitis per cow
t.test(pdata23$mastitis,pdata33$mastitis)
# hoof and limb diseases per cow
t.test(pdata23$hoofdis,pdata33$hoofdis)
# Reproductive disorders  per cow
t.test(pdata23$reprodis,pdata33$reprodis)
# Other disorders  per cow
t.test(pdata23$otherdis,pdata33$otherdis)

#### Milk quality indicators: shares ####
# Cell count category #
table(pdata23$cell1==1)
table(pdata33$cell1==1)
# Viable count category #
table(pdata23$viable1==1)
table(pdata33$viable1==1)
# Spore count category #
table(pdata23$spore1==1)
table(pdata33$spore1==1)

### Type of cow: shares ###
table(pdata23$jersey==1)
table(pdata33$jersey==1)

### Milking system: shares ###
# Milking system = AMS #
table(pdata23$AMS==1)
table(pdata33$AMS==1)
```

```
# Milking system = fishbone #
table(pdata23$fishbone==2)
table(pdata33$fishbone==2)

# Milking system = othersys #
table(pdata23$othersys==3)
table(pdata33$othersys==3)

### Organic:shares ###
table(pdata23$organic==1)
table(pdata33$organic==1)

#### Manager age: t-test ####
t.test(pdata23$managerage,pdata33$managerage)

### Production consultant: shares ###
table(pdata23$consultant==1)
table(pdata33$consultant==1)

### STEP 2 ###
# The first thing here is to deflate DKK-values using input and output specific prices
# This will allow us to compare performance over time
# Read PriceIndex.csv
pricedata <- read.csv("PriceIndex.csv", sep=";", dec = ".")
summary(pricedata)
# substract 2010 such that Year variable will have values 0:5
pricedata$Regnskabsaar <- pricedata$Regnskabsaar - 2010
pdata33$Year <- as.numeric(pdata33$Regnskabsaar)

# deflating nominal values of selected variable
# to real values with the base year = 2010
# First, create empty variable, e.g. "output_2010"

### DEFLATING THE OUTPUTS ###
### MILK OUTPUT ###
pdata33$grossmilk_2010level <- NA

# Deflate the variable (i.e. express it in constant, 2010, prices)
# For variable_2010 in year i == 1 in data assing the value of this
# variable in year i == 1 devided by the CPI in i == 1 from PriceInd
# do it for each i in 1:5 using following "for loop":
# for each i in 1:5 do the following:
for( i in 1:5 ) { # for each i in 1:5 do the following:
  pdata33$grossmilk_2010level[ pdata33$Year == i ] <-
```

```
    pdata33$grossmilk[ pdata33$Year == i ] /
    pricedata$indexmilk[ pricedata$Regnskabsaar ==  i  ]
}

### OTHER OUTPUT ###
pdata33$grossother_2010level <- NA
for( i in 1:5 ) { # for each i in 1:5 do the following:
  pdata33$grossother_2010level[ pdata33$Year == i ] <-
    pdata33$grossother[ pdata33$Year == i ] /
    pricedata$indexother[ pricedata$Regnskabsaar ==  i  ]
}

### DEFLATING THE INPUTS  ###
### FEEDEXP ###
pdata33$feedexp_2010level <- NA
for( i in 1:5 ) { # for each i in 1:5 do the following:
  pdata33$feedexp_2010level[ pdata33$Year == i ] <-
    pdata33$feedexp[ pdata33$Year == i ] /
    pricedata$indexfeed[ pricedata$Regnskabsaar ==  i  ]
}

### VETMED ###
pdata33$vetmed_2010level <- NA
for( i in 1:5 ) { # for each i in 1:5 do the following:
  pdata33$vetmed_2010level[ pdata33$Year == i ] <-
    pdata33$vetmed[ pdata33$Year == i ] /
    pricedata$indexvetmed[ pricedata$Regnskabsaar ==  i  ]
}

### WAGES ###
pdata33$totalwages_2010level <- NA
for( i in 1:5 ) { # for each i in 1:5 do the following:
  pdata33$totalwages_2010level[ pdata33$Year == i ] <-
    pdata33$totalwages[ pdata33$Year == i ] /
    pricedata$indexwages[ pricedata$Regnskabsaar ==  i  ]
}

### MATERIALS ###
pdata33$materials_2010level <- NA
for( i in 1:5 ) { # for each i in 1:5 do the following:
  pdata33$materials_2010level[ pdata33$Year == i ] <-
    pdata33$materials[ pdata33$Year == i ] /
    pricedata$indexmaterials[ pricedata$Regnskabsaar ==  i  ]
}
```

```
#### Now we make a correlation analysis ####
# we test the correlation bewteen the gross margin (GM) from milk output per cow #
# we need here to include expenditures to own feed, and thus create a new dataset #
# just for the correlation and linear regression analysis #
pdata34<-subset(pdata33, ownfeed>0)

pdata34$ownfeed_2010level <- NA
for( i in 1:5 ) { # for each i in 1:5 do the following:
  pdata34$ownfeed_2010level[ pdata34$Year == i ] <-
    pdata34$ownfeed[ pdata34$Year == i ] /
    pricedata$indexfeed[ pricedata$Regnskabsaar ==  i  ]
}

dim(pdata34)
pdata34$grossmarginmilk<-with(pdata34,grossmilk_2010level-
              (feedexp_2010level+ownfeed_2010level+vetmed_2010level))

### The gross margin from milk per cow ###
pdata34$grossmarginpercow<-(pdata34$grossmarginmilk/pdata34$yearcows)

### The influence of the health indicators has on the GM per cow  ###

# mastitis per cow
plot(pdata34$mastitis,pdata34$grossmarginpercow)
abline(lm(pdata34$grossmarginpercow~pdata34$mastitis))
cor.test(pdata34$mastitis,pdata34$grossmarginpercow,
         use="pairwise.complete.obs",method="pearson")

# hoof and limb disorders per cow
plot(pdata34$hoofdis,pdata34$grossmarginpercow)
abline(lm(pdata34$grossmarginpercow~pdata34$hoofdis))
cor.test(pdata34$hoofdis,pdata34$grossmarginpercow,
         use="pairwise.complete.obs",method="pearson")

# Reproductive disorders  per cow
plot(pdata34$reprodis,pdata34$grossmarginpercow)
abline(lm(pdata34$grossmarginpercow~pdata34$reprodis))
cor.test(pdata34$reprodis,pdata34$grossmarginpercow,
         use="pairwise.complete.obs",method="pearson")

# Other disorders  per cow
plot(pdata34$otherdis,pdata34$grossmarginpercow)
abline(lm(pdata34$grossmarginpercow~pdata34$otherdis))
```

```
cor.test(pdata34$otherdis,pdata34$grossmarginpercow,
        use="pairwise.complete.obs",method="pearson")
### Milk quality indicators ###
# par(mfrow=c(1,3))

### Cell counts ###
boxplot(pdata34$grossmarginpercow,pdata34$cell1,names=c("Cell 1","Cell 2"))
cor.test(pdata34$cell1,pdata34$grossmarginpercow,
        use="pairwise.complete.obs",method="pearson")

### Viable counts ###
boxplot(pdata34$grossmarginpercow,pdata34$viable1,names=c("Viable 1","Viable 2"))
cor.test(pdata34$viable1,pdata34$grossmarginpercow,
        use="pairwise.complete.obs",method="pearson")

### Spore counts ###
boxplot(pdata34$grossmarginpercow,pdata34$spore1==0,names=c("Spore 1","Spore 2"))
cor.test(pdata34$spore1,pdata34$grossmarginpercow,
        use="pairwise.complete.obs",method="pearson")

### Breed ###
boxplot(pdata34$grossmarginpercow~pdata34$jersey,names=c("Jersey","Large breed"))
cor.test(pdata34$grossmarginpercow,pdata34$jersey,
        use="pairwise.complete.obs",method="pearson")

### Milking system ###
boxplot(pdata34$grossmarginpercow~pdata34$milkingsystem,names=c("AMS","Fishbone","Ot
her"))
title("Gross margin per cow and the type of milking system")
cor.test(pdata34$grossmarginpercow,pdata34$AMS,
        use="pairwise.complete.obs",method="pearson")
cor.test(pdata34$grossmarginpercow,pdata34$fishbone,
        use="pairwise.complete.obs",method="pearson")
cor.test(pdata34$grossmarginpercow,pdata34$othersys,
        use="pairwise.complete.obs",method="pearson")

### Organic ###
boxplot(pdata34$grossmarginpercow~pdata34$organic==0,names=c("Organic","Conventional")
)
cor.test(pdata34$grossmarginpercow,pdata34$organic,
        use="pairwise.complete.obs",method="pearson")

### Manager age ###
cor.test(pdata34$grossmarginpercow,pdata34$managerage,
```

```
        use="pairwise.complete.obs",method="pearson")

### Consultant ###
cor.test(pdata34$grossmarginpercow,pdata34$consultant,
        use="pairwise.complete.obs",method="pearson")

### STEP 3 ###
# Estimating the GM per cow using Pooled OLS #
# The general model #
OLS<-plm(grossmarginpercow~Regnskabsaar
        + mastitis
        + hoofdis
        + reprodis
        + otherdis
        + cell1
        + viable1
        + spore1
        + jersey
        + organic
        + milkingsystem
        + consultant
        + managerage
        + I(managerage^2)
        ,model="pooling",data=pdata34)
summary(OLS)

### STEP 4 ###
### DISTANCE FUNCTIONS ###
library(foreign)
library(frontier)
library(AER)
library(plm)
library(gmp)
library(Rmpfr)

# In some model specifications, it is an advantage to use mean-scaled quantities.
# Therefore, we create new variables with mean-scaled input and output quantities:
# ouptut quantities
pdata33$grossmilkMS<-pdata33$grossmilk_2010level/mean(pdata33$grossmilk_2010level)
pdata33$grossotherMS<-pdata33$grossother_2010level/mean(pdata33$grossother_2010level)
# input quantities
pdata33$vetmedMS<-pdata33$vetmed_2010level/mean(pdata33$vetmed_2010level)
pdata33$hoursMS<-pdata33$hours/mean(pdata33$hours)
pdata33$feedexpMS<-pdata33$feedexp_2010level/mean(pdata33$feedexp_2010level)
```

```
pdata33$totalwagesMS<-pdata33$totalwages_2010level/mean(pdata33$totalwages_2010level)
pdata33$landMS<-pdata33$land/mean(pdata33$land)
pdata33$materialsMS<-pdata33$materials_2010level/mean(pdata33$materials_2010level)
pdata33$capitalMS<-pdata33$capital/mean(pdata33$capital)

# set arguments for all estimations using  sfa()
sfaSearchStep <- 1e-6
sfaSearchTol <- 1e-11
sfaTol <- 1e-11

####################################
#### INPUT DISTANCE FUNCTION ###
####################################

#### Cobb-Douglas: General model ####
#######################################
CDidfunres<-sfa( -log(materialsMS) ~ log(grossmilkMS)
            + log(grossotherMS)
            + log( feedexpMS/materialsMS )
            + log( vetmedMS/materialsMS )
            + log(totalwagesMS/materialsMS)
            + log(landMS/materialsMS)
            + log(capitalMS/materialsMS)
            + Regnskabsaar
            + milkingsystem
            + jersey
            + organic
            |
            + mastitis
            + hoofdis
            + reprodis
            + otherdis
            + cell1
            + viable1
            + spore1
            + consultant
            + managerage
            ,data=pdata33, ineffDecrease = TRUE,
            timeEffect = TRUE,
            searchStep = sfaSearchStep, searchTol = sfaSearchTol, tol = sfaTol )
summary(CDidfunres, extraPar = TRUE)

### Translog: general model  ###
########################################
```

```
TLidfunres <- sfa(-log(materialsMS) ~ log(grossmilkMS)
                 + log (grossotherMS)

                 + I( 0.5 * log( grossmilkMS )^2 )
                 + I( 0.5 * log( grossotherMS )^2 )
                 + I( log( grossmilkMS) * log ( grossotherMS ) )

                 + log (feedexpMS/materialsMS)
                 + log( vetmedMS/materialsMS )
                 + log(totalwagesMS/materialsMS)
                 + log(landMS/materialsMS)
                 + log(capitalMS/materialsMS)

                 + I(0.5 * log(feedexpMS/materialsMS)^2)
                 + I(0.5 * log(vetmedMS/materialsMS)^2)
                 + I(0.5 * log(totalwagesMS/materialsMS)^2)
                 + I(0.5 * log(landMS/materialsMS)^2)
                 + I(0.5 * log(capitalMS/materialsMS)^2)

                 + I(log(feedexpMS/materialsMS) * log(vetmedMS/materialsMS))
                 + I(log(feedexpMS/materialsMS) * log(totalwagesMS/materialsMS))
                 + I(log(feedexpMS/materialsMS) * log(landMS/materialsMS))
                 + I(log(feedexpMS/materialsMS) * log(capitalMS/materialsMS))
                 + I(log(vetmedMS/materialsMS) * log(totalwagesMS/materialsMS))
                 + I(log(vetmedMS/materialsMS) * log(landMS/materialsMS))
                 + I(log(vetmedMS/materialsMS) * log(capitalMS/materialsMS))
                 + I(log(totalwagesMS/materialsMS) * log(landMS/materialsMS))
                 + I(log(totalwagesMS/materialsMS) * log(capitalMS/materialsMS))
                 + I(log(landMS/materialsMS) * log(capitalMS/materialsMS))

                 + I(log(feedexpMS/materialsMS) * log( grossmilkMS ))
                 + I(log(feedexpMS/materialsMS) * log( grossotherMS ))
                 + I(log(vetmedMS/materialsMS) * log( grossmilkMS ))
                 + I(log(vetmedMS/materialsMS) * log( grossotherMS ))
                 + I(log(totalwagesMS/materialsMS) * log( grossmilkMS ))
                 + I(log(totalwagesMS/materialsMS) * log( grossotherMS ))
                 + I(log(landMS/materialsMS) * log( grossmilkMS ))
                 + I(log(landMS/materialsMS) * log( grossotherMS ))
                 + I(log(capitalMS/materialsMS) * log( grossmilkMS ))
                 + I(log(capitalMS/materialsMS) * log( grossotherMS ))
                 + Regnskabsaar
                 + milkingsystem
                 + jersey
                 + organic
```

```
                |
                + mastitis
                + hoofdis
                + reprodis
                + otherdis
                + cell1
                + viable1
                + spore1
                + managerage
                + consultant
                ,data = pdata33
               , ineffDecrease = TRUE ,timeEffect = TRUE
                ,searchStep = sfaSearchStep, searchTol = sfaSearchTol, tol = sfaTol)

lrtest(TLidfunres)

### Translog: restricted model with diseases ###
#################################################
TLidfres1 <- sfa(-log(materialsMS) ~ log(grossmilkMS)
           + log (grossotherMS)

           + I( 0.5 * log( grossmilkMS )^2 )
           + I( 0.5 * log( grossotherMS )^2 )
           + I( log( grossmilkMS) * log ( grossotherMS ) )

           + log (feedexpMS/materialsMS)
           + log( vetmedMS/materialsMS )
           + log(totalwagesMS/materialsMS)
           + log(landMS/materialsMS)
           + log(capitalMS/materialsMS)

           + I(0.5 * log(feedexpMS/materialsMS)^2)
           + I(0.5 * log(vetmedMS/materialsMS)^2)
           + I(0.5 * log(totalwagesMS/materialsMS)^2)
           + I(0.5 * log(landMS/materialsMS)^2)
           + I(0.5 * log(capitalMS/materialsMS)^2)

           + I(log(feedexpMS/materialsMS) * log(vetmedMS/materialsMS))
           + I(log(feedexpMS/materialsMS) * log(totalwagesMS/materialsMS))
           + I(log(feedexpMS/materialsMS) * log(landMS/materialsMS))
           + I(log(feedexpMS/materialsMS) * log(capitalMS/materialsMS))
           + I(log(vetmedMS/materialsMS) * log(totalwagesMS/materialsMS))
           + I(log(vetmedMS/materialsMS) * log(landMS/materialsMS))
           + I(log(vetmedMS/materialsMS) * log(capitalMS/materialsMS))
```

```
            + I(log(totalwagesMS/materialsMS) * log(landMS/materialsMS))
            + I(log(totalwagesMS/materialsMS) * log(capitalMS/materialsMS))
            + I(log(landMS/materialsMS) * log(capitalMS/materialsMS))

            + I(log(feedexpMS/materialsMS) * log( grossmilkMS ))
            + I(log(feedexpMS/materialsMS) * log( grossotherMS ))
            + I(log(vetmedMS/materialsMS) * log( grossmilkMS ))
            + I(log(vetmedMS/materialsMS) * log( grossotherMS ))
            + I(log(totalwagesMS/materialsMS) * log( grossmilkMS ))
            + I(log(totalwagesMS/materialsMS) * log( grossotherMS ))
            + I(log(landMS/materialsMS) * log( grossmilkMS ))
            + I(log(landMS/materialsMS) * log( grossotherMS ))
            + I(log(capitalMS/materialsMS) * log( grossmilkMS ))
            + I(log(capitalMS/materialsMS) * log( grossotherMS ))
            + Regnskabsaar
            + milkingsystem
            + jersey
            + organic
             |
             + mastitis
             + hoofdis
             + reprodis
             + otherdis
             + managerage
             + consultant
             ,data = pdata33 ,searchStep = sfaSearchStep, searchTol = sfaSearchTol, tol = sfaTol
             ,ineffDecrease = TRUE ,timeEffect = TRUE )
lrtest(TLidfres1)
lrtest(TLidfres1,TLidfunres)

### Translog: restricted model with milk quality ###
###################################################
TLidfres2 <- sfa(-log(materialsMS) ~ log(grossmilkMS)
            + log (grossotherMS)

            + I( 0.5 * log( grossmilkMS )^2 )
            + I( 0.5 * log( grossotherMS )^2 )
            + I( log( grossmilkMS) * log ( grossotherMS ) )

            + log (feedexpMS/materialsMS)
            + log( vetmedMS/materialsMS )
            + log(totalwagesMS/materialsMS)
            + log(landMS/materialsMS)
            + log(capitalMS/materialsMS)
```

```
            + I(0.5 * log(feedexpMS/materialsMS)^2)
            + I(0.5 * log(vetmedMS/materialsMS)^2)
            + I(0.5 * log(totalwagesMS/materialsMS)^2)
            + I(0.5 * log(landMS/materialsMS)^2)
            + I(0.5 * log(capitalMS/materialsMS)^2)

            + I(log(feedexpMS/materialsMS) * log(vetmedMS/materialsMS))
            + I(log(feedexpMS/materialsMS) * log(totalwagesMS/materialsMS))
            + I(log(feedexpMS/materialsMS) * log(landMS/materialsMS))
            + I(log(feedexpMS/materialsMS) * log(capitalMS/materialsMS))
            + I(log(vetmedMS/materialsMS) * log(totalwagesMS/materialsMS))
            + I(log(vetmedMS/materialsMS) * log(landMS/materialsMS))
            + I(log(vetmedMS/materialsMS) * log(capitalMS/materialsMS))
            + I(log(totalwagesMS/materialsMS) * log(landMS/materialsMS))
            + I(log(totalwagesMS/materialsMS) * log(capitalMS/materialsMS))
            + I(log(landMS/materialsMS) * log(capitalMS/materialsMS))

            + I(log(feedexpMS/materialsMS) * log( grossmilkMS ))
            + I(log(feedexpMS/materialsMS) * log( grossotherMS ))
            + I(log(vetmedMS/materialsMS) * log( grossmilkMS ))
            + I(log(vetmedMS/materialsMS) * log( grossotherMS ))
            + I(log(totalwagesMS/materialsMS) * log( grossmilkMS ))
            + I(log(totalwagesMS/materialsMS) * log( grossotherMS ))
            + I(log(landMS/materialsMS) * log( grossmilkMS ))
            + I(log(landMS/materialsMS) * log( grossotherMS ))
            + I(log(capitalMS/materialsMS) * log( grossmilkMS ))
            + I(log(capitalMS/materialsMS) * log( grossotherMS ))
            + Regnskabsaar
            + milkingsystem
            + jersey
            + organic
             |
             + cell1
             + viable1
             + spore1
             + managerage
             + consultant
             ,data = pdata33
             , ineffDecrease = TRUE ,timeEffect = TRUE
             ,searchStep = sfaSearchStep, searchTol = sfaSearchTol, tol = sfaTol)
lrtest(TLidfres2)
lrtest(TLidfunres,TLidfres2)
```

Translog: restricted model without managerage and consultant
##

```
TLidfres3 <- sfa(-log(materialsMS) ~ log(grossmilkMS)
            + log (grossotherMS)

            + I( 0.5 * log( grossmilkMS )^2 )
            + I( 0.5 * log( grossotherMS )^2 )
            + I( log( grossmilkMS) * log ( grossotherMS ) )

            + log (feedexpMS/materialsMS)
            + log( vetmedMS/materialsMS )
            + log(totalwagesMS/materialsMS)
            + log(landMS/materialsMS)
            + log(capitalMS/materialsMS)

            + I(0.5 * log(feedexpMS/materialsMS)^2)
            + I(0.5 * log(vetmedMS/materialsMS)^2)
            + I(0.5 * log(totalwagesMS/materialsMS)^2)
            + I(0.5 * log(landMS/materialsMS)^2)
            + I(0.5 * log(capitalMS/materialsMS)^2)

            + I(log(feedexpMS/materialsMS) * log(vetmedMS/materialsMS))
            + I(log(feedexpMS/materialsMS) * log(totalwagesMS/materialsMS))
            + I(log(feedexpMS/materialsMS) * log(landMS/materialsMS))
            + I(log(feedexpMS/materialsMS) * log(capitalMS/materialsMS))
            + I(log(vetmedMS/materialsMS) * log(totalwagesMS/materialsMS))
            + I(log(vetmedMS/materialsMS) * log(landMS/materialsMS))
            + I(log(vetmedMS/materialsMS) * log(capitalMS/materialsMS))
            + I(log(totalwagesMS/materialsMS) * log(landMS/materialsMS))
            + I(log(totalwagesMS/materialsMS) * log(capitalMS/materialsMS))
            + I(log(landMS/materialsMS) * log(capitalMS/materialsMS))

            + I(log(feedexpMS/materialsMS) * log( grossmilkMS ))
            + I(log(feedexpMS/materialsMS) * log( grossotherMS ))
            + I(log(vetmedMS/materialsMS) * log( grossmilkMS ))
            + I(log(vetmedMS/materialsMS) * log( grossotherMS ))
            + I(log(totalwagesMS/materialsMS) * log( grossmilkMS ))
            + I(log(totalwagesMS/materialsMS) * log( grossotherMS ))
            + I(log(landMS/materialsMS) * log( grossmilkMS ))
            + I(log(landMS/materialsMS) * log( grossotherMS ))
            + I(log(capitalMS/materialsMS) * log( grossmilkMS ))
            + I(log(capitalMS/materialsMS) * log( grossotherMS ))
            + Regnskabsaar
            + milkingsystem
```

```
            + jersey
            + organic
            |
            + mastitis
            + hoofdis
            + reprodis
            + otherdis
            + cell1
            + viable1
            + spore1
            ,data = pdata33
            , ineffDecrease = TRUE ,timeEffect = TRUE
            ,searchStep = sfaSearchStep, searchTol = sfaSearchTol, tol = sfaTol)
summary(TLidfres3)
lrtest(TLidfres3)
lrtest(TLidfunres,TLidfres3)

### Choosing the restricted model: TLidfres3  ###
TLidf <- TLidfres3
summary(TLidf)
### Skeewness  test ###
library(moments)
skewness(residuals(TLidf,asInData = TRUE))

### Coefficients of TLidf ###

amilk <- coef( TLidf )["log(grossmilkMS)"]
aother <- coef( TLidf )["log(grossotherMS)"]
amilkmilk <- - coef( TLidf )["I(0.5 * log(grossmilkMS)^2)"]
aotherother <- coef( TLidf )["I(0.5 * log(grossotherMS)^2)"]
amilkother <- aothermilk <- coef( TLidf )["I(log(grossmilkMS) * log(grossotherMS))"]

bfeed <- coef(TLidf)["log(feedexpMS/materialsMS)"]
bvet <- coef(TLidf)["log(vetmedMS/materialsMS)"]
bwages <- coef(TLidf)["log(totalwagesMS/materialsMS)"]
bland <- coef(TLidf)["log(landMS/materialsMS)"]
bcap <- coef(TLidf)["log(capitalMS/materialsMS)"]
bmat <- 1-(bfeed+bvet+bwages+bland+bcap)

bfeedfeed <- coef(TLidf)["I(0.5 * log(feedexpMS/materialsMS)^2)"]
bvetvet <- coef(TLidf)["I(0.5 * log(vetmedMS/materialsMS)^2)"]
bwageswages <- coef(TLidf)["I(0.5 * log(totalwagesMS/materialsMS)^2)"]
blandland <-coef(TLidf)["I(0.5 * log(landMS/materialsMS)^2)"]
bcapcap <- coef(TLidf)["I(0.5 * log(capitalMS/materialsMS)^2)"]
```

```
bfeedvet <- bvetfeed <-coef(TLidf)["I(log(feedexpMS/materialsMS) *
log(vetmedMS/materialsMS))"]
bfeedwages <- bwagesfeed <- coef(TLidf)["I(log(feedexpMS/materialsMS) *
log(totalwagesMS/materialsMS))"]
bfeedland <- blandfeed <- coef(TLidf)["I(log(feedexpMS/materialsMS) *
log(landMS/materialsMS))"]
bfeedcap <- bcapfeed <- coef(TLidf)["I(log(feedexpMS/materialsMS) *
log(capitalMS/materialsMS))"]
bvetwages <- bwagesvet <-coef(TLidf)["I(log(vetmedMS/materialsMS) *
log(totalwagesMS/materialsMS))"]
bvetland <- blandvet <- coef(TLidf)["I(log(vetmedMS/materialsMS) *
log(landMS/materialsMS))"]
bvetcap <- bcapvet <- coef(TLidf)["I(log(vetmedMS/materialsMS) *
log(capitalMS/materialsMS))"]
bwagesland <- blandwages <-coef(TLidf)["I(log(totalwagesMS/materialsMS) *
log(landMS/materialsMS))"]
bwagescap <- bcapwages <-coef(TLidf)["I(log(totalwagesMS/materialsMS) *
log(capitalMS/materialsMS))"]
blandcap <- bcapland <-coef(TLidf)["I(log(landMS/materialsMS) *
log(capitalMS/materialsMS))"]

bfeedmat <- bmatfeed <- -(bfeedfeed + bfeedvet + bfeedwages + bfeedland + bfeedcap)
bvetmat <- bmatvet <- - (bvetvet + bvetfeed + bvetwages + bvetland + bvetcap )
bwagesmat <- bmatwages <- -(bwageswages + bwagesfeed + bwagesvet + bwagesland +
bwagescap)
blandmat <- bmatland <- - (blandland + blandfeed + blandvet + blandwages + blandcap)
bcapmat <- bmatcap <- - (bcapcap + bcapfeed + bcapvet + bcapwages + bcapland)
bmatmat <- -(bmatfeed + bmatvet + bmatwages + bmatland + bmatcap)

zfeedmilk <- coef(TLidf)["I(log(feedexpMS/materialsMS) * log(grossmilkMS))"]
zfeedother <- coef(TLidf)["I(log(feedexpMS/materialsMS) * log(grossotherMS))"]
zvetmilk <- coef(TLidf)["I(log(vetmedMS/materialsMS) * log(grossmilkMS))"]
zvetother <- coef(TLidf)["I(log(vetmedMS/materialsMS) * log(grossotherMS))"]
zwagesmilk <- coef(TLidf)["I(log(totalwagesMS/materialsMS) * log(grossmilkMS))"]
zwagesother <- coef(TLidf)["I(log(totalwagesMS/materialsMS) * log(grossotherMS))"]
zlandmilk <- coef(TLidf)["I(log(landMS/materialsMS) * log(grossmilkMS))"]
zlandother <- coef(TLidf)["I(log(landMS/materialsMS) * log(grossotherMS))"]
zcapmilk <- coef(TLidf)["I(log(capitalMS/materialsMS) * log(grossmilkMS))"]
zcapother <- coef(TLidf)["I(log(capitalMS/materialsMS) * log(grossotherMS))"]
zmatmilk <- -( zfeedmilk + zvetmilk + zwagesmilk + zlandmilk + zcapmilk  )
zmatother <- -(zfeedother + zvetother + zwagesother + zlandother + zcapother )

### Input elasticities ###
```

```
### feed ###
pdata33$efeedTLidf <- with(pdata33, bfeed
+ bfeedfeed * I(log(feedexpMS))
+ bfeedvet * I(log(vetmedMS))
+ bfeedwages * I(log(totalwagesMS))
+ bfeedland * I(log(landMS))
+ bfeedcap* I(log(capitalMS))
+ bfeedmat * I(log(materialsMS))
+ I(zfeedmilk * log(grossmilkMS))
+ I(zfeedother * log(grossotherMS)))
mean(pdata33$efeedTLidf)
sd(pdata33$efeedTLidf)
### vet ###
pdata33$evetTLidf <- with(pdata33, bvet
+ bvetvet * I(log(vetmedMS))
+ bvetwages * I(log(totalwagesMS))
+ bvetfeed * I(log(feedexpMS))
+ bvetland * I(log(landMS))
+ bvetcap* I(log(capitalMS))
+ bvetmat * I(log(materialsMS))
+ I(zvetmilk * log(grossmilkMS))
+ I(zvetother * log(grossotherMS)))
summary(pdata33$evetTLidf)
mean(pdata33$evetTLidf)
sd(pdata33$evetTLidf)
### totalwages ###
pdata33$ewagesTLidf <- with(pdata33, bwages
+ bwageswages * I(log(totalwagesMS))
+ bwagesvet * I(log(vetmedMS))
+ bwagesfeed * I(log(feedexpMS))
+ bwagesland * I(log(landMS))
+ bwagescap* I(log(capitalMS))
+ bwagesmat * I(log(materialsMS))
+ I(zwagesmilk * log(grossmilkMS))
+ I(zwagesother * log(grossotherMS)))
mean(pdata33$ewagesTLidf)
sd(pdata33$ewagesTLidf)
### land ###
pdata33$elandTLidf <- with(pdata33, bland
+ blandland * I(log(landMS))
+ blandvet * I(log(vetmedMS))
+ blandfeed * I(log(feedexpMS))
+ blandwages * I(log(totalwagesMS))
+ blandcap* I(log(capitalMS))
```

```
+ blandmat * I(log(materialsMS))
+ I(zlandmilk * log(grossmilkMS))
+ I(zlandother * log(grossotherMS)))
mean(pdata33$elandTLidf)
sd(pdata33$elandTLidf)
### capital ###
pdata33$ecapTLidf <- with(pdata33, bcap
+ bcapcap * I(log(capitalMS))
+ bcapfeed * I(log(feedexpMS))
+ bcapvet * I(log(vetmedMS))
+ bcapwages * I(log(totalwagesMS))
+ bcapland * I(log(landMS))
+ bcapmat * I(log(materialsMS))
+ I(zcapmilk * log(grossmilkMS))
+ I(zcapother * log(grossotherMS)))
mean(pdata33$ecapTLidf)
sd(pdata33$ecapTLidf)
### materials ###
pdata33$ematTLidf <- with(pdata33, bmat
+ bmatmat * I(log(materialsMS))
+ bmatfeed * I(log(feedexpMS))
+ bmatvet * I(log(vetmedMS))
+ bmatwages * I(log(totalwagesMS))
+ bmatland * I(log(landMS))
+ bmatcap * I(log(capitalMS))
+ I(zmatmilk * log(grossmilkMS))
+ I(zmatother * log(grossotherMS)))
mean(pdata33$ematTLidf)
sd(pdata33$ematTLidf)
### Output elasticities ###
### gross milk ###
pdata33$emilkTLidf <- with(pdata33, amilk
+ amilkother * I(log(grossotherMS))
+ zfeedmilk * I(log(feedexpMS))
+ zvetmilk * I(log(vetmedMS))
+ zwagesmilk * I(log(totalwagesMS))
+ zlandmilk * I(log(landMS))
+ zcapmilk * I(log(capitalMS))
+ zmatmilk * I(log(materialsMS)))
mean(pdata33$emilkTLidf)
sd(pdata33$emilkTLidf)
### gross other ###
pdata33$eotherTLidf <- with(pdata33, aother
+ aothermilk * I(log(grossmilkMS))
```

```
+ zfeedother * I(log(feedexpMS))
+ zvetother * I(log(vetmedMS))
+ zwagesother * I(log(totalwagesMS))
+ zlandother * I(log(landMS))
+ zcapother * I(log(capitalMS))
+ zmatother * I(log(materialsMS)))
mean(pdata33$eotherTLidf)
sd(pdata33$eotherTLidf)

### Test the properties ###
######################

### summary statistics of distance elasticities ###
summary(pdata33[,c( "efeedTLidf", "evetTLidf","ewagesTLidf","elandTLidf",
          "ecapTLidf","ematTLidf","emilkTLidf","eotherTLidf")])

### check if distance elasticities of the inputs sum up to one ###
range(pdata33$efeedTLidf + pdata33$evetTLidf+pdata33$ewagesTLidf+pdata33$elandTLidf
   +pdata33$ecapTLidf + pdata33$ematTLidf)

### Elasticity of scale ###
pdata33$eScaleTLidf <- (-(pdata33$eotherTLidf + pdata33$emilkTLidf))^(-1)
summary(pdata33$eScaleTLidf)
sd(pdata33$eScaleTLidf)
sum( pdata33$eScaleTLidf> 2 | pdata33$eScaleTLidf < 0.5 )

# Monotonicity in input and output quantities.
# Di(x,y) is non-decreasing in x if its first order derivatives with
# respect to input quantities are non-negative

sum(!pdata33$efeedTLidf >= 0)
sum(!pdata33$evetTLidf  >= 0)
sum(!pdata33$ewagesTLidf >= 0)
sum(!pdata33$elandTLidf >= 0)
sum(!pdata33$ecapTLidf  >= 0)
sum(!pdata33$ematTLidf >= 0)

pdata33$monoTLidf<-with(pdata33,efeedTLidf >= 0 & evetTLidf  >= 0 & ewagesTLidf >= 0
             & elandTLidf >= 0 & ecapTLidf  >= 0 & ematTLidf >= 0)

### number that violates monotonicity assumption (input) ###
sum( pdata33$monoTLidf )
sum( !pdata33$monoTLidf )
sum( pdata33$monoTLidf )/(sum( !pdata33$monoTLidf ) +sum(pdata33$monoTLidf))
```

```
# Di(x,y) is non-increasing in y if its first order derivatives with respect
# to input quantities are non-positive.

sum(!pdata33$emilkTLidf <= 0)
sum(!pdata33$eotherTLidf  <= 0)

### number that violates monotonicity assumption (output) ###
pdata33$monooutTLidf<-with(pdata33,emilkTLidf <= 0 & eotherTLidf  <= 0)
sum( pdata33$monooutTLidf )
sum( !pdata33$monooutTLidf )
sum( pdata33$monooutTLidf)/(sum( !pdata33$monooutTLidf ) +sum(pdata33$monooutTLidf))

# We can check if estimated Translog input distance is concave in input quantities (x)
# and quasiconcave in output quantities (y) at the frontier.

### Quasi-concave in outputs ###
# first-order partial derivatives wrt outputs (with input distance measure = 1 )
pdata33$fmilkmilk <- amilk * (amilk-1) * 1 / pdata33$grossmilkMS^2
pdata33$fmilkother <- amilk * aother * 1 / (pdata33$grossmilkMS * pdata33$grossotherMS)
pdata33$fotherother <- aother * (aother-1) * 1 / pdata33$grossotherMS^2

# The following code creates a three-dimensional array with the Hessian matrices
# at all observations stacked upon each other:
bhm <- matrix( 0, nrow = 2, ncol = 2 )
bhm[ 1, 1 ] <- pdata33$fmilkmilk [1]
bhm[ 1, 2 ] <- bhm[ 2, 1 ] <- pdata33$fmilkother [1]
bhm[ 2, 2 ] <- pdata33$fotherother [1]

print(bhm)

# We check quasiconcavity in output quantities at the first observation
# by calculating the first and second leading principal minor:
det(bhm)

pdata33$quasiConv <- NA
for( obs in 1:nrow( pdata33 ) ) {
bhmLoop <- matrix( 0, nrow = 2, ncol = 2 )
bhmLoop[ 1, 1 ] <- pdata33$fmilkmilk[ 1 ]
bhmLoop[ 1, 2 ] <- bhmLoop[ 2, 1 ] <- pdata33$fmilkother[ 1 ]
bhmLoop[ 2, 2 ] <- pdata33$fotherother [ 1 ]
pdata33$quasiConv[ obs ] <- det( bhmLoop[ 1:2, 1:2 ] ) < 0
}
sum(pdata33$quasiConv)
```

```
#### Concave in inputs ####
pdata33$ffeed <- pdata33$efeedTLidf *  1 / pdata33$feedexpMS
pdata33$fvet <- pdata33$evetTLidf *  1 / pdata33$vetmedMS
pdata33$fwages <- pdata33$ewagesTLidf *  1 / pdata33$totalwagesMS
pdata33$fland <- pdata33$elandTLidf *  1 / pdata33$landMS
pdata33$fcap <- pdata33$ecapTLidf *  1 / pdata33$capitalMS
pdata33$fmat <- pdata33$ematTLidf *  1 / pdata33$materialsMS


pdata33$ffeedfeedTLidf <- with(pdata33,
                    ( bfeedfeed + efeedTLidf * efeedTLidf - 1 * efeedTLidf )
                    / + ( 1 / ( feedexpMS * feedexpMS ) ) )

pdata33$ffeedvetTLidf <- fvetfeedTLidf <- with (pdata33,
                                 ( bfeedvet + efeedTLidf * evetTLidf - 0 * efeedTLidf )
                                 / +(1/( feedexpMS*vetmedMS)))

pdata33$ffeedwagesTLidf <- fwagesfeedTLidf <- with (pdata33,
                                   ( bfeedwages + efeedTLidf * ewagesTLidf - 0 * efeedTLidf )
                                   / +(1/( feedexpMS*totalwagesMS)))

pdata33$ffeedlandTLidf <- flandfeedTLidf <- with (pdata33,
                                  ( bfeedland + efeedTLidf * elandTLidf - 0 * efeedTLidf )
                                  / +(1/( feedexpMS*landMS)))

pdata33$ffeedcapTLidf <- fcapfeedTLidf <- with (pdata33,
                                 ( bfeedcap + efeedTLidf * ecapTLidf - 0 * efeedTLidf )
                                / +(1/( feedexpMS*capitalMS)))

pdata33$ffeedmatTLidf <- fmatfeedTLidf <- with (pdata33,
                                ( bfeedmat + efeedTLidf * ematTLidf - 0 * efeedTLidf )
                                 / +(1/( feedexpMS*materialsMS)))
pdata33$fvetvetTLidf <- with(pdata33,
                    ( bvetvet + evetTLidf * evetTLidf - 1 * evetTLidf )
                    / + ( 1 / ( vetmedMS * vetmedMS ) ) )

pdata33$fvetwagesTLidf <- fwagesvetTLidf <- with (pdata33,
                                 ( bvetwages + evetTLidf * ewagesTLidf - 0 * evetTLidf )
                                 / +(1/( vetmedMS*totalwagesMS)))

pdata33$fvetlandTLidf <- flandvetTLidf <- with (pdata33,
                                   ( bvetland + evetTLidf * elandTLidf - 0 * evetTLidf )
                                   / +(1/( vetmedMS*landMS)))
```

```
pdata33$fvetcapTLidf <- fcapvetTLidf <- with (pdata33,
                              ( bvetcap + evetTLidf * ecapTLidf - 0 * evetTLidf )
                              / +(1/( vetmedMS*capitalMS)))

pdata33$fvetmatTLidf <- fmatvetTLidf <- with (pdata33,
                              ( bvetmat + evetTLidf * ematTLidf - 0 * evetTLidf )
                              / +(1/( vetmedMS*materialsMS)))

pdata33$fwageswagesTLidf <- with(pdata33,
                      ( bwageswages + ewagesTLidf * ewagesTLidf - 1 * ewagesTLidf )
                      / + ( 1 / ( totalwagesMS * totalwagesMS ) ) )

pdata33$fwageslandTLidf <- flandwagesTLidf <- with (pdata33,
                           ( bwagesland + ewagesTLidf * elandTLidf - 0 * ewagesTLidf )
                           / +(1/( totalwagesMS*landMS)))

pdata33$fwagescapTLidf <- fcapwagesTLidf <- with (pdata33,
                              ( bwagescap + ewagesTLidf * ecapTLidf - 0 * ewagesTLidf )
                                / +(1/( totalwagesMS*capitalMS)))

pdata33$fwagesmatTLidf <- fmatwagesTLidf <- with (pdata33,
                               ( bwagesmat + ewagesTLidf * ematTLidf - 0 * ewagesTLidf )
                                / +(1/( totalwagesMS*materialsMS)))

pdata33$flandlandTLidf <- with(pdata33,
                         ( blandland + elandTLidf * elandTLidf - 1 * elandTLidf )
                         / + ( 1 / ( landMS * landMS ) ) )

pdata33$flandcapTLidf <- fcaplandTLidf <- with (pdata33,
                              ( blandcap + elandTLidf * ecapTLidf - 0 * elandTLidf )
                              / +(1/( landMS*capitalMS)))

pdata33$flandmatTLidf <- fmatlandTLidf <- with (pdata33,
                              ( blandmat + elandTLidf * ematTLidf - 0 * elandTLidf )
                              / +(1/( landMS*materialsMS)))

pdata33$fcapcapTLidf <- with(pdata33,
                   ( bcapcap + ecapTLidf * ecapTLidf - 1 * ecapTLidf )
                   / + ( 1 / ( capitalMS * capitalMS ) ) )

pdata33$fcapmatTLidf <- fmatcapTLidf <- with (pdata33,
                            ( bcapmat + ecapTLidf * ematTLidf - 0 * ecapTLidf )
                            / +(1/( capitalMS*materialsMS)))
```

```
pdata33$fmatmatTLidf <- with(pdata33,
                       ( bmatmat + ematTLidf * ematTLidf - 1 * ematTLidf )
                       / + ( 1 / ( materialsMS * materialsMS ) ) )

# First, we prepare the Hessian matrix for the first observation:
hessian <- matrix( NA, nrow = 6, ncol = 6 )
hessian[ 1, 1 ] <- pdata33$ffeedfeedTLidf[1]
hessian[ 1, 2 ] <- hessian[ 2, 1 ] <- pdata33$ffeedvetTLidf[1]
hessian[ 1, 3 ] <- hessian[ 3, 1 ] <- pdata33$ffeedwagesTLidf[1]
hessian[ 1, 4 ] <- hessian[ 4, 1 ] <- pdata33$ffeedlandTLidf[1]
hessian[ 1, 5 ] <- hessian[ 5, 1 ] <- pdata33$ffeedcapTLidf[1]
hessian[ 1, 6 ] <- hessian[ 6, 1 ] <- pdata33$ffeedmatTLidf[1]
hessian[ 2, 2 ] <- pdata33$fvetvetTLidf[1]
hessian[ 2, 3 ] <- hessian[ 3, 2 ] <-pdata33$fvetwagesTLidf[1]
hessian[ 2, 4 ] <- hessian[ 4, 2 ] <-pdata33$fvetlandTLidf[1]
hessian[ 2, 5 ] <- hessian[ 5, 2 ] <-pdata33$fvetcapTLidf[1]
hessian[ 2, 6 ] <- hessian[ 6, 2 ] <-pdata33$fvetmatTLidf[1]
hessian[ 3, 3 ] <- pdata33$fwageswagesTLidf[1]
hessian[ 3, 4 ] <- hessian[ 4, 3 ] <-pdata33$fwageslandTLidf[1]
hessian[ 3, 5 ] <- hessian[ 5, 3 ] <-pdata33$fwagescapTLidf[1]
hessian[ 3, 6 ] <- hessian[ 6, 3 ] <-pdata33$fwagesmatTLidf[1]
hessian[ 4, 4 ] <- pdata33$flandlandTLidf[1]
hessian[ 4, 5 ] <- hessian[ 5, 4 ] <-pdata33$flandcapTLidf[1]
hessian[ 4, 6 ] <- hessian[ 6, 4 ] <-pdata33$flandmatTLidf[1]
hessian[ 5, 5 ] <- pdata33$fcapcapTLidf[1]
hessian[ 5, 6 ] <- hessian[ 6, 5 ] <-pdata33$fcapmatTLidf[1]
hessian[ 6, 6 ] <- pdata33$fmatmatTLidf[1]
print( hessian )

# As all diagonal elements of this Hessian matrix are zero, the necessary
# conditions for positive semidefiniteness are not fulfilled for the first observation.

det( hessian )

pdata33$concaveTLidf <- pdata33$ffeedfeed <= 0
sum(pdata33$concaveTLidf)

### Effects of production characteristics ###
(exp(coef( TLidf )[c("Regnskabsaar2012")])-1)*100
(exp(coef( TLidf )[c("Regnskabsaar2013")])-1)*100
(exp(coef( TLidf )[c("Regnskabsaar2014")])-1)*100
(exp(coef( TLidf )[c("Regnskabsaar2015")])-1)*100
```

```
(exp(coef( TLidf )[c("milkingsystem2")])-1)*100
(exp(coef( TLidf )[c("milkingsystem3")])-1)*100
(exp(coef( TLidf )[c("jersey")])-1)*100
(exp(coef( TLidf )[c("organic")])-1)*100

### Calculating marginal effects of z-variables ###
pdata33$effzvar <- efficiencies(TLidf, asInData=TRUE, margEff = TRUE)
summary(pdata33$effzvar)
METLidfz<-attr(pdata33$effzvar, "margEff")
summary(METLidfz)

hist(METLidfz,50)

### Efficiency over time ###

# 2011
summary(pdata33$Regnskabsaar==2011)
summary(pdata33$eff>=0.95 & pdata33$Regnskabsaar==2011)
summary(pdata33$eff>=0.90 & pdata33$eff<0.95 & pdata33$Regnskabsaar==2011)
summary(pdata33$eff>=0.85 & pdata33$eff<0.90 & pdata33$Regnskabsaar==2011)
summary(pdata33$eff>=0.80 & pdata33$eff<0.85 & pdata33$Regnskabsaar==2011)
summary(pdata33$eff<0.80 & pdata33$Regnskabsaar==2011)

# 2012
summary(pdata33$Regnskabsaar==2012)
summary(pdata33$eff>=0.95 & pdata33$Regnskabsaar==2012)
summary(pdata33$eff>=0.90 & pdata33$eff<0.95 & pdata33$Regnskabsaar==2012)
summary(pdata33$eff>=0.85 & pdata33$eff<0.90 & pdata33$Regnskabsaar==2012)
summary(pdata33$eff>=0.80 & pdata33$eff<0.85 & pdata33$Regnskabsaar==2012)
summary(pdata33$eff<0.80 & pdata33$Regnskabsaar==2012)

# 2013
summary(pdata33$Regnskabsaar==2013)
summary(pdata33$eff>=0.95 & pdata33$Regnskabsaar==2013)
summary(pdata33$eff>=0.90 & pdata33$eff<0.95 & pdata33$Regnskabsaar==2013)
summary(pdata33$eff>=0.85 & pdata33$eff<0.90 & pdata33$Regnskabsaar==2013)
summary(pdata33$eff>=0.80 & pdata33$eff<0.85 & pdata33$Regnskabsaar==2013)
summary(pdata33$eff<0.80 & pdata33$Regnskabsaar==2013)

# 2014
summary(pdata33$Regnskabsaar==2014)
summary(pdata33$eff>=0.95 & pdata33$Regnskabsaar==2014)
summary(pdata33$eff>=0.90 & pdata33$eff<0.95 & pdata33$Regnskabsaar==2014)
summary(pdata33$eff>=0.85 & pdata33$eff<0.90 & pdata33$Regnskabsaar==2014)
```

```
summary(pdata33$eff>=0.80 & pdata33$eff<0.85 & pdata33$Regnskabsaar==2014)
summary(pdata33$eff<0.80 & pdata33$Regnskabsaar==2014)

# 2015
summary(pdata33$Regnskabsaar==2015)
summary(pdata33$eff>=0.95 & pdata33$Regnskabsaar==2015)
summary(pdata33$eff>=0.90 & pdata33$eff<0.95 & pdata33$Regnskabsaar==2015)
summary(pdata33$eff>=0.85 & pdata33$eff<0.90 & pdata33$Regnskabsaar==2015)
summary(pdata33$eff>=0.80 & pdata33$eff<0.85 & pdata33$Regnskabsaar==2015)
summary(pdata33$eff<0.80 & pdata33$Regnskabsaar==2015)

### TE and Size ###
pdata33$totaloutput<-pdata33$grossmilk+pdata33$grossother
plot(pdata33$yearcows,pdata33$eff, log="x")
plot( pdata33$totaloutput, pdata33$eff , log="x")

################################################
#### Excluding those without monotonicity ####
################################################
pdata44<-subset(pdata33,efeedTLidf >= 0 & evetTLidf  >= 0 & ewagesTLidf >= 0
                & elandTLidf >= 0 & ecapTLidf  >= 0 & ematTLidf >= 0
                & emilkTLidf <= 0 & eotherTLidf  <= 0)

pdata44$effzvar <- efficiencies(TLidf, asInData=TRUE, margEff = TRUE)
summary(pdata44$effzvar)

### summary statistics of distance elasticities ###
summary(pdata44[,c( "efeedTLidf", "evetTLidf","ewagesTLidf","elandTLidf",
            "ecapTLidf","ematTLidf","emilkTLidf","eotherTLidf")])

# check if distance elasticities of the inputs sum up to one: they do!
range(pdata44$efeedTLidf + pdata44$evetTLidf+pdata44$ewagesTLidf+pdata44$elandTLidf
    +pdata44$ecapTLidf + pdata44$ematTLidf)

### Elasticity of scale ###
pdata44$eScaleTLidf <- (-(pdata44$eotherTLidf + pdata44$emilkTLidf))^(-1)
summary(pdata44$eScaleTLidf)
sd(pdata44$eScaleTLidf)
sum(pdata44$eScaleTLidf> 2 | pdata44$eScaleTLidf< 0.5)

sd(pdata44$efeedTLidf)
sd(pdata44$evetTLidf)
sd(pdata44$ewagesTLidf)
sd(pdata44$elandTLidf)
```

```
sd(pdata44$ecapTLidf)
sd(pdata44$ematTLidf)
sd(pdata44$emilkTLidf)
sd(pdata44$eotherTLidf)

### The input distance function with "hours" instead of "labour" (totalwages) ###

#### Using hours instead of wages ####
TLidfhourshours <- sfa(-log(materialsMS) ~ log(grossmilkMS)
                + log (grossotherMS)

                + I( 0.5 * log( grossmilkMS )^2 )
                + I( 0.5 * log( grossotherMS )^2 )
                + I( log( grossmilkMS) * log ( grossotherMS ) )

                + log (feedexpMS/materialsMS)
                + log( vetmedMS/materialsMS )
                + log(hoursMS/materialsMS)
                + log(landMS/materialsMS)
                + log(capitalMS/materialsMS)

                + I(0.5 * log(feedexpMS/materialsMS)^2)
                + I(0.5 * log(vetmedMS/materialsMS)^2)
                + I(0.5 * log(hoursMS/materialsMS)^2)
                + I(0.5 * log(landMS/materialsMS)^2)
                + I(0.5 * log(capitalMS/materialsMS)^2)

                + I(log(feedexpMS/materialsMS) * log(vetmedMS/materialsMS))
                + I(log(feedexpMS/materialsMS) * log(hoursMS/materialsMS))
                + I(log(feedexpMS/materialsMS) * log(landMS/materialsMS))
                + I(log(feedexpMS/materialsMS) * log(capitalMS/materialsMS))
                + I(log(vetmedMS/materialsMS) * log(hoursMS/materialsMS))
                + I(log(vetmedMS/materialsMS) * log(landMS/materialsMS))
                + I(log(vetmedMS/materialsMS) * log(capitalMS/materialsMS))
                + I(log(hoursMS/materialsMS) * log(landMS/materialsMS))
                + I(log(hoursMS/materialsMS) * log(capitalMS/materialsMS))
                + I(log(landMS/materialsMS) * log(capitalMS/materialsMS))

                + I(log(feedexpMS/materialsMS) * log( grossmilkMS ))
                + I(log(feedexpMS/materialsMS) * log( grossotherMS ))
                + I(log(vetmedMS/materialsMS) * log( grossmilkMS ))
                + I(log(vetmedMS/materialsMS) * log( grossotherMS ))
                + I(log(hoursMS/materialsMS) * log( grossmilkMS ))
                + I(log(hoursMS/materialsMS) * log( grossotherMS ))
```

```
                + I(log(landMS/materialsMS) * log( grossmilkMS ))
                + I(log(landMS/materialsMS) * log( grossotherMS ))
                + I(log(capitalMS/materialsMS) * log( grossmilkMS ))
                + I(log(capitalMS/materialsMS) * log( grossotherMS ))
                + Regnskabsaar
                + milkingsystem
                + jersey
                + organic
                |
                + mastitis
                + hoofdis
                + reprodis
                + otherdis
                + cell1
                + viable1
                + spore1
                ,data = pdata33
                , ineffDecrease = TRUE ,timeEffect = TRUE
                ,searchStep = sfaSearchStep, searchTol = sfaSearchTol, tol = sfaTol)
summary(TLidfhourshours, extraPar = TRUE)
lrtest(TLidfhourshours)

### Test the properties ###
###########################
### Coefficients of TLidfhourshours ###

amilk <- coef( TLidfhourshours )["log(grossmilkMS)"]
aother <- coef( TLidfhourshours )["log(grossotherMS)"]
amilkmilk <- - coef( TLidfhourshours )["I(0.5 * log(grossmilkMS)^2)"]
aotherother <- coef( TLidfhourshours )["I(0.5 * log(grossotherMS)^2)"]
amilkother <- aothermilk <- coef( TLidfhourshours )["I(log(grossmilkMS) *
log(grossotherMS))"]

bfeed <- coef(TLidfhourshours)["log(feedexpMS/materialsMS)"]
bvet <- coef(TLidfhourshours)["log(vetmedMS/materialsMS)"]
bhours <- coef(TLidfhourshours)["log(hoursMS/materialsMS)"]
bland <- coef(TLidfhourshours)["log(landMS/materialsMS)"]
bcap <- coef(TLidfhourshours)["log(capitalMS/materialsMS)"]
bmat <- 1-(bfeed+bvet+bhours+bland+bcap)

bfeedfeed <- coef(TLidfhourshours)["I(0.5 * log(feedexpMS/materialsMS)^2)"]
bvetvet <- coef(TLidfhourshours)["I(0.5 * log(vetmedMS/materialsMS)^2)"]
bhourshours <- coef(TLidfhourshours)["I(0.5 * log(hoursMS/materialsMS)^2)"]
blandland <-coef(TLidfhourshours)["I(0.5 * log(landMS/materialsMS)^2)"]
```

```
bcapcap <- coef(TLidfhourshours)["I(0.5 * log(capitalMS/materialsMS)^2)"]

bfeedvet <- bvetfeed <-coef(TLidfhourshours)["I(log(feedexpMS/materialsMS) *
log(vetmedMS/materialsMS))"]
bfeedhours <- bhoursfeed <- coef(TLidfhourshours)["I(log(feedexpMS/materialsMS) *
log(hoursMS/materialsMS))"]
bfeedland <- blandfeed <- coef(TLidfhourshours)["I(log(feedexpMS/materialsMS) *
log(landMS/materialsMS))"]
bfeedcap <- bcapfeed <- coef(TLidfhourshours)["I(log(feedexpMS/materialsMS) *
log(capitalMS/materialsMS))"]
bvethours <- bhoursvet <-coef(TLidfhourshours)["I(log(vetmedMS/materialsMS) *
log(hoursMS/materialsMS))"]
bvetland <- blandvet <- coef(TLidfhourshours)["I(log(vetmedMS/materialsMS) *
log(landMS/materialsMS))"]
bvetcap <- bcapvet <- coef(TLidfhourshours)["I(log(vetmedMS/materialsMS) *
log(capitalMS/materialsMS))"]
bhoursland <- blandhours <-coef(TLidfhourshours)["I(log(hoursMS/materialsMS) *
log(landMS/materialsMS))"]
bhourscap <- bcaphours <-coef(TLidfhourshours)["I(log(hoursMS/materialsMS) *
log(capitalMS/materialsMS))"]
blandcap <- bcapland <-coef(TLidfhourshours)["I(log(landMS/materialsMS) *
log(capitalMS/materialsMS))"]

bfeedmat <- bmatfeed <- -(bfeedfeed + bfeedvet + bfeedhours + bfeedland + bfeedcap)
bvetmat <- bmatvet <- - (bvetvet + bvetfeed + bvethours + bvetland + bvetcap )
bhoursmat <- bmathours <- -(bhourshours + bhoursfeed + bhoursvet + bhoursland + bhourscap)
blandmat <- bmatland <- - (blandland + blandfeed + blandvet + blandhours + blandcap)
bcapmat <- bmatcap <- - (bcapcap + bcapfeed + bcapvet + bcaphours + bcapland)
bmatmat <- -(bmatfeed + bmatvet + bmathours + bmatland + bmatcap)

zfeedmilk <- coef(TLidfhourshours)["I(log(feedexpMS/materialsMS) * log(grossmilkMS))"]
zfeedother <- coef(TLidfhourshours)["I(log(feedexpMS/materialsMS) * log(grossotherMS))"]
zvetmilk <- coef(TLidfhourshours)["I(log(vetmedMS/materialsMS) * log(grossmilkMS))"]
zvetother <- coef(TLidfhourshours)["I(log(vetmedMS/materialsMS) * log(grossotherMS))"]
zhoursmilk <- coef(TLidfhourshours)["I(log(hoursMS/materialsMS) * log(grossmilkMS))"]
zhoursother <- coef(TLidfhourshours)["I(log(hoursMS/materialsMS) * log(grossotherMS))"]
zlandmilk <- coef(TLidfhourshours)["I(log(landMS/materialsMS) * log(grossmilkMS))"]
zlandother <- coef(TLidfhourshours)["I(log(landMS/materialsMS) * log(grossotherMS))"]
zcapmilk <- coef(TLidfhourshours)["I(log(capitalMS/materialsMS) * log(grossmilkMS))"]
zcapother <- coef(TLidfhourshours)["I(log(capitalMS/materialsMS) * log(grossotherMS))"]
zmatmilk <- -( zfeedmilk + zvetmilk + zhoursmilk + zlandmilk + zcapmilk  )
zmatother <- -(zfeedother + zvetother + zhoursother + zlandother + zcapother )

### Input elasticities ###
```

```
### feed ###
pdata33$efeedTLidfhours <- with(pdata33, bfeed
                    + bfeedfeed * I(log(feedexpMS))
                    + bfeedvet * I(log(vetmedMS))
                    + bfeedhours * I(log(hoursMS))
                    + bfeedland * I(log(landMS))
                    + bfeedcap* I(log(capitalMS))
                    + bfeedmat * I(log(materialsMS))
                    + I(zfeedmilk * log(grossmilkMS))
                    + I(zfeedother * log(grossotherMS)))

### vet ###
pdata33$evetTLidfhours <- with(pdata33, bvet
                   + bvetvet * I(log(vetmedMS))
                   + bvethours * I(log(hoursMS))
                   + bvetfeed * I(log(feedexpMS))
                   + bvetland * I(log(landMS))
                   + bvetcap* I(log(capitalMS))
                   + bvetmat * I(log(materialsMS))
                   + I(zvetmilk * log(grossmilkMS))
                   + I(zvetother * log(grossotherMS)))
mean(pdata33$evetTLidfhours)

### totalhours ###
pdata33$ehoursTLidfhours <- with(pdata33, bhours
                     + bhourshours * I(log(hoursMS))
                     + bhoursvet * I(log(vetmedMS))
                     + bhoursfeed * I(log(feedexpMS))
                     + bhoursland * I(log(landMS))
                     + bhourscap* I(log(capitalMS))
                     + bhoursmat * I(log(materialsMS))
                     + I(zhoursmilk * log(grossmilkMS))
                     + I(zhoursother * log(grossotherMS)))
mean(pdata33$ehoursTLidfhours)

### land ###
pdata33$elandTLidfhours<- with(pdata33, bland
                    + blandland * I(log(landMS))
                    + blandvet * I(log(vetmedMS))
                    + blandfeed * I(log(feedexpMS))
                    + blandhours * I(log(hoursMS))
                    + blandcap* I(log(capitalMS))
                    + blandmat * I(log(materialsMS))
                    + I(zlandmilk * log(grossmilkMS))
```

```
                        + I(zlandother * log(grossotherMS)))

### capital ###
pdata33$ecapTLidfhours <- with(pdata33, bcap
                       + bcapcap * I(log(capitalMS))
                       + bcapfeed * I(log(feedexpMS))
                       + bcapvet * I(log(vetmedMS))
                       + bcaphours * I(log(hoursMS))
                       + bcapland * I(log(landMS))
                       + bcapmat * I(log(materialsMS))
                       + I(zcapmilk * log(grossmilkMS))
                       + I(zcapother * log(grossotherMS)))

### materials ###
pdata33$ematTLidfhours <- with(pdata33, bmat
                       + bmatmat * I(log(materialsMS))
                       + bmatfeed * I(log(feedexpMS))
                       + bmatvet * I(log(vetmedMS))
                       + bmathours * I(log(hoursMS))
                       + bmatland * I(log(landMS))
                       + bmatcap * I(log(capitalMS))
                       + I(zmatmilk * log(grossmilkMS))
                       + I(zmatother * log(grossotherMS)))
### Output elasticities ###
### gross milk ###
pdata33$emilkTLidfhours <- with(pdata33, amilk
                        + amilkother * I(log(grossotherMS))
                        + zfeedmilk * I(log(feedexpMS))
                        + zvetmilk * I(log(vetmedMS))
                        + zhoursmilk * I(log(hoursMS))
                        + zlandmilk * I(log(landMS))
                        + zcapmilk * I(log(capitalMS))
                        + zmatmilk * I(log(materialsMS)))

### gross other ###
pdata33$eotherTLidfhours <- with(pdata33, aother
                         + aothermilk * I(log(grossmilkMS))
                         + zfeedother * I(log(feedexpMS))
                         + zvetother * I(log(vetmedMS))
                         + zhoursother * I(log(hoursMS))
                         + zlandother * I(log(landMS))
                         + zcapother * I(log(capitalMS))
                         + zmatother * I(log(materialsMS)))
```

```
### summary statistics of distance elasticities ###
summary(pdata33[,c( "efeedTLidfhours",
"evetTLidfhours","ehoursTLidfhours","elandTLidfhours",
            "ecapTLidfhours","ematTLidfhours","emilkTLidfhours","eotherTLidfhours")])

# check if distance elasticities of the inputs sum up to one: they do!
range(pdata33$efeedTLidfhours +
pdata33$evetTLidfhours+pdata33$ehoursTLidfhours+pdata33$elandTLidfhours
   +pdata33$ecapTLidfhours + pdata33$ematTLidfhours )

### Elasticity of scale ###
pdata33$eScaleTLidfhours <- (-(pdata33$eotherTLidfhours + pdata33$emilkTLidfhours))^(-1)
summary(pdata33$eScaleTLidfhours)
sd(pdata33$ eScaleTLidfhours)
sum( pdata33$eScaleTLidfhours> 2 | pdata33$eScaleTLidfhours < 0.5 )
# Monotonicity in input and output quantities.
# Di(x,y) is non-decreasing in x if its first order derivatives with
# respect to input quantities are non-negative

sum(pdata33$efeedTLidfhours >= 0)
sum(pdata33$evetTLidfhours  >= 0)
sum(pdata33$ehoursTLidfhours >= 0)
sum(pdata33$elandTLidfhours >= 0)
sum(pdata33$ecapTLidfhours  >= 0)
sum(pdata33$ematTLidfhours >= 0)

pdata33$monoTLidfhourshours<-with(pdata33,efeedTLidfhours >= 0 & evetTLidfhours  >= 0 &
                 ehoursTLidfhours >= 0 & elandTLidfhours >= 0 & ecapTLidfhours  >= 0
                 & ematTLidfhours >= 0)

# number that violates monotonicity assumption (input)
sum( pdata33$monoTLidfhourshours )
sum( !pdata33$monoTLidfhourshours )
sum( pdata33$monoTLidfhourshours )/(sum( !pdata33$monoTLidfhourshours )
+sum(pdata33$monoTLidfhourshours))

# Di(x,y) is non-increasing in y if its first order derivatives with respect
# to input quantities are non-positive.

sum(pdata33$emilkTLidfhours <= 0)
sum(pdata33$eotherTLidfhours  <= 0)

# number that violates monotonicity assumption (output)
```

```
pdata33$monooutTLidfhourshours<-with(pdata33,emilkTLidfhours <= 0 & eotherTLidfhours
<= 0)
sum( pdata33$monooutTLidfhourshours )
sum( !pdata33$monooutTLidfhourshours )
sum( pdata33$monooutTLidfhourshours)/(sum( !pdata33$monooutTLidfhourshours )
+sum(pdata33$monooutTLidfhourshours))

# We can check if estimated Translog input distance is concave in input quantities (x)
# and quasiconcave in output quantities (y) at the frontier.

#### Quasi-concave in outputs ####
# first-order partial derivatives wrt outputs (with input distance measure = 1 )
pdata33$fmilkmilk <- amilk * (amilk-1) * 1 / pdata33$grossmilkMS^2
pdata33$fmilkother <- amilk * aother * 1 / (pdata33$grossmilkMS * pdata33$grossotherMS)
pdata33$fotherother <- aother * (aother-1) * 1 / pdata33$grossotherMS^2

# The following code creates a three-dimensional array with the Hessian matrices
# at all observations stacked upon each other:
bhm <- matrix( 0, nrow = 2, ncol = 2 )
bhm[ 1, 1 ] <- pdata33$fmilkmilk [1]
bhm[ 1, 2 ] <- bhm[ 2, 1 ] <- pdata33$fmilkother [1]
bhm[ 2, 2 ] <- pdata33$fotherother [1]

print(bhm)

# We check quasiconcavity in output quantities at the first observation
# by calculating the first and second leading principal minor:
det(bhm)

pdata33$quasiConv <- NA
for( obs in 1:nrow( pdata33 ) ) {
  bhmLoop <- matrix( 0, nrow = 2, ncol = 2 )
  bhmLoop[ 1, 1 ] <- pdata33$fmilkmilk[ 1 ]
  bhmLoop[ 1, 2 ] <- bhmLoop[ 2, 1 ] <- pdata33$fmilkother[ 1 ]
  bhmLoop[ 2, 2 ] <- pdata33$fotherother [ 1 ]
  pdata33$quasiConv[ obs ] <- det( bhmLoop[ 1:2, 1:2 ] ) < 0
}
sum(pdata33$quasiConv)

#### Concave in inputs ####
# second-order partial derivatives wrt output (with input distance measure = 1 )
pdata33$ffeed <- pdata33$efeedTLidfhours *  1 / pdata33$feedexpMS
pdata33$fvet <- pdata33$evetTLidfhours *  1 / pdata33$vetmedMS
pdata33$fhours <- pdata33$ehoursTLidfhours *  1 / pdata33$hoursMS
```

```
pdata33$fland <- pdata33$elandTLidfhours *  1 / pdata33$landMS
pdata33$fcap <- pdata33$ecapTLidfhours *  1 / pdata33$capitalMS
pdata33$fmat <- pdata33$ematTLidfhours *  1 / pdata33$materialsMS

pdata33$ffeedfeedTLidfhours <- with(pdata33,
                    ( bfeedfeed + efeedTLidfhours * efeedTLidfhours - 1 * efeedTLidfhours )
                    / + ( 1 / ( feedexpMS * feedexpMS ) ) )

pdata33$ffeedvetTLidfhours <- fvetfeedTLidfhours <- with (pdata33,
                               ( bfeedvet + efeedTLidfhours * evetTLidfhours - 0 *
efeedTLidfhours )
                               / +(1/( feedexpMS*vetmedMS)))

pdata33$ffeedhoursTLidfhours <- fhoursfeedTLidfhours <- with (pdata33,
                                 ( bfeedhours + efeedTLidfhours * ehoursTLidfhours - 0 *
efeedTLidfhours )
                                 / +(1/( feedexpMS*hoursMS)))

pdata33$ffeedlandTLidfhours <- flandfeedTLidfhours <- with (pdata33,
                                ( bfeedland + efeedTLidfhours * elandTLidfhours - 0 *
efeedTLidfhours )
                                / +(1/( feedexpMS*landMS)))

pdata33$ffeedcapTLidfhours <- fcapfeedTLidfhours <- with (pdata33,
                               ( bfeedcap + efeedTLidfhours * ecapTLidfhours - 0 *
efeedTLidfhours )
                               / +(1/( feedexpMS*capitalMS)))

pdata33$ffeedmatTLidfhours <- fmatfeedTLidfhours <- with (pdata33,
                               ( bfeedmat + efeedTLidfhours * ematTLidfhours - 0 *
efeedTLidfhours )
                               / +(1/( feedexpMS*materialsMS)))
pdata33$fvetvetTLidfhours <- with(pdata33,
                  ( bvetvet + evetTLidfhours * evetTLidfhours - 1 * evetTLidfhours )
                  / + ( 1 / ( vetmedMS * vetmedMS ) ) )

pdata33$fvethoursTLidfhours <- fhoursvetTLidfhours <- with (pdata33,
                                ( bvethours + evetTLidfhours * ehoursTLidfhours - 0 *
evetTLidfhours )
                                / +(1/( vetmedMS*hoursMS)))

pdata33$fvetlandTLidfhourshours <- flandvetTLidfhours <- with (pdata33,
                               ( bvetland + evetTLidfhours * elandTLidfhours - 0 *
evetTLidfhours )
```

```
                                  / +(1/( vetmedMS*landMS)))

pdata33$fvetcapTLidfhours <- fcapvetTLidfhours <- with (pdata33,
                                 ( bvetcap + evetTLidfhours * ecapTLidfhours - 0 * evetTLidfhours
)
                                 / +(1/( vetmedMS*capitalMS)))

pdata33$fvetmatTLidfhours <- fmatvetTLidfhours <- with (pdata33,
                                 ( bvetmat + evetTLidfhours * ematTLidfhours - 0 * evetTLidfhours
)
                                 / +(1/( vetmedMS*materialsMS)))

pdata33$fhourshoursTLidfhours <- with(pdata33,
                        ( bhourshours + ehoursTLidfhours * ehoursTLidfhours - 1 *
ehoursTLidfhours )
                        / + ( 1 / ( hoursMS * hoursMS ) ) )

pdata33$fhourslandTLidfhours <- flandhoursTLidfhours <- with (pdata33,
                         ( bhoursland + ehoursTLidfhours * elandTLidfhours - 0 *
ehoursTLidfhours )
                          / +(1/( hoursMS*landMS)))

pdata33$fhourscapTLidfhours <- fcaphoursTLidfhours <- with (pdata33,
                      ( bhourscap + ehoursTLidfhours * ecapTLidfhours - 0 * ehoursTLidfhours )
                      / +(1/( hoursMS*capitalMS)))

pdata33$fhoursmatTLidfhours <- fmathoursTLidfhours <- with (pdata33,
                        ( bhoursmat + ehoursTLidfhours * ematTLidfhours - 0 * ehoursTLidfhours
)
                    / +(1/( hoursMS*materialsMS)))

pdata33$flandlandTLidfhours <- with(pdata33,
                      ( blandland + elandTLidfhours * elandTLidfhours - 1 * elandTLidfhours )
                      / + ( 1 / ( landMS * landMS ) ) )

pdata33$flandcapTLidfhours <- fcaplandTLidfhours <- with (pdata33,
                                  ( blandcap + elandTLidfhours * ecapTLidfhours - 0 *
elandTLidfhours )
                                  / +(1/( landMS*capitalMS)))

pdata33$flandmatTLidfhours <- fmatlandTLidfhours <- with (pdata33,
                                  ( blandmat + elandTLidfhours * ematTLidfhours - 0 *
elandTLidfhours )
                                  / +(1/( landMS*materialsMS)))
```

```
pdata33$fcapcapTLidfhours <- with(pdata33,
                      ( bcapcap + ecapTLidfhours * ecapTLidfhours - 1 * ecapTLidfhours )
                      / + ( 1 / ( capitalMS * capitalMS ) ) )

pdata33$fcapmatTLidfhours <- fmatcapTLidfhours <- with (pdata33,
                                  ( bcapmat + ecapTLidfhours * ematTLidfhours - 0 *
ecapTLidfhours )
                                  / +(1/( capitalMS*materialsMS)))

pdata33$fmatmatTLidfhours <- with(pdata33,
                      ( bmatmat + ematTLidfhours * ematTLidfhours - 1 * ematTLidfhours )
                      / + ( 1 / ( materialsMS * materialsMS ) ) )

# First, we prepare the Hessian matrix for the first observation:
hessian <- matrix( NA, nrow = 6, ncol = 6 )
hessian[ 1, 1 ] <- pdata33$ffeedfeedTLidfhours[1]
hessian[ 1, 2 ] <- hessian[ 2, 1 ] <- pdata33$ffeedvetTLidfhours[1]
hessian[ 1, 3 ] <- hessian[ 3, 1 ] <- pdata33$ffeedhoursTLidfhours[1]
hessian[ 1, 4 ] <- hessian[ 4, 1 ] <- pdata33$ffeedlandTLidfhours[1]
hessian[ 1, 5 ] <- hessian[ 5, 1 ] <- pdata33$ffeedcapTLidfhours[1]
hessian[ 1, 6 ] <- hessian[ 6, 1 ] <- pdata33$ffeedmatTLidfhours[1]
hessian[ 2, 2 ] <- pdata33$fvetvetTLidfhours[1]
hessian[ 2, 3 ] <- hessian[ 3, 2 ] <-pdata33$fvethoursTLidfhours[1]
hessian[ 2, 4 ] <- hessian[ 4, 2 ] <-pdata33$fvetlandTLidfhours[1]
hessian[ 2, 5 ] <- hessian[ 5, 2 ] <-pdata33$fvetcapTLidfhours[1]
hessian[ 2, 6 ] <- hessian[ 6, 2 ] <-pdata33$fvetmatTLidfhours[1]
hessian[ 3, 3 ] <- pdata33$fhourshoursTLidfhours[1]
hessian[ 3, 4 ] <- hessian[ 4, 3 ] <-pdata33$fhourslandTLidfhours[1]
hessian[ 3, 5 ] <- hessian[ 5, 3 ] <-pdata33$fhourscapTLidfhours[1]
hessian[ 3, 6 ] <- hessian[ 6, 3 ] <-pdata33$fhoursmatTLidfhours[1]
hessian[ 4, 4 ] <- pdata33$flandlandTLidfhours[1]
hessian[ 4, 5 ] <- hessian[ 5, 4 ] <-pdata33$flandcapTLidfhours[1]
hessian[ 4, 6 ] <- hessian[ 6, 4 ] <-pdata33$flandmatTLidfhours[1]
hessian[ 5, 5 ] <- pdata33$fcapcapTLidfhours[1]
hessian[ 5, 6 ] <- hessian[ 6, 5 ] <-pdata33$fcapmatTLidfhours[1]
hessian[ 6, 6 ] <- pdata33$fmatmatTLidfhours[1]
print( hessian )

# As all diagonal elements of this Hessian matrix are zero, the necessary
# conditions for positive semidefiniteness are not fulfilled for the first observation.

det( hessian )
```

```
pdata33$concaveTLidfhours <- pdata33$ffeedfeedTLidfhours<= 0
sum(pdata33$concaveTLidfhours)

######### ESTIMATE ODF #############
## step 4.2 ###

library(foreign)
library(frontier)
library(AER)
library(plm)
library(gmp)
library(Rmpfr)

# We create new variables with mean-scaled input and output quantities:
# ouptut quantities
pdata33$grossmilkMS<-pdata33$grossmilk_2010level/mean(pdata33$grossmilk_2010level)
pdata33$grossotherMS<-pdata33$grossother_2010level/mean(pdata33$grossother_2010level)
# input quantities
pdata33$vetmedMS<-pdata33$vetmed_2010level/mean(pdata33$vetmed_2010level)
pdata33$hoursMS<-pdata33$hours/mean(pdata33$hours)
pdata33$feedexpMS<-pdata33$feedexp_2010level/mean(pdata33$feedexp_2010level)
pdata33$totalwagesMS<-pdata33$totalwages_2010level/mean(pdata33$totalwages_2010level)
pdata33$landMS<-pdata33$land/mean(pdata33$land)
pdata33$materialsMS<-pdata33$materials_2010level/mean(pdata33$materials_2010level)
pdata33$capitalMS<-pdata33$capital/mean(pdata33$capital)



######################################
#### OUTPUT DISTANCE FUNCTION ###
######################################

# set arguments for all estimations using  sfa()
sfaSearchStep <- 1e-6
sfaSearchTol <- 1e-11
sfaTol <- 1e-11

#### Cobb-Douglas: General model (or unrestricted) ####
########################################################
CDodfunres<-sfa( -log(grossmilkMS) ~ log(grossotherMS/grossmilkMS)
            + log( feedexpMS)
            + log( vetmedMS)
            + log(totalwagesMS)
            + log(landMS)
```

```
            + log(capitalMS)
            + Regnskabsaar
            + milkingsystem
            + jersey
            + organic
            |
            + mastitis
            + hoofdis
            + reprodis
            + otherdis
            + cell1
            + viable1
            + spore1
            + consultant
            + managerage
            ,data=pdata33, ineffDecrease = FALSE,
            timeEffect = TRUE,
            searchStep = sfaSearchStep, searchTol = sfaSearchTol, tol = sfaTol )
summary(CDodfunres)

### Translog: general model (unrestricted) ###
#########################################
TLodfunres <- sfa(-log(grossmilkMS) ~ log(grossotherMS/grossmilkMS)
             + I(0.5*log(grossotherMS/grossmilkMS)^2)
             + log(feedexpMS) + log(vetmedMS) + log(totalwagesMS)
             + log(landMS) + log(capitalMS) + log(materialsMS)
             + I(0.5*log(feedexpMS)^2)
             + I(0.5*log(vetmedMS)^2)
             + I(0.5*log(totalwagesMS)^2)
             + I(0.5*log(landMS)^2)
             + I(0.5*log(capitalMS)^2)
             + I(0.5*log(materialsMS)^2)
             + I(log(feedexpMS)*log(vetmedMS))
             + I(log(feedexpMS)*log(totalwagesMS))
             + I(log(feedexpMS)*log(landMS))
             + I(log(feedexpMS)*log(capitalMS))
             + I(log(feedexpMS)*log(materialsMS))
             + I(log(vetmedMS)*log(totalwagesMS))
             + I(log(vetmedMS)*log(landMS))
             + I(log(vetmedMS)*log(capitalMS))
             + I(log(vetmedMS)*log(materialsMS))
             + I(log(totalwagesMS)*log(landMS))
             + I(log(totalwagesMS)*log(capitalMS))
             + I(log(totalwagesMS)*log(materialsMS))
```

```
            + I(log(landMS)*log(capitalMS))
            + I(log(landMS)*log(materialsMS))
            + I(log(capitalMS)*log(materialsMS))
            + I(log(feedexpMS)*log(grossotherMS/grossmilkMS))
            + I(log(vetmedMS)*log(grossotherMS/grossmilkMS))
            + I(log(totalwagesMS)*log(grossotherMS/grossmilkMS))
            + I(log(landMS)*log(grossotherMS/grossmilkMS))
            + I(log(capitalMS)*log(grossotherMS/grossmilkMS))
            + I(log(materialsMS)*log(grossotherMS/grossmilkMS))
            + Regnskabsaar
            + milkingsystem
            + jersey
            + organic
            |
            + mastitis
            + hoofdis
            + reprodis
            + otherdis
            + cell1
            + viable1
            + spore1
            + managerage
            + consultant
            ,data = pdata33
            , ineffDecrease = FALSE ,timeEffect = TRUE
            ,searchStep = sfaSearchStep, searchTol = sfaSearchTol, tol = sfaTol)
summary(TLodfunres)

lrtest(TLodfunres) # Small p-value indicate that there is (output oriented) technical inefficiency.

### Translog: restricted model with diseases ###
################################################
TLodfres1 <- sfa(-log(grossmilkMS) ~ log(grossotherMS/grossmilkMS)
            + I(0.5*log(grossotherMS/grossmilkMS)^2)
            + log(feedexpMS) + log(vetmedMS) + log(totalwagesMS)
            + log(landMS) + log(capitalMS) + log(materialsMS)
            + I(0.5*log(feedexpMS)^2)
            + I(0.5*log(vetmedMS)^2)
            + I(0.5*log(totalwagesMS)^2)
            + I(0.5*log(landMS)^2)
            + I(0.5*log(capitalMS)^2)
            + I(0.5*log(materialsMS)^2)
            + I(log(feedexpMS)*log(vetmedMS))
            + I(log(feedexpMS)*log(totalwagesMS))
```

```
            + I(log(feedexpMS)*log(landMS))
            + I(log(feedexpMS)*log(capitalMS))
            + I(log(feedexpMS)*log(materialsMS))
            + I(log(vetmedMS)*log(totalwagesMS))
            + I(log(vetmedMS)*log(landMS))
            + I(log(vetmedMS)*log(capitalMS))
            + I(log(vetmedMS)*log(materialsMS))
            + I(log(totalwagesMS)*log(landMS))
            + I(log(totalwagesMS)*log(capitalMS))
            + I(log(totalwagesMS)*log(materialsMS))
            + I(log(landMS)*log(capitalMS))
            + I(log(landMS)*log(materialsMS))
            + I(log(capitalMS)*log(materialsMS))
            + I(log(feedexpMS)*log(grossotherMS/grossmilkMS))
            + I(log(vetmedMS)*log(grossotherMS/grossmilkMS))
            + I(log(totalwagesMS)*log(grossotherMS/grossmilkMS))
            + I(log(landMS)*log(grossotherMS/grossmilkMS))
            + I(log(capitalMS)*log(grossotherMS/grossmilkMS))
            + I(log(materialsMS)*log(grossotherMS/grossmilkMS))
            + Regnskabsaar
            + milkingsystem
            + jersey
            + organic
            |
            + mastitis
            + hoofdis
            + reprodis
            + otherdis
            + managerage
            + consultant
            ,data = pdata33
            , ineffDecrease = FALSE ,timeEffect = TRUE
            ,searchStep = sfaSearchStep, searchTol = sfaSearchTol, tol = sfaTol)
summary(TLodfres1)

lrtest(TLodfres1) # Small p-value indicate that there is (output oriented) technical inefficiency.
lrtest(TLodfunres,TLodfres1)

### Translog: restricted model with milk quality ###
####################################################
TLodfres2 <- sfa(-log(grossmilkMS) ~ log(grossotherMS/grossmilkMS)
            + I(0.5*log(grossotherMS/grossmilkMS)^2)
            + log(feedexpMS) + log(vetmedMS) + log(totalwagesMS)
            + log(landMS) + log(capitalMS) + log(materialsMS)
```

```
              + I(0.5*log(feedexpMS)^2)
              + I(0.5*log(vetmedMS)^2)
              + I(0.5*log(totalwagesMS)^2)
              + I(0.5*log(landMS)^2)
              + I(0.5*log(capitalMS)^2)
              + I(0.5*log(materialsMS)^2)
              + I(log(feedexpMS)*log(vetmedMS))
              + I(log(feedexpMS)*log(totalwagesMS))
              + I(log(feedexpMS)*log(landMS))
              + I(log(feedexpMS)*log(capitalMS))
              + I(log(feedexpMS)*log(materialsMS))
              + I(log(vetmedMS)*log(totalwagesMS))
              + I(log(vetmedMS)*log(landMS))
              + I(log(vetmedMS)*log(capitalMS))
              + I(log(vetmedMS)*log(materialsMS))
              + I(log(totalwagesMS)*log(landMS))
              + I(log(totalwagesMS)*log(capitalMS))
              + I(log(totalwagesMS)*log(materialsMS))
              + I(log(landMS)*log(capitalMS))
              + I(log(landMS)*log(materialsMS))
              + I(log(capitalMS)*log(materialsMS))
              + I(log(feedexpMS)*log(grossotherMS/grossmilkMS))
              + I(log(vetmedMS)*log(grossotherMS/grossmilkMS))
              + I(log(totalwagesMS)*log(grossotherMS/grossmilkMS))
              + I(log(landMS)*log(grossotherMS/grossmilkMS))
              + I(log(capitalMS)*log(grossotherMS/grossmilkMS))
              + I(log(materialsMS)*log(grossotherMS/grossmilkMS))
              + Regnskabsaar
              + milkingsystem
              + jersey
              + organic
              |
              + cell1
              + viable1
              + spore1
              + managerage
              + consultant
              ,data = pdata33
              , ineffDecrease = FALSE ,timeEffect = TRUE
              ,searchStep = sfaSearchStep, searchTol = sfaSearchTol, tol = sfaTol)
summary(TLodfres2)
lrtest(TLodfres2) # Small p-value indicate that there is (output oriented) technical inefficiency.
lrtest(TLodfunres,TLodfres2) # proceeding with TLodfres2
```

```
### Translog: restricted model without managerage and consultant ###
########################################################################
TLodfres3 <- sfa(-log(grossmilkMS) ~ log(grossotherMS/grossmilkMS)
          + I(0.5*log(grossotherMS/grossmilkMS)^2)
          + log(feedexpMS) + log(vetmedMS) + log(totalwagesMS)
          + log(landMS) + log(capitalMS) + log(materialsMS)
          + I(0.5*log(feedexpMS)^2)
          + I(0.5*log(vetmedMS)^2)
          + I(0.5*log(totalwagesMS)^2)
          + I(0.5*log(landMS)^2)
          + I(0.5*log(capitalMS)^2)
          + I(0.5*log(materialsMS)^2)
          + I(log(feedexpMS)*log(vetmedMS))
          + I(log(feedexpMS)*log(totalwagesMS))
          + I(log(feedexpMS)*log(landMS))
          + I(log(feedexpMS)*log(capitalMS))
          + I(log(feedexpMS)*log(materialsMS))
          + I(log(vetmedMS)*log(totalwagesMS))
          + I(log(vetmedMS)*log(landMS))
          + I(log(vetmedMS)*log(capitalMS))
          + I(log(vetmedMS)*log(materialsMS))
          + I(log(totalwagesMS)*log(landMS))
          + I(log(totalwagesMS)*log(capitalMS))
          + I(log(totalwagesMS)*log(materialsMS))
          + I(log(landMS)*log(capitalMS))
          + I(log(landMS)*log(materialsMS))
          + I(log(capitalMS)*log(materialsMS))
          + I(log(feedexpMS)*log(grossotherMS/grossmilkMS))
          + I(log(vetmedMS)*log(grossotherMS/grossmilkMS))
          + I(log(totalwagesMS)*log(grossotherMS/grossmilkMS))
          + I(log(landMS)*log(grossotherMS/grossmilkMS))
          + I(log(capitalMS)*log(grossotherMS/grossmilkMS))
          + I(log(materialsMS)*log(grossotherMS/grossmilkMS))
          + Regnskabsaar
          + milkingsystem
          + jersey
          + organic
          |
          + cell1
          + viable1
          + spore1
          ,data = pdata33
          , ineffDecrease = FALSE ,timeEffect = TRUE
          ,searchStep = sfaSearchStep, searchTol = sfaSearchTol, tol = sfaTol)
```

```
summary(TLodfres3)
lrtest(TLodfres3) # Small p-value indicate that there is (output oriented) technical inefficiency.

lrtest(TLodfres2,TLodfres3) # small p-value, can reject null-hypothesis of no difference, choose TLodfres2

### Skeewness  test ###
library(moments)
skewness(residuals(TLodfunres,asInData = TRUE)) # all are right skewed
skewness(residuals(TLodfres1,asInData = TRUE))
skewness(residuals(TLodfres21,asInData = TRUE))

hist( residuals(TLodfunres), 15 )
hist( residuals(TLodfres1), 15 )
hist( residuals(TLodfres2), 15 )

### Choosing the general model  ###
TLodf <- TLodfres2
summary(TLodf, extraPar = TRUE )

### Test the properties ###
###########################
# Coefficients of TLodf #
aothero <- coef(TLodf)["log(grossotherMS/grossmilkMS)"]
amilko <- 1-aothero
aothertohero <- coef(TLodf)["I(0.5 * log(grossotherMS/grossmilkMS)^2)"]
aothermilko <- amilkothero <- - aothertohero
amilkmilko <- - amilkothero

bfeedo <- coef(TLodf)["log(feedexpMS)"]
bveto <- coef(TLodf)["log(vetmedMS)"]
btoto <- coef(TLodf)["log(totalwagesMS)"]
blando <- coef(TLodf)["log(landMS)"]
bcapo <- coef(TLodf)["log(capitalMS)"]
bmato <- coef(TLodf)["log(materialsMS)"]

bfeedfeedo <-coef(TLodf)["I(0.5 * log(feedexpMS)^2)"]
bfeedveto <- bvetfeedo <-coef(TLodf)["I(log(feedexpMS) * log(vetmedMS))"]
bfeedtoto <- btotfeedo <-coef(TLodf)["I(log(feedexpMS) * log(totalwagesMS))"]
bfeedlando <- blandfeedo <-coef(TLodf)["I(log(feedexpMS) * log(landMS))"]
bfeedcapo <- capfeedo <-coef(TLodf)["I(log(feedexpMS) * log(capitalMS))"]
bfeedmato <- bmatfeedo <-coef(TLodf)["I(log(feedexpMS) * log(materialsMS))"]

bvetveto <-coef(TLodf)["I(0.5 * log(vetmedMS)^2)"]
```

```
bvettoto <- btotveto <-coef(TLodf)["I(log(vetmedMS) * log(totalwagesMS))"]
bvetlando <- blandveto <-coef(TLodf)["I(log(vetmedMS) * log(landMS))"]
bvetcapo <- bcapveto <-coef(TLodf)["I(log(vetmedMS) * log(capitalMS))"]
bvetmato <- bmatveto <-coef(TLodf)["I(log(vetmedMS) * log(capitalMS))"]

btottoto <-coef(TLodf)["I(0.5 * log(totalwagesMS)^2)"]
btotlando <- blandtoto <-coef(TLodf)["I(log(totalwagesMS) * log(landMS))"]
btotcapo <- bcaptoto <-coef(TLodf)["I(log(totalwagesMS) * log(capitalMS))"]
btotmato  <- bmattoto <-coef(TLodf)["I(log(totalwagesMS) * log(materialsMS))"]
blandlando <-coef(TLodf)["I(0.5 * log(landMS)^2)"]
blandcapo <- bcaplando <-coef(TLodf)["I(log(landMS) * log(capitalMS))"]
blandmato <- bmatlando <-coef(TLodf)["I(log(landMS) * log(materialsMS))"]
bcapcapo <-coef(TLodf)["I(0.5 * log(capitalMS)^2)"]
bcapmato  <- bmatcapo<-coef(TLodf)["I(log(capitalMS) * log(materialsMS))"]
bmatmato <-coef(TLodf)["I(0.5 * log(materialsMS)^2)"]

zfeedothero<- coef(TLodf)["I(log(feedexpMS) * log(grossotherMS/grossmilkMS))"]
zfeedmilko <- - zfeedothero
zvetothero <- coef(TLodf)["I(log(vetmedMS) * log(grossotherMS/grossmilkMS))"]
zvetmilko <- - zvetothero
ztotothero <- coef(TLodf)["I(log(totalwagesMS) * log(grossotherMS/grossmilkMS))"]
ztotmilko <- - ztotothero
zlandothero <- coef(TLodf)["I(log(landMS) * log(grossotherMS/grossmilkMS))"]
zlandmilko <- - zlandothero
zcapothero <- coef(TLodf)["I(log(capitalMS) * log(grossotherMS/grossmilkMS))"]
zcapmilko <- - zcapothero
zmatothero <- coef(TLodf)["I(log(materialsMS) * log(grossotherMS/grossmilkMS))"]
zmatmilko<- - zmatothero

### Input elasticities ###
pdata33$efeedTLodf <- with(pdata33, bfeedo
                     +bfeedfeedo * I(log(feedexpMS))
                     +bfeedveto * I(log(vetmedMS))
                     +bfeedtoto * I(log(totalwagesMS))
                     +bfeedlando * I(log(landMS))
                     +bfeedcapo *I(log(capitalMS))
                     +bfeedmato * I(log(materialsMS))
                     +I(zfeedothero * log(grossotherMS))
                     +I(zfeedmilko *log(grossmilkMS)))
mean(pdata33$efeedTLodf)
sd(pdata33$efeedTLodf)
pdata33$evetTLodf <- with(pdata33, bveto
                     +bvetveto * I(log(vetmedMS))
                     +bfeedveto * I(log(feedexpMS))
```

```
                    +bvettoto * I(log(totalwagesMS))
                    +bvetlando * I(log(landMS))
                    +bvetcapo *I(log(capitalMS))
                    +bvetmato * I(log(materialsMS))
                    +I(zvetothero * log(grossotherMS))
                    +I(zvetmilko *log(grossmilkMS)))
mean(pdata33$evetTLodf)
sd(pdata33$evetTLodf)
pdata33$etotTLodf <- with(pdata33, btoto
                    +btottoto * I(log(totalwagesMS))
                    +bfeedtoto * I(log(feedexpMS))
                    +bvettoto * I(log(vetmedMS))
                    +btotlando * I(log(landMS))
                    +btotcapo *I(log(capitalMS))
                    +btotmato * I(log(materialsMS))
                    +I(ztotothero * log(grossotherMS))
                    +I(ztotmilko *log(grossmilkMS)))
mean(pdata33$etotTLodf)
sd(pdata33$etotTLodf)
pdata33$elandTLodf <- with(pdata33, blando
                    +blandlando * I(log(landMS))
                    +bfeedlando * I(log(feedexpMS))
                    +bvetlando * I(log(vetmedMS))
                    +btotlando * I(log(totalwagesMS))
                    +blandcapo *I(log(capitalMS))
                    +blandmato * I(log(materialsMS))
                    +I(zlandothero * log(grossotherMS))
                    +I(zlandmilko *log(grossmilkMS)))
mean(pdata33$elandTLodf)
sd(pdata33$elandTLodf)
pdata33$ecapTLodf <- with(pdata33, bcapo
                    +bcapcapo * I(log(capitalMS))
                    +bfeedcapo * I(log(feedexpMS))
                    +bvetcapo * I(log(vetmedMS))
                    +btotcapo * I(log(totalwagesMS))
                    +blandcapo *I(log(landMS))
                    +bcapmato * I(log(materialsMS))
                    +I(zcapothero * log(grossotherMS))
                    +I(zcapmilko *log(grossmilkMS)))
mean(pdata33$ecapTLodf)
sd(pdata33$ecapTLodf)
pdata33$ematTLodf <- with(pdata33, bmato
                    +bmatmato * I(log(materialsMS))
                    +bfeedmato * I(log(feedexpMS))
```

```
                      +bvetmato * I(log(vetmedMS))
                      +btotmato * I(log(totalwagesMS))
                      +blandmato *I(log(landMS))
                      +bcapmato * I(log(capitalMS))
                      +I(zmatothero * log(grossotherMS))
                      +I(zmatmilko *log(grossmilkMS)))
mean(pdata33$ematTLodf)
sd(pdata33$ematTLodf)

### Output elasticities ###
pdata33$emilkTLodf <- with(pdata33, amilko
                      + amilkothero * I(log(grossotherMS))
                      + zfeedmilko * I(log(feedexpMS))
                      + zvetmilko * I(log(vetmedMS))
                      + ztotmilko * I(log(totalwagesMS))
                      + zlandmilko * I(log(landMS))
                      + zcapmilko * I(log(capitalMS))
                      + zmatmilko * I(log(materialsMS)))
mean(pdata33$emilkTLodf)
sd(pdata33$emilkTLodf)
pdata33$eotherTLodf <- with(pdata33, aothero
                       + aothermilko * I(log(grossmilkMS))
                       + zfeedothero * I(log(feedexpMS))
                       + zvetothero * I(log(vetmedMS))
                       + ztotothero * I(log(totalwagesMS))
                       + zlandothero * I(log(landMS))
                       + zcapothero * I(log(capitalMS))
                       + zmatothero * I(log(materialsMS)))
mean(pdata33$eotherTLodf)
sd(pdata33$eotherTLodf)

### summary statistics of distance elasticities ###
summary(pdata33[,c( "efeedTLodf", "evetTLodf", "etotTLodf", "elandTLodf",
                "ecapTLodf","ematTLodf","emilkTLodf","eotherTLodf")])

range(pdata33$emilkTLodf+pdata33$eotherTLodf)

range(pdata33$efeedTLodf + pdata33$evetTLodf+pdata33$etotTLodf+pdata33$elandTLdf
    +pdata33$ecapTLodf + pdata33$ematTLodf)

### Elasticity of scale ###
pdata33$eScaleTLodf <- - rowSums (pdata33 [,c("efeedTLodf", "evetTLodf", "etotTLodf",
"elandTLodf", "ecapTLodf","ematTLodf")])
```

```
summary(pdata33$eScaleTLodf)
table(pdata33$eScaleTLodf & pdata33$Regnskabsaar==2011)
sd(pdata33$eScaleTLodf)
sum( pdata33$eScaleTLodf> 2 | pdata33$eScaleTLodf < 0.5 )
# all elasticities are within the interval

# Monotonicity in input and output quantities.
# Di(x,y) is non-decreasing in x if its first order derivatives with
# respect to input quantities are non-negative

# Monotonicity in input and output quantities.
# Do(x,y) is non-decreasing in y  and non-increasing in x.

sum(!pdata33$efeedTLodf <= 0)
sum(!pdata33$evetTLodf  <= 0)
sum(!pdata33$etotTLodf <= 0)
sum(!pdata33$elandTLodf <= 0)
sum(!pdata33$ecapTLodf  <= 0)
sum(!pdata33$ematTLodf <= 0)

pdata33$monoTLodf<-with(pdata33,efeedTLodf <= 0 & evetTLodf  <= 0 & etotTLodf <= 0
                    & elandTLodf <= 0 & ecapTLodf  <= 0 & ematTLodf <= 0)

# number that violates monotonicity assumption (input)
sum( pdata33$monoTLodf )
sum( !pdata33$monoTLodf )
sum( pdata33$monoTLodf )/(sum( !pdata33$monoTLodf ) +sum(pdata33$monoTLodf))

# Do(x,y) is non-decreasing in y  and non-increasing in x.
sum(!pdata33$emilkTLodf >= 0)
sum(!pdata33$eotherTLodf  >= 0)

# number that violates monotonicity assumption (output)
pdata33$monooutTLodf<-with(pdata33,emilkTLodf >= 0 & eotherTLodf  >= 0)
sum( pdata33$monooutTLodf )
sum( !pdata33$monooutTLodf )
sum( pdata33$monooutTLodf)/(sum( !pdata33$monooutTLodf )
+sum(pdata33$monooutTLodf))

# We can check if estimated Translog output distance is quasi-convex in input quantities (x)
# and convex in output quantities (y) at the frontier.

#### Quasi-convex in inputs ####
# first-order partial derivatives, inputs
```

```
pdata33$ffeedTLodf <- with(pdata33, efeedTLodf * (1/I(log(feedexpMS))))
pdata33$fvetTLodf<- with(pdata33, evetTLodf * (1/I(log(vetmedMS))))
pdata33$ftotTLodf<- with(pdata33, etotTLodf * (1/I(log(totalwagesMS))))
pdata33$flandTLodf<- with(pdata33, elandTLodf * (1/I(log(landMS))))
pdata33$fcapTLodf<- with(pdata33, ecapTLodf * (1/I(log(capitalMS))))
pdata33$fmatTLodf<- with(pdata33, ematTLodf * (1/I(log(materialsMS))))

#values of elasticities with respect to output quantities for each observation
pdata33$hmilkTLodf <- with ( pdata33, emilkTLodf * ( 1 / I( log( grossmilkMS ) ) ) )
summary ( pdata33$hmilkTLodf )
sum ( ( pdata33$hmilkTLodf ) < 0 )

pdata33$hotherTLodf <- with ( pdata33, eotherTLodf * ( 1 / I( log( grossotherMS ) ) ) )
summary ( pdata33$hotherTLodf )
sum ( ( pdata33$hOtherTLodf ) < 0 )

# quasi convex in inputs, valus of elasticity with respect to input quantities
pdata33$ffeedfeedTLodf <-with(pdata33,(bfeedfeedo+efeedTLodf*efeedTLodf-1*efeedTLodf)
                /+(1/(feedexpMS*feedexpMS)))

pdata33$ffeedvetTLodf <- fvetfeedTLodf <-with(pdata33,
                (bfeedveto+efeedTLodf*evetTLodf-0*efeedTLodf)
                /+(1/(feedexpMS*vetmedMS)))

pdata33$ffeedtotTLodf <- ftotfeedTLodf <-with(pdata33,
                (bfeedtoto+efeedTLodf*etotTLodf-0*efeedTLodf)
                /+(1/(feedexpMS*totalwagesMS)))

pdata33$ffeedlandTLodf <- flandfeedTLodf <-with(pdata33,
                 (bfeedlando+efeedTLodf*elandTLodf-0*efeedTLodf)
                 /+(1/(feedexpMS*landMS)))

pdata33$ffeedcapTLodf <- fcapfeedTLodf <-with(pdata33,
                  (bfeedcapo+efeedTLodf*ecapTLodf-0*efeedTLodf)
                  /+(1/(feedexpMS*capitalMS)))

pdata33$ffeedmatTLodf <- fmatfeedTLodf <-with(pdata33,
                  (bfeedmato+efeedTLodf*ematTLodf-0*efeedTLodf)
                  /+(1/(feedexpMS*materialsMS)))

pdata33$fvetvetTLodf <- with(pdata33,
                  (bvetveto+evetTLodf*evetTLodf-1*evetTLodf)
                  /+(1/(vetmedMS*vetmedMS)))
```

```
pdata33$fvettotTLodf <- ftotvetTLodf <-with(pdata33,
                    (bvettoto+evetTLodf*etotTLodf-0*evetTLodf)
                    /+(1/(vetmedMS*totalwagesMS)))

pdata33$fvetlandTLodf <- flandvetTLodf <-with(pdata33,
                     (bvetlando+evetTLodf*elandTLodf-0*evetTLodf)
                     /+(1/(vetmedMS*landMS)))

pdata33$fvetcapTLodf <- fcapvetTLodf <-with(pdata33,
                     (bvetcapo+evetTLodf*ecapTLodf-0*evetTLodf)
                     /+(1/(vetmedMS*capitalMS)))

pdata33$fvetmatTLodf <- fmatvetTLodf <-with(pdata33,
                      (bvetmato+evetTLodf*ematTLodf-0*evetTLodf)
                      /+(1/(vetmedMS*materialsMS)))

pdata33$ftottotTLodf <- with(pdata33,
              (btottoto+etotTLodf*etotTLodf-1*etotTLodf)
              /+(1/(totalwagesMS*totalwagesMS)))

pdata33$ftotlandTLodf <- flandtotTLodf <- with(pdata33,
               (btotlando+etotTLodf*elandTLodf-0*etotTLodf)
               /+(1/(totalwagesMS*landMS)))

pdata33$ftotcapTLodf <- fcaptotTLodf <- with(pdata33,
                    (btotcapo+etotTLodf*ecapTLodf-0*etotTLodf)
                     /+(1/(totalwagesMS*capitalMS)))

pdata33$ftotmatTLodf <- fmattotTLodf <- with(pdata33,
                      (btotmato+etotTLodf*ematTLodf-0*etotTLodf)
                       /+(1/(totalwagesMS*materialsMS)))

pdata33$flandlandTLodf <- with(pdata33,
                (blandlando+elandTLodf*elandTLodf-1*elandTLodf)
                /+(1/(landMS*landMS)))

pdata33$flandcapTLodf <- fcaplandTLodf <- with(pdata33,
                       (blandcapo+elandTLodf*ecapTLodf-0*elandTLodf)
                       /+(1/(landMS*capitalMS)))

pdata33$flandmatTLodf <- fmatlandTLodf <- with(pdata33,
                       (blandmato+elandTLodf*ematTLodf-0*elandTLodf)
                           /+(1/(landMS*materialsMS)))
```

```
pdata33$fcapcapTLodf <- with(pdata33,
                        (bcapcapo+ecapTLodf*ecapTLodf-1*ecapTLodf)
                        /+(1/(capitalMS*capitalMS)))

pdata33$fcapmatTLodf <- fmatcapTLodf <- with(pdata33,
                                (bcapmato+ecapTLodf*ematTLodf-0*ecapTLodf)
                                    /+(1/(capitalMS*materialsMS)))

pdata33$fmatmatTLodf <- with(pdata33,
                       (bmatmato+ematTLodf*ematTLodf-1*ematTLodf)
                       /+(1/(materialsMS*materialsMS)))

# Hessian matrix for the first observation
hm <- array( 0, c( 7, 7, nrow( pdata33 ) ) )
dim(hm)
hm[ 1, 2, ] <- hm[ 2, 1, ] <- pdata33$ffeedTLodf#[ 1 ]
hm[ 1, 3, ] <- hm[ 3, 1, ] <- pdata33$fvetTLodf#[ 1 ]
hm[ 1, 4,] <- hm[ 4, 1, ] <- pdata33$ftotTLodf#[ 1 ]
hm[ 1, 5, ] <- hm[ 5, 1, ] <- pdata33$flandTLodf#[ 1 ]
hm[ 1, 6, ] <- hm[ 6, 1,] <- pdata33$fcapTLodf#[ 1 ]
hm[ 1, 7, ] <- hm[ 7, 1, ] <- pdata33$fmatTLodf#[ 1 ]
hm[ 2, 2, ] <- pdata33$ffeedfeedTLodf#[ 1 ]
hm[ 3, 3, ] <- pdata33$fvetvetTLodf#[ 1 ]
hm[ 4, 4, ] <- pdata33$ftottotTLodf#[ 1 ]
hm[ 5, 5, ] <- pdata33$flandlandTLodf#[ 1 ]
hm[ 6, 6, ] <- pdata33$fcapcapTLodf#[ 1 ]
hm[ 7, 7, ] <- pdata33$fmatmatTLodf#[ 1 ]
hm[ 2, 3, ] <- hm[ 3, 2, ] <- pdata33$ffeedvetTLodf#[ 1 ]
hm[ 2, 4, ] <- hm[ 4, 2, ] <- pdata33$ffeedtotTLodf#[ 1 ]
hm[ 2, 5, ] <- hm[ 5, 2, ] <- pdata33$ffeedlandTLodf#[ 1 ]
hm[ 2, 6, ] <- hm[ 6, 2, ] <- pdata33$ffeedcapTLodf#[ 1 ]
hm[ 2, 7, ] <- hm[ 7, 2, ] <- pdata33$ffeedmatTLodf#[ 1 ]
hm[ 3, 4, ] <- hm[ 4, 3, ] <- pdata33$fvettotTLodf#[ 1 ]
hm[ 3, 5, ] <- hm[ 5, 3, ] <- pdata33$fvetlandTLodf#[ 1 ]
hm[ 3, 6, ] <- hm[ 6, 3, ] <- pdata33$fvetcapTLodf#[ 1 ]
hm[ 3, 7, ] <- hm[ 7, 3, ] <- pdata33$fvetmatTLodf#[ 1 ]
hm[ 4, 5, ] <- hm[ 5, 4, ] <- pdata33$ftotlandTLodf#[ 1 ]
hm[ 4, 6, ] <- hm[ 6, 4, ] <- pdata33$ftotcapTLodf#[ 1 ]
hm[ 4, 7, ] <- hm[ 7, 4, ] <- pdata33$ftotmatTLodf#[ 1 ]
hm[ 5, 6, ] <- hm[ 6, 5, ] <- pdata33$flandcapTLodf#[ 1 ]
hm[ 5, 7, ] <- hm[ 7, 5, ] <- pdata33$flandmatTLodf#[ 1 ]
hm[ 6, 7, ] <- hm[ 7, 6, ] <- pdata33$fcapmatTLodf#[ 1 ]
print(hm)
print( hm[ , , 1 ] )
```

```
hm <- array( 0, c( 7, 7, nrow( pdata33 ) ) )
pdata33$quasiConcTLodf <- apply( hm[ 1:2, 1:2, ], 3, det ) < 0 &
  apply( hm[ 1:3, 1:3, ], 3, det ) > 0 &
  apply( hm, 3, det ) < 0

sum( pdata33$quasiConcTLodf ) # the quasi-convex assumption is violated in all observations
table( pdata33[ , c( "monoTLodf", "quasiConcTLodf" ) ] )

# second-order partial derivatives wrt inputs (with output distance measure = 1 )
pdata33$hmilkmilk <- with(pdata33, (amilkmilko + emilkTLodf^2 - emilkTLodf) /
grossmilkMS^2)
pdata33$hmilkother <- with(pdata33, (amilkothero + emilkTLodf * eotherTLodf) /
(grossmilkMS*grossotherMS))
pdata33$hotherother <- with(pdata33, (aothertohero + eotherTLodf^2 - eotherTLodf) /
grossotherMS^2)
summary( pdata33[ , c( "hmilkmilk", "hmilkother", "hotherother" ) ] )

# Hessian matrices
hessianArray <- array( NA, c( 2, 2, nrow( pdata33 ) ) )
hessianArray[ 1, 1, ] <- pdata33$hmilkmilk
hessianArray[ 2, 2, ] <- pdata33$hotherother
hessianArray[ 1, 2, ] <- hessianArray[ 2, 1, ] <- pdata33$hmilkother
print( hessianArray[ , , 1 ] )

# check convexity in outputs at the first observation
diag( hessianArray[ , , 1 ] )
det( hessianArray[ , , 1 ] ) # first principal minor is negative

# check convexity in outputs at all observations
pdata33$concaveCDodf <- pdata33$hmilkmilk <= 0
sum( pdata33$concaveCDodf )

#ratios between outputs
plot( pdata33$grossmilkMS, pdata33$grossotherMS )
plot( pdata33$grossmilkMS, pdata33$grossmilkMS, log="xy" )

#### Calculating marginal effects of production variables ###
pdata33$effo <- efficiencies(TLodf,asInData=TRUE)
summary(pdata33$effo)
hist(pdata33$effo ,50)
effovertimeo<-tapply(pdata33$effo,pdata33$Regnskabsaar,mean)
plot(effovertimeo,type="l",lty="dotted",col="blue",ylim=c(0.93,0.97),
    ylab="TE%",xlab="Year 2011-2015",main="TE over time")
```

```
pdata33$totaloutp<-pdata33$grossmilk + pdata33$grossother
plot(pdata33$yearcows, pdata33$effo, log="x")
plot(pdata33$totaloutp, pdata33$effo, log="x")

### Effects of production characteristics ###
summary(TLodf)
(exp(coef( TLodf )["Regnskabsaar2012"])-1)*100
(exp(coef( TLodf )["Regnskabsaar2013"])-1)*100
(exp(coef( TLodf )["Regnskabsaar2014"])-1)*100
(exp(coef( TLodf )["Regnskabsaar2015"])-1)*100
(exp(coef( TLodf )["milkingsystem2"])-1)*100
(exp(coef( TLodf )["milkingsystem3"])-1)*100
(exp(coef( TLodf )["jersey"])-1)*100
(exp(coef( TLodf )["organic"])-1)*100

#### Calculating marginal effects of z-variables ###
pdata33$effzvaro <- efficiencies(TLodf, asInData=TRUE, margEff = TRUE)
summary(pdata33$effzvaro)
effzovertimeo<-tapply(pdata33$effzvaro,pdata33$Regnskabsaar,mean)
plot(effzovertimeo,type="l",lty="dotted",col="blue",ylim=c(0.94,0.97),
    ylab="TE%",xlab="Year 2011-2015",main="TE over time for z variables")

METLodfz<-attr(pdata33$effzvaro, "margEff")
summary(METLodfz)

hist(METLodfz,50)
plot(METLodfz,pdata33$grossmilkMS)
s<-summary(METLodfz)

# Efficiency over time

# 2011
summary(pdata33$Regnskabsaar==2011)
summary(pdata33$effo>=0.95 & pdata33$Regnskabsaar==2011)
summary(pdata33$effo>=0.90 & pdata33$effo<0.95 & pdata33$Regnskabsaar==2011)
summary(pdata33$effo>=0.85 & pdata33$effo<0.90 & pdata33$Regnskabsaar==2011)
summary(pdata33$effo>=0.80 & pdata33$effo<0.85 & pdata33$Regnskabsaar==2011)
summary(pdata33$effo<0.80 & pdata33$Regnskabsaar==2011)

# 2012
summary(pdata33$Regnskabsaar==2012)
summary(pdata33$effo>=0.95 & pdata33$Regnskabsaar==2012)
summary(pdata33$effo>=0.90 & pdata33$effo<0.95 & pdata33$Regnskabsaar==2012)
```

```
summary(pdata33$effo>=0.85 & pdata33$effo<0.90 & pdata33$Regnskabsaar==2012)
summary(pdata33$effo>=0.80 & pdata33$effo<0.85 & pdata33$Regnskabsaar==2012)
summary(pdata33$effo<0.80 & pdata33$Regnskabsaar==2012)

# 2013
summary(pdata33$Regnskabsaar==2013)
summary(pdata33$effo>=0.95 & pdata33$Regnskabsaar==2013)
summary(pdata33$effo>=0.90 & pdata33$effo<0.95 & pdata33$Regnskabsaar==2013)
summary(pdata33$effo>=0.85 & pdata33$effo<0.90 & pdata33$Regnskabsaar==2013)
summary(pdata33$effo>=0.80 & pdata33$effo<0.85 & pdata33$Regnskabsaar==2013)
summary(pdata33$effo<0.80 & pdata33$Regnskabsaar==2013)

# 2014
summary(pdata33$Regnskabsaar==2014)
summary(pdata33$effo>=0.95 & pdata33$Regnskabsaar==2014)
summary(pdata33$effo>=0.90 & pdata33$effo<0.95 & pdata33$Regnskabsaar==2014)
summary(pdata33$effo>=0.85 & pdata33$effo<0.90 & pdata33$Regnskabsaar==2014)
summary(pdata33$effo>=0.80 & pdata33$effo<0.85 & pdata33$Regnskabsaar==2014)
summary(pdata33$effo<0.80 & pdata33$Regnskabsaar==2014)

# 2015
summary(pdata33$Regnskabsaar==2015)
summary(pdata33$effo>=0.95 & pdata33$Regnskabsaar==2015)
summary(pdata33$effo>=0.90 & pdata33$effo<0.95 & pdata33$Regnskabsaar==2015)
summary(pdata33$effo>=0.85 & pdata33$effo<0.90 & pdata33$Regnskabsaar==2015)
summary(pdata33$effo>=0.80 & pdata33$effo<0.85 & pdata33$Regnskabsaar==2015)
summary(pdata33$effo<0.80 & pdata33$Regnskabsaar==2015)

### Plotting development in efficiency for idf and odf ###
pdata33$effi2011<-ifelse(pdata33$Regnskabsaar==2011,pdata33$eff,0)
pdata33$effo2011<-ifelse(pdata33$Regnskabsaar==2011,pdata33$effo,0)
plot(pdata33$effi2011,pdata33$effo2011,col="green",ylim=c(0.60,1),xlim=c(0.60,1),
    ylab="TE% odf 2011",xlab="TE% idf 2011",main="TE% in 2011")
abline(lm(pdata33$effi2011~pdata33$effo2011))

pdata33$effi2012<-ifelse(pdata33$Regnskabsaar==2012,pdata33$eff,0)
pdata33$effo2012<-ifelse(pdata33$Regnskabsaar==2012,pdata33$effo,0)
plot(pdata33$effi2012,pdata33$effo2012,col="green",ylim=c(0.60,1),xlim=c(0.60,1),
    ylab="TE% odf 2012",xlab="TE% idf 2012",main="TE% in 2012")
abline(lm(pdata33$effi2012~pdata33$effo2012))

pdata33$effi2013<-ifelse(pdata33$Regnskabsaar==2013,pdata33$eff,0)
pdata33$effo2013<-ifelse(pdata33$Regnskabsaar==2013,pdata33$effo,0)
plot(pdata33$effi2013,pdata33$effo2013,col="green",ylim=c(0.60,1),xlim=c(0.60,1),
```

```
    ylab="TE% odf 2013",xlab="TE% idf 2013",main="TE% in 2013")
abline(lm(pdata33$effi2013~pdata33$effo2013))

pdata33$effi2014<-ifelse(pdata33$Regnskabsaar==2014,pdata33$eff,0)
pdata33$effo2014<-ifelse(pdata33$Regnskabsaar==2014,pdata33$effo,0)
plot(pdata33$effi2014,pdata33$effo2014,col="green",ylim=c(0.60,1),xlim=c(0.60,1),
    ylab="TE% odf 2014",xlab="TE% idf 2014",main="TE% in 2014")
abline(lm(pdata33$effi2014~pdata33$effo2014))

pdata33$effi2015<-ifelse(pdata33$Regnskabsaar==2015,pdata33$eff,0)
pdata33$effo2015<-ifelse(pdata33$Regnskabsaar==2015,pdata33$effo,0)
plot(pdata33$effi2015,pdata33$effo2015,col="green",ylim=c(0.60,1),xlim=c(0.60,1),
    ylab="TE% odf 2015",xlab="TE% idf 2015",main="TE% in 2015")
abline(lm(pdata33$effi2015~pdata33$effo2015))

####### PRICE INDEX ########
############################

### The different outputs within "grossother" and their shares ###
# Bru_mark:
# Korn (X100)
pdata3333$corn<-with(pdata3333,X100/grossother)
mean(pdata33$corn)
# Frøafgrøder (X105)
pdata33$seedcrops<-with(pdata33,X105/grossother)
mean(pdata33$seedcrops)
# Handelsroer (X110)
pdata33$beets<-with(pdata33,X110/grossother)
mean(pdata33$beets)
# Kartofler (X115)
pdata33$potatoes<-with(pdata33,X115/grossother)
mean(pdata33$potatoes)
# Raps (X120)
pdata33$rape<-with(pdata33,X120/grossother)
mean(pdata33$rape)
# Ærter mv.(X125)
pdata33$peas<-with(pdata33,X125/grossother)
mean(pdata33$peas)
# Andre industriafgrøder mv.(X130)
pdata33$othercrops<-with(pdata33,X130/grossother)
mean(pdata33$othercrops)
# Gartneriafgrøder (X135)
pdata33$horticulturalcrops<-with(pdata33,X135/grossother)
mean(pdata33$horticulturalcrops)
```

```
# Energiafgrøder mv. (X145)
pdata33$energycrops<-with(pdata33,X145/grossother)
mean(pdata33$energycrops)

# Bru_dyr_i_alt:
# Kvæg (X175)
pdata33$cattle<-with(pdata33,X175/grossother)
mean(pdata33$cattle)
# Svin (X180)
pdata33$pigs<-with(pdata33,X180/grossother)
mean(pdata33$pigs)
# Fjerkræ (X185)
pdata33$poultry<-with(pdata33,X185/grossother)
mean(pdata33$poultry)
# Pelsdyr (X190)
pdata33$furanimals<-with(pdata33,X190/grossother)
mean(pdata33$furanimals)
# Får (X195)
pdata33$sheep<-with(pdata33,X195/grossother)
mean(pdata33$sheep)
# Husdyr i øvrigt (X200)
pdata33$otherlivestock<-with(pdata33,X200/grossother)
mean(pdata33$otherlivestock)

# Andre_landbrugsindt
# Maskinstationsindtægter (X205)
pdata33$machine<-with(pdata33,X205/grossother)
mean(pdata33$machine)
# Andre landbrugsindtægter - del 1 + del 2 (X210 + X212)
pdata33$otheragriincome<-with(pdata33,(X210+X212)/grossother)
mean(pdata33$otheragriincome)

### The different elements within feed bought ###
pdata33$expfeedbought<-with(pdata33,(X250+X255+X260)*-1)
# Expenditure on corn
mean(pdata33$X250*-1)/mean(pdata33$expfeedbought)
# Expenditure on readymix
mean(pdata33$X255*-1)/mean(pdata33$expfeedbought)
# Expenditure on other feed bought
mean(pdata33$X260*-1)/mean(pdata33$expfeedbought)


# The different elements within materials and their shares #
pdata33$materials <- with(pdata33, (X230+X235+X240+X245+
```

```
                        X275+X280+X310+X315+X320 +X325+X330+X375+X380)*-1)
# Udsæd
mean(pdata33$X230*-1)/mean(pdata33$materials)
# Gødning
mean(pdata33$X235*-1)/mean(pdata33$materials)
# Planteværn.
mean(pdata33$X240*-1)/mean(pdata33$materials)
# Diverse vedrørende markbrug
mean(pdata33$X245*-1)/mean(pdata33$materials)
# Diverse vedrørende husdyrbrug
mean(pdata33$X275*-1)/mean(pdata33$materials)
# Skovomkostninger
mean(pdata33$X280*-1)/mean(pdata33$materials)
# Høst af korn og frøafgrøder
mean(pdata33$X310*-1)/mean(pdata33$materials)
# Optagning
mean(pdata33$X315*-1)/mean(pdata33$materials)
# Høst af grovfoder
mean(pdata33$X320*-1)/mean(pdata33$materials)
# Udbringning af husdyrgødning
mean(pdata33$X325*-1)/mean(pdata33$materials)
# Maskinstation diverse
mean(pdata33$X330*-1)/mean(pdata33$materials)
# Forsikringer
mean(pdata33$X375*-1)/mean(pdata33$materials)
# Diverse omkostninger
mean(pdata33$X380*-1)/mean(pdata33$materials)

### Input/output shares ###
pdata2011<-subset(pdata33, Regnskabsaar==2011)
pdata2012<-subset(pdata33, Regnskabsaar==2012)
pdata2013<-subset(pdata33, Regnskabsaar==2013)
pdata2014<-subset(pdata33, Regnskabsaar==2014)
pdata2015<-subset(pdata33, Regnskabsaar==2015)

# 2011 #
pdata2011$allinput<-with(pdata2011,
feedexp_2010level+vetmed_2010level+totalwages_2010level
          +land+capital+materials_2010level)
mean(pdata2011$feedexp_2010level/pdata2011$allinput)
mean(pdata2011$vetmed_2010level/pdata2011$allinput)
mean(pdata2011$totalwages_2010level/pdata2011$allinput)
mean(pdata2011$land/pdata2011$allinput)
mean(pdata2011$capital/pdata2011$allinput)
```

```
mean(pdata2011$materials_2010level/pdata2011$allinput)

mean(pdata2011$grossmilk_2010level/(pdata2011$grossmilk_2010level+pdata2011$grossother
_2010level))
mean(pdata2011$grossother_2010level/(pdata2011$grossmilk_2010level+pdata2011$grossother
_2010level))

# 2012 #
pdata2012$allinput<-with(pdata2012,
feedexp_2010level+vetmed_2010level+totalwages_2010level
                    +land+capital+materials_2010level)
mean(pdata2012$feedexp_2010level/pdata2012$allinput)
mean(pdata2012$vetmed_2010level/pdata2012$allinput)
mean(pdata2012$totalwages_2010level/pdata2012$allinput)
mean(pdata2012$land/pdata2012$allinput)
mean(pdata2012$capital/pdata2012$allinput)
mean(pdata2012$materials_2010level/pdata2012$allinput)

mean(pdata2012$grossmilk_2010level/(pdata2012$grossmilk_2010level+pdata2012$grossother
_2010level))
mean(pdata2012$grossother_2010level/(pdata2012$grossmilk_2010level+pdata2012$grossother
_2010level))

# 2013 #
pdata2013$allinput<-with(pdata2013,
feedexp_2010level+vetmed_2010level+totalwages_2010level
                    +land+capital+materials_2010level)
mean(pdata2013$feedexp_2010level/pdata2013$allinput)
mean(pdata2013$vetmed_2010level/pdata2013$allinput)
mean(pdata2013$totalwages_2010level/pdata2013$allinput)
mean(pdata2013$land/pdata2013$allinput)
mean(pdata2013$capital/pdata2013$allinput)
mean(pdata2013$materials_2010level/pdata2013$allinput)

mean(pdata2013$grossmilk_2010level/(pdata2013$grossmilk_2010level+pdata2013$grossother
_2010level))
mean(pdata2013$grossother_2010level/(pdata2013$grossmilk_2010level+pdata2013$grossother
_2010level))

# 2014 #
pdata2014$allinput<-with(pdata2014,
feedexp_2010level+vetmed_2010level+totalwages_2010level
                    +land+capital+materials_2010level)
mean(pdata2014$feedexp_2010level/pdata2014$allinput)
```

```
mean(pdata2014$vetmed_2010level/pdata2014$allinput)
mean(pdata2014$totalwages_2010level/pdata2014$allinput)
mean(pdata2014$land/pdata2014$allinput)
mean(pdata2014$capital/pdata2014$allinput)
mean(pdata2014$materials_2010level/pdata2014$allinput)

mean(pdata2014$grossmilk_2010level/(pdata2014$grossmilk_2010level+pdata2014$grossother
_2010level))
mean(pdata2014$grossother_2010level/(pdata2014$grossmilk_2010level+pdata2014$grossother
_2010level))

# 2015 #
pdata2015$allinput<-with(pdata2015,
feedexp_2010level+vetmed_2010level+totalwages_2010level
                  +land+capital+materials_2010level)
mean(pdata2015$feedexp_2010level/pdata2015$allinput)
mean(pdata2015$vetmed_2010level/pdata2015$allinput)
mean(pdata2015$totalwages_2010level/pdata2015$allinput)
mean(pdata2015$land/pdata2015$allinput)
mean(pdata2015$capital/pdata2015$allinput)
mean(pdata2015$materials_2010level/pdata2015$allinput)

mean(pdata2015$grossmilk_2010level/(pdata2015$grossmilk_2010level+pdata2015$grossother
_2010level))
mean(pdata2015$grossother_2010level/(pdata2015$grossmilk_2010level+pdata2015$grossother
_2010level))

#### Mean of actual output/input values in 2010-values ####
mean(pdata2011$grossmilk_2010level)
mean(pdata2012$grossmilk_2010level)
mean(pdata2013$grossmilk_2010level)
mean(pdata2014$grossmilk_2010level)
mean(pdata2015$grossmilk_2010level)
mean(pdata2011$grossother_2010level)
mean(pdata2012$grossother_2010level)
mean(pdata2013$grossother_2010level)
mean(pdata2014$grossother_2010level)
mean(pdata2015$grossother_2010level)
mean(pdata2011$feedexp_2010level)
mean(pdata2012$feedexp_2010level)
mean(pdata2013$feedexp_2010level)
mean(pdata2014$feedexp_2010level)
mean(pdata2015$feedexp_2010level)
mean(pdata2011$vetmed_2010level)
```

```
mean(pdata2012$vetmed_2010level)
mean(pdata2013$vetmed_2010level)
mean(pdata2014$vetmed_2010level)
mean(pdata2015$vetmed_2010level)
mean(pdata2011$totalwages_2010level)
mean(pdata2012$totalwages_2010level)
mean(pdata2013$totalwages_2010level)
mean(pdata2014$totalwages_2010level)
mean(pdata2015$totalwages_2010level)
mean(pdata2011$land)
mean(pdata2012$land)
mean(pdata2013$land)
mean(pdata2014$land)
mean(pdata2015$land)
mean(pdata2011$capital)
mean(pdata2012$capital)
mean(pdata2013$capital)
mean(pdata2014$capital)
mean(pdata2015$capital)
mean(pdata2011$materials_2010level)
mean(pdata2012$materials_2010level)
mean(pdata2013$materials_2010level)
mean(pdata2014$materials_2010level)
mean(pdata2015$materials_2010level)
```