UNIVERSITY OF COPENHAGEN
DEPARTMENT OF FOOD AND RESOURCE ECONOMICS

Master Thesis

Julie Skovgaard Hansen and

Amanda Prehn Henriksen

Estimating Abatement Costs of Greenhouse Gas Emissions in the Danish Agricultural Sector using Non-parametric Efficiency Analysis

Supervisor: Mette Asmild

Submitted on 30 September 2020

ESTIMATING ABATEMENT COSTS OF GREENHOUSE GAS EMISSIONS IN THE DANISH AGRICULTURAL SECTOR USING NON-PARAMETRIC EFFICIENCY ANALYSIS

Julie Skovgaard Hansen &
Amanda Prehn Henriksen

Master thesis in Agricultural Economics
September 2020

Name of University	University of Copenhagen
Name of Department	Department of Food and Resource Economics
Authors	Julie Skovgaard Hansen and Amanda Prehn Henriksen
Title	Estimating abatement costs of Greenhouse gas emissions in the Danish agricultural sector using non-parametric efficiency analysis
ECTS	30
Supervisor	Mette Asmild
Submitted	September 30, 2020

Abstract

This thesis investigates the costs related to reducing greenhouse gas emissions within the Danish dairy sector and the diversity in the abatement costs across the sector. The results can contribute to finding a cost-efficient way towards the national 70 pct. reduction target implying minimum costs for both the agricultural sector and the society.

The first part of this thesis concerns estimating greenhouse gas (GHG) emissions, as suitable farm- specific emissions are not currently available.
The second part contains an efficiency analysis, carried out by using a modified Data Envelopment Analysis model including the estimated GHG emissions as an undesirable output. The model assumes weak disposability and applies a directional distance function. *Frontier shadow prices* are extracted from the model, measuring the marginal abatement costs for GHG emissions given the best technology available in the sector today. Furthermore, this thesis proposes a new method for measuring abatement costs for inefficient farms, where the technological lags, i.e. efficiency potentials, found in the model are taken into account. These abatement costs represent the *average opportunity costs* and are calculated by examining the trade-off in the potentials found for either focusing on maximizing revenue or reducing GHG emissions.

The model defines best practice for the sector and through this the current technological lag for inefficient farms. Thereby, the model determines the improvement potential for farms not operating at best practice. The models find that if inefficient farms only focus on improving economic performance, there is a potential within the included sample to increase the aggregated revenue with approximately 3 bill. DKK (corresponding to an increase of 28 pct.). Contrary if they only focus on reducing GHG emissions, there is a potential reduction of approximately 639.000 ton CO_2e (corresponding to a decrease of 35 pct.).

The thesis finds that it is relatively costly to reduce GHG emissions for Danish dairy farms compared to the general cost of abating GHG emissions in Denmark, which is estimated by the Danish Council of Climate Change to be 1.500 DKK. The results show that the average abatement cost of a ton CO_2e is around 4.500 DKK for conventional farms and 6.400 DKK for organic farms

Table of content

List of abbreviations

APE	Average Partial Effect
CH_4	Methane
CO_2	Carbon Dioxide
CO_2e	Carbon Dioxide equivalent
CRS	Constant Returns to Scale
DCE	Danish Center for Energy and Environment
DDF	Directional Distance Function
DE	Digestibility of feed expressed as a fraction of gross energy
DEA	Data Envelopment Analysis
DM	Dry matter
DMU	Decision Making Unit
EF	Emission Factor
ENT	Enteric Fermentation
EU	European Union
GE	Gross Energy intake
GHG	Greenhouse gas
GWP	Global Warming Potential
IPCC	International Panel on Climate Change
LCA	Life Cycle Assessment
Man	Manure Management
MCF	Methane conversion factor
N	Nitrogen
N_2O	Nitrous Oxide
NE_a	Net energy required for activity in MJ per day
NE_g	Net energy required for growth
NE_l	Net energy required for lactating
NE_m	Net energy required for maintenance
NE_p	Net energy required for pregnancy
Nex	Annual excretion of N

OLS	Ordinary Least squares
PEA	Partial Effect of the Average
PPF	Production Possibility Frontier
PPS	Production Possibility Set
REG	Ration of net energy available for growth in a diet to digestible energy consumed
REM	Ratio of net energy available in a diet for maintenance to digestible energy
VRS	Variable Returns to scale
VS	Volatile solids

List of figures

List of tables

Introduction

Greenhouse gases in the atmosphere are essential for the earth's regulation of the temperature. In the last century, the temperature has increased due to an increase in greenhouse gases in the atmosphere leading to global warming. A study from 2013 finds that 97 pct. of researchers endorse that the global warming is caused by humans as a result of higher greenhouse gas (GHG) emissions partly due to more intense agriculture and burning of fossil fuels (Cook, et al., 2013). The temperature will continuously increase, if the global emissions of greenhouse gases are not reduced.

To mitigate the increasing temperatures, the Paris Agreement was adopted at the 2015 United Nations Climate Change Conference (COP21). With the Paris Agreement, 196 countries settled on a common goal within the global challenges of climate change. One of the main and most discussed objectives of the Paris Agreement, is the worldwide goal of keeping the temperature increase below 2°C, also known as the global temperature target, and to further initiate actions to limit the rise in temperature to only 1,5°C.

National states are to implement actions that will enable the objectives set out in the Paris Agreement. Furthermore, for the non-quota sector, which in Denmark mainly includes agriculture and transport, there is a common target set out by the European Union (EU) of reducing the greenhouse gas emissions with 39 pct. in 2030 compared to the level in 2005 (The Ministry of Environment and Food, 2020).
In Denmark, the objectives are partially implemented through a new climate law, which was agreed upon by the government and a majority of the Danish parliament in December 2019 and adopted in 2020. The Danish climate law concerns a 70 pct. reduction of the greenhouse gas emissions by the year 2030 compared to the 1990 level (Lov om klima no. 965/2020). The 70 pct. reduction in GHG emissions is an intermediate target of the overall objective of Denmark being climate neutral by the year 2050. The goal is to be reached as cost efficient as possible, with a specific focus on factors such as the green transition in the long run, sustainable business development and the Danish competitiveness internationally. Furthermore, it is explicitly stated that actions must lead to real inland reduction of the GHG emissions.

In 2017, the agricultural sector accounted for 22 pct. of the emissions of greenhouse gasses in Denmark, which makes the agricultural sector one of the main sources of GHG emissions in Denmark. Therefore, the sector will eventually face regulation dealing with the climate impact from the sector in the following years due to the national and international reduction targets.
To reduce the Danish GHG emissions with 70 pct. by the year 2030, it is important to gain knowledge about the interaction between GHG emissions and economic performance in the agricultural sector. This knowledge can contribute as to how the GHG emissions from agriculture can be reduced at minimum cost, without lowering the productivity of the Danish agricultural sector significantly, so that the sector maintains its competitiveness on the international market.

Currently, GHG emissions from the agricultural sector are only reported on an aggregated level. Therefore, it is only possible to impose indirect regulations - such as nitrate regulation or subsidies to environmental or climate actions - on the entire sector as a whole. By regulating the entire sector indirectly as a whole, the regulation cannot be designed in the most cost-efficient way (The Danish Council on Climate Change, 2016). Therefore, in order to pave the way for regulating the agricultural sector most cost-efficiently, the GHG emissions should optimally be estimated on farm level. Farm specific GHG emissions can be useful in order to detect the diversity across farms regarding the specific emissions and the costs of reducing them.

The aim of this thesis is to develop an applied framework where GHG emissions are implemented in an economic benchmarking of the agricultural sector in order to calculate farm-specific abatement costs of reducing GHG emissions. In addition to estimating the farm-specific abatement costs, the empirical analysis defines the current improvement potential of reducing GHG emissions and/or increasing economic performance. This framework could provide a contribution to a possible future cost-efficient regulation of the agricultural sector, as it reveals the diversity in the cost of abating GHG emissions and thereby where abating GHG emissions is carried out at lowest cost.
The empirical analysis of this thesis is carried out for Danish dairy farms in the year 2017. Nevertheless, the benchmarking framework could potentially be expanded to cover various parts of the agricultural sector.

As data on GHG emissions on farm level is currently not existing, Part A of this thesis focuses on estimating these for Danish dairy farms to be able to include the estimated farm-specific GHG emissions in the benchmarking model presented in Part B. Part A is therefore essential for being able to conduct the benchmarking analysis in part B.

The empirical analysis in Part B concerns estimating a non-parametric benchmarking model with the assumption of weak disposability and including a directional distance function. This is done in order to be able to integrate an undesirable output (GHG emissions) in an economic efficiency analysis of the Danish dairy farms. The method used is a variation of the Data Envelopment Analysis (DEA) method first developed by Charnes et al. (1978).
DEA is a non-parametric benchmarking method that firstly defines a performance standard and secondly evaluates the performance of Decision Making Units (DMUs), such as farms, relative to best practice (Bogetoft & Otto, 2011, p. 81). DEA is based on linear programming, creating an efficient frontier used as comparison for other DMUs to estimate their relative efficiency.
It is becoming increasingly more common to also include undesirable outputs such as externalities from production in such benchmarking models. With an undesirable output, the estimated frontier can be used to derive frontier shadow prices of the undesirable output using the trade-off that occurs along the frontier.

This thesis is structured as follows:
Part A concerns estimating farm-specific GHG emissions for Danish dairy farms. The section covers the *background* for GHG emissions from the agricultural sector as well as the *methodology* for estimating GHG emissions from enteric fermentation and manure management using national and international guidelines. The *results* display the estimated GHG emissions which will be used in the benchmarking model in Part B. The estimated GHG emissions are *validated* by comparing them to national estimates.
Part B concerns the benchmarking analysis of the dairy farms. This section consists of a b*ackground* describing the concept of abatement costs as well as the basic method of Data Envelopment Analysis. The *methodology* section elaborates on the theory behind including an undesirable output in a benchmarking model. Furthermore, *Part B* contains a section presenting the *empirical model*, used in this thesis. The *results* from part B displays the abatement costs for both efficient and inefficient farms, as well as the existing potentials within the sector for improving both revenue and/or climate. Lastly, the *results* display which characteristics of the farms that are associated with low abatement costs of GHG emissions and higher efficiency.

In the *Discussion and perspective for future research*, the implication of the estimation of farm-specific GHG emissions for the benchmarking model is discussed. Furthermore, it is discussed how the results can be used in future regulation, as well as how to improve the results through future research.

Part A: Estimating farm-specific greenhouse gas emissions

A.1 Background

A.1.1 Greenhouse gases in the agricultural sector

The total Danish GHG emissions have decreased from approximately 74 mill. tons CO_2e in 1990 to approximately 51 mill. tons CO_2e in 2017. This is including LULUCF and corresponds to a decrease of approximately 32 pct. The agricultural sector was the third largest source of GHG emissions in 2017.

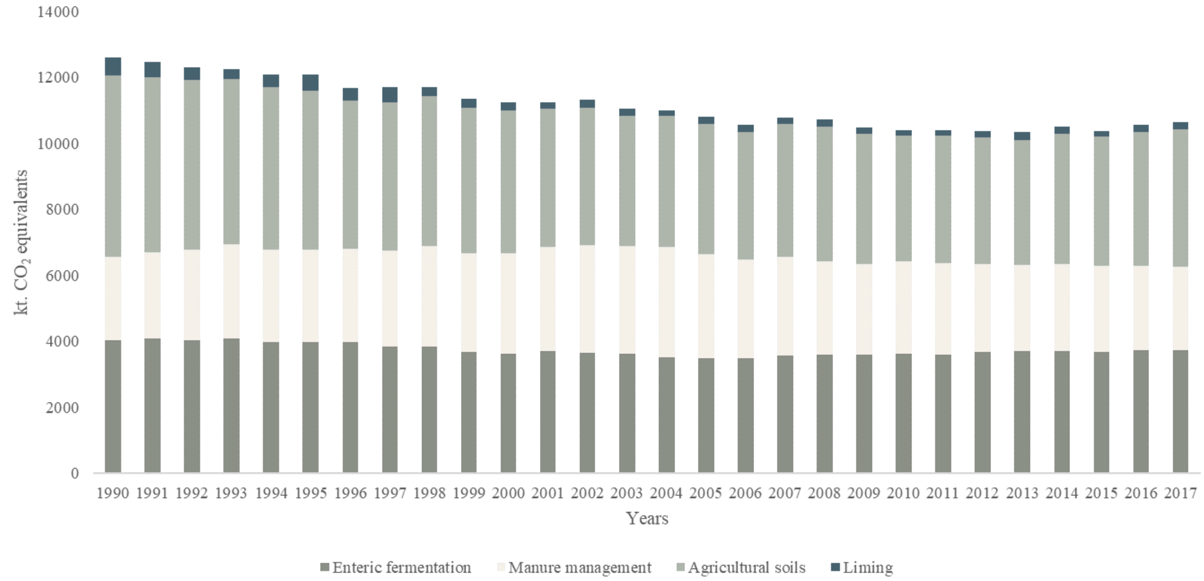
In 2017 the GHG emissions from the agricultural sector, excluding LULUCF, contributed to 22 pct. of the total emissions in Denmark. These are primarily due to livestock, where cattle are a particularly large source of the total emissions in the agricultural sector (DCE, 2019). The aggregated GHG emissions from the agricultural sector was approximately 11 mill. tons CO_2e in 2017 (excluding LULUCF). The emissions from the agricultural sector has decreased 16 pct. since 1990, which is mainly due to a decrease in emissions of nitrous oxide (N_2O). The reduction in N_2O emissions are highly due to a derived effect of an increased focus on the aquatic environment (DCE, 2019), such as the implementation of the EU's Water framework Directive (2000/60/EC) (WFD).

However, there are numerous sources of GHG emissions in the agricultural sector. Following DCE (2019), these cover the categories:

- CH_4 emissions from enteric fermentation
- CH_4 and N_2O emissions from manure management
- Direct and indirect N_2O emissions from agricultural soils
- CO_2 emissions from liming, urea and other carbon-containing fertilizers

The distribution of the different greenhouse gases over time is displayed in Figure A.1-1.

Figure A.1-1: Time-series of GHG emissions within the Danish agricultural sector

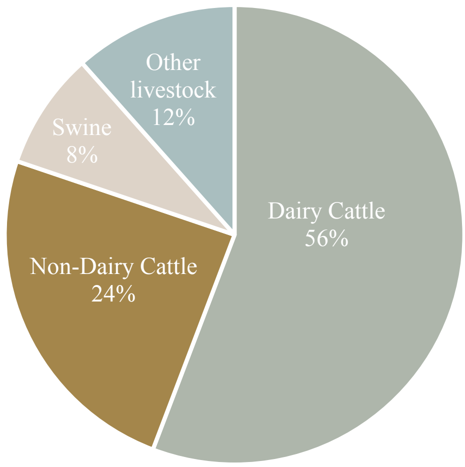


Source: Authors construction, based on numbers from DCE (2019).
Note: Greenhouse gas emissions from the agricultural sector also cover "Field burning of agricultural residues", "Urea application" and "Other carbon-containing fertilizers". However, as these two categories only represent 0,03, 0,04 and 0,09 pct. respectively of the total emissions from 1990 to 2017, these are not included in the illustration.

The figure shows the different GHG emissions in CO_2 equivalents (CO_2e), as the effect on global warming, from the different greenhouse gases listed above, are not directly comparable. The effect from the greenhouse gases can be compared by using their global warming potentials (GWP). The individual GWP of a greenhouse gas is dependent on the specific lifespan of the gas in the atmosphere. The GWP of a given GHG represents the effect on climate for this gas relative to the effect on climate from CO_2 over a given time period, typically 100 years. To be able to compare these with each other, the different greenhouse gasses are converted into CO_2 equivalents and referred to as greenhouse gas emissions. The GWP for methane is 25 and 298 for nitrous oxide. This means that 1 ton of methane corresponds to 25 CO_2e and 1 ton of Nitrous Oxide corresponds to 298 tons CO_2e.

From Figure A.1-1, it can be seen that GHG emissions from enteric fermentation represent a large part of the aggregated emissions. Enteric fermentation is a digestive process, which is a process where carbohydrates are broken down to simple molecules that can be obtained in the blood of the livestock. Methane (CH_4) is a byproduct of this process and the majority is primarily released as burps and exhalation. The methane emissions from this process is dependent on factors such as the livestock's individual digestive tract as well as the age, weight, activity level and the feed composition (IPCC, 2019). Ruminants, such as cattle, produce a higher level of methane from enteric fermentation compared to non-ruminant livestock such as pigs. Furthermore, the emission of methane from cattle is dependent on the milk productivity where a higher milk yield leads to increased methane emissions (IPCC, 2019). In 2017, dairy cattle contributed with 56 pct. of the overall GHG emissions stemming from enteric fermentation whereas non-dairy cattle contributed with 24 pct. cf. Figure A.1-2.

Figure A.1-2: Sources of GHG emissions from enteric fermentation distributed for livestock categories



Source: Authors construction, based on numbers from DCE (2019).

Emissions from enteric fermentation have decreased slightly over the period from 1990-2017 as can be seen in Figure A.1-1. The reduction is mainly due to a decrease in the total number of dairy cattle (DCE, 2019). The methodological framework for estimating the methane emission from enteric fermentation is further explained in section A.3.1.

From Figure A.1-1 it can be seen that GHG emissions from manure management also constitute a large part of the sectors aggregated emissions. The aggregated GHG emissions from manure management have been relatively stable over the period from 1990-2017 (cf. Figure A.1-1).

The GHG emissions from manure management consist of two byproducts; methane (CH_4) and Nitrous oxide (N_2O).
CH_4 emissions from manure management is the methane released due to anaerobic processes when manure is managed and stored. The methane emissions depend on the amount of manure deposited, the number of animals and the handling of the manure, which is directly related to the housing system (IPCC, 2019). Swine is the livestock category which contributes mostly to the methane emissions related to manure management. Swine cover approximately 60 pct. of the total emissions. Cattle are the second largest contributors where dairy cattle contribute with 19 pct. and non-dairy with 17 pct. of the total emissions related to manure management.

N_2O emissions from manure management are related to the manure handling in housing and storage. Emissions related to manure deposited on agricultural soils are not measured as part of the manure management. These emissions are included in the category for agricultural soils and is thereby not included in the estimated GHG emissions from livestock categories such as cattle (DCE, 2019). As can be seen in Figure A.1-1, emission from agricultural represents a relative high share of the total GHG emissions from the agricultural sector as a whole. However, these emissions are not included in the empirical analysis, as GHG emissions from agricultural soils only represent a minor share of the GHG emissions from dairy farms.

N_2O from handling manure can be divided into indirect and direct N_2O emissions. The direct emissions are emitted directly from handling manure, whereas the indirect N_2O emissions are associated with the emissions of NH_3 (ammonia) and NO_X (nitrogen oxides) related to the manure handling.

A.1.2 Farm-specific greenhouse gas emissions

Farm-specific GHG emissions for the Danish agricultural sector, are not currently available. In the climate agreement made in June 2020 (The Danish Ministry of Finance, 2020), 5 mill. DKK have been set aside to develop climate accounting at farm level. The climate accounts will contribute to a more precise climate regulation of the Danish agricultural sector (The Ministry of Environment and food, 2020). Finding a method for properly estimating GHG emissions on farm level, as well as implementing the data from the climate accounting in economic modelling is therefore more relevant than ever.

There exist different sources describing how to estimate GHG emissions on both detailed and aggregated levels. The Intergovernmental Panel on Climate Change (IPPC) has developed a guidance for how to estimate GHG emissions on a national level (IPCC, 2019)l. This guidance is the generally applied methodology for estimating GHG emissions on sectoral level. Another approach to measure GHG emissions on a more disaggregated level is the Life Cycle Assessment (LCA). Nevertheless, this approach measures GHG emissions through the entire production chain and does therefore not distinguish between national and international emissions (The Danish Council on Climate Change, 2016). The IPPC guidelines only include inland emissions and are thereby more directly applicable to use in national regulation of climate. This is the case as the national reduction target is based on numbers for inland emissions.

The overall emissions from different sectors are estimated by the Danish Center for Environment and Energy (DCE) in a National Inventory Report (DCE, 2019). The annual National Inventory Report for Denmark covers the national GHG emissions across different sectors. The report follows the general guidelines provided by IPCC. However, the guidelines are modified using national standards and methods where possible. The estimations of GHG emissions are aggregated for the relevant sectors and are thereby reported on a relatively undetailed level.

The farm-specific GHG emissions are in this thesis also estimated using the IPPC guidelines (IPCC, 2019, pp. 10.33-10.99). In order to obtain as precise estimates as possible, the IPCC guidelines have been modified by using national standards from various sources. These include:

- The National Inventory Report (DCE, 2019)
- The Danish Council on Climate Change – climate accounting on farm level (The Danish Council on Climate Change, 2016),
- Norm figures for cattle (Lund & Aaes, 2016/2017)

The Danish accounting on farm level, developed by the Danish Council on Climate Change, is a prototype tool to estimate the GHG emissions on farm level for a single farm (The Danish Council on Climate Change, 2016). The tool is developed in order to provide a single farm with a tool to monitor the given GHG emissions from the specific farm. The tool is thereby not directly applicable as an instrument to estimate farm-specific data for a range of farms with the aim of executing analyses across a large dataset. Furthermore, the calculations of GHG emissions from cattle that originates from enteric fermentation, can be calculated more precisely than what the Danish Council on Climate Changes tool allows for today. The estimation of GHG emissions from enteric fermentation can thereby be done more detailed to obtain a higher variation across farms. However, in this thesis, parts of the calculations found in the tool provided by the Danish Council on Climate Change are used to calculate farm-specific GHG emissions from manure management. Given the current data availability, the standardized values provided by the Danish Council on Climate Change concerning manure management, still serves as the most detailed way of estimating GHG emissions from manure management.

A.2 Data

The dataset for this analysis consist of data from the economic database Ø90 provided by SEGES, norm figures from Lund and Aaes (2016/2017), Poulsen (2017), IPPC (2019, pp. 10.33-10.90) and the National Inventory Report (DCE, 2019) and lastly of data from the fertilizer accounts provided by the Danish Agricultural Agency.

The data from the economic database Ø90 is provided by SEGES and is obtained from the financial management and bookkeeping tool supplied by the branch-collaboration DLBR (Lillethorup, 2017). This data consists of accounting data reported by dairy farms in Denmark. It is voluntary to use the Ø90 management tools, so not all Danish dairy farms are represented in the database.

For this thesis, only data from 2017 has been include. This is mainly due to the fact that the latest year present in the database is 2018, which was an extremely dry year in Denmark leading to an extraordinary decrease in the income for the Danish agricultural sector (The Danish Agriculture and Food Council, 2018). Thereby, 2018 is not found suitable for the analysis as this year might not be representative. Therefore, 2017 is the latest year, where the required data for this analysis is present, and is thereby used in the empirical analysis. The raw data from 2017 originally consist of 2.201 observations.

The data cleaning has been carried out following the methodology developed by Lillethorup (2017). The data cleaning help secures that the dataset is representative for the sector and only represents comparable farms by excluding potential outliers and misspecified data. Different criteria of which farms to excluded are made e.g. what makes a farm specialized in dairy production. Furthermore, the data cleaning ensures that only full-time farms are included in the sample and only farms with a minimum of 100 cattle are included. (Lillethorup, 2017)

In this analysis, the amount of milk produced is used to estimate Greenhouse gas emissions. Therefore, farms with a milk production below 5.000 ECM (energy corrected milk) per dairy cattle per year, have been removed from the dataset. This has been carried out as 5.000 ECM is less than half of the average production of ECM per dairy cattle per year. Less than an average of 5.000 ECM might be an unreasonable low production of ECM and can potentially lead to a smaller GHG emission, thus making the farm an outlier in the efficiency analysis. 12 observations are removed due to this criterion.

The data from the fertilizer accounts is provided by the Danish Agricultural Agency and gives information about which housing systems the different farms use for their cattle, which type of manure the cattle produces, and how many cattle is in each housing system on a given farm.

The dataset from the economic database and from the fertilizer account have been combined for the analysis. Observations which are not present in both datasets have been removed from the sample. Table A.2-1 displays the distribution of the Danish population of cattle in 2017 and how many of these cattle that are included in the final sample.

Table A.2-1: Distribution of cattle types in population and sample

Type	**Population**	**Sample**	**Share of population**
Dairy cattle	571.115	292.088	51 pct.
Heifers (>6 months)	426.810	131.502	31 pct.
Heifers (0-6 months)	161.788	66.237	41 pct.
Bulls and studs (>6 months)	41.941	6.749	16 pct.
Bulls (0-6 months)	124.221	15.193	12 pct.

The final sample for the empirical analysis covers 292.088 dairy cattle in 2017, corresponding to approximately 51 pct. of the total number of Danish dairy cattle. Furthermore, data covers a total of 1.254 specialized dairy farms, corresponding to approximately the same share of the population of dairy farms in Denmark. Of the 1.254 farms, 204 are categorized as organic farms whereas 1050 are categorized as conventional.

A.3 Methodology

The IPCC guidelines contain different approaches for estimating the GHG emissions, depending on the detail level of the analysis. The different levels range from the lowest Tier 1 to the highest Tier 3 approach. Tier 1 represents the approach with the lowest level of details, primarily applying standardized values for the basic characterization of the different sources of emissions.
Tier 2 contains a more detailed approach, making it possible to incorporate information on e.g. specific livestock subcategories or the diversity in feed intake. Furthermore, Tier 2 expands the Tier 1 approach by applying country-specific estimates for the gross energy intake (GE) and the methane conversion factor (Y_m) for specific livestock categories (IPCC, 2019, p. 10.35).

The highest Tier method, Tier 3, further expands the detail level beyond the diversity of the Tier 2 approach. The IPCC do not provide comprehensive instructions on the Tier 3 method, but encourage country-specific improvements, where possible. This could include default values for e.g. the feed digestibility and the chemical composition of feed, leading to a higher diversity in the methane conversion factor.

The estimation of GHG emissions in this thesis, is carried out using a mixture of the three Tiers. Given the data available for the estimation, the analysis is carried out as detailed as possible within the different sub-elements of both emissions from enteric fermentation and manure management. A higher detail-level, than what is currently available, will depend on data being reported on a highly disaggregated level from farms.

The aggregated GHG emissions from each farm are calculated by applying the specific emission factors related to enteric fermentation and manure management with the number of cattle within each subcategory for all farms in the dataset. The CH_4 emissions for enteric fermentation vary across each farm, according to characteristics such as the breed, composition of cattle, fat and protein content in the milk and the milk yield.
CH_4 emission factors related to enteric fermentation ($EF_{CH_4,Ent}$) are only estimated specifically for dairy cattle and do not vary across breed and activity level for non-dairy cattle as norm figures are used for the emission factors of these.

The CH_4 and N_2O emissions for manure management varies across breed, composition of cattle and the specific housing type for each of the farms in the dataset.

The methodology provided by IPPC concerns estimating the emission factors for a specific livestock. Nevertheless, as data is on farm level, the emission factors in this thesis, are estimated as an average emission factor for a specific farm. The farm-specific emission factor for e.g. dairy cattle will thereby represent the emission factor of the average dairy cattle on the specific farm.

IPCC's framework involves estimating emission factors relating to the individual livestock and thereafter aggregating the total emissions by multiplying the emission factors with the total number of livestock within the specific category. The method is originally developed to estimate emissions on a national level, taking a countries total number of livestock into account.
In the GHG estimation of this report, the method for estimating emissions is adjusted and applied to individual dairy farms in Denmark in order to estimate the GHG emissions from dairy cattle.

The following sections describe the methodology for calculating farm-specific emission factors for CH_4 related to enteric fermentation and manure management respectively and emission factors for N_2O related to manure management for different types of cattle.

A.3.1 Enteric fermentation

The emissions related to enteric fermentation is the greenhouse gas, methane (CH_4). The amount of methane stemming from enteric fermentation is dependent on a range of determinants, hereunder the age and weight of the dairy cattle, along with the quality and quantity of the feed (IPCC, 2019, p. 10.33).

Figure A.3-1 displays an overview of the method for calculating the emission factors of CH_4 associated with enteric fermentation $(EF_{CH_4,Ent})$. $EF_{CH_4,Ent}$ is calculated by using the gross energy intake (GE) of the individual dairy cattle. GE varies across the individual cattle according to factors such as the general activity level as well as the lactating and growth of the cattle.

Figure A.3-1: Overview of equations used in the calculating of $EF_{CH_4,Ent}$

$$EF_{CH_4,ent} = \frac{GE * \frac{Y_m}{100} * 365}{55,65}$$

$$GE = \frac{\frac{NE_m + NE_a + NE_l + NE_p}{REM} + \frac{NE_g}{REG}}{\frac{DE}{100}}$$

$$NE_m = Cf_l * (Weight)^{0,75}$$

$$NE_a = C_a * NE_m$$

$$NE_l = Milk * (1,47 + 0,40 * Fat)$$

$$NE_{work} = 0,10 * NE_m * hours$$

$$NE_g = 22,02 * \left(\frac{BW}{C * MW}\right)^{0,75} * WG^{1,097}$$

$$REM = 1,123 - (4,092 * 10^{-3} * DE) + (1,126 * 10^{-5} * (DE)^2) - \left(\frac{25,4}{DE}\right)$$

$$REG = 1,164 - (5,16 * 10^{-3} * DE) + (1,308 * 10^{-5} * (DE)^2) - \left(\frac{37,4}{DE}\right)$$

Source: Authors construction, based on theoretical framework from IPCC (2019).

The following section covers the estimation of $EF_{CH_4,Ent}$, hereunder GE as well as the net energy requirements that goes into the calculation of GE.

The Methane emission factors from enteric fermentation for dairy cattle ($EF_{CH_4,Ent}$) are calculated according to equation A.3-1. Equation A.3-1 converts the gross energy intake of each dairy cattle into an emission factor, expressing how many kilos of methane is emitted from the individual dairy cattle per year.

$$(EF_{CH_4,Ent}). = \frac{GE * \frac{Y_m}{100} * 365}{55,65} \qquad \text{A.3-1}$$

The methane conversion factor, Y_m, represents the fraction of GE that is converted into methane in the process of enteric fermentation. The country specific Y_m is estimated to be 6 pct. for dairy cattle in Denmark according to the Danish national inventory report by DCE (2019). This implies that on average, 6 pct. of the gross energy intake from feed is converted into methane for cattle in Denmark.

The factor 55,65 represents the energy content of methane measured in MJ/kg CH_4 (IPCC, 2019, p. 10.46). The numerator of equation A.3-1 is expressed as energy in MJ, where the denominator is used to expressed the emission factor in kg. methane (CH_4).

The gross energy intake in MJ per day per cattle (GE) is the total energy need for cattle. This covers the net energy required for maintenance, activity, lactating processes, pregnancy and growth, taking the availability of energy in feed (REM and REG) as well as the digestibility of the feed (DE) into account (IPCC, 2019, p. 10.29).
GE is calculated according to equation A.3-2.

$$GE = \frac{\frac{NE_m + NE_a + NE_l + NE_p}{REM} + \frac{NE_g}{REG}}{\frac{DE}{100}} \qquad \text{A.3-2}$$

Where:
GE = Gross energy, MJ per day
NE_m= Net energy required by the animal for maintenance, MJ per day
NE_a = Net energy for animal activity, MJ per day
NE_l = Net energy required for lactation, MJ per day
NE_{work} = Net energy for work, MJ per day
Ne_p = Net energy required for pregnancy, MJ per day
NE_g = Net energy needed for growth, MJ per day
REM = Ratio of net energy available in a diet for maintenance to digestible energy consumed
REG = Ratio of net energy available for growth in a diet to digestible energy consumed
DE% = Digestible energy expressed as a percentage of gross energy

The digestibility of feed (DE) is given as the fraction of gross energy that is used for the digestion process. In the estimation of GHG, the IPCC default value for Western Europe of 71 pct. is applied for DE. A standard value of 71 pct. for DE is generally used in the context of determining gross energy from cattle in a Danish context (see e.g. Lund & Aaes (2016/2017)).

The net energy requirements for maintenance, activity, lactating, pregnancy and growth respectively, are calculated separately following equations A.3-3-A.3-7.

$$NE_m = Cf_i * (Weight)^{0,75} \qquad \text{A.3-3}$$

NE_m represents the net energy required for maintenance in MJ per day i.e. the net energy required to keep a stable weight without neither weight gains nor losses. NE_m is calculated according to equation A.3-3, using the weight of the cattle and a coefficient determining the energy required for maintenance per kilo metabolic weight. The metabolic weight is calculated as the weight raised to the power of 0,75 and is used as the net energy is only required to maintain the active tissue of the cattle.

The coefficient, Cf_i, is the energy need in MJ per day per kg metabolic weight for cattle. Cf_i varies according to the livestock category. The IPCC default value for dairy cattle is 0,386 (table 10.4) (IPCC, 2019). However, in a Danish context, the Institute of Animal Science at Aarhus University applies a value of 0,293, developed specifically for Nordic countries (Volden, 2011). A value of Cf_i for cattle of $0{,}293$ is therefore also applied in this estimation.

The data for this estimation does not contain specific information regarding the weight of the cattle on each individual farm. It is thereby not possible to vary the net energy required for maintenance based on actual data for the respective farms. Standard values for the average weight of dairy cattle, varying across breed, are thereby applied in the empirical analysis. Data for the average weight of dairy cattle is obtained from the normative figures developed by the Institute of Animal Science at Aarhus University (Lund & Aaes, 2016/2017). These can be seen in Appendix A. The variables are diversified according to whether the herd is heavy breed, jersey or mixed breed.
In the data provided by SEGES, it is not possible to detect the exact composition of breeds for farms categorized as mixed breed. The mix between heavy breed and Jersey are in these cases based on the composition of these types of breeds provided by the fertilizer accounts from the Danish Agricultural Agency. A similar approach is used when calculating the emissions from manure management in section 0.

$$NE_a = C_a * NE_m \qquad \text{A.3-4}$$

NE_a is the net energy required for activity in MJ per day, which is linked to the feeding situation on each individual farm as the feeding situation affects the activity level of cattle.

C_a is determined by whether the dairy cattle are fed in stalls or on pasture lands. The coefficient is determined by an IPCC default value of 0 for dairy cattle in stalls and 0,17 for pasture land (IPCC, 2019, p. 10.24). The coefficient covers the fact that dairy cattle use little or no energy to feed in stalls and modest energy to feed on pasture land.[1]

Following standard assumptions from the Danish National Inventory Reports, all cattle are assumed to spend an average of 18 days on pasture land during a year. This number is as well implemented in the tool developed by the Danish Council on Climate Change. The days spend on pasture land for dairy cattle, will in reality differ considerably across each farm. However, as data is not available on farm level, it will not be possible to estimate a precise number of days for each farm. The standard number of 18 days from the National Inventory Report is applied to ensure consistency in the calculations of the net energy required for activity across all farms. Furthermore, the average of 18 days on grass for dairy cattle is consistent with the assumptions used in the calculations of methane emissions from manure management in section A.3.2.1.

$$NE_l = Milk * (1{,}47 + 0{,}40 * Fat) \quad \text{A.3-5}$$

NE_l is the net energy required for lactating. NE_l is determined by the amount of milk produced as well as the fat content of the milk. A higher fat content implies a higher net energy required for lactating. The net energy required for lactating is calculated according to equation A.3-5.

$$NE_p = C_{pregnancy} * NE_m \quad \text{A.3-6}$$

NE_p is the net energy required for pregnancy in MJ per day. NE_p is calculated according to equation A.3-6. The pregnancy coefficient $C_{pregnancy}$ has an IPCC default value for dairy cattle of 0,1, implying that pregnancy requires 10 pct. additional net energy for cattle relative to the net energy required for maintenance.

[1] The category pasture land covers livestock fed in restricted areas. IPCC also provide a default value for livestock on large grazing areas, where significant energy is required to feed. However, this coefficient is not relevant for the empirical analysis, as it covers open range land or hilly terrain which is not present in the Danish agricultural sector.

The gestation period for cows is 284 days (Volden, 2011) corresponding to 78 pct. of a year. Furthermore, it is assumed that all cows give birth to an average of 0,6 calves a year (Lund & Aaes, 2016/2017). This means that in average 60 pct. of the dairy cows on a farm give birth every year. This correspond to a situation where every cow in the herd is pregnant an average of 47 pct. (78 pct. × 60 pct.) of the days in a year.

When aggregating the emission factors across the individual herd of the farms, the net energy required for pregnancy in MJ per day is thereby only included for 47 pct. of the days in one year. This corresponds to multiplying NE_P with 47 pct. when computing the Gross energy intake (GE).

$$NE_g = 22{,}02 * \left(\frac{Weight}{C * MW}\right)^{0{,}75} * WG^{1{,}097} \qquad \text{A.3-7}$$

NE_g is the net energy required for growth in MJ per day. The net energy required for growth takes the weight and mature weight (MW) of the cattle as well as the average daily weight gain (WG) into account.
As data for the weight and weight gain of the herd is not available at farm level, standard numbers from Lund & Aaes (2016/2017) are used in the empirical analysis for the three weight variables corrected for the mixture of breed in a specific farm (cf. Appendix A).
The coefficient C is dependent on whether the NE_g is estimated for dairy cattle, heifers, bulls or studs, having a lower value for female cattle than male cattle. According to IPPC the coefficient C takes the value 0,8 as the estimation is for dairy cattle (IPCC, 2019, p. 10.25).
In cases where NE_g is estimated for nondairy the value for C should vary according to cattle type.

REG is the amount of net energy available for growth relative to the total digestible energy consumed. REG is calculated according to equation A.3-8.

$$REG = 1{,}164 - (5{,}16 * 10^{-3} * DE) + (1{,}308 * 10^{-5} * (DE)^2) - \left(\frac{37{,}4}{DE}\right) \qquad \text{A.3-8}$$

REM is the amount of net energy available for maintenance, activity, lactating and pregnancy relative to the total digestible energy consumed. REM is calculated according to equation A.3-9.

$$REM = 1{,}123 - (4{,}092 * 10^{-3} * DE) + (1{,}126 * 10^{-5} * (DE)^2) - \left(\frac{25{,}4}{DE}\right) \qquad \text{A.3-9}$$

After estimating the farm-specific methane emission factors for dairy cattle related to enteric fermentation ($EF_{CH_4,Ent}$), these are converted in to CO_2e by using the GWP of methane. These emissions are then multiplied with the number of dairy cattle on the farms in order to obtain the aggregated GHG emissions for each farm.

A.3.2 Manure management

The emissions from manure management can be divided into two parts: methane (CH_4) emissions and nitrous oxide (N_2O) emissions. The emissions from manure management varies across the type of cattle, breed and housing system. The breed is either defined as heavy, jersey or mixed. The composition between heavy breed and jersey cattle for farms with mixed breed is based on the composition from the fertilizer accounts following the same procedure as the calculations of GHG emissions from enteric fermentation.

It should be noted that it is possible for a single farm to operate with different housing systems for their herd of dairy cattle. This has been taken into account in the calculations of the GHG emissions from manure, such that the emissions for each farm is based on the number of cattle in each housing system.
The data on the housing systems were for this thesis only provided for dairy cattle, and therefore it is assumed that non-dairy cattle at the same farm have the same housing system as the dairy cattle. However, it is assumed that calves up to six months lives in deep litter housing systems.

A.3.2.1 Methane emissions

Figure A.3-2 displays an overview of the equations and variables used to calculate the methane emission factor related to manure management ($EF_{CH_4,man}$) for a specific type of cattle. The emission factor consists of two parts; methane emitted in the stable, and methane emitted from time spend on grass.

Figure A.3-2: Overview of equations used in the calculation of $EF_{CH_4,man}$

$$EF_{CH_4,man} = (CH_{4Stable} + CH_{4Grass})$$

$$CH_{4Stable} = VS_{Stable} * MCF * 0,67 * \beta_0$$

$$VS_{Stable} = \frac{M}{365} * DM_M * VS_{DM} * (365 - g_1) + \frac{S}{365} * DM_S * \left(1 - \frac{\%ash}{100}\right) * (365 - g_2)$$

$$CH_{4\,Grass} = VS_{Grass} * MCF * 0,67 * B_0$$

$$VS_{Grass} = \frac{M}{365} * DM_M * VS_{DM} * g_1$$

Source: Authors construction, based on theoretical framework from IPCC (2019).

Where:
VS = volatile solids, kg per animal per year
M = amount of manure excreted, kg per animal per year
S = amount of deep litter, kg per animal per year
DM = dry matter of M manure or S straw, %
VSDM = volatile solids of dry matter, %
g1 = feeding days on grass, days per year
g2 = actual days on grass, days per year
s = amount of straw, kg per animal per year
% ash = ash content in straw

The emission factors from both time spend in stable and on grass depends on the methane conversion factor (MCF) times the maximum methane producing capacity (β_0) and a density factor for CH_4 (0,67). The methane conversion factor is the share of the maximum amount of methane that diffuses under certain circumstances depended on temperature and how the manure is stored. In Denmark the MCF is relatively small, as the average temperature is relatively low. The MCF varies between different type of cattle, breeds and housing systems. The methane conversion factor used in these calculations are specific values from Denmark (2017) and are given by the standard norm tables from DCE (2019) (Annex 3D-15). β_0 varies between the type of cattle and are given by the IPPC guidelines (2019, p. 10.66) (table 10.16). Lastly, the methane emission factors from manure management depend on the volatile solids for respectively housing (VS_{stable}) and grassing (VS_{grass}), which will be further explained in the following.

The volatile solids are created from the manure. These are emitted both in the stable and on pastureland. However, the emissions of volatile solids in the stable are dependent on the amount of manure and the amount of deep litter from a stable. Furthermore, the emissions of volatile solids

from stable are dependent on the days the cattle spend in the stable. The emissions of volatile solids in the stable are dependent on the dry matter content in manure and straw. The volatile solids from grass are only estimated for feeding days on grass and do thereby not include the volatile solids created from deep litter.
The volatile solids are based on national specific data for Denmark and vary across housing type and type of cattle for each farm.

The ash content for straws are assumed to be 4,5 pct. (DCE, 2019) and days on grass are assumed to be 18 for all type of cattle besides heifers, who has 132 days on grass (DCE, 2019). A value of 80 pct. volatile solids of dry matter are used in line with the DCE standard normative tables (2019).

A.3.2.2 Nitrous oxide emissions

Figure A.3-3 displays an overview of the calculation of N_2O emissions related to manure management. The total emission of N_2O from manure management consist of direct ($N_2O_{D,man}$) and indirect N_2O ($N_2O_{I,man}$) emission, shown in the figure below.

Figure A.3-3: Overview of equations used in the calculation of $EF_{N_2O,man}$

$$EF_{N_2O,man} = N_2O_{Dman} + N_2O_{I_{man}}$$

$$N_2O_{D,man} = Nex_T * EF_{3(S)} * \frac{44}{28}$$

$$N_2O_{I,man} = N_{volitalization-mms} * EF_{4(s)} * \frac{44}{28}$$

$$N_{volitalization,mm} = Nex_T * \left(\frac{Frac_{Gas}}{100}\right)$$

Source: Authors construction, based on theoretical framework from IPCC (2019).

Where:

S: Management system

T: Cattle type

Nex_T: Annual excretion of N per cattle type T per year expressed in N_2O-N

$EF_{(3)s}$: Emission factor for direct emissions in given manure management system

44/28: Conversion factor from N_2O-N emissions to N_2O emissions

$N_{volatilization,\ mms}$: Annual N lost as volatilization of NH_3 and NO_x measured in kg. N

$EF_{4(s)}$: Emission factor volatilized nitrous oxide from agricultural soils

$Frac_{gas}$: % of nitrogen that volatilize as NH_3 and NO_x

The direct N_2O emissions consist of a measure of annual excretion of N (kg.) per year for cattle (Nex). The value used in this estimation is from the standard norm tables estimated by Poulsen (2017) and is country specific for Denmark. Nex varies across cattle types, housing systems and breeds. Lastly an emission factor (EF_4) (kg N_2O-N per. kg, Nex) is multiplied with the annual excretion of N expressed as N_2O-N. In order to transform this into N_2O the conversion factor is used. The emission factor various across different housing systems and is a standard value from IPCC (2019) (table 10.21) following the same procedure as the National Inventory Report.

The indirect N_2O emissions are the amount of NO_x and NH_3 that is volatilized from the manure management system measured in kg. nitrogen (N). This is multiplied with an emission factor for the volatilization of N and is converted by the value 28/44 to measure it in N_2O. The emission factor is a standard value from IPPC (2019, p. 10.90) (table 10.21) and is the same value as used in the National Inventory Report.

The Nitrogen volatilization is calculated as the annual excretion of N per cattle per year times the percentage share of nitrogen that volatilize as NH_3 and NO_x in a given manure management system.

The two different emission factors for methane and nitrous oxide respectively found above varies across manure management system and cattle types. Each of these different factors have been converted into CO_2e and been multiplied with the specific number of a cattle type living in a specific housing system. These are then aggregated to calculate the total GHG emissions for a specific farm expressed in CO_2e.

A.4 Results

In this section the aggregated emission factors across farms are represented for both enteric fermentation, manure management and in total. Firstly, the results from enteric fermentation are presented, secondly the results from manure management and lastly the total aggregated GHG emissions from the farms consisting of both is presented.

A.4.1 Emission factors for enteric fermentation

The emission factors for dairy cattle vary across farms according to the individual variation of gross energy intake (GE) for the average dairy cattle on a given farm. In order to obtain the aggregated methane emissions for the total number of dairy cattle of each individual farm, the methane emission factors ($EF_{CH_4,ent}$) for each subcategory of cattle is multiplied with the farms number of dairy cattle.
A summary of $EF_{CH_4,ent}$ for dairy cattle is shown in Table A.4-1.

Table A.4-1: Summary of estimated $EF_{CH_4,ent}$ for dairy cattle

Breed	**Min**	**Q_1**	**Mean**	**Q_3**	**Max.**	**Sd**
Heavy	101,95	147,39	155,88	164,59	216,95	13,28
Jersey	102,28	124,26	129,20	134,95	146,42	8,36
Mixed	116,86	147,93	155,85	164,55	191,50	14,22

There are differences in emission factors across the three types of breed presented in Table A.4-1. Farms with dairy cattle of heavy breed have on average a higher emission factor for methane related to enteric fermentation relative to farms with jersey cattle. This is as expected, as dairy cattle of heavy breed are in general larger animals with a higher milk yield which drives the emission factors upwards. For farms with a mixture of heavy breed and jersey, the emission factors are in general relatively similar to those for farms with only heavy breed. This is caused by the fact that the mixed farms tend to have a relatively low share of jersey cattle in general.

The emission factors for methane related to enteric fermentation have only been estimated for dairy cattle. For the other subcategories: bulls, heifers, studs and calves, emission factors from norm tables have been used (table 5.7) (DCE, 2019). This is done as specific data for non-dairy cattle were not available for the analysis. Nevertheless, non-dairy cattle represent a relatively small share of the sample. Furthermore, the emission factors for non-dairy cattle are generally significantly smaller than for dairy cattle. As the aggregated emissions from non-dairy cattle only represents a relatively small part of the total GHG emissions from enteric fermentation it is assessed not to impact the overall results significantly.

As norm figures are used for the emission factors for the subgroups of non-dairy cattle, these will not vary according to breed and general activity level on the specific farm. The emission factors for the subcategories of nondairy cattle are shown in Table A.4-2. Nevertheless, the calculation methods and assumptions behind the estimation of emission factors for nondairy cattle are similar to those applied for the estimation of emission factors for dairy cattle in this analysis, as both methods are based on the national normative values as well as the IPCC guidelines.

Table A.4-2: $EF_{CH_4,ent}$ for non-dairy cattle

Cattle type	Emission factor (kg. CH_4 /head/ year)
Heifers (>6 months)	55,51
Heifer (0-6 months)	43,62
Bulls and studs (>6 months)	21,38
Bulls (0-6 months)	13,05

Source: DCE (2019) (table 5.7)

A.4.2 Emission factors from manure management

The estimated emission factors from manure management are shown in Table A.4-3. The table only contains emission factors from dairy cattle, but these also varies across different types of non-dairy cattle (see Appendix A). The N_2O emission factors represent the sum of indirect and direct emissions. In order to aggregate the methane emissions factors ($\mathrm{EF}_{\mathrm{CH_4,man}}$) and the nitrous oxide emission factors ($\mathrm{EF}_{\mathrm{N_2O,man}}$), the emissions should be converted into CO_2e as they are presented in different units.

Table A.4-3: Estimated $EF_{CH_4,man}$ and $EF_{N_2O,man}$ across stable systems for dairy cattle

Housing systems	Share of dairy cattle	**N_2O Emission factor** $EF_{CH_4,man}$ **(kg N_2O catle^{-1} year^{-1})**		**CH_4 Emission factor** $EF_{N_2O,man}$ **(kg CH_4 catle^{-1} year^{-1})**	
		Heavy breed	*Jersey*	*Heavy breed*	*Jersey*
Tethered with urine and solid manure	0,2 pct.	1,195	0,993	11,702	8,749
Tethered with slurry	0,9 pct.	1,328	1,103	27,908	19,899
Loose-holding with beds, solid floor	15,6 pct.	1,232	1,023	23,377	18,133
Loose-holding with beds, slatted floor	23,8 pct.	1,280	1,063	23,377	18,133
Loose-holding with beds, slatted floor, scrape	48,9 pct.	1,256	1,043	23,377	18,133
Loose-holding with beds, solid floor with tilt	3,9 pct.	1,304	1,083	23,377	18,133
Deep litter (all)	3,2 pct.	2,638	2,192	118,980	98,728
Deep litter, long eating space, solid floor	0,6 pct.	2,145	1,492	97,912	79,124
Deep litter, slatted floor	1,1 pct.	2,164	1,522	97,912	79,124
Deep litter, slatted floor, scrape	1,8 pct.	2,155	1,507	97,912	79,124

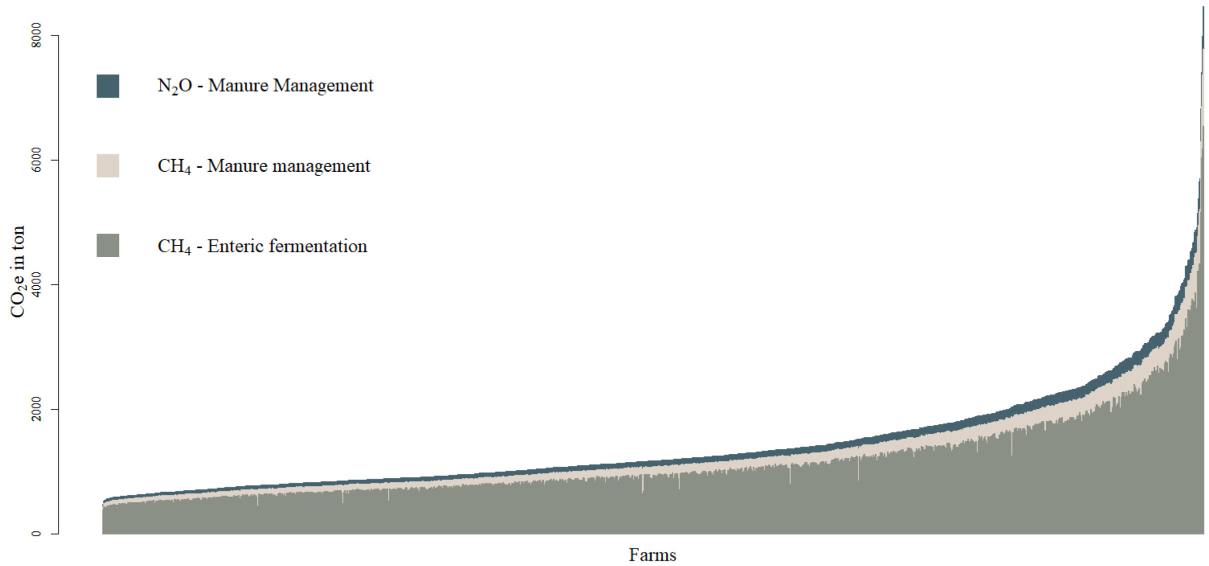
The table illustrates the difference in emission factors, which are primarily determined by the manure management system which is primarily given by the housing system. The most frequently used housing system in the sample is different loose-holding with beds, as they represent approximately 88 pct. of the sample. These type of housing systems emit less CH_4 and N_2O in relation to deep litter systems, which are mostly used for calves.

A.4.3 Aggregated GHG emissions from enteric fermentation and manure management

In order to estimate the aggregated GHG emissions across enteric fermentation and manure management, the emission factors are converted into CO_2e to ensure comparability. Emission factors are thereafter multiplied with the corresponding number of dairy and non-dairy cattle for each subcategory on the specific farm to obtain the total GHG emissions at farm level.

Figure A.4-1 contains an overview of the total GHG emissions for each farm in the dataset, measured in CO_2e.

Figure A.4-1: Composition of GHG emissions for each farm in the dataset (CO_{2e})



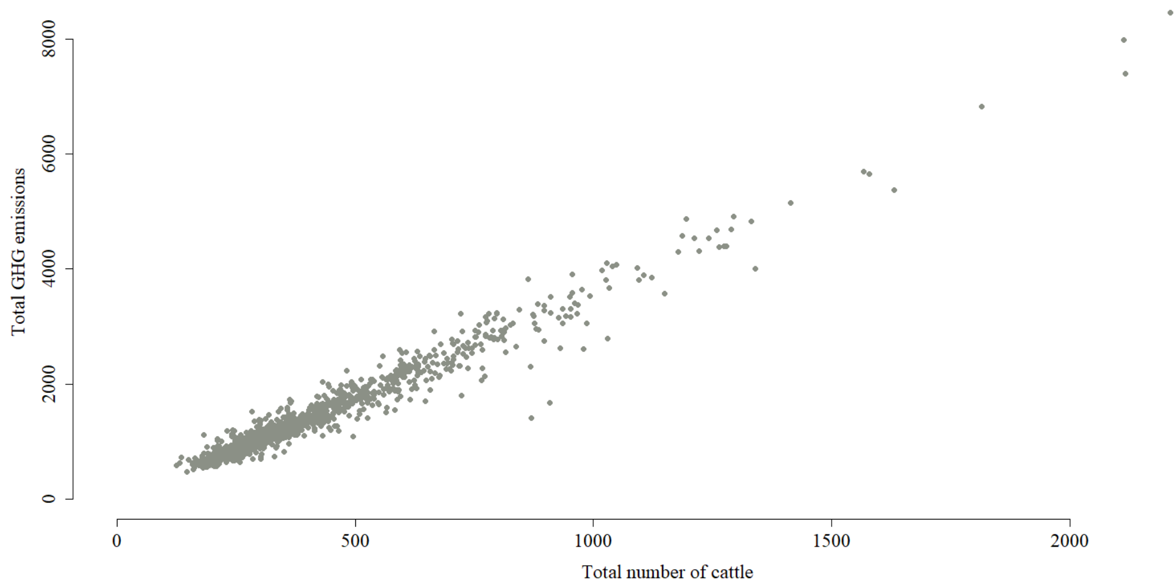
The figure illustrates the estimated contribution from enteric fermentation and manure management respectively to the total GHG emission of each farm in the dataset. Not surprisingly, CH_4 emissions from enteric fermentation represents the majority of the GHG emissions for all farms. These are followed by CH_4 emissions from manure management, where N_2O emissions from manure management represents the smallest share of the total GHG emissions.

The composition of the three sources of GHG for cattle, should according to both DCE (2019) and Mikkelsen (2020) be approximately 76 pct. CH_4 emissions related to enteric fermentation, 15 pct. CH_4 emissions related to manure management and the remaining 9 pct. N_2O emissions from manure management. This is relatively in line with what is found in this empirical analysis as illustrated in the figure above. The specific composition of the sources of greenhouse gas emissions vary to some extend between the farms. Nevertheless, the composition is relatively homogenous for the majority of the sample. It is also shown how a few farms have relatively large aggregated greenhouse gas emissions compared to the majority of the sample.

The total GHG emissions are highly dependent on the number of cattle for each farm. The relationship between total GHG emissions and the number of cattle is illustrated in Figure A.4-2.

The figure illustrates a relatively high correlation between total GHG emissions and the number of cattle, as expected. However, it can also be seen that there is some variation in the total GHG emissions for farms with similar number of cattle. Furthermore, it is seen that a handful of farms differs from the rest, by having a relatively large number of cattle and correspondingly high aggregated GHG emissions.

Figure A.4-2: Correlation between total GHG emissions and number of cattle



The aggregated GHG emissions across the total sample are presented in Table A.4-4 both expressed in CH_4, N_2O as well as in CO_2e.

The overall GHG emissions used in the following benchmarking analysis in Part B consists of an aggregation of the three components listed in the table below.

Table A.4-4: Summary of estimated GHG emissions across all farms

	Emissions from enteric fermentation		**Emissions from manure management**				**Aggregated GHG emissions**
	CH_4	CO_2e	CH_4	CO_2e	N_2O	CO_2e	CO_2e
Aggregated GHG emissions (tons)	57.011	1.425.270	9.553	238.822	474	141.246	1.805.338
	(78,9 pct.)		*(13,3 pct.)*		*(7,8 pct.)*		*(100 pct.)*

A.5 Validation of the estimated GHG emissions

In order to validate the estimated GHG emissions from enteric fermentation and manure management, the aggregated emissions are compared to those from the Danish National Inventory report. The sample in the empirical analysis might not be directly representative of the population of dairy cattle on all measures. Therefore, the share of the population of dairy cattle from the sample might not represent the exact same share of the total estimated GHG emissions from the National Inventory Reports. However, comparing the estimated GHG emissions from the sample to the corresponding share of total GHG emissions from the National Inventory Report can be a guide as to whether the estimated emissions are consistent with the national estimations.
Furthermore, in the process of calculating CH_4 emissions from enteric fermentation, calculations of sub elements such as the net energy requirements have been compared to standard numbers used in the National Inventory report from Lund & Aaes (2016/2017) to ensure validity of the estimation.

The results are only presented for dairy cattle in the specific validation of emissions from enteric fermentation and manure management. Non-dairy cattle are estimated to only contribute with approximately 18 pct. of the aggregated estimated emissions and are not the dominant type of cattle for the dairy farms in the empirical analysis. Specific validation for the subcategories of non-dairy cattle are thereby not presented in the same way as for dairy cattle but is included in the validation of the final aggregated GHG emissions.

The gross energy intake (GE) of dairy cattle is the determinant factor for the estimation of the CH_4 emission factors for enteric fermentation ($EF_{CH_4,ent}$). Therefore, the purpose of the following is to validate that the estimated gross energy intake for the dairy cattle in the sample are in line with the national estimations.

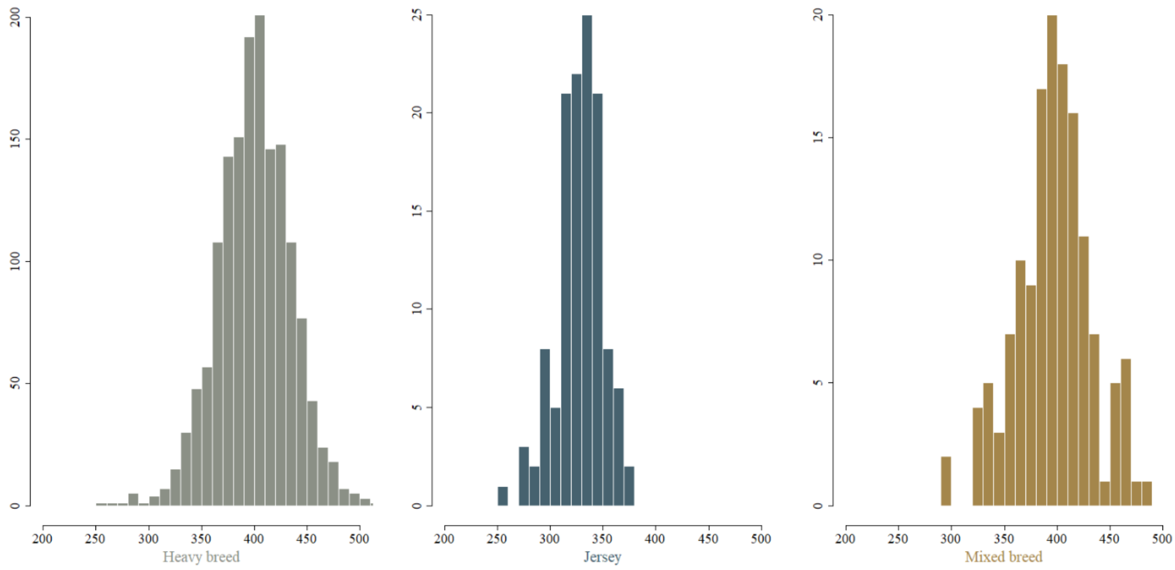
Table A.5-1: Estimated average GE for dairy cattle compared to national estimates

Gross Energy intake (MJ per head per year)	**Heavy breed**	**Jersey**	**Mixed breed**
National Inventory report	401,9	334,9	392,2
Average from own estimation	411,2	328,3	396

Table A.5-1 shows a comparison of the estimated average gross energy intake (GE) and the national estimates from the National Inventory Report. This shows that the estimations of GE are on average relatively similar to those estimated on a national level.

Figure A.5-1 contains an overview of the distribution for the estimated GE for dairy cattle across different breeds. It can be seen that for all three types of breeds, the estimated GE tends to be rather normally distributed around the average GE. This implies that GE is, throughout the sample, is estimated relatively similar to what could be expected.

Figure A.5-1: Distribution of estimated GE for dairy cattle across breeds



Note: The range of the y-axes varies across the three plots as heavy breed represents the largest share of the sample.

In order to validate the overall emissions from both enteric fermentation and manure management, the estimated GHG emissions from the two categories are compared to the corresponding national estimations.

According to the National Inventory Reports, enteric fermentation accounts for 35 pct. of the total emissions in the agricultural sector in 2017, corresponding to 3,731 mill. tons CO_2e. Of these, dairy-cattle in particular contributes with a total of approximately 2,268 mill. tons CO_2e (DCE, 2019).

The sample used in this thesis, includes approximately 51pct. of the total population of dairy cattle in Denmark for the year 2017. Comparing this to the estimations from the national inventory report, 51 pct. of the total greenhouse gas emissions from enteric fermentation for dairy cattle corresponds to **1,157 mill. tons CO_2e**. In comparison, the total CO_2e emissions stemming from enteric fermentation estimated in this thesis, are estimated to be **1,162 mill. tons CO_2e** in 2017 across all farms in the sample. This shows that overall, the aggregated GHG emissions from enteric fermentation are estimated in line with what is estimated on a national level.

In 2017 CH_4 emissions from manure management contributed with a total of 1,812 mill. tons CO_2e following the National inventory report. CH_4 from Manure management accounts for approximately 17 pct. of the total CO_2e emissions in the agricultural sector, where 20 pct. of these emissions originate from dairy cattle in 2017. Thus, the total emissions from manure management stemming from dairy cattle is estimated by the DCE (2019) to be 0,348 mill. tons CO_2e for 2017. 51 pct. of this corresponds to **0,178 mill. tons CO_2e**. In comparison, the total CO_2e emissions stemming from manure management, are estimated to be **0,21 mill. tons CO_2e** in 2017 across all farms in the sample.

CH_4 emissions from manure management thereby seems to be slightly overestimated in the thesis compared to the numbers estimated on national level in the national inventories. However, it might not be the case that the 51 pct. of the population of dairy cattle, which are represented in the sample, corresponds to 51 pct. of the total emissions. Furthermore, CH_4 from manure management contribute less to the total GHG emissions from dairy cattle compared to emissions stemming from enteric fermentation. The consequences for the aggregated GHG emissions are thereby not necessarily substantial for the final estimations even if the CH_4 emissions are slightly overestimated.

N_2O emissions cover both direct and indirect emissions. Data, for the emissions of N_2O stemming from dairy cattle, is not directly extractable from the National inventory reports. Therefore, emissions from 2018 estimated by Mikkelsen (2020) is used in the comparison of the N_2O emissions from manure management. Numbers from 2018 might deviate from 2017 estimations but are still assumed to be relatively comparable with the estimations from the empirical analysis.

The aggregated N_2O emissions from dairy cattle are estimated to be 750 tons N_2O by Mikkelsen (2020), corresponding to 223.500 tons CO_2e. As the sample in the empirical analysis covers approximately 51 pct. of the Danish population of dairy cattle, emissions from this share of the total number of dairy cattle corresponds to approximately **114.306 tons CO_2e**.

In the empirical analysis, the total N_2O emissions from dairy cattle are estimated to be **114.625 CO_2e** which is very close to the corresponding estimation by Mikkelsen (2020).

Part B: Benchmarking analysis

B.1 Background

According to the Danish Council on Climate Change, among others, there exist a need for a tool which creates the best possible conditions for a cost-efficient climate regulation. To ensure the effectiveness of any future regulation it is necessary to thoroughly consider and develop methods which do not neglect the interaction between GHG emissions and economic performance. It is in general assumed that it is costly to reduce GHG emissions. Therefore, it is assumed that there exists a trade-off where increasing the climate performance will lead to lower economic performance. However, the trade-off might not be similar throughout an entire sector and may vary from producer to producer. A cost-efficient regulation of climate in the agricultural sector must thereby take this into account as it paves the way to legislate such that GHG emissions are reduced at minimum cost i.e. at an optimal trade-off between climate and economic performance.

GHG emissions can be considered as an undesirable output in the production of dairy products. Currently, GHG emissions occur as a negative externality in the production as the emissions are costless for the farm but bares a cost to society.
Externalities is in economic literature referred to as a spillover effect from a production. A negative externality will thereby affect either other firms' production or people's utility level negatively (Sørensen, et al., 2016). The producer itself, does not bear the cost of the externality and will thereby not act according to the actual socioeconomic cost of the negative externality. In order to ensure that producers act in a socioeconomic optimal way, the externality must therefore be internalized. Internalizing an externality means that the socioeconomic cost is included in the incurred cost of the producer which could be done by e.g. taxing the externality according to the socioeconomic cost.

GHG emissions in the Danish dairy production have an additional cost to society, which the farm does not incur. The additional cost for society of emitting one more unit of GHG is in economic theory referred to as the marginal damage cost (Kolstad, 2011, pp. 234-251). If a farm reduces its GHG emissions with one ton, the marginal benefit for society will thereby be that the marginal damage of that unit has been avoided. The marginal abatement cost of reducing this ton of GHG emissions for the farm must be equal to the marginal benefit for society in order to ensure a socioeconomic optimal production.

Both the marginal abatement cost and the marginal benefit to society is not observed directly. These can be very difficult to measure, and the measures will often be uncertain. The current price of CO_2 is 214 DKK per ton (The Danish Ministry of Finance, 2019). The price is currently given by the quota sector but is also used as the price of CO_2 in areas outside the quota sector for estimations such as cost benefit analyses. Nevertheless, it is currently debated whether this price is too low to reflect the actual socioeconomic damage and abatement costs of CO_2. Therefore, the Danish Council on Climate Change has suggested that the price of CO_2 emissions must increase from the current 214 DKK per ton to 1.500 DKK per ton (The Danish Council on Climate Change, 2020). The Danish Council on Climate Change has proposed this as the volume of a tax on CO_2 both within and outside the quota sector (The Danish Council on Climate Change, 2020). The 1.500 DKK are estimated by using the 70-pct. reduction target as a measure of how much GHG emission that must be abated to reach the target. This is then used to calculate the price per ton of CO_2 of abating this amount of CO_2 (The Danish Council on Climate Change, 2020). The 1.500 DKK thereby reflects the marginal abatement cost of this reduction. However, there might be relatively great variance in the actual abatement costs for different sectors and the empirical analysis of this thesis seeks to estimate farm-specific abatement costs for the agricultural sector.

The marginal abatement costs can be referred to as shadow prices. Shadow prices are often used to compare different alternatives of abating CO_2 emissions, as it is more cost-effective to reduce the CO_2 emissions with the lowest shadow prices (The Danish Ministry of finance, 2017). The empirical analysis covers estimating the shadow price of GHG emissions for Danish dairy farms. This is done by estimating a non-parametric benchmarking model with two outputs and a single input. The farm-specific GHG emissions, estimated in Part A, are used as an undesirable output, revenue a desirable output and total costs as the input. The model thereby includes both the economic performance as well as the climate performance of the farms. The model is based on linear programming, and in this specific framework, weak disposability of GHG emissions is assumed, and a directional distance function is imposed to handle this output as an undesirable output. From this model, the estimated farm-specific marginal abatement costs (shadow prices) for

GHG emissions are estimated by considering the trade-off between GHG emissions and revenue. The shadow price of GHG for a specific farm will thereby equal the estimated decrease in revenue that occurs when a farm decrease GHG emissions by one ton.

The benchmarking framework presented in section B.2 can be applied to all parts of the agricultural sector and will ultimately also be able to handle more variables than presented in this analysis. These could include parameters describing animal welfare, environmental performance etc. where there is also assumed to be a trade-off with economic performance or between the given variables.

The foundation of the benchmarking framework is the non-parametric efficiency analysis, Data Envelopment Analysis (DEA). DEA is developed based on Farrells (1957) proportional measure of efficiency (Bogetoft & Otto, 2011, p. 15). The output-oriented Farrell efficiency for DMU$^{k'}$ is defined as $\frac{Y^*}{Y^{k'}}$, where Y^* represents a vector of the maximum possible outputs, and $Y^{k'}$ represents the realized output quantities of $DMU^{k'}$ (Bogetoft & Otto, 2011, p. 15). The interpretation of the output-oriented Farrell efficiency score thereby becomes how much $DMU^{k'}$ should increase its output in order to become efficient. The Farrell output-oriented efficiency score is thereby always greater than one. An output-oriented Farrell efficiency score of 1,2 would indicate that the DMU should produce 120 pct. of its current output quantity in order to become efficient.

In the same matter, the input-oriented Farrell efficiency score is defined as $\frac{X^*}{X^{k'}}$, where X^* represents the maximum possible input, and $X^{k'}$ the observed input of $DMU^{k'}$. The input-oriented Farrell efficiency can thereby be interpreted as how much of $DMU^{k'}'s$ input quantity is actually necessary to operate. The input-oriented efficiency score range from zero to one, where a score of e.g. 0,8 indicates that $DMU^{k'}$ should reduce its inputs proportionally (i.e. in the same ratio) to 80 pct. of its currents inputs in order to become efficient.

Charnes et al. (1978) developed the non-parametric efficiency analysis “Data Envelopment Analysis” (DEA), which is based on the concepts of the proportional Farrell efficiency as well as linear programming. The Charnes et al. DEA model from 1978 assumes constant returns to scale (CRS), and was followed by a DEA model assuming variable returns to scale (VRS), developed by Banker (1984).

The DEA approach involves estimating a production possibility frontier (PPF), representing best practice within the area of analysis. The frontier is formed by enveloping the DMUs, where the frontier is then constituted of the efficient DMUs and convex combinations of these. Each DMU in the dataset is projected onto the frontier in order to measure their relative efficiency when assessing these against best practice. DMUs are in this way assigned an efficiency score, ϕ, by evaluating their performance relative to the frontier.

The DEA approach applies a set of essential axioms for defining the production possibility set. These are that *all observed DMUs are in the production possibility set, free disposability, convexity* and lastly an assumption of *returns to scale* for the specific area of analysis.
The production possibility set, $PPS(x)$, can be formally defined by considering a dataset with K DMUs ($k = (1, \dots K)$) and for DMU^k letting $x^k = \left(x_1^k, \dots, x_M^k\right) \in \Re_+^M$ denote the input vector used to produce the output vector denoted by $y^k = \left(y_1^k, \dots, y_S^k\right) \in \Re_+^S$.

Firstly, it is assumed that *all observed DMUs are in the production possibility set*, indicating that it is possible for any DMU in the dataset to operate as any other DMU. This defines the production possibility set as seen in equation B.1-2.

$$PPS = \{(x, y) \in \Re_+^{M+S} | x = x^k, y = y^k, k = 1, \dots, K\} \qquad \text{B.1-1}$$

This assumption implies that there are no measurement errors in the estimation. Thus, the non-parametric DEA approach is deterministic and assumes no noise in the dataset, which makes the DEA model relatively sensitive to outliers.

The second axiom is the assumption of *free disposability* (also referred to as strong disposability). In the standard DEA approach, it is assumed that any DMU can freely dispose of both inputs and outputs. This indicates that a DMU can always use more input to produce a fixed amount of output and always produce less output given any fixed input quantity. This assumption indicates that the production possibility set can be expressed following equation B.1-2.

$$PPS = \{(x, y) \in \Re_+^{M+S} | x \geq x^k, y \leq y^k, k = 1, \dots, K\} \qquad \text{B.1-2}$$

This assumption is essential to the empirical analysis of this thesis. The assumption of free disposability is revised for the undesirable output, as weak rather than strong disposability is assumed for this output. This is discussed further in section B.2.

The assumption of *convexity* implies that any convex combination of DMUs is feasible. This indicates that the frontier, which constitute best practice and the benchmark for inefficient DMUs, consists of both observed efficient DMUs and convex combinations of these. Adding the assumption of convexity the production possibility set, can be expressed as shown in equation B.1-3.

$$PPS = \{(x,y)\in \Re_{+}^{M+S} | x \geq \sum_{k=1}^{K} \lambda^k x^k \ , y \leq \sum_{k=1}^{K} \lambda^k y^k \, , k = 1, \dots, K\} \qquad \text{B.1-3}$$

Furthermore, it is necessary to specify the returns to scale, which is dependent on the specific assumption regarding the subject of analysis. In the empirical analysis, the model is specified with the assumption of constant returns to scale (CRS). Assuming CRS implies that scaled feasible points are also assumed to be feasible. With CRS, intensity vectors ($\lambda's$) are constrained to $\lambda^k \geq 0$ and represent the scaling of DMU^k when specifying the production possibility set.

Benchmarking DMUs with DEA implies measuring the efficiency of each DMU relative to the frontier of the production possibility set. In an output-oriented CRS problem, the mathematic optimization problem is presented as in equation B.1-4, where ϕ represents the output-oriented Farrell efficiency score.

$$\begin{aligned}
&\max_{\lambda^k, \phi} \phi \\
&\sum_{k=1}^{K} \lambda^k y^k \geq \phi y^{k'} \\
&\sum_{k=1}^{K} \lambda^k x^k \ \leq x^{k'} \\
&\lambda^k \geq 0, k = 1, \dots, K
\end{aligned} \qquad \text{B.1-4}$$

One of the advantages of the non-parametric DEA approach is that it takes a minimum of assumptions regarding the production technology. A minimum prior knowledge of the production technology is thereby needed, as opposed to certain parametric analyzes where it is necessary to specify e.g. a specific production or cost function. However, as the method is deterministic, assuming no error term in the estimation, the method is relatively sensitive to outliers as oppose to a range of statistical methods, which to a greater extend is able to account for measurement errors.

The standard DEA approach is either input-oriented – minimizing all inputs proportionally keeping outputs fixed, or output-oriented – maximizing all outputs proportionally keeping inputs fixed.
The original DEA method thereby assumes that all outputs are desirable outputs, which each DMU is interested in maximizing. However, there has been a growing need for the possibility of modelling undesirable outputs, such as GHG emissions, as part of the production. By the nature of the undesirable outputs, each DMU should not seek to maximize but instead minimize these specific outputs, which is contradicting to the general modelling of desirable outputs.

Modelling undesirable outputs are especially relevant in areas concerning environmental and climate performance, as a method which does not neglecting society's cost of the externality in a given performance benchmarking.

The following section describes the methodology behind a benchmarking model where GHG emissions are incorporated as an undesirable output by assuming weak disposability between this and the desirable output and using a directional distance function.

The results from the benchmarking model can be split into two steps. The first step defines the shadow prices of GHG emissions along the frontier. These are hereafter referred to as "frontier shadow prices". The frontier shadow prices represent the marginal trade-offs between GHG emissions and revenue when farms operate at best practice, representing the best possible technology currently available.
The second step defines the current potentials for the inefficient farms of reducing GHG emissions and/or maximizing revenue. Through the potentials, it is possible to calculate the average opportunity costs of reducing GHG emissions for inefficient farms by comparing scenarios where the farms either only focus on reducing GHG emissions or increasing revenue. The average

opportunity costs thereby represent how much potential revenue an inefficient farm must give up, in order to move in a direction where only GHG emissions are reduced.

Both the estimated frontier shadow prices and the average opportunity costs are so called abatement costs. However, they differ in interpretation. Frontier shadow prices represent the marginal cost of moving along the frontier for efficient farms i.e. the marginal abatement cost when operating at best practice. The average opportunity costs represent the average cost of reducing GHG emissions, measured in loss of potential revenue. The loss in potential revenue is estimated by comparing the potentials when farms are benchmarked towards efficient farms with low GHG emissions instead of farms with high revenue.

B.2 Methodology

B.2.1 Modelling undesirable outputs

A range of methods have been developed in order to incorporate undesirable outputs in benchmarking. Overall, the methods can be divided into two overarching approaches; the direct and indirect methods (Scheel, 2001).

The indirect methods – transforming the undesirable output

The indirect methods all concern transforming the specific variable for the undesirable output (denoted w) with a function $f(w)$ *before* integrating these in a given benchmarking model (Scheel, 2001). The transformed variable is then used as a regular desirable output in the model.

The most basic indirect method is the transformation referred to as *the additive inverse*. The method is originally developed by Koopmans (1951), and simply reverse the undesirable output by multiplying the variable with -1 so that $f(w) = -w$. In this way, the variable now acts as a desirable output, with the standard "more is better" assumption, and can thereby enter as an output that is sought to be maximized. However, it should be noted that e.g. DEA and other non-parametric benchmarking approaches are not able to handle negative variables, and the method is thereby more suitable for e.g. additive models such as the one developed by Cooper et al. (1999). An option which corresponds to the additive inverse is instead modelling the undesirable output as an input in the benchmarking model. This method is theoretically similar to the additive inverse, however as the undesirable output is in reality an output of production rather than an input, this method distorts the real production process, as the input-output relation is changed using this method (Scheel, 2001).

Another way of dealing with the fact that the additive inverse leads to negative values for the transformed undesirable output is adding a constant, C, to the negative of the variable, choosing C such that the translated variable becomes positive for all DMUs (Ali & Seiford, 1990). The method can be formulated as $f(w) = -w + C$. However, this approach is relatively sensitive to the choice of the constant C. Even though the ranking of the undesirable output of each DMU does not change

using this approach, the relative differences between them can be highly affected by choosing either a relatively large or small value of C.
An alternative direct approach is the so called *multiplicative inverse* developed by (Golany & Roll, 1989), where the undesirable output is transformed by taking the inverse of the original $f(w) = 1/w$. However, as this is a non-linear transformation, the transformation tends to distort the relationship between DMUs in relation to the altered variable.

The direct method – weak disposability with a directional distance function

Common for the direct approaches of handling an undesirable output is that these all assume free disposability of this undesirable output, as the transformed variable is incorporated in the benchmarking model as a regular output. The indirect methods seek to deal with this problem by not transforming the variable for the undesirable output itself, but instead changing the benchmarking model to cope with the nature of the undesirable output.

The standard DEA assumption of free disposability implies that it is always possible to produce more output given any fixed input quantity and that it is always possible to use more input given any output quantity. However, this might not be the case with certain undesirable outputs such as the undesirable output of this empirical analysis, GHG emissions. The production of dairy is naturally linked to GHG emissions through both enteric fermentation and manure management of cattle. Furthermore, it is in reality assumed that it is costly for a farm to reduce GHG emissions i.e. the marginal abatement cost of GHG emissions is positive. Thereby, it requires resources to reduce GHG emissions and these resources can therefore not be used to e.g. make investments to increase revenue. It is thereby not possible to decrease GHG emissions without also reducing revenue, as the amount of capital available for creating revenue is reduced.

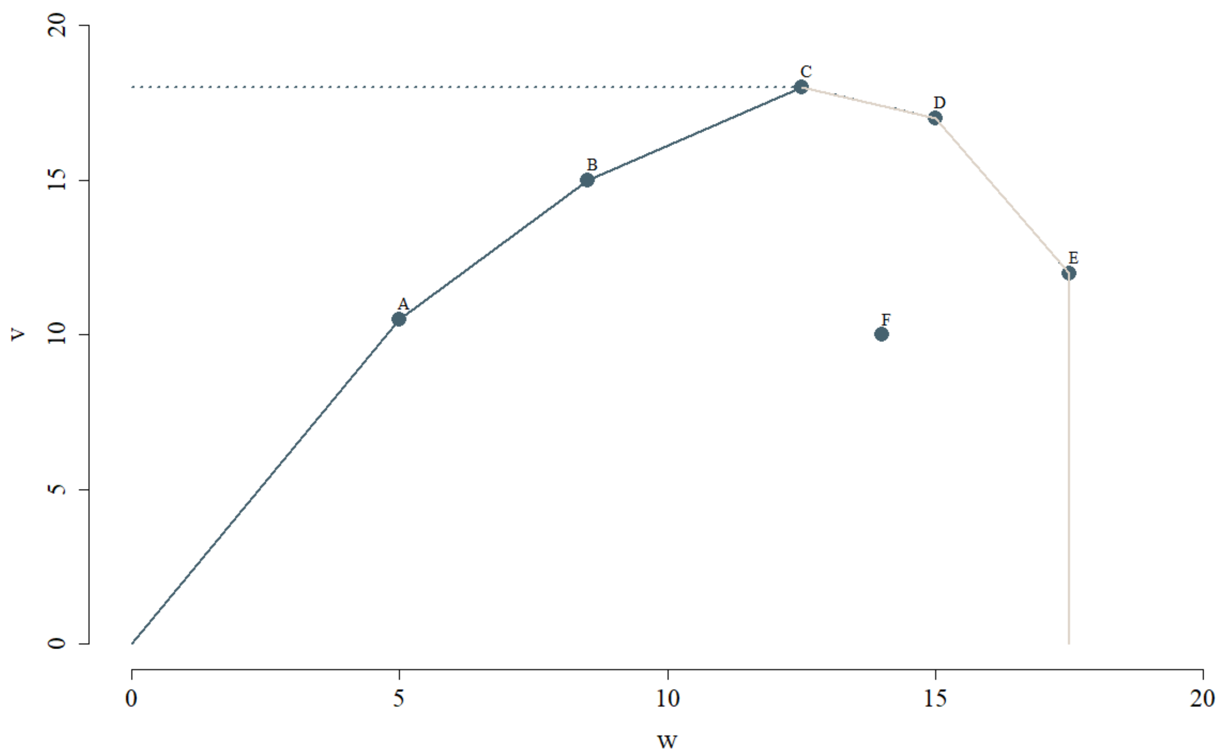
The indirect method, which is applied for handling GHG emissions in the following empirical analysis, relaxes the basic assumption of free disposability, imposing weak disposability of the undesirable output(s) but still assuming strong disposability of the desirable output(s).

B.2.2 The production possibility set with weak disposability of the undesirable output

The concept of weak disposability can formally be expressed by letting the output matrix be denoted by $y = (v, w)$, where v represents a desirable output and w represents an undesirable output, following the example from Scheel (2001).

The undesirable output, w, exhibit weak disposability, defined as $y \in PPS(x) \Longrightarrow \theta y \in P(x)$, $\forall\ 0 \leq \theta \leq 1$. In order to still belong to the production possibility set, it is thereby necessary to reduce both the undesirable *and* the desirable output proportionally/radially in order to reduce the undesirable output, given a fixed level of input (x).
For the desirable output, v, it is the case that $(v, w) \in P(x) \Longrightarrow (v', w) \in P(x),\ for\ v' \leq v$, Implying that it is possible to reduce the production of the desirable outputs without changing the input level.
Figure B.2-1 illustrates how the production possibility set changes when estimating a model where weak disposability is imposed for the undesirable output w.

Figure B.2-1: The production possibility set with a weak disposable output



For the standard DEA approach, presented in section B.1, it is assumed that all outputs exhibit strong disposability. In the standard DEA approach, the frontier will follow the horizontal dotted line from point C and to the left. This implies that it is possible to reduce output w, keeping the other output, v, constant.

However, when w is assumed to exhibit weak disposability, this part of the frontier will not be feasible, as it is necessary to also reduce the desirable output when reducing the undesirable output. This part of the frontier is therefore instead estimated by reducing both outputs simultaneously. The frontier is thereby created by first following the assumption of convexity between point A, B and C and thereafter radially contracting both outputs from point A towards the origin. The desirable output, v, still exhibit strong disposability, as can be seen following the line from point E and vertically down, indicating that it is still possible to reduce the desirable output, v, without reducing the undesirable output, w.

With an undesirable output, the production possibility set, $PPS(x)$, can be formally defined by considering a dataset with K DMUs and for DMU^k letting $x^k = (x_1^k, \dots, x_M^k) \in \Re_+^M$ denote the input vector, $v^k = (v_1^k, \dots, v_N^k) \in \Re_+^N$ denote the output vector of the desirable outputs and $w^k = (w_1^k, \dots, w_J^k) \in \Re_+^J$ denote the output vector of undesirable outputs. The production possibility set is thereby constructed according to B.2-1.

$$
\begin{aligned}
PPS(x) = \{v, w :\ & \sum_{k=1}^{K} \lambda^k v_m^k \geq v_m^{k'}\ , && m = 1, \dots, M \\
& \sum_{k=1}^{K} \lambda^k w_j^k = w_j^{k'}, && j = 1, \dots, J \\
& \sum_{k=1}^{K} \lambda^k x_n^k \leq x_n^{k'}, && n = 1, \dots, N \\
& \lambda^k \geq 0, && k = 1, \dots, K\ \}
\end{aligned} \qquad \text{B.2-1}
$$

From B.2-1 it can be seen that the undesirable outputs are treated differently from the regular outputs as the undesirable outputs exhibit weak rather than strong disposability. This leads to the change in the frontier as seen in Figure B.2-1. The undesirable output is in this formulation determined directly by the convex combination of the desirable outputs and input quantities. As it is not possible to reduce the undesirable output without also reducing the desirable, the undesirable output is in the creation of the production possibility set, constrained to follow the convex combination given by the intensity vectors $\lambda's$ directly by having $\sum_{k=1}^{K} \lambda^k w_j^k = w_j^{k'}$.

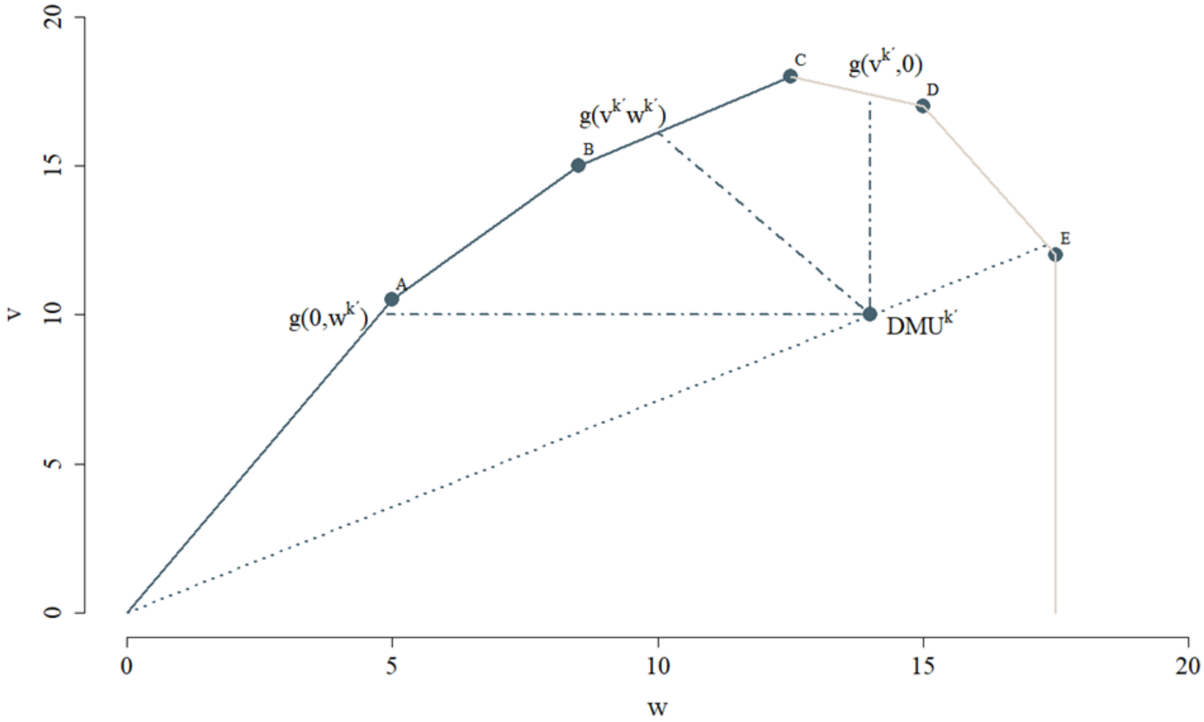
In standard DEA, having two *desirable* outputs, the dominating DMUs would be located towards the north-east in the production possibility set. An implication of including an undesirable output is that a part of the frontier is no longer dominating for the rest of the DMUs. This is the case as it is now preferable to be located towards the north-west of the production possibility set with a higher amount of the desirable output, and less of the undesirable output. The inefficient part of the frontier is illustrated by the light part of the frontier in Figure B.2-1.
The inefficient part of the frontier is no longer superior to locations horizontally to the left within the production possibility set. A location horizontally to the left of the inefficient part of the frontier indicate the same amount of the desirable output but with less of the undesirable output, which is preferred in this case.
The interpretation of a projection of DMUs to the inefficient part of the frontier is thereby questionable, as the DMU is no longer benchmarked against best practice.

B.2.3 Directional distance functions

Modelling an undesirable output with the use of the original variable directly within the benchmarking model, requires the use of a directional distance function (DDF). This is necessary in order to steer the benchmarking in a direction where desirable outputs are maximized and/or undesirable outputs are minimized.
The use of a standard distance function from DEA implies minimizing all inputs or all outputs. However, with the introduction of an undesirable output, the distance function will have to be able to cope with an output that needs to be minimized instead of maximized.

Figure B.2-2 illustrates how the projection towards the frontier (the benchmark) changes when adjusting the directional distance function.

Figure B.2-2: Illustration of the directional distance function



If both v and w are assumed to be desirable outputs, the standard DEA radial distance function for $DMU^{k'}$ follows the dotted line from the origin through $DMU^{k'}$ and continues towards the benchmark at the frontier. This implies that more of both outputs is preferred and the $DMU^{k'}$ is thereby projected north-east onto the frontier. However, as w in this case is an undesirable output, the distance function must be specified towards the north-west in order to ensure a desirable benchmark for each DMU.

The directional distance function was firstly introduced by Chambers et al. (1998), and is widely used to cope with models where it is necessary to specify a direction towards the frontier (see e.g. Lim et al. (2019)). Elaborating on the definition from Chambers et al. (1998), the output-oriented directional distance function with both desirable and undesirable outputs can be formulated as follows:

$$\vec{D}_T(x, v, w, g) = \max \{ \beta \in \Re : (x, v + \beta g_v, w - \beta g_w) \in PPS(x) \qquad \text{B.2-2}$$

The directional vector $g = (g_v, g_w)$ determines the specific direction of the distance function. In this example, input is held constant whereas the desirable output is maximized while the undesirable output is minimized proportionally. This can be seen as the vector g_w is in fact subtracted from the bad output, w.

The directional distance function can be specified in numerous ways. As illustrated on Figure B.2-2, the specification of the directional vector, g, determines the direction in which each DMU is projected on to the frontier. For the directional distance function to cope with the characteristics of both the desirable and undesirable output, it is possible to specify the direction anywhere north-west to each DMU in order to ensure that w is minimized, and v is maximized.

One extreme specification of the directional distance function for $DMU^{k\acute{}}$ is $g(0, w^{k\prime})$, indicating that only the undesirable output is minimized. This specification is equivalent to projecting $DMU^{k\acute{}}$ horizontally to the left as illustrated in Figure B.2-2.
As the inefficient part of the frontier is always located at the right side of the production possibility frontier (illustrated by the light line in Figure B.2-2) a directional distance function which only minimize the undesirable output will always ensure that DMUs are always projected onto the strongly efficient part of the frontier.

The opposite extreme is specifying a directional distance function where only the desirable output is maximized, given by a directional vector of $g(v^{k\acute{}}, 0)$. This direction is illustrated in Figure B.2-2 where $DMU^{k'}$ is projected vertically upwards towards the frontier.

In this specific example, with the particular directional vector of $g(v^{k\acute{}}, 0)$, $DMU^{k\acute{}}$ is projected onto the so-called inefficient part of the frontier, where the benchmark itself is questionable. Whether DMUs are projected on to the inefficient part of the frontier will depend on the specific model and thereby the shape of the frontier as well as the location of the given DMU. This characteristic of the model is a relatively large drawback of the method. Nevertheless, it is possible to either exclude the specific DMUs which are projected to the inefficient part of the frontier, or alternatively specify a direction which ensures that $DMU's$ are only projected onto the strongly efficient part of the frontier.

It is possible to defines the DDF such that both the desirable output is maximized, and the undesirable output is minimized. An example could be to use a directional vector such as $g(v^{k'}, w^{k'})$, leading to a benchmark somewhere in between the two extremes.

Equation B.2-3 defines for DMU$^{k'}$= $(v^{k'}, w^{k'}, x^{k'})$ the LP problem combining a production possibility set assuming weak disposability of the undesirable output(s) with a radial distance function. The model imposes CRS and is output-oriented with a directional vector defined by $g = (g_v^{k'}, g_w^{k'}) \in \Re_0^N, \Re_0^J$.

$$
\begin{aligned}
&\max_{\lambda^k, \beta} \beta \\
s.t. \quad &\sum_{k=1}^{K} \lambda^k v_m^k \geq v_m^{k'} + \beta g_{v,m}^{k'} \quad , m = 1, \ldots, M \\
&\sum_{k=1}^{K} \lambda^k w_j^k = w_j^{k'} - \beta g_{w,j}^{k'} \quad , j = 1, \ldots, J \\
&\sum_{k=1}^{K} \lambda^k x_n^k \leq x_n^{k'} \quad , n = 1, \ldots, N \\
&\lambda^k \geq 0 \; k = 1, \ldots, K
\end{aligned}
\qquad \text{B.2-3}
$$

β represents the inefficiency term. For an output oriented model, keeping input fixed, the optimal β value for each DMU represents the potential reduction in the undesirable output corresponding directly to the potential increase of the desirable output, which is a consequence of the model being radial (Bogetoft & Otto, 2011, p. 33).

The β value can both be interpreted as an *excess* or *shortage* function. It corresponds to the number of times the directional vector g has been produced in *excess* of what is necessary of the undesirable output and in *shortage* of what could have been produced of the desirable output.

When specifying the directional vector, it is important to notice that the interpretation of the $\beta's$ across DMUs are dependent on both the direction and length of the specific directional vector. If e.g. the directional vector is doubled in length, the relative potential in form of the $\beta's$ is halved

(Bogetoft & Otto, 2011, p. 33). However, in the specific case where the directional vector, g, is specified using the input and output vectors of DMU' itself ($g = \left(v^{k'}, w^{k'}\right)$), the excess/shortage function, β, measures (relative) inefficiency similar to that known from the Farrell method (Bogetoft & Otto, 2011, p. 33).
A β value of 0,2 thereby implies that the given DMU is able to reduce its production of undesirable outputs by 20 pct. while increasing its desirable outputs by 20 pct. when comparing the DMU to best practice.
The βs are thereby measures of *inefficiency.* The higher the value of β for any given DMU, the more improvement potential is found for the DMU in the model, indicating a higher level of inefficiency.

In order to obtain an *efficiency* score it is therefore necessary to transform the β measure of *inefficiency*. In the literature of output oriented models with directional distance functions, efficiency scores are typically reported using the Shephard output distance function[2] (see e.g. Yuan et al. (2011) or Chung et al. (1997)). The calculation of *efficiency* scores for desirable output is shown equation B.2-4.

$$eff_v = \frac{1}{1+\beta} \qquad \text{B.2-4}$$

However, it is worth noting that the direct interpretation of this efficiency score is related to the desirable output. A β of 0,2 and a corresponding efficiency score of 0,83 implies that a DMU is currently producing 83 pct. of the obtainable desirable output defined by best practice.
A similar interpretation of the efficiency score, directly related to the undesirable output, is calculated by taking $eff_w = \frac{1}{1-\beta}$. With a β of 0,2 the DMU would thereby be producing 125 pct. of the optimal amount of the undesirable output.

[2] The notion of efficiency obtained from the Shephard output distance function corresponds to the inverse of that from the Farrell efficiency (Bogetoft and Otto 2011, 30).

B.2.4 Duality and frontier shadow prices

The LP program from Equation B.2-3 is presented as the primal problem, which in a DEA context is referred to as the envelopment space. This exact same program can also be expressed as the dual problem referred to as the multiplier space.
The duality of the LP problem i.e. the multiplier space, is essential to the concept of estimating shadow prices of an undesirable output. The primal and the dual formulation measures efficiency similar using different approaches.

The envelopment formulation of the LP problem is specified with the use of intensity vectors $\lambda's$. The primal formulation thereby contributes with knowledge regarding peers and is suitable for analysis of the efficient DMUs and specific targets for each DMU. The primal formulation can be illustrated by enveloping the production possibility set as seen in previous figures, hence the name envelopment space.

In multiplier space, the productivity vector of each DMU is maximized in order to measure efficiency. The productivity vector is defined by the weighted sum of outputs divided by the weighted sum of inputs. The point of interest in the dual is thereby maximizing the total value of the outputs divided by the total costs. The weights of each input and output are referred to as multipliers, hence the name multiplier space. When prices are known, these are obvious multipliers for each variable. However, with DEA and similar non-parametric methods, prices are assumed to be unknown or uncertain. Weights are therefore chosen such that the individual productivity ratio for each DMU is maximized. Each DMU is, through the LP problem, thereby given the benefit of the doubt when estimating the efficiency through weights (prices) of each output and input.

However, when providing each DMU with the benefit of the doubt by choosing weights without constraints, the productivity ratio will be chosen to be infinitely large. Without constraints, DMUs can e.g. choose to let the output weights be infinitely large or the input weights zero leading to an infinitely large productivity ratio. Therefore, the productivity ratio is maximized under the constraint that no other DMUs productivity ratio can exceed 1, indicating that no other DMU can have a net profit (Bogetoft & Otto, 2011, p. 135). As in the envelopment formulation, the efficient

DMUs get assigned a score of 1, as these can maximize their own productivity ratio to 1 without being limited to the constraint by other DMUs productivity ratio.

The individual weights of each DMU can be used to point out the performance within the different variables for each DMU. A relative high weight for a given variable, indicates that the DMU is performing relatively well within that certain area, as the DMU performs best when weighting this variable relatively high compared to other variables.

As variables can be measured in different units, as for GHG emissions and revenue, it can be useful to examine so called *virtuals* instead of the specific weights. Virtuals are defined by the product of the given input or output variable and the corresponding weight. The interpretation of the virtuals are thereby comparable across variables measured in different units as the virtuals represent the contribution of each variable to the DMUs estimated efficiency.

With multiple outputs and/or inputs, it is possible for a DMU to assign zero weights to a given variable, thereby excluding this variable when assessing the DMUs individual performance. Depending on the scope of the analysis, it may be necessary to include weight restrictions to limit the possibility of a DMU weighting a variable zero. Weight restrictions can e.g. also be imposed to restrict the relative weights between two variables. As pointed out by Podinovski (2004), introducing weight restrictions in multiplier space leads to changes in the trade-offs and thereby the shape of the frontier in envelopment space. For the empirical analysis of this report, weight restrictions are therefore not suitable, as the shape of the frontier is essential to the analysis of shadow prices.

Weights in multiplier space can be interpreted as trade-offs in envelopment space. Weights do not occur directly in the envelopment formulation, as is the case with the multiplier formulation. However, as these are each other's dual, weights are still meaningful when assessing each DMU in the envelopment space. The relative weights between two outputs (or inputs) represent the trade-off between the two outputs. When assessing a model in two dimensions as seen in e.g. Figure B.2-1 or Figure B.2-2, the trade-off graphically corresponds to the slope of the frontier. As can be seen from the figures, the slope and thereby the trade-off between the desirable and undesirable output, changes along the frontier. This implies that the frontier shadow prices change dependent on the specific projection point on the frontier for a given DMU.

Weights from the dual formulation of the LP problem can therefore be used to derive frontier shadow prices. The frontier shadow price for an undesirable output w is calculated according to equation B.2-5

$$p_w = p_v \frac{\pi_w}{\pi_v} \qquad \text{B.2-5}$$

In equation B.2-5, p_v is the price of the desirable output v, π_w is the weight of the undesirable output and π_v is the weight of the desirable output. In the empirical analysis, the desirable output is revenue. This implies that p_v=1 as the desirable output is measured in DKK.

The frontier shadow prices thereby represent the trade-off between GHG emissions and revenue at the frontier i.e. the cost of CO_2e given the best available technology. The shadow price of CO_2e is therefore a marginal abatement cost for an efficient DMU, indicating the cost of lost revenue when reducing the CO_2e emissions with one unit.

The method to estimate a model for calculating shadow prices from an undesirable output using a directional distance function and weak disposability of the undesirable output is originally developed by Färe et al. (1993).
Färe et al. (1993) utilizes the characteristics of the trade-offs to derive firm specific shadow prices of pollution from the Canadian pulp and paper industry, and the method has since then been applied by a range of scholars estimating shadow prices (see e.g. Harkness (2006) and Fukuyama (2008)).

The dual formulation of the weak disposability technology using a directional distance function is still widely discussed. The method is still developing and there exists a body of literature which seeks to formulate the dual program of the envelopment formulation presented in equation B.2-3 (see e.g. Kuosmanen et al. (2010) , Leleu (2012), Leleu et al. (2016)).
Leleu et al. (2016) provide a dual formulation which is in line with the models presented in this empirical analysis. In order to deduct the dual formulation of the problem presented in equation B.2-3, the problem is rewritten to equation B.2-6.

The rewriting is done by defining a new variable $\sigma = 1 - \sum_{k=1}^{K} \lambda^k$, which implies that $1 = \sum_{k=1}^{K} \lambda^k + \sigma$. Using this, $v_m^{k'}$ can be rewritten so that $v_m^{k'} = \sum_{k=1}^{K} \lambda^k v_m^{k'} + \sigma v_m^{k'}$, and similar for the inputs $\left(x_n^{k'}\right)$ and undesirable outputs $\left(w_j^{k'}\right)$ for the DMU under observation.

The constraint $\sum_{k=1}^{K} \lambda^k v_m^k \geq v_m^{k'} + \beta g_{v,m}^{k'}$ from B.2-3 can thereby be defined as $\sum_{k=1}^{K} \lambda^k v_m^k \geq \sum_{k=1}^{K} \lambda^k v_m^{k'} + \sigma v_m^{k'} + \beta g_{v,m}^{k'}$ which can be rearranged to the constraint showed in B.2-7.

The dual variables for the deduction are presented together with the constraints in the envelopment formulation in B.2-6. The color coding in B.2-6 and B.2-7 serve to illustrate the relationship between the primal and the dual problem. As the program is output oriented, it is a maximization problem in envelopment space, and a minimization problem in multiplier space.

The primal formulation

$$\max_{\lambda^k, \beta} \beta$$

$$s.t.$$

$$-\sum_{k=1}^{K} \lambda^k (v_m^k - v_m^{k'}) + \sigma v_m^{k'} \leq -\beta g_{v,m}^{k'} \qquad \pi_{v,m}$$

$$-\sum_{k=1}^{K} \lambda^k (w_j^k - w_j^{k'}) + \sigma w_j^{k'} = \beta g_{w,j}^{k'} \qquad \pi_{w,j}$$

$$\sum_{k=1}^{K} \lambda^k (x_n^k - x_n^{k'}) - \sigma x_n^{k'} \leq 0 \qquad \pi_{x,n} \qquad \text{B.2-6}$$

$$\sum_{k=1}^{K} \lambda^k + \sigma = 1 \qquad \gamma$$

$$\lambda^k \geq 0$$

$$\sigma \geq 0$$

The dual formulation

$$\min_{\pi^v, \pi^w, \pi^x, \gamma} \gamma$$

$$s.t.$$

$$\left(\sum_{m=1}^{M} \pi_{v,m}\, v_m^k + \sum_{j=1}^{J} \pi_{w,j}\, w_j^k - \sum_{n=1}^{N} \pi_{x,n}\, x_n^k\right) - \left(\sum_{m=1}^{M} \pi_{v,m}\, v_m^{k'} + \sum_{j=1}^{J} \pi_{w,j}\, w_j^{k'} - \sum_{n=1}^{N} \pi_{x,n}\, x_n^{k'}\right) - \gamma \leq 0$$

$$-\sum_{m=1}^{M} \pi_{v,m} g_{v,m}^{k'} + \sum_{j=1}^{J} \pi_{w,j} g_{w,j}^{k'} - 1 = 0 \qquad \text{B.2-7}$$

$$\sum_{m=1}^{M} \pi_{v,m}\, v_m^{k'} + \sum_{j=1}^{J} \pi_{w,j}\, w_j^{k'} - \sum_{n=1}^{N} \pi_{x,n}\, x_n^{k'} + \gamma \geq 0$$

$$\pi_{v,m} \geq 0$$

$$\pi_{w,j} \leq 0$$

$$\pi_{x,n} \geq 0$$

In the dual formulation presented in equation B.2-7, γ represents the shadow profit inefficiency and thereby equals zero for efficient DMUs and is positive for inefficient DMUs. $\pi_{v,m}$ represents the weight of the m'th desirable output, $\pi_{w,j}$ represents the weight of the j'th undesirable output and $\pi_{x,n}$ represents the weight of the n'th input.

The term $\sum_{m=1}^{M} \pi_{v,m}\, v_m^{k'} + \sum_{j=1}^{J} \pi_{w,j}\, w_j^{k'} - \sum_{n=1}^{N} \pi_{x,n}\, x_n^{k'}$ represents the shadow profit for the DMU under observation, whereas $\sum_{m=1}^{M} \pi_{v,m}\, v_m^{k'} + \sum_{j=1}^{J} \pi_{w,j}\, w_j^{k'} - \sum_{n=1}^{N} \pi_{x,n}\, x_n^{k'} + \gamma$ represents the shadow profit for the DMU under observation's benchmark at the frontier. Therefore, the first condition implies that if the DMU is efficient, these two shadow profits will be equal to one another forcing $\gamma = 0$.

With an arbitrary set of weights chosen for a DMU, the difference between the obtainable profit at the frontier and the DMU's current profit is equal to the shadow profit inefficiency (γ). The aim of the program is thereby finding the optimal set of shadow prices (π^v, π^w, π^x) in order to minimize the shadow profit inefficiency for each DMU (Leleu, Mixing DEA and FDH models together, 2009).

In B.2-7 the constraint $\sum_{m=1}^{M} \pi_m^v \, v_m^{k'} + \sum_{j=1}^{J} \pi_j^w \, w_j^{k'} - \sum_{n=1}^{N} \pi_n^x \, x_n^{k'} + \gamma = 0$ serves as a "*zero shadow profit condition*" imposed under CRS. All fully efficient DMUs under CRS operate at the most optimal scale with the highest output/input ratio and thereby highest possible profit. The shadow profit at the frontier must therefore be zero, as moving along the frontier will not increase shadow profit. Under VRS the term $\sum_{m=1}^{M} \pi_{v,m} \, v_m^{k'} + \sum_{j=1}^{J} \pi_{w,j} \, w_j^{k'} - \sum_{n=1}^{N} \pi_{x,n} \, x_n^{k'} + \gamma$ is unconstrained as it allows for differences in economies of scale.

In the formulation in equation B.2-7 proposed by Leleu, et al. (2016) the weight of the undesirable output ($\pi_{w,j}$) is constrained to be non-positive. It is in the referenced literature discussed whether $\pi_{w,j}$ should be unconstrained or must be constrained to be non-positive. Whether weights of undesirable outputs are positive or negative in multiplier space can be interpreted as whether the slope of the frontier in envelopment space is allowed to be both positive and negative. i.e. whether negative shadow prices for the undesirable output should be possible. When the weight of the undesirable output is constrained to be non-positive, this implies that negative shadow revenue is generated from the undesirable output. Positive weights will, on the contrary, indicate that undesirable outputs are able to generate revenue in the same manner as is the case for desirable outputs. Having $\pi_{w,j}$ unconstrained is for this reason criticized in both Leleu et al. (2016), Hailu et al. (2001) and Hailu (2003).

Due to limitations of the software used to compute the empirical analysis in this report, the weight for the undesirable output is not constrained to be non-positive. In the formulation used to compute the models in this empirical analysis, the constraint $\pi_{w,j} \leq 0$ presented in equation B.2-7 is thereby not present. It should nevertheless be noted that observations for which $\pi_{w,j}$ is positive (corresponding to the DMUs which are projected onto the inefficient part of the frontier) are removed from the presentation of the results of the empirical analysis in section B.4.

The problem used in the following empirical analysis in multiplier space for a CRS model with weak disposability for the undesirable output and a directional distance function can be presented as follows from equation B.2-8.

$$
\begin{aligned}
&\min_{\pi^v, \pi^w, \pi^x, \gamma} \gamma \\
&s.t. \\
&\left(\sum_{m=1}^{M} \pi_{v,m}\, v_m^k + \sum_{j=1}^{J} \pi_{w,j}\, w_j^k - \sum_{n=1}^{N} \pi_{x,n}\, x_n^k\right) - \left(\sum_{m=1}^{M} \pi_{v,m}\, v_m^{k'} + \sum_{j=1}^{J} \pi_{w,j}\, w_j^{k'} - \sum_{n=1}^{N} \pi_{x,n}\, x_n^{k'}\right) \leq \gamma \\
&\sum_{m=1}^{M} \pi_{v,m} g_{v,m}^{k'} - \sum_{j=1}^{J} \pi_{w,j} g_{w,j}^{k'} = 1 \\
&\sum_{m=1}^{M} \pi_{v,m}\, v_m^{k'} + \sum_{j=1}^{J} \pi_{w,j}\, w_j^{k'} - \sum_{n=1}^{N} \pi_{x,n}\, x_n^{k'} + \gamma = 0 \\
&\pi_{v,m} \geq 0 \\
&\pi_{w,j}\ unconstrained \\
&\pi_{x,n} \geq 0
\end{aligned}
\qquad \text{B.2-8}
$$

For an output-oriented model, the program is a maximization problem in envelopment space, and a minimization problem in multiplier space. For such models in multiplier space, the sum of output virtuals are held equal by restricting these to a constant (often 1), while the sum of the input virtuals are minimized. This is the normalization constraint.

With $g = \left(v_m^{k'}, w_j^{k'}\right)$ the normalization can be rewritten as: $\sum_{m=1}^{M} \pi_{v,m} v_m^{k'} - \sum_{j=1}^{J} \pi_{w,j} w_j^{k'} = 1$, which is in line with the normalization of a standard output oriented DEA problem where the sum of the output virtuals is 1, given that the weights on the undesirable output should ideally be non-positive.

B.2.5 Second stage analysis

The benchmarking model presented in the following empirical analysis contains various information. In addition to information regarding the trade-off between GHG emissions and revenue, the benchmarking analysis of dairy farms also provide information as to which farms constitute best practice within the sector. In the following empirical analysis, a second stage analysis of the inefficiency term (β) is conducted in order to extract knowledge of the characteristics from the efficient farms. This information can ultimately help inefficient farm as how to manage their production in order to reach the frontier. The second stage analysis can provide information on the optimal structure of a farm given the characteristics of efficient farms. Information regarding what characterizes the efficient farms could also be used in future regulation of the dairy farms. The information could be used to legislate such that inefficient farms are steered towards what define best practice within the sector today. This could thereby be a way of catching up with the current technological lag found within the sector.

There exist different methods for second stage analyses. The inefficiency term β is the dependent variable of this second stage efficiency analysis. As this variable is limited to a value between 0-1, the dependent variable of the second stage analysis is censored around 0. It is thereby suitable to use a Tobit regression. The Tobit regression allows for regressing β on multiple variables.

A Tobit regression should be used when the dependent variable is censored as is the case with an (in)efficiency score (Bogetoft & Otto, 2011, p. 189). The range of an (in)efficiency score varies between the left censored value 0 to the right censored value 1, and the Tobit regression is thereby often used in the literature to conduct second stage analyses for DEA models (see e.g. Ahmad et al. (2017), McDonald (2008), Saglam, (2018) and Liu et al. (2017)).

The Tobit regression with the inefficiency score having a range between 0-1 can formally be expressed by equation B.2-9 following a modified example from McDonald (2008):

$$\beta_i^* = xb_i + e_i, e_i|x_i \sim Normal(0, \sigma^2)$$

$$\begin{gathered} if\ \beta_i^* \leq 0,\ \ \beta_i = 0 \\ if\ 0 < \beta_i^* < 1,\ \ \beta_i = \beta_i{}^* \\ if\ \beta_i^* \geq 1,\ \ \beta_i = 1 \end{gathered} \qquad \text{B.2-9}$$

In equation B.2-9, β_i is the observed inefficiency scores which are the censored values of the latent variable β_i^*. The latent response variable is estimated by the explanatory variables x and an error term which is normally distributed with a mean value of 0 and variance of σ^2. The latent variable for inefficiency can be interpreted as an underlying measure of inefficiency which is not restricted to certain values by the implications of the estimation in the benchmarking model.

The interpretation of the b_i's are similar to the interpretation of OLS, where the coefficients (b) can be interpreted as the partial effect on the latent variable. The indicator function from equation B.2-9, converts the latent variable into the response variable (in this case the inefficiency score). This conversion is non-linear (Wooldridge, 2016, pp. 536-547) implying that the coefficients from the Tobit estimation cannot be interpreted directly on the response variable (β_i). Nevertheless, the sign of the coefficients can be used to see in which direction the explanatory variable is affecting the response variable.

The partial effects on the response variable in a Tobit regression are dependent on the specific values of the explanatory values and can therefore only be calculated given the specific values of the explanatory variables. However, it is possible to calculate some general partial effects of the response variable such as the average partial effect (APE) or partial effect of the average (PEA) (Wooldridge, 2016, pp. 536-547).

For the empirical analysis the average partial effects are calculated and presented for the Tobit models in the second stage analysis. In the following second stage analysis the average partial effects are calculated. The average partial effects are calculated for expected values of the depend variable greater than 0. In this thesis this implies that the regression of the inefficiency score is only conducted for inefficient farms, as efficient farms have an inefficiency score of 0.

The calculation of APE for a given explanatory variable varies dependent on whether the explanatory variable is continuous or discrete. The partial effects for continuous variables are calculated following equation B.2-10 (Wooldridge, 2016, pp. 536-547). For discrete variables, such as a binary variable the calculation is shown in equation B.2-11 (Wooldridge, 2016, pp. 536-547). Estimating APE concerns firstly estimating the partial effects and thereafter taking the mean of these to obtain the *average* partial effect for both continuous and discrete variables.

In equation B.2-10 and B.2-11, b_i represents the coefficient from the Tobit regression for variable i. Φ is the normal probability density function, ϕ is the normal cumulative density function. $\boldsymbol{X}\mathbf{b}$ is the predicted latent variable from the Tobit model and σ is the standard deviation of the error term for the latent variable.

$$\frac{\partial E(y|y>0)}{\partial x_i} = b_i\left(1-\left(\frac{\phi\left(\frac{\mathbf{Xb}}{\sigma}\right)}{\Phi\left(\frac{\boldsymbol{X}\mathbf{b}}{\sigma}\right)}\right)\left(\frac{\mathbf{Xb}}{\sigma}+\frac{\phi\left(\frac{\mathbf{Xb}}{\sigma}\right)}{\Phi\left(\frac{\boldsymbol{X}\mathbf{b}}{\sigma}\right)}\right)\right) \qquad \text{B.2-10}$$

$$E(y|y>0) = \mathbf{Xb} + \sigma\left(\frac{\phi\left(\frac{\mathbf{Xb}}{\sigma}\right)}{\Phi\left(\frac{\boldsymbol{X}\mathbf{b}}{\sigma}\right)}\right) \qquad \text{B.2-11}$$

The partial effects for a dummy variable are calculated by estimating the partial effects twice according to equation B.2-11. First it is necessary to estimate the predicted values having the dummy variable take the value 0 and again having the dummy variable take the value 1. The partial effects for the two scenarios are hereafter subtracted from one another to obtain the partial effect of going from a dummy value of 0 to 1. The average partial effect of this dummy variable is hereafter calculated as the average of the partial effects.

B.3 Empirical benchmarking analysis

The empirical analysis seeks to measure the efficiency potential within Danish dairy farms in relation to both revenue and GHG emissions, and to estimate abatement costs of reducing GHG emissions. The analysis will be conducted by benchmarking dairy farms using a non-parametric efficiency analysis with a technology set exposing weak disposability of the undesirable output, GHG emissions, and using an output oriented directional distance function.

The analysis initially explores the differences in organic versus conventional farming. As a consequence of this analysis, the remaining empirical analysis is conducted by treating organic and conventional farms in two separate benchmarking models.

Especially, the frontier shadow prices of GHG emissions, which will be estimated, are relatively sensitive to the specific direction of the distance function, as mentioned in section B.2.4. Therefore, three different directional vectors have been used in both the conventional and organic benchmarking, to illustrate the impact of these to the results. By applying the three different directional vectors, it is possible to examine the variation in frontier shadow prices of GHG emissions as well as calculating average opportunity costs by utilizing the differences across the models.

The six models shown in Table B.3-1will be presented throughout the empirical analysis. The models cover different choices of the directional vector g. The GHG models (*"GHG conventional"* and *"GHG organic"*) are estimated with a directional vector which only seeks to minimize GHG emissions for each farm keeping both input (total costs) and revenue fixed. The revenue models (*"Revenue conventional"* and *"Revenue organic"*) are estimated with a directional vector which only seeks to maximize revenue, keeping both input and GHG emissions fixed. The mixed models (*"Mix conventional"* and *"Mix organic"*) are estimated with a directional vector where both GHG emissions are minimized while revenue is maximized simultaneously, still keeping input fixed. In the mixed models, the reduction of GHG emissions and the increase in revenue is proportional to each other, as a radial distance function is used.

Table B.3-1: Overview of benchmarking models estimated in the empirical analysis

		Farming Type	
		Conventional	**Organic**
Orientation	**Radial distance function, $g(v^k,w^k)$**	*Mix conventional*	*Mix organic*
	Only minimizing GHG emissions $g(0,w^k)$	*GHG conventional*	*GHG organic*
	Only maximizing revenue $g(v^k,0)$	*Revenue conventional*	*Revenue organic*

B.3.1 Input and output variables

Table B.3-2 displays the composition of the three variables used in this empirical analysis: **Total revenue** (desirable output (v)), **GHG emissions** (undesirable output (w)) and **Total costs** (input (x)). The variables for revenue and total costs are constructed based on Lillethorup (2017).

Table B.3-2: Overview of sub elements included in the input and output variables

		Mean	**Min**	**Max**	**St. Dev.**
Inputs	Feed (DKK)	1.930.622	225.914	14.220.617	1.466.139
	Total labor cost (DKK)	1.245.177	325.044	7.543.761	727.992
	Other variable costs (DKK)	1.851.062	475.695	8.991.549	977.783
	Fixed costs (DKK)	1.673.815	395.369	8.673.713	939.089
	Capital costs (DKK)	1.405.580	111.572	6.290.664	736.326
	Total costs (DKK)	**8.106.255**	**2.564.402**	**38.007.383**	**4.382.512**
Outputs	Revenue from milk (DKK)	6.826.507	1.893.486	38.301.378	4.166.906
	Revenue from Other outputs (DKK)	1.882.169	197.822	11.457.263	1.221.275
	Total Revenue (DKK)	**8.708.676**	**2.825.628**	**45.158.160**	**5.093.578**
	GHG emissions (tons CO_2e)	**1.440**	**481**	**8.456**	**866**

The output variable *revenue* is the sum of revenue from milk and other outputs constructed by Lillethorup (2017).

The input variable *total costs* in this empirical analysis is the sum of the five inputs: feed, total labor costs, other variable costs, and capital costs also constructed by Lillethorup (2017).

The variable cost to feed is defined by Lillethorup (2017) and covers the costs of buying grains and fodder. Labor costs are estimated as: "*the value of family labor plus paid labor"*, other variable costs cover costs to *"fuel, fertilizer, veterinary costs etc."* Fixed costs are estimated to "*include various costs for maintenance, taxes, insurances etc."* Capital costs consist of *"various costs for maintenance, taxes, insurances etc."* (Lillethorup, 2017). The undesirable output, GHG emissions, is the variable estimated in Part A of this thesis.

In Table B.3-2 it can be seen that the different sub elements of the input variable contribute on average relatively equally to the aggregated input variable "total costs". For the output variable "total revenue", the revenue from milk represents the largest share, which is preferable, as the main focus of the empirical analysis is the climate impact of dairy production in particular.

Total costs and total revenue are used as aggregated measures as the sub elements of these variables are all measured in the same unit (DKK). Therefore, there is not necessarily a need for using the sub elements separately in the benchmarking model. However, Asmild (2019) argues that it could be preferable to estimate variable and fixed costs as separate inputs, as fixed costs are long term investments which can be bound in soils, buildings etc. It might therefore not be possible to reduce these as easily as the variable costs which represents short term investments such as labor and feed. Furthermore, the capital costs are calculated as 4 pct.[3]of the capital stock consisting of e.g. buildings, machinery and land, indicating that the costs of having these are measured as the opportunity costs of not having invested the capital. This measure might be uncertain due to the difficult choice of choosing the right interest rate, and the cost is not an incurred cost in the same way as the remaining costs. Therefore, it might make sense to split the inputs up as the capital costs are not directly comparable to the remaining costs. However, as the model estimated in the empirical analysis is not straight forward to use and interpret, and as the R package used for the construction of the model is relatively new and unused, it is convenient to only include three variables in the model. This makes it more straightforward to graphically illustrate the results and to examine whether the model behave as expected.

[3]The ministry of finance suggests using an interest of 4 pct. in the first 35 years when evaluating projects. The interest rate consists of the risk-free real interest rate corrected for risks that are not systematic or diverse (The Danish Ministry of finance 2017).

B.3.2 Comparing production conditions between organic and conventional farms

A basic assumption of the benchmarking model is that DMUs are in general comparable. This indicate that DMUs must operate under the same framework conditions and thereby have the same opportunities within the area of analysis. As this empirical analysis only covers specialized dairy farms, and as the data cleaning has sought to create a relatively homogenized dataset in relation to the relevant variables, the farms are assumed to be comparable across the dataset. However, it might not be the case that organic and conventional farms are directly comparable and it might not be suitable to model these types of farms within the same benchmarking model.

Figure B.3-1 illustrates how organic and conventional models are located in the production possibility set. To be able to illustrate the three variables in a two-dimensional plot, the two outputs: revenue and GHG emissions have been divided with the input: total costs.
The plot indicates that organic farms have a tendency to be located towards the left, with a lower GHG emission/Costs ratio than conventional farms, but with approximately the same variance in the Revenue/Costs ratio.

Figure B.3-1: The production technology for organic and conventional farms

The sample of conventional farms includes a total of 1.049 farms, whereas the sample for organic farms include 204 farms. To further investigate the determination of the differences between conventional and organic farms, and to ensure that it is not driven by the estimation of GHG emissions from Part A, the two production types are compared across a range of variables as seen in Table B.3-3. The variables chosen for comparison are specified "per cattle" or "per liter of milk" in order to eliminate any structural differences in scale between the two production types.

Table B.3-3: Comparison of the two production types: conventional and organic farms

	Mean		**Standard deviation**		***P-value from wilcoxon test***
	Organic	***Conventional***	***Organic***	***Conventional***	
Share of Jersey cattle	0,11	0,10	0,30	0,29	*6,08E-01*
GHG from enteric fermentation per cattle (Tons)	4,62	4,91	0,42	0,45	*3,04E-20****
GHG from manure management per cattle (Tons)	1,34	1,30	0,52	0,38	*6,91E-01*
Total GHG per cattle (Tons)	5,96	6,20	0,66	0,60	*2,95E-12****
Total GHG per liter of milk (Tons)	37,93	41,57	18,21	24,41	*1,46E-01*
Revenue from milk per cattle (DKK)	32.953,94	28.357,46	4.021,57	2.740,80	*1,03E-52****
Revenue from Other outputs per cattle (DKK)	9.268,83	8.325,73	3.365,12	3.945,09	*1,37E-08****
Costs per cattle (DKK)	33.416,14	29.903,01	4.798,35	4.398,19	*5,50E-24****
Note:			*p<0,1** p<0,05 ***p<0,01		

The significance in the difference for the relevant variables across the two production types are tested using a Wilcoxon signed-rank test. From Table B.3-3 it can be seen that the two production types are shown to be significantly different on a range of parameters. The total GHG emissions per cattle are on average lower for organic farms than for the conventional farms. This seem to be driven by the fact that the specific GHG emissions stemming from enteric fermentation, which contributes with the majority of the total GHG emissions, is lower for the organic farms.

Overall, both the absolute difference in GHG emissions from enteric fermentation and total GHG emissions between organic and conventional farms are relatively small. The average GHG emission stemming from manure management is on average slightly higher for organic farms in the dataset, which can be due to the fact that more organic farms have a tendency to have deep litter stables, which are associated with significantly higher emission factors than other stable systems. Based on the Wilcoxon test, the difference is not significant. However, the p-value for the test is not highly insignificant with a p-value of 0,15.

There are no structural differences in how the GHG emission variable is estimated for organic versus conventional farms. However, it is still worth noticing that there might be factors, between organic and conventional farms, that could alter the difference in the GHG estimation across the two production types. This could be structural differences in feed composition and the number of days each cattle spend on grass which the estimation of GHG emissions on farm level does not capture.

The organic farms have significantly higher revenue from both milk and other outputs per cattle, but also higher total cost per cattle. This indicate that organic farms generally produce more cost intensively but at the same time manage to achieve a relatively higher level of revenue. This implies that it is not only the GHG emissions which drives the difference between the two types of production, but that it is also affected by different cost/revenue structures.

Other benchmarking analyses have included both organic and conventional Danish dairy farms within the same benchmarking model (see e.g. Asmild (2019)). An argument for keeping organic and conventional dairy farms in the same benchmarking model could be that if organic farms are in general shown to be relatively more efficient, then there should be an incentive for other farms to restructure their farms to an organic production. However, there might not be a socioeconomic interest in having a sector consisting of either only organic or conventional farms. Because of the natural constraint in land, it might not be possible to convert all farms to organic farms with the existing technology of today. This is especially because organic livestock require a relatively larger area for feed production relative to conventional farms (Paalberg, 2013).

Furthermore, keeping both conventional and organic farms in the same model implies that convex combinations of these are feasible, which in reality might not be the case. Both required standards and consumer preferences vary across the two types of production, and a convex combination of these is thereby hard to interpret.

For the remaining analysis organic and conventional farms are therefore handled in separate models.

B.3.3 The benchmarking model

A weak disposability model with a directional distance function assuming constant returns to scale (CRS) is used in the empirical analysis. CRS is assumed as it gives incentives to operate at the optimal scale. With a competitive market, such as that for dairy, each farm should be operating at the optimal scale and should thereby not be compensated in terms of efficiency for operating at a non-optimal scale. Furthermore, the VRS formulation of this specific model is not yet well-defined in the literature (Leleu, 2013) and could thereby potentially be estimated incorrectly.

The model for the empirical analysis is specified as illustrated in B.3-1, following the notation from section B.2.3.

$$
\begin{aligned}
&\max_{\lambda^k,\beta} \beta \\
&s.t. \\
&\sum_{k=1}^{K} \lambda^k Revenue^k \geq Revenue^{k'} + \beta g^{k'}_{Revenue} \\
&\sum_{k=1}^{K} \lambda^k GHG^k = GHG^{k'} - \beta g^{k'}_{GHG} \\
&\sum_{k=1}^{K} \lambda^k Total\ cost^k \leq Total\ cost^{k'} - \beta g^{k'}_{Total\ cost} \\
&\lambda^k \geq 0 \quad k = 1, \dots, K
\end{aligned}
\qquad \text{B.3-1}
$$

For the efficiency analysis, the output-oriented radial vector is defined using the farms' own output-quantities. The directional vector takes three different forms as illustrated in Table B.3-1.

In the following it is explored whether any of the efficient farms from the models that should be treated as outliers.

B.3.3.1 Outlier detection of the model

It has been pointed out by several researchers that one of the weaknesses of the non parametric framework Data Envelopment Analysis is that it is very sensitiv to outliers see e.g. Khezrimotlagh (2013). Various of tools have been developed to identify outliers in a DEA set-up. The most widespread method to identify outliers is the so-called super efficiency measures proposed by Bankers and Chang (2005). However, these measures do not have a straightforward interpretation for the weak disposability model with the directional distance function and are not easily implemented in the available software. Therefore, for this empirical analysis the efficient units which constitute the frontier, have been excluded from the model one by one to examine how they affect both the average inefficiency scores and the average frontier shadow prices i.e. the shape of the frontier.

As the dataset has already been cleaned with the purpose of obtaining a homogenized sample, it should not contain any outliers. However, as efficiency analyses are very sensitive to outliers, an outlier detection of the model itself has been carried out.
To ensure that observations are not wrongly identified as outliers, four criteria for being an outlier in the model has been defined:

1) *When removing the observation from the model, it should have a relatively large impact on the mean inefficiency ($\beta's$) of the model.*
2) *When removing the observation, it should have a relatively large impact on the mean frontier shadow price.*
3) *The observation removed should be a peer for a relatively large number of observations.*
4) *It should be possible to see graphically that the observation might be an outlier.*

If a farm affects the model within all four criteria, the farm is assessed to have a too large impact on the overall results of the model and is thereby categorized as an outlier and excluded from the final model estimation. The outlier identification is in this analysis a rather restrictive process, where the farm under analysis must influence the model to a relative great extent, as the farms following the data cleaning should be comparable and relatively homogeneous.

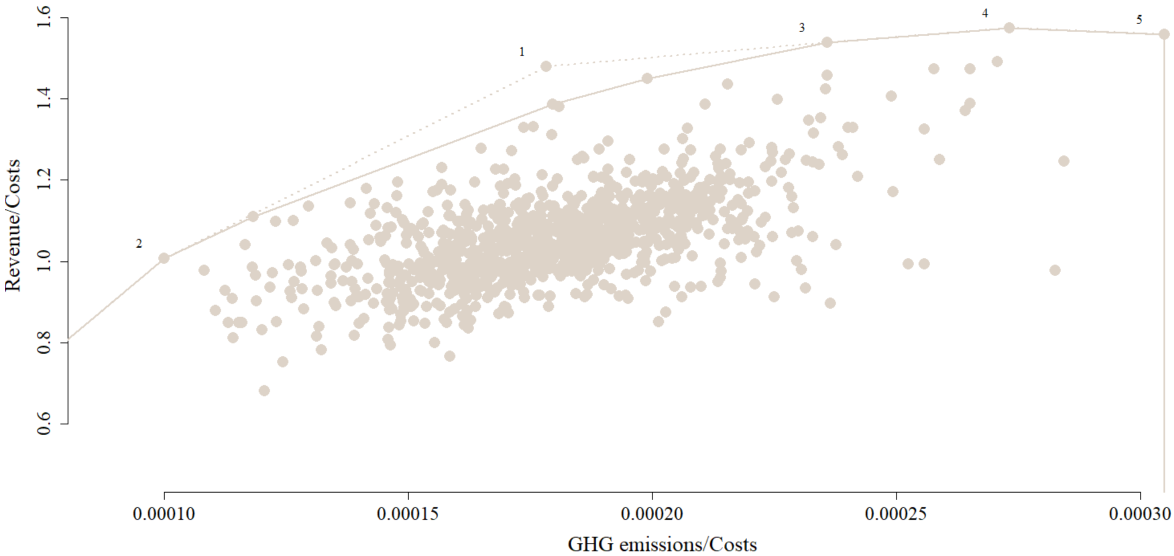
For simplicity the outlier detection has been carried out for the two mixed models for conventional and organic farms respectively. Nevertheless, the specific directional vector of the mixed models can be seen an average of the two extreme directions, and the results from the outlier detection using the mixed models do thereby to some extent capture the diversity of the two extreme models. If the two extreme models, only reducing GHG emissions or only increasing revenue were used in the outlier detection, the peers which influence the models would to a great extend be the peers located with either a relatively high revenue/cost ratio or low GHG/cost ratio. There are relatively few farms located at these areas and it is thereby harder to detect whether these are in fact outliers or representative for the sector when assessing the extremes.
Thus, the mixed models for conventional and organic farms respectively have been estimated excluding the peers one at a time. The effects on both inefficiency and frontier shadow prices of excluding the given peer from the two models are reported in Table B.3-4 and Table B.3-5.

Table B.3-4:Outlier detection for conventional farms

Farm	Mean inefficiency	Change in mean inefficiency (pct.-points)	Mean frontier shadow price (DKK)	Change in mean frontier shadow price (DKK)	Change in mean frontier shadow price (pct.)	Time used as peer
1	0,18	0,02	4.522	-1.321	23 pct.	1.033
2	0,20	0,00	6.171	328	-6 pct.	1.007
3	0,20	0,00	5.847	-0,73	0 pct.	40
4	0,20	0,00	5.846	-3	0 pct.	5
5	0,20	0,00	5.849	0	0 pct.	1

From the table it can be seen that farm number 1, has a higher impact on both mean inefficiency and mean frontier shadow prices than the remaining efficient farms. The mean inefficiency decreases by approximately 2 pct.-points when excluding farm 1 from the estimation, whereas the mean inefficiency decreases by far less than 1 pct.-point when excluding one of the remaining efficient farms. For the frontier shadow prices, the mean shadow price increases by 22,6 pct. when excluding farm 1 from the estimation. The results are partly driven by the fact that this farm is a peer for the majority of the observations.

Figure B.3-2: Illustration of the frontier with and without outlier (conventional model)



From the table it can be seen that farm number 1, has a higher impact on both mean inefficiency and mean frontier shadow prices than the remaining efficient farms. The mean inefficiency decreases by approximately 2 pct.-points when excluding farm 1 from the estimation, whereas the mean inefficiency decreases by far less than 1 pct.-point when excluding one of the remaining efficient farms. For the frontier shadow prices, the mean shadow price increases by 22,6 pct. when excluding farm 1 from the estimation. The results are partly driven by the fact that this farm is a peer for the majority of the observations.

Figure B.3-2 displays the frontier for conventional farms with and without farm 1. The shape of the frontier changes relatively much by excluding farm 1, and it can be seen that it is located relatively isolated towards the north west. Therefore, as all four criteria for removing outliers are fulfilled for farm 1, the farm is excluded from the final conventional models presented in the following.

When excluding farm 1 from the estimation new peers will now constitute the frontier. To ensure that the peers in a model excluding farm 1do not have the characteristics of being outliers, the outlier detection has been carried out once again following the same procedure as the initial outlier detection. The results from this second step, after farm 1 has already been excluded from the estimation, do not give a reason to exclude more peers. The results from this outlier detection is presented in Appendix B.

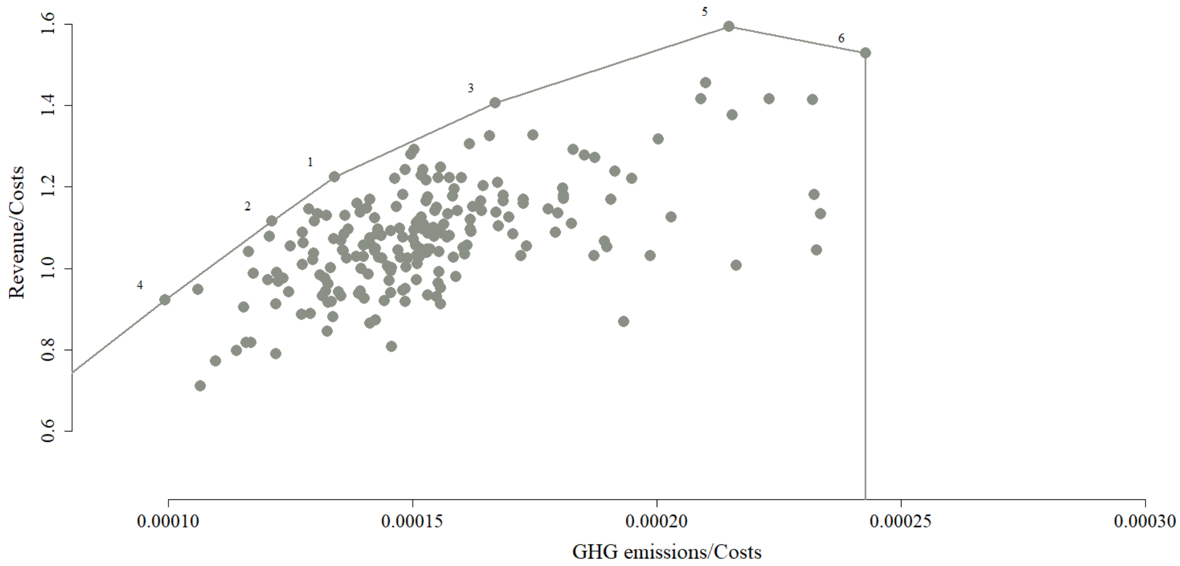
Table B.3-5 and Figure B.3-3 displays an overview of the outlier detection for organic farms in the same way as for the conventional farms. From Table B.3-5 it can be seen that organic farm 1, and 3 have the largest impact on the average frontier shadow prices. However, none of these farms affect the average inefficiency score greatly.

Table B.3-5: Outlier detection for organic farms

Farm	Mean inefficiency	Change in mean inefficiency	Mean frontier shadow price (DKK)	Change in mean frontier shadow price (DKK)	Change in mean frontier shadow price (pct.)	Time used as peer
1	0,11	0,005	6.755	-393	5 pct.	144
2	0,12	0,001	7.209	61	-1 pct.	112
3	0,12	0,003	6.837	-333	5 pct.	85
4	0,12	0,000	7.212	69	-1 pct.	47
5	0,12	0,001	7.107	-93	1 pct.	10
6	0,12	0,000	7.231	31	0 pct.	1

From the graphical illustration of the organic frontier, farm 5 could potentially look like an outlier, as it is located with a relatively large distance to the surrounding farms. However, there are relatively few observations in the organic model in general, and especially few observations located in the area around farm 5. It is therefore difficult to determine whether this specific farm is an outlier, and as both the average inefficiency score and frontier shadow price do not change much when excluding the farm, this observation is not identified as an outlier.

Figure B.3-3: Illustration of frontier with and without outlier (organic model)



The outlier detection has been carried out using a simple method which seeks to define how much each peer affect the results from the mixed models. However, it would be preferable to apply a more thorough method which could take differences in the directional vector into account. Such a method is currently not available for the specific model estimation of this thesis and it is outside the scope of this thesis to develop such a method. Nevertheless, outlier detection in this model specification could be a perspective for future research.

B.3.3.2 Validation of the DJL package

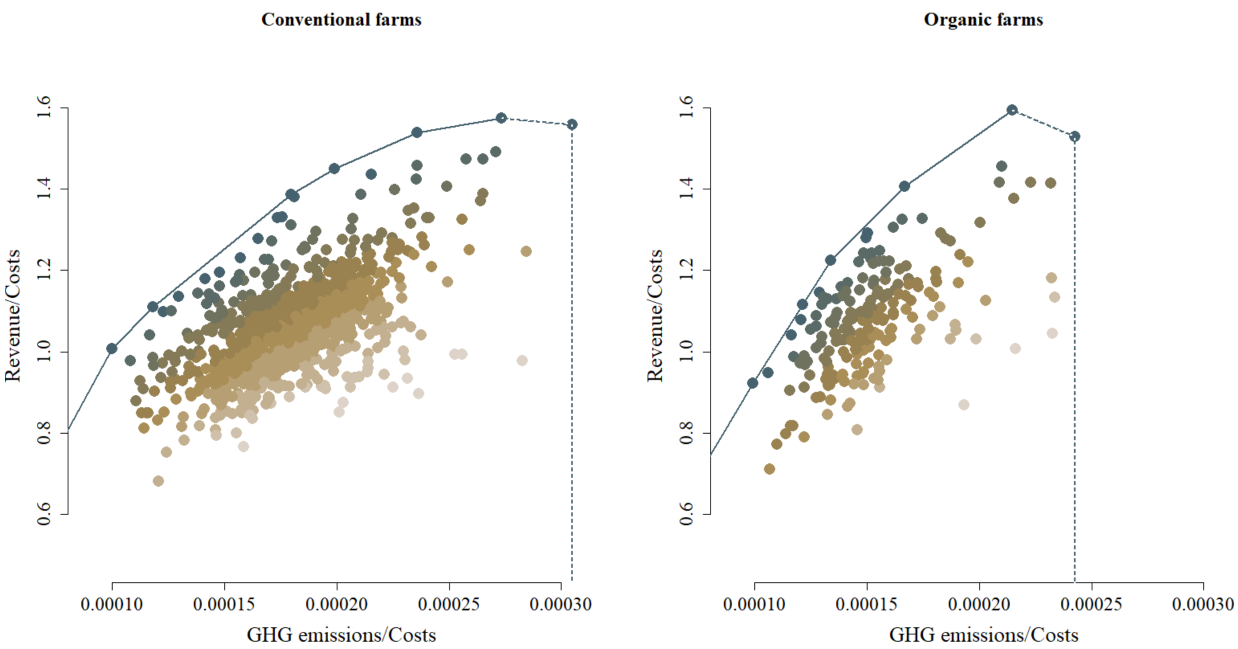
The specific models for the empirical analysis have been computed with the open sourced software R, using the source code of the shortage function from the DJL package (Lim, Package "DJL", 2020) and adapting it to this empirical analysis. The DJL package is used as it is outside the scope of this thesis to program the specific model from scratch.

There is generally not a wide range of software available for computing models such as those applied in this empirical analysis. This is likely due to the fact that this specific method for modeling undesirable outputs is still developing and widely discussed as referred to in section B.2.4. The DJL package used to estimate the models in this thesis is relatively new and has been released this year. The package is thereby not yet widely used and reviewed. Therefore, the following section explores whether the package works as intended by validating the results in regard to parameters such as inefficiency scores and frontier shadow prices.

The specific models for the empirical analysis have therefore been limited to only include three variables. This simplifies the model, making it more straightforward to verify the results both graphically and with simple mathematical calculations in order to ensure that the software package works as intended.

Figure B.3-4 graphically displays the mixed models constructed with the DJL package in R plotted in two dimensions for both conventional and organic farms. From the figure it can be seen that the model is behaving as expected as the more efficient farms are placed north west, where both revenue is relatively high, and the GHG emissions are relatively low compared to the south east part of the plot.

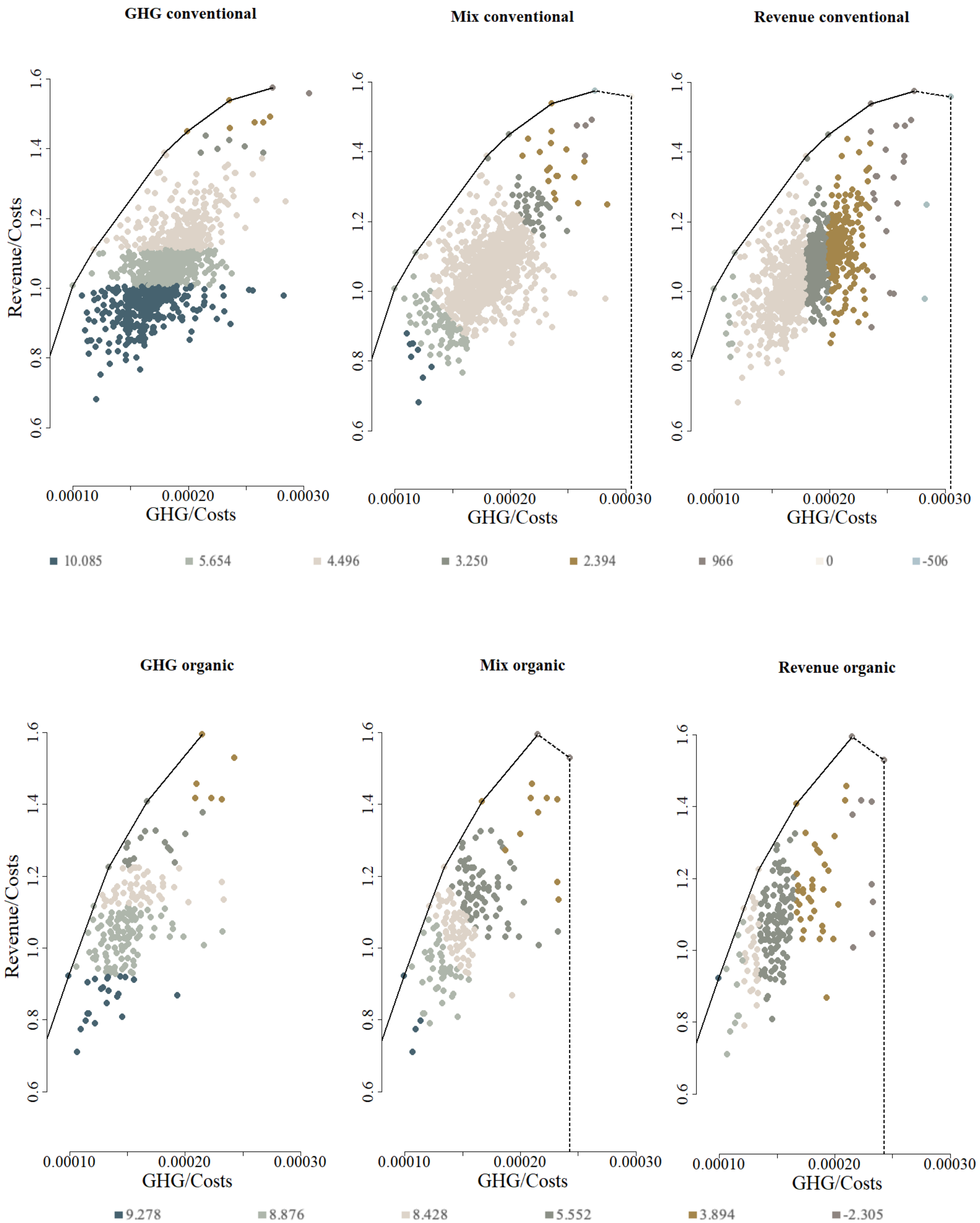
Figure B.3-4: Weak disposability technology with a directional distance function



Note: The color gradients show the level in efficiency, where darker colors imply a higher efficiency level.

Figure B.3-5 shows the weak disposability model with different directional distances (cf. Table B.3-1) for conventional and organic farms respectively. The colors of the farms correspond to the respective frontier shadow prices for each farm. This implies that farms with the same color are projected onto the same part of the frontier having the same peers and frontier shadow prices. It can therefore be seen from the figures, that the farms are being projected in the predefined direction. In the GHG models farms are projected horizontally towards the frontier, implying that revenue is held fixed while only GHG emission are reduced. In the mixed models farms are projected diagonally towards north west, and in the revenue models vertically towards the frontier.

Figure B.3-5: The weak disposability model with three different directional distance functions: colored in relation to different frontier shadow prices



Note: The farm observations are colored in relation to different frontier shadow prices

Furthermore, to ensure that the frontier shadow prices are correctly estimated in the models as $\frac{\pi^k_{GHG}}{\pi^k_{revenue}}$, the unique slopes along the frontier have been calculated manually. For a specific facet in two dimensions (part of the frontier created by a convex combination of two efficient farms (peers))[4] created by peer 1 and peer 2, the slope has been manually calculated as: $\frac{\Delta revenue/cost}{\Delta GHG/cost} = \frac{(revenue_1/cost_1)-(revenue_2/cost_2)}{(GHG_1/cost_1)-(GHG_2/cost_2)}$.

It is found that the slope calculated manually in two dimensions corresponds exactly to the frontier shadow prices found by calculating the relative weights from the model estimation. This implies that the frontier shadow prices estimated from the DJL package for each farm by using the weights of GHG emissions (π^k_{GHG}) and the weights of revenue ($\pi^k_{revenue}$) do in fact correspond to the trade-off found at the frontier and thereby the actual frontier shadow prices.

Furthermore, as described in section B.2.4 the normalization constraint is given by equation B.2-8: $\sum_{m=1}^{M} \pi_{v,m} g^{k'}_{v,m} - \sum_{j=1}^{J} \pi_{w,j} g^{k'}_{w,j} = 1$. In this empirical analysis, this constraint is fulfilled for all farms projected onto the efficient part of the frontier. Nevertheless, for the few farms projected onto the inefficient part of the frontier, this constraint is not fulfilled, which is an implication of having $\pi_{w,j}$ unconstrained.

For both models, an inefficient part of the frontier is defined, illustrated by the dotted lines in Figure B.3-5. However, the majority of the frontier is strongly efficient and only a minority of observations are projected onto the inefficient part. Nevertheless, (in)efficiency measures and frontier shadow prices for the specific farms which are projected onto the inefficient part of the frontier do not have a straightforward interpretation. These specific farms will thereby be excluded from the remaining empirical analysis in order to ensure that these farms do not affect the overall findings. For the conventional models, 2 farms have been removed from the mixed model and 3 from the revenue model. For the organic models, 2 farms have been removed from the mixed model and 7 from the revenue model. No farms are excluded from the GHG models as the horizontal projection in this model ensures that farms are only projected on to the strongly efficient part of the frontier.

[4] The facet with more than two variables might in reality consist of a convex combination of more than two peers. However, when illustrating it in 2 dimensions, there will only be two peers creating the facet.

B.4 Results

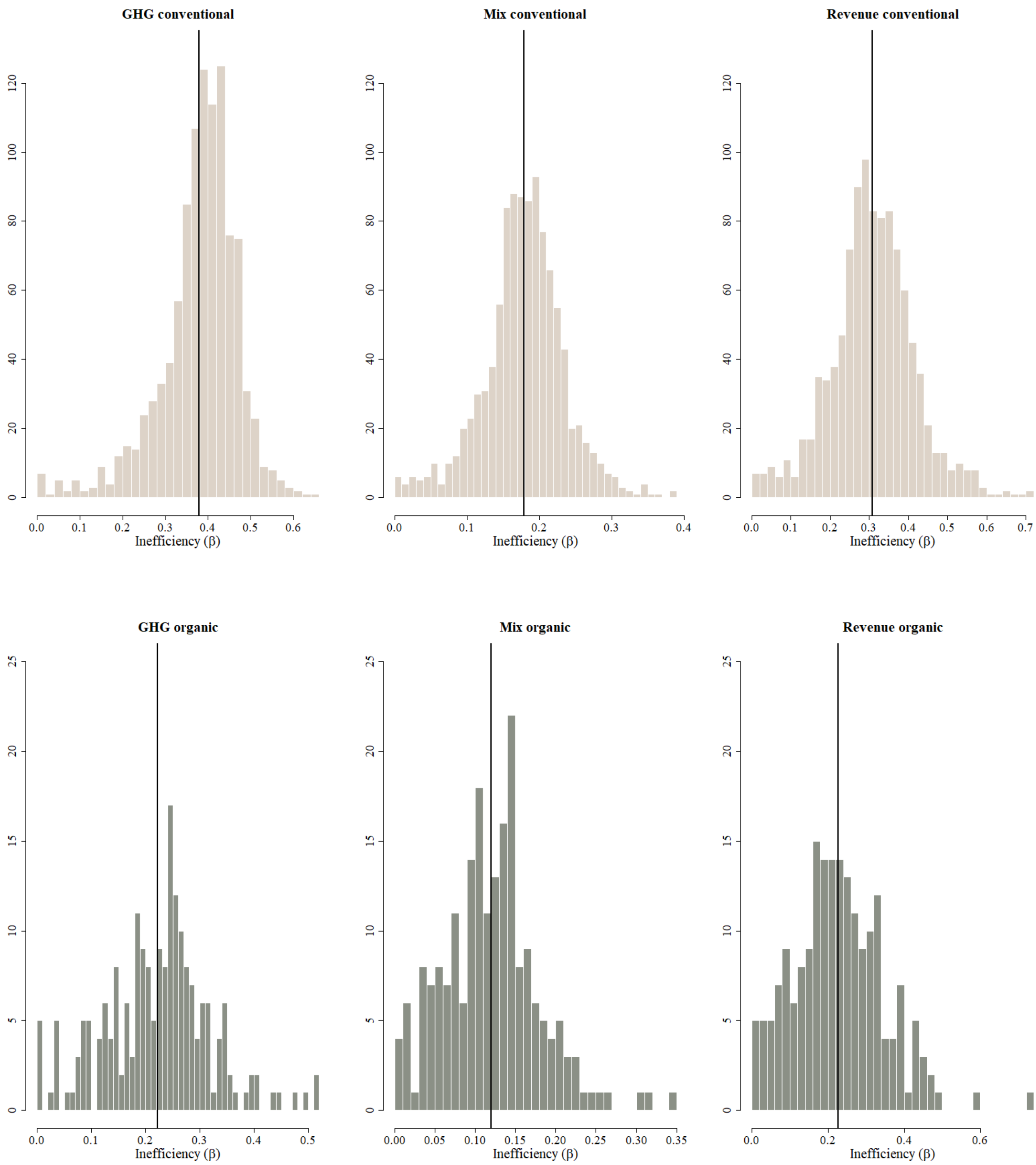
This section covers the results from the empirical benchmarking analysis. The first part of the analysis includes results from the three different models: *the GHG model, the mix model* and *the revenue model* for both conventional and organic farms. The section is structured as follows. Firstly, the individual inefficiency terms (β) for the farms, estimated from the different models, are presented. Secondly, the frontier shadow prices from both the conventional and organic models are presented, focusing on the variation of these across the different models. The variation in frontier shadow prices across the models, shows how the frontier shadow prices for the specific farms are very sensitive to the direction of the projection towards the frontier.
The third part of the results presents the existing potentials within the dairy sector, given by the technological lag found through the different models. The potentials found from the revenue and GHG models are used to detect the average opportunity costs of abating GHG emissions for inefficient farms. The average opportunity costs are further analyzed by giving an example of a second stage analysis regarding which characteristics of a farm that are associated with relatively low abatement costs.
The last part of the analysis shows an example of a second stage analysis of how to detect the characteristics of the farms performing at best practice.

B.4.1 Inefficiency detected through the different models

Figure B.4-1displays the distribution of inefficiency scores (β) and the average inefficiency of the farms in the different models for conventional and organic farms respectively.
The histogram of the inefficiency scores for the conventional model seems to be rather normally distributed around the average inefficiency score. The average inefficiency score for the *GHG conventional* model is 0,379 indicating that on average each farm can reduce their current level of GHG emissions with 37,9 pct. Similarly, this indicates that the average *efficiency score* from this model is $eff_{GHG} = \frac{1}{1-0,379}$ = 1,6 implying that farms on average produce 160 pct. of the efficient level of GHG.
The average inefficiency for the *Revenue conventional* model is 0,308 i.e. farms can on average increase revenue by 30,8 pct. In this model, the efficiency score is dependent on the desirable output, revenue, and should thereby be calculated as $eff_{Revenue} = \frac{1}{1+0,308} = 76,4$ indicating that on

average farms produce 76,4 pct. of the efficient level of revenue. Thus, the efficiency terms are not directly comparable across the two models as the calculation is dependent on whether the focus is a desirable or undesirable output. Nevertheless, the inefficiency term β represents the potential of either reducing GHG emissions or increasing revenue in both models and is therefore directly comparable across these models.

Figure B.4-1: Distribution of inefficiency scores for conventional and organic farms



Note: the black line is the average inefficiency score

Conventional farms are on average more inefficient when measuring their potential in the *GHG conventional* model, compared to the *Revenue conventional* model. This indicates that farms are in general located further away from the frontier when assessing their level of GHG emissions, whereas farms are in general relatively closer to the frontier in regard to their level of revenue. Assessing inefficiency of the conventional farms in the *Mix conventional* model, farms have an average inefficiency score of 0,179 which implies that when the focus is simultaneously reducing GHG emissions and increasing revenue, farms have an average potential for doing both of 17,9 pct.

When assessing the inefficiency scores for the organic farms, it can be seen that the inefficiency scores for organic farms are not as normally distributed as is the case for conventional farms. However, this is likely due to the fact that there are significantly fewer organic farms. The average inefficiency for the two extreme models *GHG organic* and *Revenue organic* are relatively similar with average inefficiency scores of 0,222 and 0,227 respectively. However, the average inefficiency for the *Mix organic* model is relatively lower with an average inefficiency score of 0,12. Similar to the models for the conventional farms, organic farms are relatively less inefficient when assessing inefficiency in a model which seeks to both reduce GHG emissions and increase revenue simultaneously.

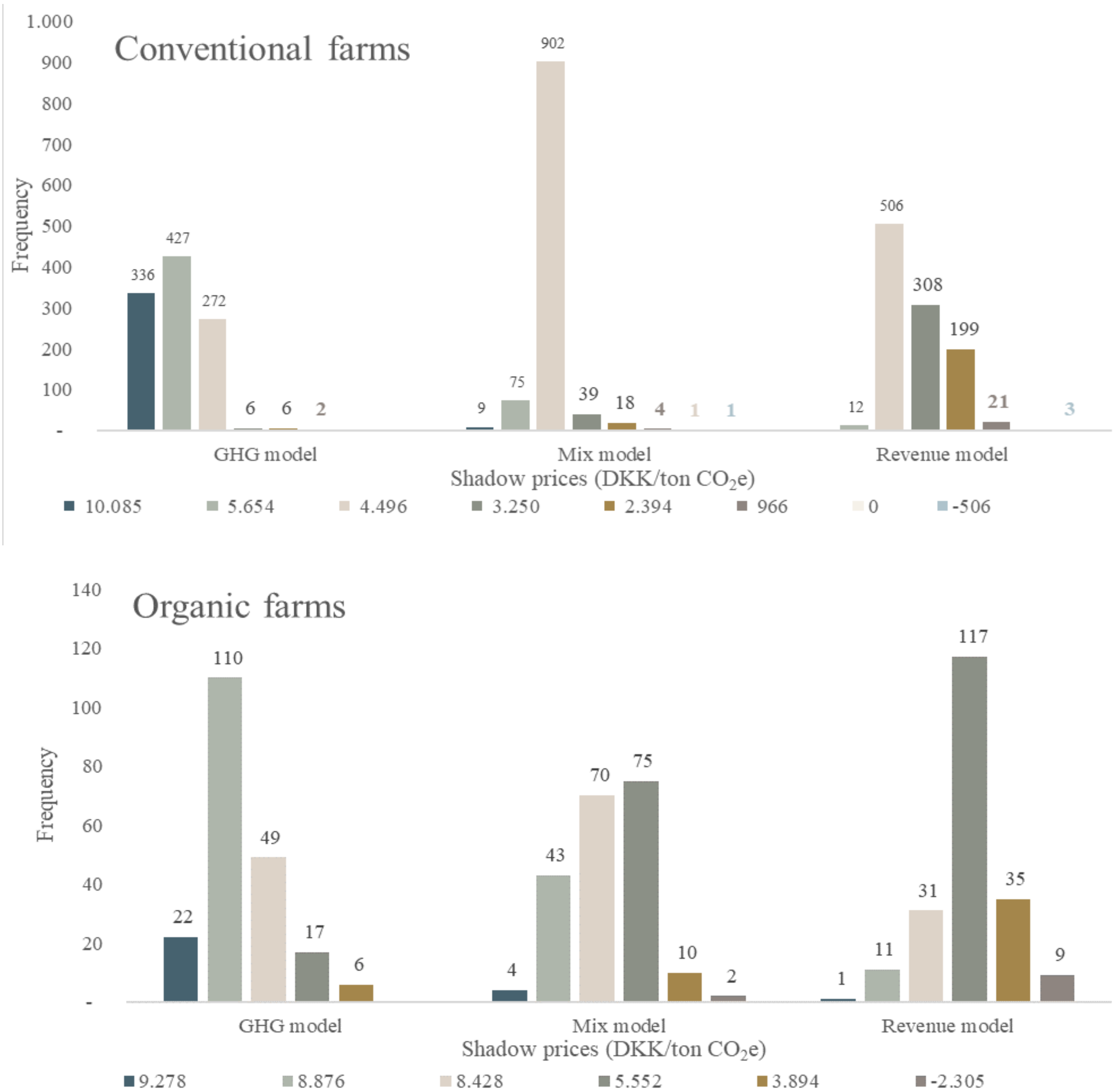
It should be noted that the inefficiency scores are not directly comparable across the conventional and organic models. The inefficiency for the different farms is measured relative to the frontier for either the conventional or organic farms. As the two frontiers differ across the two type of farms, the relative measure of inefficiency is not comparable between the two.

B.4.2 Frontier shadow prices

As described in section B.2.4, the slope of the frontier can be interpreted as a frontier shadow price. When a farm moves along the frontier, the slope describes the trade-off between the level of revenue and GHG emissions. That is, a slope of 4.496 which is found on a part of the frontier for conventional farms, implies moving to the left along the frontier and thereby reducing 1 ton of GHG will cause a loss of 4.496 DKK in revenue. This means that if a farm becomes efficient and thereby located on the frontier, the farm will have the frontier shadow price corresponding to the slope of the part of the frontier where it is projected onto, as illustrated by the different colors on the figure.

The frontier shadow prices thereby represent the trade-off between GHG and revenue for the farms operating at best practice i.e. the efficient farms and the current inefficient farms when they reach the frontier. Figure B.4-2 illustrates the potential distribution in frontier shadow prices for farms if they catch up with the frontier in models with different directional distance functions. The frontier shadow prices for the conventional and organic farms correspond to the graphical illustration of the models presented in Figure B.3-5.

Figure B.4-2: Distribution of the frontier shadow prices with different directional distance functions



Overall, the results of the frontier shadow prices show that the distribution of these is highly dependent on the specific model. There are eight unique frontier shadow prices across the three conventional models. However, only six of them have a straightforward interpretation, as two unique frontier shadow prices (0 DKK/ton and -506 DKK/ton) represent the inefficient part of the frontier.

The six remaining frontier shadow prices show that reducing one ton of CO_2e is relatively costly with a minimum price of 966 DKK per ton and a maximum price of 10.085 DKK per ton. Assuming it is possible for all farms to operate at best practice, there will be a large difference in the distribution of the frontier shadow prices across the three models. In the GHG model, the three frontier shadow prices of 10.085 DKK per ton, 5.654 DKK per ton and 4.496 DKK per ton are the most frequent. In the mixed model, the majority of the farms are estimated to have a frontier shadow price of 4.496 DKK per ton. In the revenue, the three most frequent frontier shadow prices are 5.654 DKK per ton, 4.496 DKK per ton and 2.394 DKK per ton.

The pattern is relatively similar for the organic farms. However, the organic farms do in general have relatively higher frontier shadow prices than those estimated for the conventional farms, even though the maximum frontier shadow price for conventional farms is higher. The maximum frontier shadow price for organic farms is estimated to be 9.278 DKK per ton of CO_2e whereas the minimum frontier shadow price estimated at the fully efficient part of the frontier is 3.894 DKK. As were the case for the conventional farms, the frontier shadow prices are on average higher in the GHG model compared to the remaining two models.

This implies that when farms become efficient by only reducing GHG emissions and keeping revenue fixed, this will lead to the highest cost of reducing 1 ton of CO_2e. This is in line with what is normally assumed regarding increasing marginal abatement costs. When farms reach the frontier in the scenario where only GHG emissions are reduced, farms obtain a relatively low level of GHG emissions i.e. the farms have already abated a relatively large amount of GHG emissions. It is therefore evident that it will be more costly for farms to reduce the level of GHG emissions further in this scenario.

The different models, presented in Table B.3-1 have been estimated, with three different directional distance function with the purpose of showing the results' sensitivity to the directional distance. In particular, the distribution in frontier shadow prices are highly dependent on the specific directional distance function (cf. Figure B.4-2). It is therefore crucial for the results, that the direction is chosen correctly in regard to the purpose of the analysis. If the only purpose is to find the maximum reduction potential of GHG emissions, a directional distance function only reducing GHG emissions is suitable. However, if the main purpose of the analysis is to maximize revenue without neglecting the negative effect of GHG emissions, a directional distance function only maximizing revenue is suitable. Lastly, by choosing a directional distance function both maximizing revenue and minimizing GHG, it is possible to examine to what extend farms are able to both reduce GHG emissions while simultaneously increasing revenue.

B.4.3 Average opportunity costs for inefficient farms

If the direction of projection is not necessarily given by the aim of the analysis, it could potentially be a relatively large drawback that the frontier shadow prices are so sensitive to the directional distance function. However, it is possible to eliminate the uncertainty of the frontier shadow prices for the inefficient farms by calculating *average opportunity costs* of abating GHG emissions.

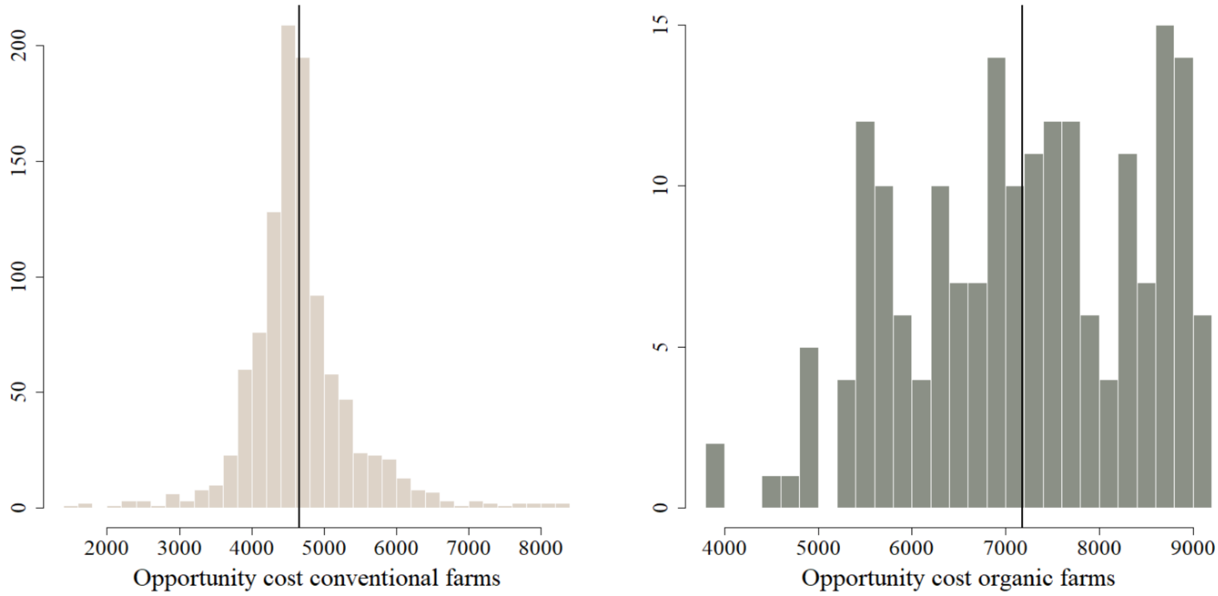
The average opportunity costs are calculated by comparing the maximum potential reduction of GHG with the maximum potential increase in revenue found in the two extreme models.

In the revenue model, the inefficient farms can reach the highest possible economic benefit as only revenue is maximized while keeping GHG emissions fixed. This corresponds to being projected vertically upwards in Figure B.3-5. In order to reduce GHG emissions, it is necessary to give up some of this maximum possible increase in revenue. Therefore, the maximum economic potential gain for each farm in the revenue model is compared with the maximum possible reduction in GHG emissions estimated through the GHG model. This is done to see the potential trade-off by moving horizontally to the left rather than vertically upwards i.e. the average opportunity costs indicate how much potential revenue an inefficient farm must on average give up in order to reduce its GHG emissions as much as possible.

The maximum potentials are calculated by using the inefficiency terms β from both the revenue and GHG models. The potential of increasing revenue is for each farm calculated by using the inefficiency term estimated from the model only increasing revenue. The potential of reducing GHG emissions is estimated by using the inefficiency term from the model only reducing GHG emissions. Thereby, the maximum potential is obtained by multiplying the different $\beta's$ from each of the model with the farms level of revenue and GHG respectively. Finding the farm-specific average opportunity cost is thereby done by dividing the maximum potential from the revenue model in DKK by the maximum reduction in GHG emissions from the GHG model. This leads to an opportunity cost expressing the average cost of reducing 1 ton of GHG in DKK as lost revenue. The average opportunity costs thereby represent an *average* abatement cost rather than the *marginal* shadow price, as the opportunity costs are calculated as the average reduction in revenue for the total possible decrease of CO_2.

Figure B.4-3 displays the distribution of the average opportunity costs for all farms both conventional and organic. For the conventional farms the range varies between less than 2.000 DKK to more than 8.000 DKK and is centered around 4.500 DKK. There are significantly fewer organic than conventional farms. This can also be seen from the distribution of the average opportunity costs for organic farms in Figure B.4-3, where the estimated average opportunity costs do not display a normal distribution as is the case for the conventional farms. However, the average opportunity cost is generally higher for the organic farms, as almost all estimated average opportunity costs are above 4.000 DKK and the average cost is around 7.000 DKK per ton GHG.

Figure B.4-3: Average opportunity cost of reducing GHG emissions in DKK for inefficient farms



Note: The black vertical lines represent the simple mean for each of the two models.

By using the individual potentials for each farm in the dataset it is possible to estimate aggregated potentials across the entire dairy sector. The aggregated potentials from the different models represent the total possible reduction in GHG emissions and/or total possible increase in revenue across all farms given that each farm reach their full potential i.e. reach the frontier and thereby operate at best practice.

In Table B.4-1 the total GHG and revenue potential for the entire sector is presented. The GHG potential is defined by the aggregated maximum potential reduction of GHG in tons. The maximum revenue potential is defined by the aggregated potential increase of revenue in DKK. The potentials represent the case where all farms produce at best practice which corresponds to catching up with the current technological lag defined in the model. The table shows the two type of potentials for the different models.

Table B.4-1: The potentials found from the different benchmarking models

	Potential reduction in GHG emissions (tons)			Potential Increase in revenue (1.000 DKK)		
	Conventional farms	**Organic farms**	**Total**	**Conventional farms**	**Organic farms**	**Total**
GHG model	584.003 *(37,9 pct.)*	54.596 *(22,2 pct.)*	638.599 *(35,4 pct.)*			
Revenue model				2.633.737 *(30,8 pct.)*	351.302 *(22,7 pct.)*	2.985.039 *(27,6 pct.)*

When the model only increases revenue, the aggregated potential for the entire sample is an increase in the total revenue of the sector of 2,99 billion DKK. This corresponds to each farm increasing their revenue with an average of 27,6 pct. This maximum potential increase in revenue, can only be reached by accepting the current level of GHG emissions and only focusing on revenue. On the contrary, by only focusing on reducing the GHG emissions from the farms, the reduction potential for the entire sample is reducing 638.599 tons of GHG corresponding to an average decrease of GHG emissions of 35,4 pct.

For conventional farms, only reducing GHG emissions means that the maximum potential revenue increase of 2,63 bill. can no longer be obtained within the current technology of the sector. Thereby, the average opportunity cost for each reduced ton of GHG when only reducing GHG emissions will on average be $\frac{2.633.737.000 \text{ DKK}}{584.003 \text{ tons GHG}} = 4.510$ DKK per ton GHG given the existing technology.

The corresponding average opportunity cost for the organic farms by using the GHG model in relation to the revenue model is 6.435 DKK. The estimated average opportunity cost of reducing GHG is generally higher for the organic farms than for the conventional farms. However, it should be noted that the organic farms tend to have a slightly lower GHG emission, which could be a reason why it is more expensive for organic farms to reduce GHG emissions.

An OLS estimation has been conducted to see which characteristics that affects the farm-specific average opportunity costs for both conventional and organic farms. In the estimation, the average opportunity costs for the different farms have been regressed on certain farm characteristics (cf. Table B.4-2). The results can be used to see if there are some specific characteristics which are determinant for the volume of the average opportunity costs which varies across each farm.

Nevertheless, the explanatory variables in the estimation, are not necessarily representative for what should ideally be included in such analysis. However, this analysis serves as an example of how to conduct a second stage analysis using the average opportunity costs for the inefficient farms.

Table B.4-2: OLS regression for average opportunity costs

	Dependent variable: Average opportunity costs	
	Conventional farms	**Organic farms**
Number of cattle (100)	-0,096	-0,677
	p = 0,451	*p = 0,390*
Milk (DKK)/Total outputs (DKK)	-28,100***	-47,966***
	p = 0,000	*p = 0,005*
Fixed costs/Total costs (pct.)	43,374***	40,957***
	p = 0,000	*p = 0,010*
Dairy cattle / All cattle (pct.)	9,373***	6,223
	p = 0,006	*p = 0,718*
Ownership (other than private)	-352,527***	-327,999
	p = 0,000	*p = 0,245*
Share of jersey cattle (pct.)	4,757***	13,329***
	p = 0,000	*p = 0,0001*
Share of cattle in a deep litter housing system (pct.)	-8,182***	-20,889***
	p = 0,000	*p = 0,00002*
Milk yield (liter/cow)	17,138***	22,666
	p = 0,00005	*p = 0,282*
Years since converted to organic		9,656
		p = 0,354
Constant	-28,100***	9.227,933***
	p = 0,000	*p = 0,00001*
R2	*0,351*	*0,259*
Adjusted R2	*0,346*	*0,222*
Residual Std. Error	*589,303 (df = 1031)*	*1.091,185 (df = 181)*
F Statistic	*69,783*** (df = 8; 1031)*	*7,012*** (df = 9; 181)*
Note:	**p<0,1; **p<0,05; ***p<0,01*	

There are a range of farm characteristics affecting the average opportunity costs. For conventional farms, the regression indicates that the following variables are negatively correlated with the average opportunity costs for the inefficient farms: the share of revenue originated from milk, having another ownership than private and the share of cattle in a deep litter housing system. This implies that farms with e.g. a high share of deep litter housing systems tend to have a lower average opportunity cost i.e. a lower average abatement cost of reducing GHG emissions. It is thereby cheaper for these types of farms to reduce their GHG emissions. In contrast, the regression indicates that inefficient farms with a higher share of fixed costs, dairy cattle and jersey cattle as well as a higher yield (technical efficiency) tend to have higher average opportunity costs.

In the OLS regression for the organic farms, less of the characteristics have a significant effect on the average opportunity costs. This might be due to the fact that there are fewer observations for organic farms. Nevertheless, from the regression it can be seen that farms with a higher share of revenue originating from the milk production and a higher share of deep litter housing systems tend to have lower average opportunity costs. This implies that for inefficient farms with these characteristics it is cheaper to reduce GHG emissions. Contrary, inefficient farms with a higher share of fixed costs and a higher share of jersey cattle seems to have a higher average opportunity cost.

Gathering this information makes it possible to detect whether there are specific parts of the dairy sector where it is less costly to reduce GHG emissions than others. Dependent on the type of regulation, that will be adopted to ensure the desired reduction in GHG emissions in the agricultural sector, it can be crucial to obtain knowledge about which specific type of farms that are able to reduce its GHG emissions at the lowest cost. With this knowledge, it is possible to design a future regulation such that the incentives for reducing the emissions are targeted towards farms where the cost of doing so is lowest.

B.4.4 Reaching the potentials - a second stage analysis using Tobit regression

The Danish agricultural sector experiences a decreasing ratio between output and input prices, increasing international competitiveness and an external pressure from consumers, institutions and governments worldwide to produce more climate friendly. Therefore, there is a need for the sector to constantly increase productivity while simultaneously decreasing the climate footprint from the agricultural sector.

The benchmarking model estimated in this empirical analysis captures the interaction between economic and climate performance within the dairy sector and provides information of the relative efficiency for each farm in regard to both areas. A second stage analysis of what drives the level of (in)efficiency has been conducted using a Tobit regression. The Tobit model regresses the inefficiency term β for the different estimated models, on variables describing different characteristics of the farms. The variables are similar to those used in the second-stage regression using average opportunity costs. In addition to these, three variables describing costs to consulting has also been included in this second stage with inefficiency scores as the dependent variable.

The second stage analysis serves as an example of how the estimated measures of relative (in)efficiency from the benchmarking models can be used to extract information regarding what drives (in)efficiency in the Danish dairy sector when benchmarking both economic and climate performance. This provides information of how the inefficient farms can catch up with the current technological lag. The estimation might not be fulfilling for which characteristics are the most relevant to include in such an analysis, and other relevant measures could potentially be included.

Table B.4-3 and Table B.4-4 show the Tobit regression for inefficiency scores ($\beta's$) for conventional and organic farms respectively. The analysis is carried out using the two extreme models where the directional distance function only seeks to either reduce GHG emissions or increase revenue. Both of the extreme models are included to examine whether there are differences in what drives (in)efficiencies dependent on which extreme direction farms are projected onto the frontier.

The estimates in the tables are reported as the average partial effects (APE) of the given variables. The average partial effects are calculated as described in section B.2.5. The original coefficients from the Tobit models are presented in Appendix C. The p-values of the original coefficients are reported alongside the APE estimates in Table B.4-3 and Table B.4-4.

Table B.4-3: Tobit regression for inefficiency scores for conventional farms

	Dependent variable: Inefficiency score			
	GHG		**Rev**	
	APE	P-value	APE	P-value
Number of cattle (100)	-0,0001	*p = 0,000****	-0,0001	*p = 0,000****
Milk (DKK)/Total outputs (DKK)	0,0090	*p = 0,000****	0,0085	*p = 0,000****
Fixed costs/Total costs (pct.)#	0,0020	*p = 0,000****	0,0038	*p = 0,000****
Dairy cattle / All cattle (pct.)#	-0,0046	*p = 0,000****	-0,0051	*p = 0,000****
Ownership (other than private)##	-0,0399	*p = 0,000****	-0,0453	*p = 0,000****
Cost to consulting - production (1.000 DKK)	0,0000	*p = 0,970*	-0,0001	*p = 0,643*
Cost to consulting - cattle (1.000DKK)	0,0001	*p = 0,107*	0,0000	*p = 0,772*
Cost to consulting - economic (1.000 DKK)	0,0000	*p = 0,350*	-0,0001	*p = 0,044****
Share of jersey cattle (pct.)#	-0,0010	*p = 0,000****	-0,0009	*p = 0,000****
Share of cattle having a deep litter housing system#	0,0022	*p = 0,000****	0,0025	*p = 0,000****
Milk yield (liter/cow/day)#	-0,0014	*p = 0,002****	-0,0009	*p = 0,098**
Observations	*1.049*		*1.046*	
Note:		**p<0,1** p<0,05 ***p<0,01*		

*#: Variables describing a percentage/share are specified as the percentage number. For a percentage of e.g. 20 pct. the variable is thereby 0,2*100=20.*
##: The variable ownership is a dummy variable taking the value 0 for personally ownership, and 1 for either partnership or limited ownership. Thereby the baseline is private ownership and the two variables in the table is the effect of another ownership relative to private ownership.

Table B.4-4: Tobit regression for inefficiency scores for organic farms

	Dependent variable: Inefficiency score			
	GHG		**REV**	
	APE	P-value	APE	P-value
Number of cattle (100)	-0,0002	*p = 0,00002****	-0,0002	*p = 0,00003****
Milk (DKK)/Total outputs (DKK)#	0,0081	*p = 0,000****	0,0081	*p = 0,000****
Fixed costs/Total costs (pct.) #	0,0008	*p = 0,274*	0,0015	*p = 0,179*
Dairy cattle / all cattle (pct.) #	-0,0077	*p = 0,000****	-0,0096	*p = 0,000****
Ownership (other than private) ##	-0,0144	*p = 0,272*	-0,0198	*p = 0,338*
Cost to consulting - production (1.000 DKK)	0,0003	*p = 0,343*	0,0004	*p = 0,341*
Cost to consulting - cattle (1.000DKK)	0,0003	*p = 0,010***	0,0003	*p = 0,068**
Cost to consulting - economic (1.000 DKK)	-0,0001	*p = 0,551*	-0,0001	*p = 0,346*
Share of jersey cattle (pct.) #	-0,0016	*p = 0,000****	-0,0017	*p = 0,000****
Share of cattle having a deep litter housing system (pct.) #	0,0028	*p = 0,000****	0,0032	*p = 0,000****
Milk yield (liter/cow/day) #	-0,0049	*p = 0,00001****	-0,0057	*p = 0,0001****
Year since converted to organic	-0,0020	*p = 0,0001****	-0,0025	*p = 0,0004****
Observations	*204*		*195*	
Note:	*p<0,1** p<0,05 ***p<0,01			

*#: Variables describing a percentage/share are specified as the percentage number. For a percentage of e.g. 20 pct. the variable is thereby 0,2*100=20.*
##: The variable ownership is a dummy variable taking the value 0 for personally ownership, and 1 for either partnership or limited ownership. Thereby the baseline is private ownership and the two variables in the table is the effect of another ownership relative to private ownership.

Overall, the second stage analysis of what drives efficiency for dairy farms, indicates that there are not great differences across the two extreme models and across organic and conventional farms. Furthermore, a large part of what drives efficiency in the benchmarking model, where climate performance is evaluated alongside economic performance, is in line with what generally drives the economic performance of the sector. Factors such as increasing the size and technical performance of farms as well as changing the ownership structure away from personally owned, seems to be factors that are both relevant when assessing economic and climate performance.

In the following, the effect of all variables included in the Tobit models will be commented and compared across the two extreme models to get an indication of what drives (in)efficiency in the different models.

The number of dairy cattle (100 cattle) is estimated to have a negative and significant effect on the inefficiency score β in both of the models for both the conventional and organic farms. This indicates that increasing the herd size leads to lower inefficiency. However, even though the estimates are significant, the volume of the APE of number of dairy cattle is small.
The APE estimate for the conventional farms in the GHG model is -0,0001 implying that increasing the number of dairy cattle with 100 will on average decrease the inefficiency of a farm with 0,01 pct.-points. The specific estimate of APE is relatively similar across all the estimated models, implying that the tendency is present for both conventional and organic farms independent of which direction farms are projected in.

This is in line with what is found when assessing only the economic performance of the sector without taking climate into consideration. The Danish Agricultural and Food Council (DAFC) state that dairy farms generally utilize economies of scale and are specializing to become competitive on the international market. Therefore, the general tendency within the sector is that the number of farms is decreasing while the size of the farms increases (Danish Agriculture and Food Council, 2018).

The variable **Milk (DKK)/total outputs (DKK)** represents the ratio between revenue generated from milk relative to total revenue from both milk and other outputs. The higher the ratio, the more of a farm's total revenue is generated from milk. This variable is positive and significant in all models, implying that the more revenue that is generated from milk relative to other outputs, the more inefficient the farm is. The average partial effect from the GHG model for conventional farms is 0,009 and for the revenue model 0,0085. This implies that increasing the share of total revenue generated from milk with one pct.-point, leads to an average increase in inefficiency of 0,09 pct.-points and 0,85 pct.-points in the GHG and revenue model respectively.

The pattern is quite similar for the organic farms with an average partial effect of approximately 0,0081 in both models. It was expected to find a positive relationship between this variable and inefficiency, as the GHG emissions in the model are only estimated based on the dairy production. Revenue generated from other outputs are not related to any GHG emissions in the estimation from Part A and farms are thereby rewarded unintendedly for producing other outputs than dairy. It would therefore have been optimal to estimate a model where either GHG emissions related to other outputs were also included in the model, or where farms were only benchmarked on the specific inputs and outputs related to the milk production.
However, it has not been possible to specify which costs that are specific to the production of milk and which costs that are specific to other outputs. Excluding other outputs from the models without excluding the corresponding costs for these outputs could lead to farms being evaluated wrongly as inefficient.
The significance of the estimate does thereby not necessarily mean that dairy farms should increase their production of other outputs but is rather an indication of the drawback of the model in relation to the GHG estimations.

The estimate for the variable **fixed costs/total costs** is positive in both regressions for the conventional farms. This implies having more capital tied in long term investment relative to variable costs tends to decrease the level of efficiency. The effect of increasing the share of fixed costs is higher in the revenue models for both conventional and organic farms. However, the effect is not significant in both organic models. For conventional farms, the average partial effect of increasing the share of fixed costs by 1 pct.-point is an average increase in inefficiency of 0,2 pct.-points in the GHG model and 0,4 pct.-points in the revenue model. A relatively high share of fixed costs indicate that a farm has a relatively high share of long-term investments such as buildings, machinery and land, which cannot be adjusted due to change in demand in the short run.

The agricultural sector is generally characterized by having a relatively high share of fixed costs as the production requires a great amount of land and machinery. Therefore, it is often seen that that larger farms have increased their performance by having a high financial gearing which can be necessary in order to invest in the necessary land and machinery in general. Farms with more than 400 cows have in average a solvency ratio on 1,6 pct. meaning that only 1,6 pct.- of the assets are financed with equity and the rest is financed with debt. Smaller farms with less than 100 cows have in contrast a solvency ratio on 40,7 pct. (SEGES, 2018). This indicates that larger farms generally operate with higher risk than small farms.

The estimate for **Dairy cattle/all cattle** represents the composition of cattle on the specific farm. This estimate is negative, thus increasing the number of dairy cattle in relation to all cattle decreases the inefficiency score. This can be driven by the fact that dairy cattle must be assumed to create more revenue than other cattle. However, the GHG emissions estimated in this analysis are closely correlated with the amount of dairy cattle per farm, which could imply that having a larger share of dairy cattle would lead to higher inefficiency. Nevertheless, even though a relatively large part of a farms GHG emissions originates from dairy cattle, the corresponding revenue generated from dairy cattle seems to outweigh the damage from GHG emissions.

The average partial effect of the share of dairy cattle in relation to all cattle for conventional farms is -0,0046 and -0,0051 in the GHG and the revenue model respectively. This implies that increasing the share of dairy cattle relative to all cattle with 1 pct.-point leads to an average decrease in inefficiency of 0,46 pct.-points in the model only reducing GHG. In the revenue model, increasing the dairy cattle share with 1 pct.-point leads to an average decrease in inefficiency of 0,51 pct.-points. The APE estimates for organic farms are -0,0077 and -0,0096 for the GHG model and the revenue model respectively.

Overall, changing the composition of cattle might create a potential increase in efficiency. Nevertheless, it must still be acknowledged that in order to sustain a dairy production there is a need for having other cattle types such as calves, heifers, studs and bulls.

The variable **Ownership** is a binary variable taking the value 0 if the farm is personally owned and 1 if the farms has another ownership structure such as grouped partnership and limited ownership. The base for this variable is thereby personally ownership and the estimate are the effect on inefficiency of having another ownership. This estimate is not significant for the organic farms but is significant on a 5 pct.-level for the conventional farms. This implies that having another ownership than personal leads to a decrease in inefficiency in both the GHG model and the revenue model for the conventional farms. For these farms, having another form of ownership than personal

ownership leads to an average decrease in inefficiency of approximately 4 pct.-points in the GHG model and approximately 4,5 pct.-points in the revenue model.
This is in line with what SEGES finds, that when assessing the operating profit, the least profitable ownership structure is personal ownership (SEGES, 2018), which is in line with the results found in this analysis.

Other ownership structures than personal might be more flexible to take on larger investments in e.g. new technology, increasing the scale etc. as the risk is diversified to more than just the personal owner. This could indicate that giving incentives to change the general ownership structure could be a tool for increasing efficiency within the Danish dairy sector and thereby catch up with some of the current technological lag for inefficient farms.

Three variables describing the **cost to consulting** are included in the regressions, as it was assumed that these variables could have an impact on the level of inefficiency for the different farms. The three consulting variables describe the cost to production consulting, cost to consulting regarding cattle and cost to economic consulting. For the conventional farms the only variable that is significant on a 5 pct. level is cost to consulting in the revenue model. The average partial effect of the estimate is -0,001 implying that spending 1.000 DKK more on economic consulting will in the revenue model for conventional farms lead to an average decrease in inefficiency on 0,1 pct.-points. The estimate itself is relatively small but must be reviewed in relation to the aggregated spending on cost to economic consulting which is on average around 36.000 DKK in the sample of dairy farms.
Therefore, there is an indication that if a farm wants to become more efficient, the farm should focus on economic consultancy rather than consultancy regarding the production which is not necessarily directly related to the profitability of the farms. It is also generally the picture that in the recent years there has been more focus on the profitability rather than the technical performance for the Danish agricultural sector (see e.g. Asmild (2019) & Danish Agriculture and Food Council (2018)).

For the organic farms the estimates for cost to cattle consulting is significant on a 5 pct. level in the GHG model and a 10 pct. significance level in the revenue model. Both estimates are positive implying that using money on consultancy regarding cattle will on average lead to higher inefficiency. This might be due to the fact that this type of consultancy is more related to technical efficiency than economic or climate efficiency.

Consultancy for production is associated with consultancy regarding the technical efficiency i.e. the production itself and not necessarily the profitability of the farms. The technical efficiency is to a certain degree linked to the economic efficiency. However, it might also be the case that a too large focus on technical efficiency can be costly for the farms.
Contrary, specific *economic consultancy* can be a driver for farms to become more profitable i.e. more efficient in their economic performance.

The **share of jersey cattle** represents the share of all cattle that is of the breed jersey. The estimates for the variable are negative and significant in both the GHG and the revenue model. This implies that increasing the share of jersey cattle will on average lead to a decrease in inefficiency for both conventional and organic farms. By increasing the share of jersey cattle with 1 pct.-point, the inefficiency will on average decrease with approximately 0,1 pct.-points in both the GHG model and the revenue model for the conventional farms. Another way to interpret the estimate is that having only jersey and no heavy breed cattle will result in an average decrease in inefficiency of 10 pct.-points in relation to only have heavy breed. For organic farms the equivalent estimates are -0,16 pct.-points and -0,17 pct.-points. These estimates are also significant.

The tendency of having a higher efficiency with more jersey cattle might be due to the fact that jersey cattle produces milk with higher protein and fat content and the milk can thereby be sold to a higher price. The jersey cattle have a lower emission factor per cattle compared to heavy breed. Therefore, when looking both at raising revenue and reducing greenhouse gas emissions, there could be an incentive to increase the share of jersey cattle. However, it should be noted that if all cattle are replaced with jersey cattle, it is not necessarily possible to satisfy the demand for milk. Furthermore, there might be other factors not accounted for in this analysis, which are relevant to look for e.g. does it require more land to produce the same amount of milk with only jersey cattle, and if that is the case, which kind of land is then used. If deforesting is taken place to release more land for cattle production, it might end up being a worse solution for reducing the GHG emissions to the atmosphere.

The variable **Share of cattle having a deep litter housing system** is included as the deep litter housing systems are related to the highest emission of GHG per cattle. Therefore, it is expected that having a higher share of these housing systems in relation to other housing system might decrease the efficiency level. The Tobit regression shows that this is indeed the case for both conventional and organic farms. The estimates for the deep litter variable show that going from a stable without any deep litter, to a housing system only including deep litter, will increase inefficiency with approximately 22 pct.-points and 25 pct.-points for the GHG and revenue model respectively for the conventional farms. For organic farms the same estimates are approximately 30 pct.-points in both of the extreme models.
This implies, that farms can replace deep litter housing systems to become more efficient. However, animal welfare could be taken into consideration when replacing housing systems. Animal welfare has not been included in the estimated benchmarking models of this empirical analysis. However, it would be possible to integrate one or more variables expressing the level of animal welfare for a given farm. By including animal welfare as a desirable output in the benchmarking analysis, it might be possible that changing housing system would not decrease efficiency, as there exists a trade-off with animal welfare.

The variable **Milk yield** is an indication of the technical efficiency of the dairy farms. The estimation shows that a higher milk yield leads to increased efficiency in general.
As oppose to economic efficiency the technical efficiency gives an indication of the productivity of the farm. The technical efficiency is linked to the economic performance, but a high productivity is not always determined for a high economic efficiency. The estimates are significant across all models and have a negative relation with the inefficiency score. The average partial effect is -0,0014 in the GHG model and -0,009 in the revenue model for conventional farms. For organic farms these average partial effects are -0,0049 and- 0,0057 respectively. This implies that e.g. increasing the milk yield with 1 liter per dairy cattle per day for organic farms will on average result in a decrease in inefficiency of 0,49 pct.-points in the GHG model.

In the estimation for the organic farms, the variable **years since converted to organic farming** is included. The estimation indicates a relationship showing that the longer time a farm has been organic, the higher efficiency level. The estimate from the GHG model is -0,002 indicating that when the years since a farm has converted to organic farming increase with one, the inefficiency in average decrease with 0,20 pct.-points in the GHG model and 0,25 pct.-points in the revenue model. This indicates that organic farms become better at increasing revenue and reducing GHG emissions as times goes. The transition to organic farming might be a costly and long-lasting process. It might be that the newer farms lack knowledge of how to produce as efficiently as those who have been in the production for a long time. Therefore, knowledge sharing between the newer and those who has been in organic farming for many years, could increase the overall efficiency level.

Discussion and perspectives for future research

This thesis shows how benchmarking analysis can act as a tool for evaluating the effect of externalities alongside economic performance and provide abatement costs of these externalities. This can potentially be an important contribution as how to cost efficiently reduce GHG emissions in the agricultural sector and ultimately reach the national 70 pct. reduction target, which is widely discussed in the public and political debate.

The analysis is conducted by firstly estimating farm-specific GHG emissions in order to implement these in an economic benchmarking model suitable for handling undesirable outputs. Due to the scope of the thesis, data availability and the fact that the literature, which provide the theoretical background for the benchmarking model itself, is still developing, there has been some limitations of the empirical analysis. Furthermore, the estimation of GHG emissions could potentially be improved by an interdisciplinary cooperation with e.g. engineers and biologists.

The limitations, entailed by the above, and the results of the analysis are discussed in the following section. Furthermore, perspectives for future research to improve the method and analysis is suggested. The section is split in to two subsections regarding limitations and perspectives for; *The benchmarking model* and *the GHG emissions*. As the limitations of the GHG estimation from Part A influence the benchmarking model from Part B, the interaction between these are also addressed in each of the subsections.

The benchmarking model

According to the results found in the empirical benchmarking analysis, the abatement cost of reducing 1 ton of CO_2e is much higher for dairy farms than the proposed tax on CO_2e of 1.500 DKK/ton. The size of the tax is estimated by the Danish Council on Climate Change as what the abatement cost will evidently be in order to reach the 70 pct. reduction target (The Danish Council on Climate Change, 2020). According to the results of this thesis, a tax of 1.500 DKK would not necessarily work as intended for the dairy sector. As reducing a ton of GHG is estimated to have a cost of approximately 4.000 DKK, or more dependent on the specific farm, farms would not have an incentive to reduce the GHG emissions. For the farms it will be less costly to pay a tax of 1.500 DKK than to reduce a ton of GHG emissions.

In the estimation of farm-specific GHG emissions, GHG reducing technologies that might be implemented on different farms, are not accounted for. These could include measures such as the different technologies proposed in the Climate Partner ships for the agricultural sector (The Danish Agriculture and Food Council; COWI, 2020) i.e. better manure management in the form of automatic slurry leaching and/or covering slurry tanks. Unfortunately, data is not available on which reduction technologies that might be implemented on different farms. By including the effect of different technologies to reduce GHG the estimates might look different.
As these technologies are not implemented in the GHG estimation, not all technology available today is necessary captured within the benchmarking model. However, It has not been possible to include the effect of specific actions, machinery etc. which could potentially already be in use for reducing the GHG emissions on a specific farm.

Furthermore, if new technology for reducing GHG emissions becomes available, the abatement costs might decrease to a level lower than the proposed tax, giving the farms an incentive to invest in new technology if this tax is adopted. However, with relatively high abatement costs for the Danish dairy sector, the approach towards reducing GHG emissions could be research and development of new technology in order to reduce the abatement costs.
It should also be noted, that independent of the level of the tax, the tax itself could create an incentive to develop or invest in new technology for reducing the GHG emissions and thereby the possible tax payments.

The empirical analysis illustrates that there exists a large potential for reducing GHG emissions within the sector today, given that the inefficient dairy farms catch up with the current technological lag in the sector. Usually in benchmarking analyses, catching up with the technological lag will be thought of as being "costless" for the farms, as it indicates that the inefficient farms can, without increasing their costs, restructure their production in order to operate as the efficient farms. As it is assumed that it is feasible for all farms to operate at the frontier, this implies that inefficient dairy farms are able to increase revenue and/or decrease GHG emissions until they reach the frontier. However, as the empirical analysis also shows, reaching a specific point of the frontier will, for inefficient farms, always be associated with some sort of trade-off between potentially reducing GHG emissions or increasing revenue. Reducing GHG emissions are thereby not assumed to be costless even for the inefficient farms, as actions taking towards reducing GHG emissions will always be done *as oppose* to initiating actions to increasing revenue.

Two types of second stage analyses have been conducted using the results from the estimated benchmarking models. Firstly, a second stage analysis examining the difference in the farm-specific average opportunity costs and secondly a second stage analysis examining the relative inefficiency. However, the focus in this thesis has been to develop and test the model and its ability to estimate abatement costs of GHG emissions. Therefore, the second stages analyses serve more as an example on how to utilize the information obtained from the benchmarking model.

By having the inefficient average opportunity costs for each farm, it has been possible to estimate an OLS regression to examine what is determinant for the average opportunity costs and which characteristics that tend to decrease these for the inefficient farms. This analysis only includes a few explanatory variables and could potentially be more complex to cover more characteristics of the farms. These characteristics should optimally be chosen such that they reflect parameters which are possible and desirable to change. The analysis could thereby detect where it is possible to pick the low hanging fruits within the sector by examining which types of farms, machinery, technology etc. that is associated with low abatement costs.

The second stage analysis for inefficiency, gives an indication of what characterize the efficient farms and what actions can be taken for the inefficient farms to reach the frontier. The findings are generally similar across both the GHG and revenue model. The second stage analysis does therefore not provide specific information on how to structure an inefficient farm in order to move towards a specific direction. This is most likely since for both of the two models, inefficient farms are generally located similar i.e. towards the south-east in the two-dimensional illustration of the production possibility set. Even though there might be some differences in the specific (in)efficiency score of a given farm across the two models, it is generally the same farms which are located close to the frontier. This is thereby rather independent of which specific direction, towards the north-west, the farms are projected in. It would be interesting to use a method which provides different characteristics of the farms, dependent on their location (i.e. if they are located with high revenue and low GHG emissions, or with low revenue and high GHG emissions). Such an analysis could potentially provide even more specific knowledge of how farms should structure their production in order to reach a specific part of the frontier.

The benchmarking model for the empirical analysis is output oriented, indicating that input for the dairy farms (total costs) are kept fixed while maximizing the desirable output (revenue) and minimizing the undesirable output (GHG emissions). However, as the different parts of the agricultural sector (including dairy farms) are generally characterized by being price takers, the possibility of maximizing revenue might be a questionable assumption. For price takers, output prices are fixed. This implies that a single dairy farm cannot itself influence the price of milk but is bound to the price determined on the world market. For price takers, profit maximization would thereby usually be done by minimizing costs rather than maximizing revenue. However, the benchmarking method applied in the empirical analysis allows for choosing the direction specifically. The method could even have been conducted non-oriented, meaning that the desirable output is maximized, the undesirable output is minimized *and* the input is minimized simultaneously. Nevertheless, for the specific empirical analysis of this report, an output-orientation allows for a clearer discussion of the trade-off between GHG emissions and revenue, and through that an estimation of the abatement cost of GHG emissions.

The frontier of the benchmarking model is estimated by assuming that the undesirable output, GHG emissions, is weakly disposable. However, this assumption implies that a part of the frontier is no longer strongly efficient. The inefficient part of the frontier is not dominant for locations within the production possibility set. This part of the frontier represents a technology where there is a positive relationship between increasing the desirable output and decreasing the undesirable output. This implies that the shadow price of reducing the undesirable output is negative at this part of the frontier. Therefore, the benchmark for some observations is no longer intuitive in its interpretation. In the empirical analysis, the farms, which are projected onto the inefficient part of the frontier, are excluded from the results. However, only a minority of the dairy farms are projected onto the inefficient part of the frontier. The effect on the results from this drawback of the model is thereby not necessarily determinant for the results of this specific analysis. Nevertheless, it could be a potentially large limitation of the model, especially if the model includes more variables which could cause a larger part of the frontier to be inefficient. Choosing a direction which only seeks to minimize the undesirable output, ensures that no DMUs are projected onto the inefficient part of the frontier. However, this projection might not always be desirable. It could therefore be necessary to research this area further in order to be able to apply the method in practice on different areas of analysis.

A basic assumption of the benchmarking method used in the empirical analysis is comparability across the analyzed observations - in this case dairy farms. For an inefficient farm to reach their full potential, they must perform similar to what is done at the point where the farm is projected onto the frontier. This implies, that it must be possible for the inefficient farm to operate exactly like the specific combination of efficient farms it is measured relative to. The dataset for the empirical analysis is constructed such that only specialized dairy farms are included in the analysis. However, there might be reasons why it is either not possible for an inefficient farm to restructure their production. There might also be cases where it is not desirable for society that farms act accordingly. If the inefficiency of certain dairy farms can be explained by differences in framework conditions, such as geography, which can affect their maximum possible performance, it might not be reasonable to assume that these farms can perform at the same level as the efficient farms. Furthermore, if efficient farms are characterized by e.g. a specific livestock composition it might not be desirable (or possible) that all inefficient farms restructure their production according to this. In the analysis, having a higher share of dairy cattle leads to a higher level of efficiency. However, it might not be realistic for all farms to act as the efficient farms if it means reducing the general number of non-dairy cattle in the sector as a whole.

The benchmarking model in the empirical analysis assumes constant return to scale for Danish dairy farms. As dairy farms operate on a competitive market, they should all be operating at the optimal scale. The measure of efficiency in the benchmarking model should therefore reflect this by not compensating farms based on their scale size. The assumption of CRS thereby gives all farms an incentive to operate at the most productive scale. However, this assumption could be discussed, if it is desirable to maintain a sector with farms of varying sizes, and the assumption of CRS could therefore be relaxed if more suitable for the aim of a given analysis.

It should be noted that the estimation of the benchmarking model only covers the year 2017. However, a standard benchmarking model is not necessarily able to handle observations for more than one period, unless the model is created with the purpose of accounting for the special condition

of having multiple periods. This is the case, as the standard benchmarking model assumes that the observations can produce as convex combinations of other observations. However, including the same observations for two or more periods, might result in a frontier created by convex combinations of the same observation across multiple periods. This benchmark can be hard to interpret and might in reality not necessarily be feasible.
However, it could still be possible to conduct a robustness check of the results found in this empirical analysis by computing the model using data for other years than 2017. It could be useful to conduct the analysis separately for several years, in order to verify the tendencies found from the estimation. Furthermore, for the results to act as a contribution to political decision making, it might be suitable to continuously estimate the model in order to follow the development in both efficiency, abatement costs as well as the technological development within the sector.

The model could potentially also be conducted for other parts of the agricultural sector than just dairy. Estimating abatement costs across the entire agricultural sector could be useful in examining the diversity across the sector. Furthermore, the model could potentially also be used outside the agricultural sector as a tool for calculating abatement costs. By doing this, it is possible to compare abatement costs across sectors and detect which sectors that are generally characterized by having relatively low abatement costs.

This thesis aims to verify the benchmarking method used in the empirical analysis, in order to make the method more reliable if it is expanded to be a more detailed benchmarking analysis, including more variables. As discussed earlier, it also possible to include factors such as animal welfare or environmental actions. By including animal welfare, it would be possible to detect the trade-off between improving animal welfare and increasing revenue, and thereby get a frontier shadow price, as well as an average opportunity cost of improving animal welfare. The trade-off between animal welfare and climate could also potentially be considered.

GHG emissions

The agricultural sector is characterized by being limited to the biological processes for both livestock and crops. This is a special characteristic of the agricultural sector which affects the development within the sector. There is a political goal for Denmark to be climate neutral by 2050. However, given the biological processes of livestock it is hard to imagine that livestock themselves will not produce any greenhouse gases. Even by developing technology that can reduce the methane emissions from enteric fermentation, it might only be possible to reduce emissions, but not entirely eliminate the emissions. However, climate neutrality can still be reached within the agricultural sector if the emissions from e.g. livestock is offset by processes that capture CO_2 such as afforestation and reforestation.

Currently, the data foundation of GHG emissions on farm level is weak. However, the focus on estimating GHG emissions on a more disaggregated level is widely discussed. Most recently it has been discussed in relation to the proposal of imposing a general Danish CO_2 tax and actions have been initiated to develop a tool for estimating farm-specific CO_2 accounts in the agricultural sector. It is also currently debated to implement climate accounts across EU, where it will be required by law to conduct climate accounts on every farm in all EU countries in a harmonized way. This law is currently negotiated and is expected to be adopted at the end of 2020 (Kristensen, 2020).

In this thesis, the IPPC guidelines have been used to estimate GHG emissions from different farms to include in the benchmarking. These estimations have been limited by the data availability and the relatively aggregated estimates of different components in the estimation. An example of this is the average value of digestibility of feed (DE) which regards the energy used for digesting feed. Ideally, values like this should vary per cattle, or alternatively simply per farm, to catch the variation across each farm. By having information on the different farm-specific feed schemes, it would be possible to estimate farm-specific DE.
Another estimate, which could have been more precise and farm-specific, is the actual days cattle spend on grass. This most likely differs across farms and is especially assumed to vary between conventional and organic farms. Nevertheless, with the current data availability, the best practice is to use the average days on grass, which is used by both by DCE in the National Inventory Reports and by the Danish Council on Climate Change. Furthermore, the methane emissions from enteric fermentation are dependent on the weight, as well as the daily weight gain of the cattle. In the estimation of GHG emissions in this thesis, breed specific averages are used. However, in reality there might be differences both across farms, but also across the specific dairy cattle within a farm. Specific data for the weight of each cattle is not currently available and would require an additional data collection from the farms.

This specific data on the herd could improve the GHG emission estimates and provide further knowledge of the relative difference across the farms. Nevertheless, it should be noted that collecting more data would require the farms to measure and submit more data, which can be costly as it is time consuming. As there always exists a trade-off between the complexity of the farm-specific data and the cost of collecting it, the added value of estimating more farm-specific GHG emissions should be weighed against the additional costs for the farms.

In this thesis, the GHG emissions of the dairy farms have only been estimated taking the emissions stemming from enteric fermentation and manure management into account. These sources of GHG emissions represent the largest share of the total GHG emissions from dairy farms. However, there are also other sources of GHG emissions which could have been included in the estimation. Farms may also have GHG emissions related to e.g. agricultural soils which is linked to any possible cultivated soils the farm might have. However, as these are rather complex to estimate and have a relatively small effect on the aggregated GHG emissions, these are not included. Nevertheless, as emissions from agricultural soils only represent a minority of the aggregated GHG emissions of dairy farms, it would likely not have changed the pattern found in this analysis.

Furthermore, the IPPC guidelines provided for estimating the GHG emissions from dairy farms do not consider the emission effects of the entire supply chain. Considering feed, some farms may import feed transported across a long distance, and thereby have a relatively larger carbon footprint. Land use and land use change and forestry (LULUCF) is neither included in the estimation. Therefore, the GHG estimations do not necessarily include the actual effect of using a large area of land for producing feed, or the alternative land use in the area. There are many components which could be taken into account and might be relevant in the estimation of GHG emissions on a disaggregated level, and the complexity of what to include is high. Furthermore, the results and the picture can change quite a lot dependent on the choice of what to include in the estimations of GHG.

Another approach to estimate the GHG emissions is to use a Life Cycle Assessment (LCA) where the climate and environmental effects are measured throughout the entire life cycle of a given product e.g. feed. This type of estimation might give a more precise estimate of the actual emitted GHG to the atmosphere of a given product or production. However, with this measure it is difficult to measure the share of the footprint actually reduced inside the national borders and LCA estimations are thereby not directly applicable in relation to the Danish reduction target of 70 pct. Nevertheless, the increasing greenhouse gases in the atmosphere is a global problem, and therefore all emissions around the world should ideally be included when estimating GHG emissions from Danish dairy farms. With that being said, measuring and reducing GHG worldwide might be too complex, and therefore it might be better to use a simpler and less precise approach than not being able to act because the issue it is too complex.

Following the formulation of the Danish climate law, agreed on in 2020, the reductions of GHG must be carried out in a way where the GHG emissions are not merely transferred outside the Danish boarders, a problem also referred to as CO_2e leakage. The wording in the objectives of the legislative text is, that Denmark is to be a pioneer for the international climate effort and thereby inspire the rest of the world.

When a country introduces actions to reduce GHG emissions, the inland reduction in emissions might not correspond 1:1 to the global reduction caused by the given actions. The leakage rate express how large a share of the decrease in inland CO_2e emissions are replaced by an increase of emissions in foreign countries (The Environmental Economic Council, 2019).

The leakage rate is relatively high for the Danish agricultural sector (The Environmental Economic Council, 2019). This indicates that a relatively large share of the inland GHG reductions within agriculture, will be offset by an increase in emissions outside Danish boarders. This is highly due to the fact that a decrease in the consumption of Danish produced food goods, will not necessarily lead to an overall reduction in the food consumption within Denmark, as agricultural products are relatively inelastic. On the contrary, it might lead to an increase in the import and consumption of food produced in foreign countries with similar or even higher GHG emissions within the production.

However, reducing GHG emissions in the agricultural sector will have a range of positive spillover effects, such as reducing the ammonia and nitrate leakage to the aquatic system (The Environmental Economic Council, 2019). Furthermore, Danish legislation to reduce GHG emissions from the agricultural sector might be complemented by similar international legislation. This could potentially reduce the GHG emissions from foreign agricultural producers, and thereby reduce the leakage rate for the Danish agricultural sector.

Conclusion

The aim of this thesis was to develop an applied benchmarking framework where GHG emissions are implemented in an economic efficiency analysis of the agricultural sector with the purpose of calculating farm-specific abatement costs of reducing GHG emissions.

The framework is based on the theoretical non-parametric framework of Data Envelopment Analysis. A key element of the applied benchmarking model is its ability to handle the undesirable output GHG emissions by assuming weak disposability and using a directional distance function to measure the relative efficiency of Danish dairy farms. In order to be able to conduct this analysis, it has been necessary to dedicate a part of the thesis to estimate the farm-specific GHG emissions based on various data sources and suggested national and international guidelines. However, there is still a large potential and need for improving the farm-specific GHG estimations both in relation to method and data limitations, as farm-specific estimations of GHG emissions are crucial to any future regulation of the agricultural sector.

The estimation of the GHG emissions could especially be improved by improving the data foundation for the estimation. This includes better use of the data already available today, as well as further collecting additional data on farm level. The interaction between climate accounts and its role in future regulation should therefore optimally be further developed in a corporation between multiple professional disciplines to ensure validity of the estimations of GHG emissions and its further use in regulation.

The theoretical framework and empirical applications of the benchmarking model applied in this thesis is still developing and there is only limited software available for computing the specific model. The method is in the literature generally used to calculate frontier shadow prices of an undesirable output. This thesis proposes an additional way of estimating the abatement costs of inefficient farms by calculating the average opportunity costs of reducing the undesirable output. The frontier shadow prices, determined by the slope of the frontier, represent the trade-off between increasing economic performance or climate performance for the efficient farms. These frontier shadow prices can thereby be interpreted as the marginal price of abating a ton of CO_2e when operating at best practice of what is currently technologically possible within the sector. As the

marginal frontier shadow prices changes along the frontier, the frontier shadow prices of inefficient farms are sensitive to the direction in which these are projected onto the frontier.
The average opportunity costs can serve as a measure of the average abatement cost of GHG emissions for inefficient farms. The benefit of the average opportunity costs is that these are not as sensitive to the specific directional vector as is the case for the frontier shadow prices.

The empirical analysis indicates that there currently exist improvement potentials on both climate- and economic performance for Danish dairy farms. There is a maximum potential in the sample of either reducing the aggregated GHG emissions with approximately 35 pct. corresponding to 638.600 tons CO_2e or increasing the aggregated revenue with approximately 28 pct. corresponding to 2,99 bill. DKK. These potentials are found by evaluating the current technological lag of the sector and assuming that all inefficient farms will eventually be able to catch up with this lag.

The average opportunity cost of reducing one ton of CO_2e is estimated to be 4.510 DKK across the conventional farms and 6.435 DKK across the organic farms. This indicates that on average the abatement costs of reducing GHG emissions are high within the Danish dairy sector compared to what is normally found as the abatement cost for GHG emissions in general. When assessing best practice within the sector, the frontier shadow prices also indicate that the reduction is very costly. The frontier shadow prices range between 966 DKK and 10.085 DKK for the conventional farms and between 3.894 DKK and 9.278 DKK for organic farms. However, it should be noted that only a minority of conventional farms are estimated to have a frontier shadow price below 2.394 DKK.

Both the estimated frontier shadow prices and the average opportunity costs indicate that it is costly for dairy farms to reduce GHG emissions. However, given the national and international agreements on reducing GHG emissions, all sectors must eventually reduce their current level of emissions. A tool to accommodate the current national reduction target of 70 pct. is the proposed CO_2 tax of 1.500 DKK per ton of CO_2. A tax following the polluters pay principle, is a cost-efficient way of reaching a socioeconomic optimal level of GHG emissions. The results found in this thesis implies that a tax of 1.500 DKK will not necessarily lead to the intended incentive for the Danish dairy farms, as the abatement cost is much higher than this tax. For the dairy sector to reduce their GHG emissions, it must therefore be acknowledged that this might be more costly than generally assumed.

If future regulation is not designed around a tax on CO_2, the results from the second stage analysis can be used to detect in which areas of the sector, the abatement costs are in general relatively low. By utilizing this information, it is possible to target incentives such that the reduction of GHG emissions within the dairy sector is done where it is less costly.

Bibliography

Ahmad, F. M., Ishtiaq, M., Hamid, K., Khurram, M. U., & Nawaz, A. (2017). Data Envelopment Analysis and Tobit Analysis for Firm Efficiency in Perspectivee of working Capital Management in Manufacturing Sector of Parkistan. *7(2)*, pp. 1-8. Retrieved from http: www.econjournals.com

Ali, A., & Seiford, L. M. (1990). Translation Invariance in Data Envelopment Analysis. *Operations Research Letters, 9*, pp. 403-405.

Asmild, M. (2019). *Landbrugsbarometer 2019.* Copenhagen: University of Copenhagen.

Banker, R. D. (1984). Estimating most productive scale size using Data Envelopment Analysis. *European Journal of Operational Research, 17*(1), pp. 35-44.

Banker, R. D., & Chang, H. (2005). The super-efficiency procedure for outlier identification, not for ranking efficient units. *European Journal of Operational Research*(175), pp. 1311–1320.

Bogetoft, P., & Otto, L. (2011). *Benchmarking with DEA, SFA and R* (Vol. 157). Springer Science & Business Media.

Chambers, R. G., Chung, Y., & Färe, R. (1998). Profit, directional distance functions, and Nerlovian efficiency. *Journal of Optimization Theory and Applications, 98*(2), pp. 351-364.

Charnes, A., Cooper, W. W., & Edwardo, R. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, pp. 429-444.

Chung, Y. H., Färe, R., & Grosskopf, S. (1997). Productivity and undesirable outputs: a directional distance function approach. *Journal of Environmental Management, 51*(3), pp. 229-240.

Cook, J., Nuccitelli, D., Green, S. A., Richardson, M., Winkler, B., Painting, R., . . . Skuce, A. (2013, May 15). Quantifying the consensous on anthropogenic global warming in the scientific litterature. *Environmental Research Letter, 8*(2).

Cooper, W. W., Kyung, S. P., & Jesus, P. T. (1999). AM: a range adjusted measure of inefficiency for use with additive models, and relations to other models and measures in DEA. *Journal of Productivity analysis, 11*(1), pp. 5-42.

Danish Agriculture and Food Council. (2018). Fakta om Fødevarerklyngen: Bæredygtig Udvikling.

DCE. (2019). *Denmark's National Inventory Report 2019, Emission Inventories 1990-2017 - Submitted under the United Nations Framework.* Danish Centre for Environment and Energy. Aarhus University.

Fare, R., Grosskopf, S., Knox, C. K., & Yaisawarng, S. (1993, May). Derivation of Shadow Prices for undesirable Outputs: A Distance Function Approach. *The MIT Press*, pp. 374-380.

Farrell, M. J. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society, 120*(3).

Fukuyama, H. a. (2008). Japanese banking inefficiency and shadow pricing. *Mathematical and Computer Modelling, 48*(11-12), pp. 1854-1867.

Golany, B., & Roll, Y. (1989). An application procedure for DEA. *Omega, 17*(3), pp. 237-250.

Hailu, A. (2003, November). Nonparametric Productivity Analysis with Undesirable Outputs: Reply. *American Journal of Agricultural Economics, 85*(4), pp. 1075-1077.

Hailu, A., & Veeman, T. S. (2001, August). Non-parametric Productivity Analysis with Undesirable Outputs: An Application to the Canadian Pulp and Paper Industry. *American Journal of Agricultural Economics, 83*(3), pp. 605-616.

Harkness, E. (2006). *CO2 Shadow Prices in the U.S. Electric Utility Industry: Calculating the Costs of Reducing CO2 emissions.* Tennessee: University of Tennessee, Knoxville.

IPCC. (2019). *Refinement to the 2006 IPCC Guidelines For National Greenhouse Gas Inventories.* Intergovernmental Panel on Climate Change.

Khezrimotlagh, D. (2013, July). How to Detect Outliers in Data Envelopment Analysis by Kourosh and Arash Method.

Kolstad, C. D. (2011). *Intermediate Environemntal Economics* (Vol. Second edition). New York: Oxford University Press.

Koopmans, T. C. (1951). Analysis of production as an efficient combination of activities. *Activity analysis of production and allocation*, pp. 33-97.

Kristensen, W. R. (2020). Klimaregnskab for samtlige landmænd i EU kan snart blive en realitet. Retrieved from https://ctwatch.dk/nyheder/klima/article12386674.ece

Kuosmanen, T., & Martin, R. K. (2010). *Duality of weakly disposable technology.* Omega.

Leleu, H. (2009). Mixing DEA and FDH models together. *Journal of the Operational Research Society, 60*(12), pp. 1730-1737.

Leleu, H. (2012). *Shadow pricing of undesirable outputs in nonparamtric analysis.* European Journal of Operational Research.

Leleu, H. (2013, May 29). Shadow pricing of undesirable outputs in nonparametric analysis. *European Journal of Operational Research, 231*, pp. 474-480.

Leleu, H., & Tarnaud, C. A. (2016, March). The duality of Shepard's weakly disposable technology. *Working paper series, No. 2016-EQM-06*.

Lillethorup, T. R. (2017). *Linking animal welfare and yields to economic performance in Danish dairy production A non-parametric approach using data envelopment analysis.* Master Thesis, University of Copenhagen, Copenhagen.

Lim, D.-J. (2020). Package "DJL".

Lim, D.-J., & Yang, D.-H. (2019). Lim, Dong-Joon, and Dong-Hyuk Yang. "Assessment of Regulatory Requirements on Technological Changes: The Increasing Dominance of Downsized Turbo/Super-Charged Engines Over Naturally Aspirated Engines. *IEEE Access, 7*, pp. 84839-84848.

Liu, J., & Junfei Zhang, Z. F. (2017). Tourism eco-efficiency of Chinese coastal cities e Analysis based on the DEA-Tobit model. *Ocean & Coastal Management, 148*, 164-170.

Lov om klima no. 965/2020. (2020).

Lund, P., & Aaes, O. (2016/2017). *Normtal for mængde og sammensætning af fæces og urin samt udskillelse af N, P og K i fæces og urin hos kvæg.* Retrieved from Department of Animal Science, Aarhus University: anis.au.dk

McDonald, J. (2008, July 28). Using least squares and tobit in second stage DEA efficiency analyses. *197*, pp. 792–798.

Mikkelsen, M. H. (2020). *Bestilling om rådgivning - data for metan/CO2 udledning fra kødkvæg kontra malkekvæg.* Aarhus University, Danish Centre For Food and Agriculture.

Podinovski, V. V. (2004). Production trade-offs and weight restrictions in data envelopment analysis. *Journal of the Operational Research Society, 55*(12), pp. 1311-1322.

Poulsen, H. D. (2017). *Normtal 2017.* Aarhus University, Department of Animal Science.

Paalberg, R. (2013). *Food Politics: What everyone needs to know* (Vol. Second Edition). Oxford, New York: Oxford University Press.

Saglam, Ü. (2018). A two-stage performance assessment of utility-scale wind farms in Texas using data envelopment analysis and Tobit models. *201*, pp. 580-598.

Scheel, H. (2001). Undesirable outputs in efficiency valuations. *European Journal of Operational Research, 132*(2), pp. 400-410.

SEGES. (2018). *Produktionsøkonomi, Kvæg 2018.* Aarhus: SEGES, Danish Agriculture and Food Council.

Sørensen, P. B., Elmeskov, J., Frederiksen, P., Jacobsen, J. B., Kristensen, N. B., Morthorst, P. E., & Richardson, K. (2016). *Afgifter der forandrer: Forslag til klimavenlige afgiftsomlægninger.* The Danish Council on Climate Change.

The Danish Agriculture and Food Council. (2018, October 11). Retrieved March 15, 2020, from Dyb lavkonjuktur i landbruget i 2018: https://lf.dk/aktuelt/nyheder/2018/oktober/dyb-lavkonjunktur-i-landbruget-i-2018

The Danish Agriculture and Food Council; COWI. (2020). Klimapartnerskabet for landbrug og fødevarersektoren.

The Danish Council on Climate Change. (2016). *Dokumentationsnotat: Klimarådets værktøj til beregning af drivhusgasudledningen fra landbrugsbedrifter.* The Danish Council on Climate Change.

The Danish Council on Climate Change. (2016). *Effektive veje til drivhusgasreduktion i landbruget.* The Danish Council on Climate Change.

The Danish Council on Climate Change. (2020). *Kendte veje og nye spor til 70 procents reduktion: Retning og tiltag for de næste ti års klimaindsats i Danmark.*

The Danish Ministry of finance. (2017). *Vejledning i samfundsøkonomiske konsekvensvurderinger.* Copenhagen: The Danish Ministry of Finance, Center for Competition, State-owned Enterprises and Utilities.

The Danish Ministry of Finance. (2019, December 18). Notat: Nøgletalskatalog.

The Danish Ministry of Finance. (2020). *Klimaaftale for energi og industri mv. 2020 af 22. juni 2020.* The Danish Ministry of Finance.

The Environmental Economic Council. (2019). *Economy and Environment.* The Economic Council.

The Ministry of Environment and Food. (2020). *Klima og landbrug*. Retrieved August 05, 2020, from https://mfvm.dk/landbrug/klima-og-landbrug/

The Ministry of Environment and food. (2020, June 23). *Landbruget får hjælp til at blive klogere på klimaet*. Retrieved August 2020, from mfvm.dk: https://mfvm.dk/nyheder/nyhed/nyhed/landbruget-faar-hjaelp-til-at-blive-klogere-paa-klimaet/#

Volden, H. (2011). *NorFor - The Nordic feed evaluation system.* Wageningen, The Netherlands: Wageningen Academic Publishers.

Wooldridge, J. M. (2016). *Introductory Economtrics* (Sixth edition ed.). Boston: Cengage learning.

Yuan, P., Cheng, S., Sun, J., & Liang, W. (2011, October 10). Measuringe the environmental efficiency of the Chinese industrial sector: A directional distance function approach. *Mathematical and Computer Modelling, 58*, pp. 936-947.

Appendix A Norm figures for GHG estimation

Table app. A 1: Average weight for dairy cattle

	Average Weight (kg)	Mature weight (kg)	Growth per day (kg)
Heavy breed	600	640	0,109
Jersey	420	440	0,068

Table app. A 2: Emission factors for manure management

	Housing Systems	CH_4	N_2O Direct	N_2O Indirect	Total GHG (CO_2e)
Dairy cattle - Heavy breed	Tethered with urine and solid manure	11,702	1,034	0,161	649
	Tethered with slurry	27,908	1,149	0,179	1.093
	Loose-holding with beds, solid floor	23,377	1,066	0,166	952
	Loose-holding with beds, slatted floor	23,377	1,108	0,173	966
	Loose-holding with beds, slatted floor, scrape	23,377	1,087	0,169	959
	Loose-holding with beds, drained floor	23,377	1,129	0,176	973
	Deep litter (all)	118,980	2,448	0,191	3.761
	Deep litter, long eating space, solid floor	97,912	1,959	0,186	3.087
	Deep litter, slatted floor	97,912	1,976	0,188	3.093
	Deep litter, slatted floor, scrape	97,912	1,968	0,187	3.090
Dairy cattle - Jersey	Tethered with urine and solid manure	8,749	0,859	0,134	514
	Tethered with slurry	19,899	0,954	0,149	826
	Loose-holding with beds, solid floor	18,133	0,885	0,138	758
	Loose-holding with beds, slatted floor	18,133	0,920	0,143	770
	Loose-holding with beds, slatted floor, scrape	18,133	0,902	0,141	764
	Loose-holding with beds, drained floor	18,133	0,937	0,146	776
	Deep litter (all)	98,728	2,033	0,158	3.121
	Deep litter, long eating space, solid floor	79,124	1,339	0,154	2.423
	Deep litter, slatted floor	79,124	1,367	0,156	2.432
	Deep litter, slatted floor, scrape	79,124	1,353	0,155	2.427
Heifer 0-6 months - Heavy breed	Deep litter (all)	1,959	0,419	0,032	183
	Deep litter, solid floor	1,959	0,419	0,032	183
Heifer 0-6 months - Jersey	Deep litter (all)	1,534	0,316	0,024	140
	Deep litter, solid floor	3,564	0,316	0,024	191
Heifer and stud 6-27 months - Heavy breed	Slatted floor-boxes	4,896	0,346	0,052	241
	Tethered with urine and solid manure	2,818	0,349	0,053	190
	Tethered with slurry	7,748	0,378	0,057	323

	Loose-holding with beds, solid floor	6,037	0,338	0,051	267
	Loose-holding with beds, slatted floor	6,037	0,359	0,054	274
	Loose-holding with beds, slatted floor, scrape	6,037	0,349	0,053	271
	Loose-holding with beds, drained floor	6,037	0,369	0,056	278
	Deep litter (all)	30,825	0,827	0,063	1.036
	Deep litter, long eating space, solid floor	25,414	0,655	0,060	848
	Deep litter, solid floor	27,251	0,810	0,061	941
	Deep litter, slatted floor	25,414	0,663	0,061	851
	Deep litter, slatted floor, scrape	25,414	0,659	0,060	850
Heifer and stud 6-27 months - Jersey	Slatted floor-boxes	4,999	0,260	0,039	214
	Tethered with urine and solid manure	2,369	0,262	0,040	149
	Tethered with slurry	6,175	0,284	0,043	252
	Loose-holding with beds, solid floor	5,396	0,254	0,038	222
	Loose-holding with beds, slatted floor	5,396	0,269	0,041	227
	Loose-holding with beds, slatted floor, scrape	5,396	0,262	0,040	225
	Loose-holding with beds, drained floor	5,396	0,277	0,042	230
	Deep litter (all)	26,139	0,630	0,048	855
	Deep litter, long eating space, solid floor	21,442	0,496	0,045	697
	Deep litter, solid floor	22,438	0,612	0,046	757
	Deep litter, slatted floor	21,442	0,502	0,046	699
	Deep litter, slatted floor, scrape	21,442	0,499	0,046	698
	Slatted floor-boxes	4,999	0,260	0,039	214
Bull 0-6 months - Heavy breed	Deep litter (all)	1,016	0,198	0,015	89
	Deep litter, solid floor	2,360	0,198	0,015	123
Bull 0-6 months - Jersey	Deep litter (all)	0,777	0,145	0,011	66
	Deep litter, solid floor	1,806	0,145	0,011	92
Bull > 6 months - Heavy breed	Slatted floor-boxes	2,038	0,159	0,024	106
	Tethered with urine and solid manure	1,583	0,164	0,025	96
	Tethered with slurry	4,717	0,178	0,027	179
	Loose-holding with beds, solid floor	3,104	0,159	0,024	132
	Loose-holding with beds, slatted floor	3,104	0,168	0,026	135
	Loose-holding with beds, slatted floor, scrape	3,104	0,164	0,025	134
	Loose-holding with beds, drained floor	3,104	0,173	0,026	137
	Deep litter (all)	14,861	0,387	0,029	495
	Deep litter, long eating space, solid floor	12,033	0,305	0,028	400
	Deep litter, solid floor	13,157	0,379	0,029	450
	Deep litter, slatted floor	12,033	0,309	0,028	401
	Deep litter, slatted floor, scrape	12,033	0,307	0,028	401
Bull > 6 months - Jersey	Slatted floor-boxes	1,615	0,127	0,019	84
	Tethered with urine and solid manure	0,828	0,128	0,019	65

Tethered with slurry	2,266	0,140	0,021	105
Loose-holding with beds, solid floor	1,828	0,124	0,019	88
Loose-holding with beds, slatted floor	1,828	0,132	0,020	91
Loose-holding with beds, slatted floor, scrape	1,828	0,128	0,019	90
Loose-holding with beds, drained floor	1,828	0,135	0,021	92
Deep litter (all)	11,454	0,302	0,023	383
Deep litter, long eating space, solid floor	9,102	0,237	0,022	305
Deep litter, solid floor	4,118	0,295	0,022	198
Deep litter, slatted floor	8,162	0,240	0,022	282
Deep litter, slatted floor, scrape	9,102	0,239	0,022	305

Appendix B Outlier detection

Table app. B 1: Second-step outlier detection for conventional farms

Mean inefficiency	Change in mean inefficiency (pct.-points)	Mean frontier shadow price (DKK)	Change in mean frontier shadow price (DKK)	Change in mean frontier shadow price (pct.)	Time used as peer
0,17	0,004	4.686	165	-3,7 pct.	975
0,18	0,003	4.372	-150	3,3 pct.	940
0,18	0,001	4.779	259	-5,7 pct.	83
0,18	0,000	4.506	-17	0,4 pct.	56
0,18	0,000	4.515	-9	0,2 pct.	22
0,18	0,000	4.522	-4	0,1 pct.	5
0,18	0,000	4.526	0	0,0 pct.	1

Appendix C Estimates for Tobit regression

Table app. c 1: Tobit regression for conventional farms

	Dependent variable: Inefficiency score	
	GHG	**Rev**
Number of cattle (100)	-0,0001***	-0,0001***
	p = 0,000	*p = 0,000*
Milk (DKK)/Total outputs (DKK)	0,009***	0,009***
	p = 0,000	*p = 0,000*
Fixed costs/Total costs (pct.)	0,002***	0,004***
	p = 0,000	*p = 0,000*
Dairy cattle / All cattle (pct.)	-0,005***	-0,005***
	p = 0,000	*p = 0,000*
Ownership (other than private)	-0,040***	-0,046***
	p = 0,000	*p = 0,000*
Cost to consulting - production (1.000 DKK)	0,00001	-0,0001
	p = 0,970	*p = 0,643*
Cost to consulting - cattle (1.000DKK)	0,0001	0,00002
	p = 0,107	*p = 0,772*
Cost to consulting - economic (1.000 DKK)	-0,00005	-0,0001**
	p = 0,350	*p = 0,044*
Share of jersey cattle (pct.)	-0,001***	-0,001***
	p = 0,000	*p = 0,000*
Share of cattle having a deep litter housing system (pct.)	0,002***	0,003***
	p = 0,000	*p = 0,000*
Milk production per dairy cattle (liter/cow)	-0,001***	-0,001*
	p = 0,002	*p = 0,098*
Constant	-0,045	-0,099**
	p = 0,152	*p = 0,021*
Squarred correlation	0,51	0,44
Observations	1.049	1.046
Log Likelihood	1.449,6	1.138
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$	

Table app. c 2: Tobit regression for organic farms

	Dependent variable: Inefficiency score	
	GHG	**Revenue**
Number of cattle	-0,0002***	-0,0003***
	p = 0,00002	*p = 0,00003*
Milk/Total outputs (DKK)	0,008***	0,009***
	p = 0,000	*p = 0,000*
Fixed costs/Total costs (DKK)	0,001	0,002
	p = 0,274	*p = 0,179*
Dairy cattle / all cattle	-0,008***	-0,010***
	p = 0,000	*p = 0,000*
Ownership - partnership	-0,015	-0,02
	p = 0,272	*p = 0,338*
Cost to consulting - production (DKK)	0,0003	0,0004
	p = 0,343	*p = 0,341*
Cost to consulting - cattle (DKK)	0,0003***	0,0004*
	p = 0,010	*p = 0,068*
Cost to consulting - economic (DKK)	-0,0001	-0,0001
	p = 0,551	*p = 0,346*
Share of jersey cattle	-0,002***	-0,002***
	p = 0,000	*p = 0,000*
Share of cattle having a deep litter housing system	0,003***	0,003***
	p = 0,000	*p = 0,000*
Milk production per dairy cattle	-0,005***	-0,006***
	p = 0,00001	*p = 0,0001*
Years since converted to organic	-0,002***	-0,003***
	p = 0,0001	*p = 0,0004*
Intercept	0,210**	0,371***
	p = 0,024	*p = 0,010*
Observations	204	195
Log Likelihood	283	201

Squared correlation	0,68	0,53
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$	

Appendix D R-script

```
#================================================#
####---- Part A: Estimating GHG Emissions ----####
#================================================#

####---- Clearing environment ----####
# rm(list=ls())

final<- read_excel("farm_data") # farm specific data from Ø90 after datacleaning following Lillthorup 2017

####---- Formatting variables ----#####
final$X5120 <-as.numeric(final$X5120)
final$X5121 <-as.numeric(final$X5121)
final$X5110 <-as.numeric(final$X5110)
final$X5120<-as.numeric(final$X5120)
final$X5154<-as.numeric(final$X5154)
final$X5152<-as.numeric(final$X5152)
final$X5110<-as.numeric(final$X5110)
final$X5152<-as.numeric(final$X5152)
final$X5154<-as.numeric(final$X5154)

####---- loading packages ----####
library(readxl)
library(extrafont)

#### Loading data ####
Man <- read_excel("Data from the fertilizer accounts") # Data for housing systems related to manure management
Cattle_5 <- read_excel("Data from Statistics Denmark") # Data regarding the population of cattle in Denmark 2017

####---- Customizing fonts ----####
# library(extrafont)
# windowsFonts()
library(extrafont)
# font_import()
# y
# loadfonts(device = "win")

#----------------------------------------------------#
#####----------- Enteric fermentation -----------#####
#----------------------------------------------------#

####---- Weight figures for Dairy cattle ----####
# Standard values for the weight of dairy cattle
weight_heavy<-600 # Lund & Aaes (2016/2017)
weight_jersey<-420 # Lund & Aaes (2016/2017)

Weights <-  cbind(c(600,420,510),c(640,440,540),c(40/365,25/365,(40/365+25/365)/2))
colnames(Weights)<-c("Average weight",  "Mature weight",        "Growth per day.  (kg)")
row.names(Weights)<-c("Heavy breed","Jersey","Mixed")

final$weight_mixed<-final$Andel_1*weight_heavy+final$Andel_3*weight_jersey
final$weight_mature_mixed<-final$Andel_1*Weights[1,2]+final$Andel_3*Weights[2,2]
final$Growth_mixed<-final$Andel_1*Weights[1,3]+final$Andel_3*Weights[2,3]

####---- NE_M: NE required for maintanence ----####

# Coefficient
CF_i<-0.29256 # Volden

final$NE_m<-0
for(i in 1:dim(final)[1]){
  if(final$X5100[i] == 1){
    final$NE_m[i]<-CF_i*(weight_heavy^0.75)}
  if(final$X5100[i] == 2){
    final$NE_m[i]<-CF_i*(weight_heavy^0.75)}
  if(final$X5100[i] == 3){
    final$NE_m[i]<-CF_i*(weight_jersey^0.75)}
  if(final$X5100[i] == 4){
    final$NE_m[i]<-CF_i*(final$weight_mixed[i]^0.75)}
}

####--- NE_a: Net Energy required for activity ---####

# Coefficients  from IPCC table 10.5
Ca_pasture<-0.17
Ca_stable<-0

final$NE_a<- final$NE_m*Ca_pasture*(18/365) # Assumption: 18 days on grass on average

####--- NE_g: Net Energy required for growth ---####

# Coefficient from IPCC for dairy cattle
C_cows <- 0.8

final$NE_g_cows <- 0
for( i in 1:dim(final)[1]){
  if(final$X5100[i] == 1){
    final$NE_g_cows[i]<- 22.02*((Weights[1,1]/(C_cows*Weights[1,2]))^0.75)*(Weights[1,3])^1.097
  }
  if(final$X5100[i] == 2){
    final$NE_g_cows[i]<- 22.02*((Weights[1,1]/(C_cows*Weights[1,2]))^0.75)*(Weights[1,3])^1.097
  }
  if(final$X5100[i] == 3){
    final$NE_g_cows[i]<- 22.02*((Weights[2,1]/(C_cows*Weights[2,2]))^0.75)*(Weights[2,3])^1.097
  }
  if(final$X5100[i] == 4){
    final$NE_g_cows[i]<- 22.02*((final$weight_mixed[i]/(C_cows*final$weight_mature_mixed[i]))^0.75)*(final$Growth_mixed[i])^1.097
  }
}

####--- NE_l: Net Energy required for lactating ---####
# Converting the quantity of energy corrected milk to quantity of regular milk
final$Milk_kv <- (((3140*final$X5118)/(383*final$X5120+242*final$X5121))/final$X5110)/365
final$NE_l <- final$Milk_kv*(1.47+0.4*final$X5120)

####---NE_p: Net Energy pregnancy--####
C_p<-0.1 # Coefficient from IPCC table 10.7
final$NE_p <- C_p*final$NE_m*(0.6*(284/365)
```

```r
                                )
####---REM---####
DE<-71 # From IPCC
final$REM  <-  1.123 -(4.092*10^(-3)*DE)+(1.126*10^(-5)*DE^2)-(25.4/DE)

####----REG----####
final$REG<- 1.164-(5.16*(10^(-3))*DE)+(1.308*(10^(-5))*(DE^2)-(37.4/DE))

####----GE----####
final$GE_Cows<- (((final$NE_m+final$NE_a+final$NE_l+final$NE_p)/final$REM) + (final$NE_g_cows/final$REG)) /(DE/100)

# Histograms of GE for dairy cattle
par(mfrow=c(1,3))
hist(final$GE_Cows[which(final$X5100==1|final$X5100==2 & final$year==2017)],breaks=20,main=NULL,xlim=c(200,500),xlab="Heavy breed",
     col="#8b9086",border="white",ylab="",
     col.lab="#8b9086",family="Times New Roman",cex.lab=2,cex.axis=1.7)
#abline(v = mean(final$GE_Cows[which(final$X5100==1|final$X5100==2 & final$year==2017)]), col="black", lwd=3, lty=1)
hist(final$GE_Cows[which(final$X5100==3 & final$year==2017)],breaks=10,main=NULL,xlim=c(200,500),ylab="",xlab="Jersey",
     col="#46626f",border="white",col.lab="#46626f",
     family="Times New Roman",cex.lab=2,cex.axis=1.7)
#abline(v = mean(final$GE_Cows[which(final$X5100==3 & final$year==2017)]), col="black", lwd=3, lty=1)
hist(final$GE_Cows[which(final$X5100==4& final$year==2017)],breaks=20,main=NULL,xlim=c(200,500),ylab="",xlab="Mixed breed",
     col="#a4864b",border="white",col.lab="#a4864b",
     family="Times New Roman",cex.lab=2,cex.axis=1.7)
#abline(v = mean(final$GE_Cows[which(final$X5100==4 & final$year==2017)]), col="black", lwd=3, lty=1)

#####---- Number of cattle types in the dataset ----####
sum(final$X5154[which(final$year==2017)])
sum(final$Heifers[which(final$year==2017)])
sum(final$Baby_heifer[which(final$year==2017)])
sum(final$Baby_bull[which(final$year==2017)])
sum(final$X5152[which(final$year==2017)])

#-------------------------------------------------------------#
####---- CH4 Emission factors for enteric fermentation ----####
#-------------------------------------------------------------#

# Emission factors for dairy cattle
Ym<-6 # From national inventory
final$EF_CH4_Ent_Cow<- (final$GE_Cows*(Ym/100)*365)/55.65
mean(final$EF_CH4_Ent_Cow)

# Emission factors for non-dairy cattle - from National inventory report Table 5.7 - same across breeds
final$EF_CH4_Ent_Baby_Bull<-13.05
final$EF_CH4_Ent_Baby_Heifer<-43.62
final$EF_CH4_Ent_Bull<-21.38
final$EF_CH4_Ent_Stud<-21.38
final$EF_CH4_Ent_Heifer<-55.51

####---- Total CH4 emissions from enteric fermentation for each farm ----####
final$CH4_Ent_Cow <- final$EF_CH4_Ent_Cow *final$X5110
final$CH4_Ent_Baby_Bull<-final$EF_CH4_Ent_Baby_Bull*final$Baby_bull
final$CH4_Ent_Baby_Heifer<-final$EF_CH4_Ent_Baby_Heifer*final$Baby_heifer
final$CH4_Ent_Bull<-final$EF_CH4_Ent_Bull*final$X5154
final$CH4_Ent_Stud<-final$EF_CH4_Ent_Stud*final$X5152
final$CH4_Ent_Heifer<-final$EF_CH4_Ent_Heifer*final$Heifers

final$CH4_Ent_total<-(   final$CH4_Ent_Cow
                     +final$CH4_Ent_Baby_Bull
                     +final$CH4_Ent_Baby_Heifer
                     +final$CH4_Ent_Bull
                     +final$CH4_Ent_Stud
                     +final$CH4_Ent_Heifer
                     )/1000

final$GHG_Ent<-final$CH4_Ent_total*25

####---- GHG emissions from enteric fermentation in CO2e - only dairy cattle ----####
final$GHG_Only_cows<-(final$CH4_Ent_Cow/1000)*25

#####---- Total GHG emissions from enteric fermentation in CO2e ----#####
# across all farms for all cattle
sum(final$GHG_Ent[which(final$year=="2017")])

# Table for summary of CH4 emission factors - enteric fermentation
Ef_Ent_Table<-rbind(c(summary(final$EF_CH4_Ent_Cow[which(final$year==2017 & final$X5100==1|final$X5100==2 )])[-3],
                      sd=sd(final$EF_CH4_Ent_Cow[which(final$year==2017 & final$X5100==1|final$X5100==2 )]))
,c(summary(final$EF_CH4_Ent_Cow[which(final$year==2017 & final$X5100==3 )])[-3],
   sd=sd(final$EF_CH4_Ent_Cow[which(final$year==2017 & final$X5100==3)]))
,c(summary(final$EF_CH4_Ent_Cow[which(final$year==2017 & final$X5100==4 )])[-3],
   sd=sd(final$EF_CH4_Ent_Cow[which(final$year==2017 & final$X5100==4)])))

rownames(Ef_Ent_Table)<-c("Heavy","Jersey","Mixed")
write.csv(Ef_Ent_Table, file = "Ef_Ent_Table.csv")

#-------------------------------------------------#
#####----------- Manure management -----------#####
#-------------------------------------------------#

# preparing data for housing systems
Man$Stable<-as.character(Man$Stable)
final$breed<-final$X5100
final$breed<-ifelse(final$breed==2,1,final$breed)

# Calculate individual CH4 emission factors for each farm
# based on breed, housing system and composition of dairy vs. non dairy cattle types

Types<-c("Cow","Bull","Studs","Heifiers")

#----------------------------------------------------#
####---- CH4 emissions from manure management ----####
#----------------------------------------------------#

final$CH4_Cow<-0
final$CH4_Bull<-0
final$CH4_Studs<-0
final$CH4_Heifiers<-0
CH4_obs<-0
breed_1<-0
breed_3<-0

vektor=as.character(1:10)

####---- Ch4 emissions for Dairy cattle, bulls, studs and heifers ----####
for (i in 1:dim(final)[1]){

  if(final$breed[i]==1|final$breed[i]==3){

    for(k in Types){
      CH4_obs<-0
    for(j in vektor){

    CH4_obs<-final[,j][i]*Man$CH4[which( Man$Stable == j & Man[,2]==k & Man$Breed==final$breed[i])]
    final[,paste("CH4_",k,sep="")][i]<-final[,paste("CH4_",k,sep="")][i]+CH4_obs

     }}
  }
```

```
  if(final$breed[i]==4){

    for(k in Types){
      breed_1<-0
      breed_3<-0
    for(j in vektor){
    breed_1<-final[,j][i]*final$Andel_1[i] *Man$CH4[which( Man$Stable == j & Man[,2]==k & Man$Breed==1)]
    breed_3<-final[,j][i]*final$Andel_3[i] *Man$CH4[which( Man$Stable == j & Man[,2]==k & Man$Breed==3)]
    final[,paste("CH4_",k,sep="")][i]<-final[,paste("CH4_",k,sep="")][i]+breed_1+breed_3
        }
    }
  }}

#### Heifers < 6 months

final$CH4_Baby_heifer<-0

for(i in 1:dim(final)[1]){
  if(final$breed[i]==1|final$breed[i]==3){
    final$CH4_Baby_heifer[i]<-Man$CH4[which(Man$Breed==final$breed[i] & Man$Type=="Baby_heifer" & Man$Stable ==7 )]
  }
  if(final$breed[i]==4){
    final$CH4_Baby_heifer[i]<-final$Andel_1[i]*Man$CH4[which(Man$Breed==1 & Man$Type=="Baby_heifer" & Man$Stable ==7 )]+
      final$Andel_3[i]*Man$CH4[which(Man$Breed==3 & Man$Type=="Baby_heifer" & Man$Stable ==7 )]
  }
}

## Bulls < 6 months

final$CH4_Baby_bull<-0

for(i in 1:dim(final)[1]){
  if(final$breed[i]==1|final$breed[i]==3){
    final$CH4_Baby_bull[i]<-Man$CH4[which(Man$Breed==final$breed[i] & Man$Type=="Baby_bull" & Man$Stable ==7 )]
  }
  if(final$breed[i]==4){
    final$CH4_Baby_bull[i]<-final$Andel_1[i]*Man$CH4[which(Man$Breed==1 & Man$Type=="Baby_bull" & Man$Stable ==7 )]+
      final$Andel_3[i]*Man$CH4[which(Man$Breed==3 & Man$Type=="Baby_bull" & Man$Stable ==7 )]
  }
}

####---- N2O direct emissions from manure management ----####

final$N2O_direct_Cow<-0
final$N2O_direct_Bull<-0
final$N2O_direct_Studs<-0
final$N2O_direct_Heifiers<-0

N2O_direct_obs<-0

####---- Direct N2O emissions for Dairy cattle, bulls, studs and heifers ----####

for (i in 1:dim(final)[1]){

  if(final$breed[i]==1|final$breed[i]==3){

    for(k in Types){
      N2O_direct_obs<-0
      for(j in vektor){

        N2O_direct_obs<-final[,j][i]*Man$N2O_direct[which( Man$Stable == j & Man[,2]==k & Man$Breed==final$breed[i])]
        final[,paste("N2O_direct_",k,sep="")][i]<-final[,paste("N2O_direct_",k,sep="")][i]+N2O_direct_obs

      }}
  }

  if(final$breed[i]==4){

    for(k in Types){
      breed_1<-0
      breed_3<-0
      for(j in vektor){
        breed_1<-final[,j][i]*final$Andel_1[i] *Man$N2O_direct[which( Man$Stable == j & Man[,2]==k & Man$Breed==1)]
        breed_3<-final[,j][i]*final$Andel_3[i] *Man$N2O_direct[which( Man$Stable == j & Man[,2]==k & Man$Breed==3)]
        final[,paste("N2O_direct_",k,sep="")][i]<-final[,paste("N2O_direct_",k,sep="")][i]+breed_1+breed_3
      }
    }
  }}

## Heifers < 6 months

final$N2O_direct_Baby_heifer<-0

for(i in 1:dim(final)[1]){
  if(final$breed[i]==1|final$breed[i]==3){
    final$N2O_direct_Baby_heifer[i]<-Man$N2O_direct[which(Man$Breed==final$breed[i] & Man$Type=="Baby_heifer" & Man$Stable ==7 )]
  }
  if(final$breed[i]==4){
    final$N2O_direct_Baby_heifer[i]<-final$Andel_1[i]*Man$N2O_direct[which(Man$Breed==1 & Man$Type=="Baby_heifer" & Man$Stable ==7 )]
    + final$Andel_3[i]*Man$N2O_direct[which(Man$Breed==3 &  Man$Type=="Baby_heifer" & Man$Stable ==7 )]
  }
}

## Bulls < 6 months

final$N2O_direct_Baby_bull<-0

for(i in 1:dim(final)[1]){
  if(final$breed[i]==1|final$breed[i]==3){
    final$N2O_direct_Baby_bull[i]<-Man$N2O_direct[which(Man$Breed==final$breed[i] & Man$Type=="Baby_bull" & Man$Stable ==7 )]
  }
  if(final$breed[i]==4){
    final$N2O_direct_Baby_bull[i]<-final$Andel_1[i]*Man$N2O_direct[which(Man$Breed==1 & Man$Type=="Baby_bull" & Man$Stable ==7 )]
    + final$Andel_3[i]*Man$N2O_direct[which(Man$Breed==3  & Man$Type=="Baby_bull" & Man$Stable ==7 )]
  }
}

####---- N2O indirect emissions ----####

final$N2O_indirect_Cow<-0
final$N2O_indirect_Bull<-0
final$N2O_indirect_Studs<-0
final$N2O_indirect_Heifiers<-0
N2O_indirect_obs<-0

####---- Indirect N2O emissions for Dairy cattle, bulls, studs and heifers ----####

for (i in 1:dim(final)[1]){

  if(final$breed[i]==1|final$breed[i]==3){

    for(k in Types){
      N2O_indirect_obs<-0
      for(j in vektor){

        N2O_indirect_obs<-final[,j][i]*Man$N2O_indirect[which( Man$Stable == j & Man[,2]==k & Man$Breed==final$breed[i])]
        final[,paste("N2O_indirect_",k,sep="")][i]<-final[,paste("N2O_indirect_",k,sep="")][i]+N2O_indirect_obs
```

```
      }}
  }

  if(final$breed[i]==4){

    for(k in Types){
      breed_1<-0
      breed_3<-0
      for(j in vektor){
        breed_1<-final[,j][i]*final$Andel_1[i] *Man$N2O_indirect[which( Man$Stable == j & Man[,2]==k & Man$Breed==1)]
        breed_3<-final[,j][i]*final$Andel_3[i] *Man$N2O_indirect[which( Man$Stable == j & Man[,2]==k & Man$Breed==3)]
        final[,paste("N2O_indirect_",k,sep="")][i]<-final[,paste("N2O_indirect_",k,sep="")][i]+breed_1+breed_3
      }
    }
  }}

## Heifiers < 6 months

final$N2O_indirect_Baby_heifer<-0

for(i in 1:dim(final)[1]){
  if(final$breed[i]==1|final$breed[i]==3){
    final$N2O_indirect_Baby_heifer[i]<-Man$N2O_indirect[which(Man$Breed==final$breed[i] & Man$Type=="Baby_heifer" & Man$Stable ==7 )]
  }
  if(final$breed[i]==4){
    final$N2O_indirect_Baby_heifer[i]<-final$Andel_1[i]*Man$N2O_indirect[which(Man$Breed==1 & Man$Type=="Baby_heifer" & Man$Stable ==7 )]
    + final$Andel_3[i]*Man$N2O_indirect[which(Man$Breed==3  & Man$Type=="Baby_heifer" & Man$Stable ==7 )]
  }
}

## Bulls < 6 months

final$N2O_indirect_Baby_bull<-0

for(i in 1:dim(final)[1]){
  if(final$breed[i]==1|final$breed[i]==3){
    final$N2O_indirect_Baby_bull[i]<-Man$N2O_indirect[which(Man$Breed==final$breed[i] & Man$Type=="Baby_bull" & Man$Stable ==7 )]
  }
  if(final$breed[i]==4){
    final$N2O_indirect_Baby_bull[i]<-final$Andel_1[i]*Man$N2O_indirect[which(Man$Breed==1 & Man$Type=="Baby_bull" & Man$Stable ==7 )]+
      final$Andel_3[i]*Man$N2O_indirect[which(Man$Breed==3 & Man$Type=="Baby_bull" & Man$Stable ==7 )]
  }
}

####---- CH4 and N2O emissions aggregated on farm level ----####

####---- CH4 emissions ----####

## CH4 emissions from manure management - for each cattle type
final$CH4_Cow_farm<-final$CH4_Cow*as.numeric(final$X5110)
final$CH4_Bull_farm<-final$CH4_Bull*as.numeric(final$X5154)
final$CH4_Stud_farm<-final$CH4_Studs*as.numeric(final$X5152)
final$CH4_Heifers_farm<-final$CH4_Heifiers*as.numeric(final$Heifers)
final$CH4_Baby_Heifers_farm<-final$CH4_Baby_heifer*as.numeric(final$Baby_heifer)
final$CH4_Baby_Bull_farm<-final$CH4_Baby_bull*as.numeric(final$Baby_bull)

## CH4 emissions from manure management in tons - across all cattle types
final$CH4_total<-
  (final$CH4_Cow_farm+
  final$CH4_Bull_farm+
  final$CH4_Stud_farm+
  final$CH4_Heifers_farm+
  final$CH4_Baby_Heifers_farm+
  final$CH4_Baby_Bull_farm)/1000

## CH4 emissions from manure management in tons CO2e - across all cattle types
final$CO2e_CH4<-final$CH4_total*25

## CH4 emissions from manure management in tons CO2e - only dairy cattle
final$GHG_Ch4_man_Only_Cows<-final$CH4_Cow_farm*25/1000
sum(final$GHG_Ch4_man_Only_Cows[which(final$year==2017)])

####---- N2O direct emissions ----####

## N2O_direct emissions from manure management - for each cattle type
final$N2O_direct_Cow_farm<-final$N2O_direct_Cow*as.numeric(final$X5110)
final$N2O_direct_Bull_farm<-final$N2O_direct_Bull*as.numeric(final$X5154)
final$N2O_direct_Stud_farm<-final$N2O_direct_Studs*as.numeric(final$X5152)
final$N2O_direct_Heifers_farm<-final$N2O_direct_Heifiers*as.numeric(final$Heifers)
final$N2O_direct_Baby_Heifers_farm<-final$N2O_direct_Baby_heifer*as.numeric(final$Baby_heifer)
final$N2O_direct_Baby_Bull_farm<-final$N2O_direct_Baby_bull*as.numeric(final$Baby_bull)

## N2O direct emissions from manure management in tons - across all cattle types
final$N2O_direct_total<-
  (final$N2O_direct_Cow_farm+
     final$N2O_direct_Bull_farm+
     final$N2O_direct_Stud_farm+
     final$N2O_direct_Heifers_farm+
     final$N2O_direct_Baby_Heifers_farm+
     final$N2O_direct_Baby_Bull_farm)/1000

## N2O direct emissions from manure management in tons CO2e - across all cattle types
final$CO2e_N2O_direct<-final$N2O_direct_total*298
sum(final$CO2e_N2O_direct[which(final$year=="2017")])

## N2O direct emissions from manure management in tons CO2e - Only dairy cattle
final$N2O_direct_Cow_farm*298

#### N2O_indirect emissions ####

## N2O indirect emissions from manure management - for each cattle type
final$N2O_indirect_Cow_farm<-final$N2O_indirect_Cow*as.numeric(final$X5110)
final$N2O_indirect_Bull_farm<-final$N2O_indirect_Bull*as.numeric(final$X5154)
final$N2O_indirect_Stud_farm<-final$N2O_indirect_Studs*as.numeric(final$X5152)
final$N2O_indirect_Heifers_farm<-final$N2O_indirect_Heifiers*as.numeric(final$Heifers)
final$N2O_indirect_Baby_Heifers_farm<-final$N2O_indirect_Baby_heifer*as.numeric(final$Baby_heifer)
final$N2O_indirect_Baby_Bull_farm<-final$N2O_indirect_Baby_bull*as.numeric(final$Baby_bull)

## N2O indirect emissions from manure management in tons - across all cattle types
final$N2O_indirect_total<-
  (final$N2O_indirect_Cow_farm+
     final$N2O_indirect_Bull_farm+
     final$N2O_indirect_Stud_farm+
     final$N2O_indirect_Heifers_farm+
     final$N2O_indirect_Baby_Heifers_farm+
     final$N2O_indirect_Baby_Bull_farm)/1000

## N2O indirect emissions from manure management in tons CO2e - across all cattle types
final$CO2e_N2O_indirect<-final$N2O_indirect_total*298
sum(final$CO2e_N2O_indirect[which(final$year=="2017")])

#### N2O total emissions: direct+indirect ####
final$N2O_total<-final$N2O_direct_total+final$N2O_indirect_total
final$C02e_N2O_total<-final$N2O_total*298

####---- Total CO2e in tons from manure management: CH4 + N2O ----####
final$GHG_man<-final$CO2e_CH4+final$C02e_N2O_total
```

```
#-------------------------------------------------------#
#####---- The GHG variable used for benchmarking ----####
#-------------------------------------------------------#

####---- Aggregating emissions from enteric fermentation and manure management ----####
final$GHG<-final$GHG_Ent+final$GHG_man

####---- validation of GHG emission estimates ----####

####---- Comparison of population and sample ----####

3730069 # Total emissions in CO2e from Enteric fermentation from the agricultural sector - National Inventory Report
0.87 # Cattle represent 87% of total GHG emissions from enteric fermentation - National Inventory Report
0.7 # Dairy cattle represent 70% of all emissions from enteric fermentation from cattle - National inventory Annex 13
571114.5  # Total number of dairy cattle in 2017 - Statistics Denmark (table "Kvæg 5")

# Total emissions from enteric fermentation in the population - only dairy cattle
3730069*0.87*0.7*0.51

#Entiric fermentation emissions - from the population of cattle
3730069*0.87/(72490*25*0.36+3730069*0.87+351000)
72490*25*0.36/(72490*25*0.36+3730069*0.87+351000)
351000/(72490*25*0.36+3730069*0.87+351000)

# Share of dairy cattle in  data sample vs. population
Andel_cows<-sum(final$X5110[which(final$year==2017)])/571114.5

# What the total emissions from enteric fermentation should be for the sample
3730069*0.87*0.7*Andel_cows

# What the total emissions from enteric fermentation are for the sample
sum(final$GHG_Only_cows[which(final$year=="2017")])
sum(final$GHG_Only_cows[which(final$year=="2017")])/(3730069*0.87*0.7*Andel_cows)

### manure - sammenlign med populationen
72490*25 # Total CH4 emissions measured in CO2e for the agricultural sector - National Inventory Report
0.36 # The share which cattle contributes with - National Inventory Report
0.534 # The share of the total emissions from cattle which originate from adiry cattle - National Inventory Report

# Aggregated emissions of CH4 from manure management for the population of dairy cattle
72490*25*0.36*0.534

# What the sample should correspond to
72490*25*0.36*0.534*Andel_cows
72490*25*0.36*0.534/25

# Estmation from the sample
sum(final$GHG_Ch4_man_Only_Cows[which(final$year==2017)])
sum(final$GHG_Ch4_man_Only_Cows[which(final$year==2017)])/(72490*25*0.36*0.534*Andel_cows)
sum(final$GHG_Ch4_man_Only_Cows[which(final$year==2017)])/sum(final$CO2e_CH4[which(final$year==2017)])

# The total number of cattle in the sample
sum(final$X5110[which(final$year==2017)]
,final$Baby_bull[which(final$year==2017)]
,final$Baby_heifer[which(final$year==2017)]
,final$X5154[which(final$year==2017)]
,final$X5152[which(final$year==2017)]
,final$Heifers[which(final$year==2017)])

# The Danish population of cattle
1561147.25

# The sample share of the total population
andel_samlet_kvaeg<-sum(final$X5110[which(final$year==2017)]
    ,final$Baby_bull[which(final$year==2017)]
    ,final$Baby_heifer[which(final$year==2017)]
    ,final$X5154[which(final$year==2017)]
    ,final$X5152[which(final$year==2017)]
    ,final$Heifers[which(final$year==2017)])/1561147.25

72490*25*0.36*andel_samlet_kvaeg

### N2O Emissions in sample vs. population
# What the emissions should be for the sample
750000*Andel_cows
750*298*Andel_cows

# The aggregated estimated N2O emissions - in N2O
sum(final$N2O_direct_Cow_farm[which(final$year==2017)])+sum(final$N2O_indirect_Cow_farm[which(final$year==2017)])

####---- Distribution of housing types in the data ----####

for (i in as.character(1:10)){
assign(paste("Housing_",i,sep=""),final[,i]*final$X5110)
}

Housing_all<-sum(sum(Housing_1)
,sum(Housing_2)
,sum(Housing_3)
,sum(Housing_4)
,sum(Housing_5)
,sum(Housing_6)
,sum(Housing_7)
,sum(Housing_8)
,sum(Housing_9)
,sum(Housing_10))

Stable_share<-c(
sum(Housing_1)/Housing_all
,sum(Housing_2)/Housing_all
,sum(Housing_3)/Housing_all
,sum(Housing_4)/Housing_all
,sum(Housing_5)/Housing_all
,sum(Housing_6)/Housing_all
,sum(Housing_7)/Housing_all
,sum(Housing_8)/Housing_all
,sum(Housing_9)/Housing_all
,sum(Housing_10)/Housing_all
)

# write.csv(Stable_share, file = "Stable_share.csv")

# Table for share of cattle in dataset vs. population
No_cattle_types<-c(
  sum(final$X5110[which(final$year==2017)])
  ,sum(final$Heifers[which(final$year==2017)])
  ,sum(final$Baby_heifer[which(final$year==2017)])
  ,sum(sum(final$X5154[which(final$year==2017)]),sum(final$X5152[which(final$year==2017)]))
  ,sum(final$Baby_bull[which(final$year==2017)]))

Table_cattle<-cbind(Cattle_5,No_cattle_types,No_cattle_types/Cattle_5[,2])
names(Table_cattle)<-c("Type","Population","Our sample","Share of population")

# write.csv(Table_cattle, file = "Table_cattle.csv")

####---- Plot for composition of GHG emissions for each farm in the dataset ----####

Matrix_for_barplot<-cbind(final$GHG
```

```
                          ,final$GHG-final$C02e_N2O_total
                          ,final$GHG-final$C02e_N2O_total-final$CO2e_CH4,final$year)

Matrix_for_barplot<-Matrix_for_barplot[order(final$GHG),]

par(mfrow=c(1,1))
par(family = "Times New Roman")
barplot(Matrix_for_barplot[which(Matrix_for_barplot[,4]==2017),1], border="#46626f",ylab="",cex.lab=1.4,family="Times New Roman",xlab="")

mtext(text = expression(paste(CO[2],"e in ton")),
      side = 2, #side 2 = left
      line = 1.9,cex = 1.5,family="Times New Roman")

mtext(text = "Farms",
      side = 1, #side 2 = left
      line = 0.6,cex = 1.5,family="Times New Roman")

barplot(Matrix_for_barplot[which(Matrix_for_barplot[,4]==2017),2], border="#ddd3c8",add=TRUE,family="Times New Roman")
barplot(Matrix_for_barplot[which(Matrix_for_barplot[,4]==2017),3], border="#8b9086",add=TRUE,family="Times New Roman")

legend("topleft", c(
                   expression(paste(N[2],"O - Manure Management")),expression(paste(CH[4]," - Manure management")),
                   expression(paste(CH[4]," - Enteric fermentation"))), col = c("#46626f", "#ddd3c8", "#8b9086"),
       text.col = "black", pch = 15,
       bg = "white",cex = 1.5,pt.cex=4,bty="n")


final$All_cattle<-final$X5110 +final$Baby_bull+final$Baby_heifer+final$X5154+final$X5152+final$Heifers

####---- plot for correlation between total number of cattle on a farm and total GHG emissions ----####
par(mfrow=c(1,1))
plot(final$All_cattle[which(final$year==2017)],final$GHG[which(final$year==2017)],
     xlab="Total number of cattle", ylab="Total GHG emissions",col="#8b9086",pch=16,frame.plot = FALSE
     ,xlim=c(0,max(final$All_cattle[which(final$year==2017)])+100),ylim=c(0,max(final$GHG[which(final$year==2017)])+100),
     cex.lab=1.5,cex.axis=1.5)
```

```r
#===========================================#
####---- Part B: Benchmarking analysis ----####
#===========================================#

####---- Clearing environment ----####
rm(list=ls())

####---- Installing packages ----####
library(lpSolveAPI)
library(readxl)
library(plotly)
library(Benchmarking)
library(lmtest)
library(car)
library(AER)
library(mfx)
library(nlme)
library(stargazer)

####---- defining function for geometrical mean ----####
geo <- function(x) {prod(x)^(1/length(x))}

####---- Customizing fonts ----####
# library(extrafont)
# windowsFonts()
library(extrafont)
# font_import()
# y
# loadfonts(device = "win")

####---- Loading data ----####
final_dea <- read_excel("Data")

# Defining new variables

final_dea$All_cattle<-final_dea$X5110+final_dea$Baby_bull+final_dea$Baby_heifer+final_dea$X5154+final_dea$X5152+final_dea$Heifers
final_dea$Year_org<-0
final_dea$Year_org[which(final_dea$type==1)]<-2020-final_dea$X6408[which(final_dea$type==1)]
final_dea$deep_litter_share<-(final_dea$`7`+final_dea$`8`+final_dea$`9`+final_dea$`10`)
final_dea$counseling_total<-final_dea$X7825+final_dea$X7826+final_dea$X7841
final_dea$X6404<-as.numeric(final_dea$X6404)
final_dea$X6404[which(final_dea$X6404==0)]<-0
final_dea$X6404[which(final_dea$X6404==1)]<-0
final_dea$X6404[which(final_dea$X6404==2)]<-1
final_dea$X6404[which(final_dea$X6404==3)]<-1
final_dea$X6404[which(final_dea$X6404==4)]<-1
final_dea$X6404[which(final_dea$X6404==5)]<-1


####---- Subsetting for the year 2017 ----####
Data_2017<-final_dea[which(final_dea$year==2017),]

####---- Removing outliers ----####
Data_2017<-Data_2017[-which(Data_2017$SysNr==44085),]
# The outlier found from the outlier detection is not removed when running the section for outlier detection in ths script

Data_2017<-data.frame(Data_2017)

#-------------------------------------------------------------------------------------------#
####---- Defining the shortage function - modified source code from the DJL package ----####
#-------------------------------------------------------------------------------------------#

dm.sf.new <-function(xdata, ydata, rts = "crs", g = NULL,
                     wd = NULL, se = FALSE, sg = "ssm", date = NULL, cv = "convex", o = NULL){

  # Initial checks
  if(is.na(match(rts, c("crs", "vrs", "irs", "drs")))) stop('rts must be "crs", "vrs", "irs", or "drs".')
  if(is.na(match(se,  c(0, 1, FALSE, TRUE))))          stop('se must be either 0(FALSE) or 1(TRUE).')
  if(is.na(match(sg,  c("ssm", "max", "min"))))        stop('sg must be "ssm", "max", or "min".')
  if(is.na(match(cv,  c("convex", "fdh"))))            stop('cv must be "convex" or "fdh".')
  if(!is.null(o) && !all(o <= nrow(xdata)))            stop('o must be element(s) of n.')

  # Load library
  # library(lpSolveAPI)

  # Parameters
  xdata <- as.matrix(xdata)
  ydata <- as.matrix(ydata)
  g     <- if(is.null(g)) cbind(xdata, ydata) else as.matrix(g)
  date  <- if(!is.null(date)) as.matrix(date)
  n     <- nrow(xdata)
  m     <- ncol(xdata)
  s     <- ncol(ydata)
  wd    <- if(is.null(wd)) matrix(c(0), ncol = s) else as.matrix(wd)
  se    <- ifelse(is.logical(se), ifelse(isTRUE(se), 1, 0), se)
  rts   <- ifelse(cv == "fdh", "vrs", rts)
  o     <- if(is.null(o)) c(1:n) else as.vector(o)

  # Data frames
  results.efficiency <- matrix(NA, nrow = n, ncol = 1)
  results.lambda     <- matrix(NA, nrow = n, ncol = n)
  results.mu         <- matrix(NA, nrow = n, ncol = n)
  results.xslack     <- matrix(NA, nrow = n, ncol = m)
  results.yslack     <- matrix(NA, nrow = n, ncol = s)
  results.w          <- matrix(NA, nrow = n, ncol = m)
  results.p          <- matrix(NA, nrow = n, ncol = s)
  results.u          <- matrix(NA, nrow = n, ncol = 1)

  # LP
  for (k in o){
    # Declare LP
    lp.sf <- make.lp(0, n + n + 1 + m + s) # lambda+mu+efficiency+xslack+yslack

    # Set objective
    set.objfn(lp.sf, c(-1), indices = c(n + n + 1))

    # RTS
    if(rts == "vrs") add.constraint(lp.sf, c(rep(1, n*2)), indices = c(1:(n*2)), "=", 1)
    if(rts == "crs") set.constr.type(lp.sf, 0, 1)
    if(rts == "irs") add.constraint(lp.sf, c(rep(1, n*2)), indices = c(1:(n*2)), ">=", 1)
    if(rts == "drs") add.constraint(lp.sf, c(rep(1, n*2)), indices = c(1:(n*2)), "<=", 1)

    # Set type
    if(cv == "fdh") set.type(lp.sf, 1:n, "binary")

    # Mu
    if(rts == "crs" || rts == "drs" || sum(wd) == 0) add.constraint(lp.sf, c(rep(1, n)),
                                                                    indices = c((n + 1):(n + n)), "=", 0)

    # Input constraints
    for(i in 1:m) add.constraint(lp.sf, c(xdata[, i], xdata[, i], g[k, i], 1),
                                 indices = c(1:n, (n + 1):(n + n), n + n + 1, n + n + 1 + i), "=", xdata[k, i])

    # Output constraints
    for(r in 1:s){
      if(wd[1, r] == 1){
        add.constraint(lp.sf, c(ydata[, r], g[k, m + r]),
```

```r
                        indices = c(1:n, n + n + 1), "=", ydata[k, r])
        add.constraint(lp.sf, c(1), indices = c(n + n + 1 + m + r), "=", 0)
      }else{
        add.constraint(lp.sf, c(ydata[, r], -g[k, m + r], -1),
                        indices = c(1:n, n + n + 1, n + n + 1 + m + r), "=", ydata[k, r])
      }
    }

    # PPS for Super
    if(se == 1) add.constraint(lp.sf, c(1, 1), indices = c(k, n + k), "=", 0)

    # Bounds
    set.bounds(lp.sf, lower = c(rep(0, n + n), -Inf, rep(0, m + s)))

    # Solve
    solve.lpExtPtr(lp.sf)

    # Get results
    results.efficiency[k] <- -1 * get.objective(lp.sf)
    temp.p                <- get.variables(lp.sf)
    results.lambda[k,]    <- temp.p[1:n]
    results.mu[k,]        <- temp.p[(n + 1):(n + n)]
    results.xslack[k,]    <- temp.p[(n + n + 2):(n + n + 1 + m)]
    results.yslack[k,]    <- temp.p[(n + n + 1 + m + 1):(n + n + 1 + m + s)]
    temp.d                <- -1 * get.dual.solution(lp.sf)
    results.u[k,]         <- temp.d[2]
    results.w[k,]         <- temp.d[4:(3 + m)]
    results.p[k,]         <- temp.d[(3 + m + 1):(3 + m + s)]

  }
  results <- list(eff = results.efficiency, lambda = results.lambda, mu = results.mu,
                  xslack = results.xslack, yslack = results.yslack, w = results.w, p = results.p, u = results.u)
  return(results)
}

####---- Creating variables for benchmarking ----####
## Input
Data_2017$X<-Data_2017$feed + Data_2017$labour_tot + Data_2017$ovc + Data_2017$fix +  Data_2017$cap
X<-Data_2017$X

# Subsetting for conventional and organic
X_con<-Data_2017$X[which(Data_2017$type==0)]
X_org<-Data_2017$X[which(Data_2017$type==1)]

## Outputs
y_REV<-Data_2017$milk + Data_2017$oo
y_GHG<-Data_2017$GHG
Data_2017$Y<-cbind(y_REV,y_GHG)
Y<-Data_2017$Y

# For conventional and organic farms
Y_con<-Data_2017$Y[which(Data_2017$type==0),]
Y_org<-Data_2017$Y[which(Data_2017$type==1),]

####---- Defining RTS ----####
rts = "crs"

####---- Defining direction ----####
# The specific direction is chosen given the specific model under analysis

g <- cbind(0.000000000000000000000000000000000000000001,Y[,1], Y[,2]) # output oriented
# g <- cbind(0.000000000000000000000000000000000000000001,0.000000000000000000000000000000000000000001, Y[,2]) # Only reducing GHG
# g <- cbind(0.000000000000000000000000000000000000000001, Y[,1],0.000000000000000000000000000000000000000001) # Only increasing revenue

# Subsetting for conventional and organic
g_con<-g[which(Data_2017$type==0),]
g_org<-g[which(Data_2017$type==1),]

# Defining the weak disposable output
wd <- matrix(c(0, 1), ncol = 2)

#----------------------------#
####---- Benchmarking ----####
#----------------------------#

# sf_solved<-dm.sf.new(xdata = X,ydata = Y,rts="crs",g=g,wd=wd)
sf_solved_con<-dm.sf.new(xdata = X_con,ydata = Y_con,rts="crs",g=g_con,wd=wd)
sf_solved_org<-dm.sf.new(xdata = X_org,ydata = Y_org,rts="crs",g=g_org,wd=wd)

####---- Inefficiency scores ----####
# Conventional inefficiency scores
beta_con<-sf_solved_con$eff

# Organic inefficiency scores
beta_org<-sf_solved_org$eff

####---- Efficiency scores ----####
# Conventional
eff_con<-1/(1+sf_solved_con$eff)

geo(eff_con)
mean(eff_con)
summary(eff_con)

# Organic
eff_org<-1/(1+sf_solved_org$eff)

geo(eff_org)
mean(eff_org)
summary(eff_org)

####---- Frontier shadow prices ----####
#### Frontier shadow prices conventional ####
Shadow_con<-round(sf_solved_con$p[,2]/sf_solved_con$p[,1],4)
Shadow_con

####---- Unique shadow prices ----####
Unique_shadow_con<-cbind(Shadow_price_con=unique(Shadow_con), Number_shadow_con=0)

for(i in unique(round(Shadow_con,4))){
  if(i==unique(round(Shadow_con,4))[1]){
    number=1}
  Unique_shadow_con[number,2]<-length(which(Shadow_con==i))

  number=number+1
}

Unique_shadow_con<-Unique_shadow_con[order(Unique_shadow_con[,1]),]

# Table for unique shadow prices of conventional farms
Unique_shadow_con

write.csv(Unique_shadow_con,file = "Unique_shadow_con_REV.csv")

#### Frontier shadow prices organic ####
Shadow_org<-round(sf_solved_org$p[,2]/sf_solved_org$p[,1],4)
Shadow_org
```

```
####---- Unique shadow prices ----####
Unique_shadow_org<-cbind(Shadow_price_org=unique(Shadow_org), Number_shadow_org=0)

for(i in unique(round(Shadow_org,4))){
  if(i==unique(round(Shadow_org,4))[1]){
    number=1}
  Unique_shadow_org[number,2]<-length(which(Shadow_org==i))

  number=number+1
}


Unique_shadow_org<-Unique_shadow_org[order(Unique_shadow_org[,1]),]

# Table for frontier shadow prices of organic farms
Unique_shadow_org

write.csv(Unique_shadow_org,file = "Unique_shadow_org_REV.csv")

####---- Peers ----####

####---- Peers for organic farms ----####
Peers_org<-which(!colSums(sf_solved_org$lambda)==0)
Peers_table_org<-cbind(Peers_org,number_peers=0)
lambda_org_table<-sf_solved_org$lambda[-Peers_org,]

for( i in Peers_org){
  if(i==Peers_org[1]){
    number=1}
  lambda_org_table<-sf_solved_org$lambda[-Peers_org[-number],]
  Peers_table_org[number,2]<-  length(which(lambda_org_table[,i]>0))
  number<-number+1
}

Peers_table_org

####---- Peers for conventional farms ----####
Peers_con<-which(!colSums(sf_solved_con$lambda)==0)
Peers_table_con<-cbind(Peers_con,number_peers=0)
lambda_con_table<-sf_solved_con$lambda[-Peers_con,]

for( i in Peers_con){
  if(i==Peers_con[1]){
    number=1}
  lambda_con_table<-sf_solved_con$lambda[-Peers_con[-number],]
  Peers_table_con[number,2]<-  length(which(  lambda_con_table[,i]>0))
  number<-number+1
}

Peers_table_con

####---- Plot of organic vs. conventional farms ----####
par(mfrow=c(1,2))

rbPal_type <- colorRampPalette(c("#ddd3c8","#aeb6ab"))
Data_2017$Col_type <- rbPal_type(10)[as.numeric(cut(Data_2017$type,breaks = 10))]
par(mfrow=c(1,1))
plot(Y[,2]/X, Y[,1]/X,col=Data_2017$Col_type,pch=16,ylab="Revenue/Costs",xlab="GHG emissions/Costs",family="Times New Roman",cex.lab=1.7,cex.axis=1.5,frame =
FALSE,xlim=c(min(Y[,2]/X)-0.00001,max(Y[,2]/X)+0.00001),ylim=c(min(Y[,1]/X)-0.1,max(Y[,1]/X)+0.1),cex=2)

text(c(0.000122,0.000132),c(1.6,1.5),c("Organic farms","Conventional farms"),col=c("#aeb6ab","#ddd3c8"),cex=1.5,family="Times New Roman")

#--------------------------------------------------------------------#
####---------------------- Outlier detection --------------------####
#--------------------------------------------------------------------#

####---- Outlier detection for organic farms ----####
Frontier_org<-which(sf_solved_org$eff==0)
Outlier_detection_org<-cbind(Frontier_org,0,0,0,0,0,Peers_table_org[,2])
colnames(Outlier_detection_org)<-c("Farm","mean eff","Diff to normal mean eff","Shadow mean","mean diff of sp","mean diff of sp/original mean","Time used as peer")

for(i in Frontier_org){

  X_org_outlier<-X_org[-i]
  Y_org_outlier<-Y_org[-i,]
  g_org_outlier<-g_org[-i,]

  sf_solved_org_outlier<-dm.sf.new(xdata = X_org_outlier,ydata = Y_org_outlier,rts="crs",g=g_org_outlier,wd=wd)

  Outlier_detection_org[which(Outlier_detection_org[,1]==i),2]<-mean(sf_solved_org_outlier$eff)
  Outlier_detection_org[which(Outlier_detection_org[,1]==i),3]<-mean(sf_solved_org$eff[-i])-mean(sf_solved_org_outlier$eff)
  Outlier_detection_org[which(Outlier_detection_org[,1]==i),4]<-mean(round(sf_solved_org_outlier$p[,2]/sf_solved_org_outlier$p[,1],4))
  Outlier_detection_org[which(Outlier_detection_org[,1]==i),5]<-mean(round(sf_solved_org_outlier$p[,2]/sf_solved_org_outlier$p[,1],4))-mean(Shadow_org[-i])
  Outlier_detection_org[which(Outlier_detection_org[,1]==i),6]<-(mean(round(sf_solved_org_outlier$p[,2]/sf_solved_org_outlier$p[,1],4))-mean(Shadow_org[-i]))/mean(Shadow_org[-
i])

}

Outlier_detection_org<-cbind(Sys=Data_2017$SysNr[which(Data_2017$type==0)][Frontier_org],Outlier_detection_org)
Outlier_detection_org

####---- Outlier detection for conventional farms ----#####
Frontier_con<-which(sf_solved_con$eff==0)
Outlier_detection_con<-cbind(Frontier_con,0,0,0,0,0,Peers_table_con[,2])
colnames(Outlier_detection_con)<-c("Farm","mean eff","Diff to normal mean eff","Shadow mean","mean diff of sp","mean diff of sp/original mean","Time used as peer")

for(i in Frontier_con){

  X_con_outlier<-X_con[-i]
  Y_con_outlier<-Y_con[-i,]
  g_con_outlier<-g_con[-i,]

  sf_solved_con_outlier<-dm.sf.new(xdata = X_con_outlier,ydata = Y_con_outlier,rts="crs",g=g_con_outlier,wd=wd)

  Outlier_detection_con[which(Outlier_detection_con[,1]==i),2]<-mean(sf_solved_con_outlier$eff)
  Outlier_detection_con[which(Outlier_detection_con[,1]==i),3]<-mean(sf_solved_con$eff[-i])-mean(sf_solved_con_outlier$eff)
  Outlier_detection_con[which(Outlier_detection_con[,1]==i),4]<-mean(round(sf_solved_con_outlier$p[,2]/sf_solved_con_outlier$p[,1],4))
  Outlier_detection_con[which(Outlier_detection_con[,1]==i),5]<-mean(round(sf_solved_con_outlier$p[,2]/sf_solved_con_outlier$p[,1],4))-mean(Shadow_con[-i])
  Outlier_detection_con[which(Outlier_detection_con[,1]==i),6]<-(mean(round(sf_solved_con_outlier$p[,2]/sf_solved_con_outlier$p[,1],4))-mean(Shadow_con[-i]))/mean(Shadow_con[-
i])

}

Outlier_detection_con<-cbind(Sys=Data_2017$SysNr[which(Data_2017$type==0)][Frontier_con],Outlier_detection_con)
Outlier_detection_con

####---- Plot with and without outlier ----####

####---- organic ----####

Outlier<-NULL # No outliers are detected in the organic model

####---- organic model without outlier ----####
X_org_outlier<-X_org#[-Outlier]
Y_org_outlier<-Y_org#[-Outlier,]
g_org_outlier<-g_org#[-Outlier,]

sf_solved_org_outlier<-dm.sf.new(xdata = X_org_outlier,ydata = Y_org_outlier,rts="crs",g=g_org_outlier,wd=wd)
```

```
eff_org_outlier<-1/(1+sf_solved_org_outlier$eff)

par(mfrow=c(1,1))
plot(Y[which(Data_2017$type==1),2]/X[which(Data_2017$type==1)], Y[which(Data_2017$type==1),1]/X[which(Data_2017$type==1)],
     col="#8b9086" ,pch=16,ylab="Revenue/Costs",xlab="GHG emissions/Costs",family="Times New Roman",cex.lab=1.7,cex.axis=1.5,frame =
FALSE,xlim=c(min(Y[,2]/X)-0.00001,max(Y[,2]/X)+0.00001),ylim=c(min(Y[,1]/X)-0.2,max(Y[,1]/X)+0.1),cex=2, main="Model only including organic farms")

# Plotting the frontier including potential outlier
x=Y_org[which(sf_solved_org$eff==0),2]/X_org[which(sf_solved_org$eff==0)]
y=Y_org[which(sf_solved_org$eff==0),1]/X_org[which(sf_solved_org$eff==0)]

lines(c(0,x[order(x)],max(x[order(x)])), c(0,y[order(x)],0), xlim=range(x), ylim=range(y), pch=16,col="#8b9086",lwd=2,lty=1)
text(x-0.000005,y+0.04,c(1,4,3,5,6,2), family="Times New Roman")

# Plotting the frontier excluding potential outlier
x=Y_org_outlier[which(sf_solved_org_outlier$eff==0),2]/X_org_outlier[which(sf_solved_org_outlier$eff==0)]
y=Y_org_outlier[which(sf_solved_org_outlier$eff==0),1]/X_org_outlier[which(sf_solved_org_outlier$eff==0)]

lines(c(0,x[order(x)],max(x[order(x)])), c(0,y[order(x)],0), xlim=range(x), ylim=range(y), pch=16,col="#8b9086",lwd=2)

####---- conventional ----####

Outlier<-721

####---- conventional model without outlier ----####
X_con_outlier<-X_con[-Outlier]
Y_con_outlier<-Y_con[-Outlier,]
g_con_outlier<-g_con[-Outlier,]

sf_solved_con_outlier<-dm.sf.new(xdata = X_con_outlier,ydata = Y_con_outlier,rts="crs",g=g_con_outlier,wd=wd)
eff_con_outlier<-1/(1+sf_solved_con_outlier$eff)

plot(Y[which(Data_2017$type==0),2]/X[which(Data_2017$type==0)], Y[which(Data_2017$type==0),1]/X[which(Data_2017$type==0)],
     col="#ddd3c8" ,pch=16,ylab="Revenue/Costs",xlab="GHG emissions/Costs",family="Times New Roman",cex.lab=1.7,cex.axis=1.5,frame =
FALSE,xlim=c(min(Y[,2]/X)-0.00001,max(Y[,2]/X)+0.00001),ylim=c(min(Y[,1]/X)-0.2,max(Y[,1]/X)+0.1),cex=2, main="Model only including conventional farms")

# Plotting the frontier including the outlier
x=Y_con[which(sf_solved_con$eff==0),2]/X_con[which(sf_solved_con$eff==0)]
y=Y_con[which(sf_solved_con$eff==0),1]/X_con[which(sf_solved_con$eff==0)]

lines(c(0,x[order(x)],max(x[order(x)])), c(0,y[order(x)],0), xlim=range(x), ylim=range(y), pch=16,col="#ddd3c8",lwd=2,lty=3)
text(x-0.000005,y+0.04,rank(-Outlier_detection_con[,7]),family="Times New Roman")
# Plotting the frontier excluding the outlier
x=Y_con_outlier[which(sf_solved_con_outlier$eff==0),2]/X_con_outlier[which(sf_solved_con_outlier$eff==0)]
y=Y_con_outlier[which(sf_solved_con_outlier$eff==0),1]/X_con_outlier[which(sf_solved_con_outlier$eff==0)]

lines(c(0,x[order(x)],max(x[order(x)])), c(0,y[order(x)],0), xlim=range(x), ylim=range(y), pch=16,col="#ddd3c8",lwd=2)


#------------------------------------------------------------------#
####------------- Results after outlier detection -------------####
#------------------------------------------------------------------#

####---- Histograms of inefficiency scores  ----####
# DMUs with negative frontier shadow prices are excluded

par(mfrow=c(1,2))
# Histogram for conventional farms
hist(sf_solved_con$eff[which(Shadow_con<0)],xlab=expression(paste("Inefficiency (",beta,")")),ylab="",breaks = 40,family="Times New
Roman",col="#ddd3c8",border="white",cex.lab=2,cex.axis=1.7,cex.main=2,main="conventional inefficiency scores",ylim=c(0,100))
abline(v = mean(sf_solved_con$eff[which(Shadow_con<0)]), col="black", lwd=2, lty=1)

# Histogram for organic farms
hist(sf_solved_org$eff[which(Shadow_org<0)],xlab=expression(paste("Inefficiency (",beta,")")),ylab="",breaks = 40,family="Times New
Roman",col="#8b9086",border="white",cex.lab=2,cex.axis=1.7,cex.main=2,main="organic inefficiency scores",ylim=c(0,25))
abline(v = mean(sf_solved_con$eff[which(Shadow_con<0)]), col="black", lwd=2, lty=1)

####---- Plotting inefficiency for organic farms ----####
# Colour palette for 2d scatterplot
rbPal_org <- colorRampPalette(c("#ddd3c8","#a4864b","#46626f"))
col_org <- rbPal_org(10)[as.numeric(cut(eff_org,breaks = 10))]

# Plotting the model
plot(Y[which(Data_2017$type==1),2]/X[which(Data_2017$type==1)], Y[which(Data_2017$type==1),1]/X[which(Data_2017$type==1)],
     col=col_org ,pch=16,ylab="Revenue/Costs",xlab="GHG emissions/Costs",family="Times New Roman",cex.lab=1.7,cex.axis=1.5,frame =
FALSE,xlim=c(min(Y[,2]/X)-0.00001,max(Y[,2]/X)+0.00001),ylim=c(min(Y[,1]/X)-0.2,max(Y[,1]/X)+0.1),cex=2, main="Model only including organic farms")

# Colouring the entire frontier
x=Y[which(Data_2017$type==1),2]/X[which(Data_2017$type==1)]
y=Y[which(Data_2017$type==1),1]/X[which(Data_2017$type==1)]
x<-x[which(eff_org==1)]
y<-y[which(eff_org==1)]

lines(c(0,x[order(x)],max(x[order(x)])), c(0,y[order(x)],0), xlim=range(x), ylim=range(y), pch=16,col="#46626f",lwd=2)

# Colouring the Inefficient part of the frontier
y<-y[order(x)][c(5,6)]
x<-x[order(x)][c(5,6)]

lines(c(x[order(x)],max(x[order(x)])), c(y[order(x)],0), xlim=range(x), ylim=range(y), pch=16,col="white",lwd=2,lty=3)

####---- Plotting inefficiency for conventional farms ----####

# Colour palette for 2d scatterplot
rbPal_conv <- colorRampPalette(c("#ddd3c8","#a4864b","#46626f"))
col_conv <- rbPal_conv(10)[as.numeric(cut(eff_con,breaks = 10))]

# Plotting the model
plot(Y[which(Data_2017$type==0),2]/X[which(Data_2017$type==0)], Y[which(Data_2017$type==0),1]/X[which(Data_2017$type==0)],
     col=col_conv ,pch=16,ylab="Revenue/Costs",xlab="GHG emissions/Costs",family="Times New Roman",cex.lab=1.7,cex.axis=1.5,frame =
FALSE,xlim=c(min(Y[,2]/X)-0.00001,max(Y[,2]/X)+0.00001),ylim=c(min(Y[,1]/X)-0.2,max(Y[,1]/X)+0.1),cex=2, main="Model only including conventional farms")

# Plotting the entire frontier
x=Y[which(Data_2017$type==0),2]/X[which(Data_2017$type==0)]
y=Y[which(Data_2017$type==0),1]/X[which(Data_2017$type==0)]
x<-x[which(eff_con==1)]
y<-y[which(eff_con==1)]

lines(c(0,x[order(x)],max(x[order(x)])), c(0,y[order(x)],0), xlim=range(x), ylim=range(y), pch=16,col="#46626f",lwd=2)

# Colouring the inefficient part of the frontier
y<-y[order(x)][c(6,7)]
x<-x[order(x)][c(6,7)]

lines(c(x[order(x)],max(x[order(x)])), c(y[order(x)],0), xlim=range(x), ylim=range(y), pch=16,col="white",lwd=2,lty=3)

####---- Plotting frontier shadow prices for conventional farms ----####
# Colour palette
palette(c("#46626f","#aeb6ab","#ddd3c8","#8b9086","#a4864b","#8E8580","#a9bebf","#f5f1e9"))

par(mfrow=c(1,1))
plot(Y[which(Data_2017$type==0),2]/X[which(Data_2017$type==0)], Y[which(Data_2017$type==0),1]/X[which(Data_2017$type==0)],
     col=as.factor(Shadow_con) ,pch=16,ylab="",xlab="",family="Times New Roman",cex.lab=3,cex.axis=2.5,frame =
FALSE,xlim=c(min(Y[,2]/X)-0.00001,max(Y[,2]/X)+0.00001),ylim=c(min(Y[,1]/X)-0.2,max(Y[,1]/X)+0.1),cex=2, main="conventional",cex.main=2.5)

mtext(text = "GHG/Costs",
      side = 1, #side 2 = left
      line = 3.5,cex = 2,family="Times New Roman")
```

```
# Plotting the entire frontier
x=Y[which(Data_2017$type==0),2]/X[which(Data_2017$type==0)]
y=Y[which(Data_2017$type==0),1]/X[which(Data_2017$type==0)]
x<-x[which(eff_con==1)]
y<-y[which(eff_con==1)]

lines(c(0,x[order(x)],max(x[order(x)])), c(0,y[order(x)],0), xlim=range(x), ylim=range(y), pch=16,col="black",lwd=2)

# Colouring the inefficient part of the frontier
y<-y[order(x)][c(6,7)]
x<-x[order(x)][c(6,7)]

lines(c(x[order(x)],max(x[order(x)])), c(y[order(x)],0), xlim=range(x), ylim=range(y), pch=16,col="white",lwd=2,lty=3)

####---- Plotting frontier shadow prices for organic farms ----####

# Colour palette
palette(c("#46626f","#aeb6ab","#ddd3c8","#8b9086","#a4864b","#8E8580","#a9bebf","#f5f1e9"))

plot(Y[which(Data_2017$type==1),2]/X[which(Data_2017$type==1)], Y[which(Data_2017$type==1),1]/X[which(Data_2017$type==1)],
     col=as.factor(Shadow_org) ,pch=16,ylab="",xlab="",family="Times New Roman",cex.lab=3,cex.axis=2.5,frame =
FALSE,xlim=c(min(Y[,2]/X)-0.00001,max(Y[,2]/X)+0.00001),ylim=c(min(Y[,1]/X)-0.2,max(Y[,1]/X)+0.1),cex=2, main="Organic",cex.main=2.5)

mtext(text = "GHG/Costs",
      side = 1, #side 2 = left
      line = 3.5,cex = 2,family="Times New Roman")

# Plotting the entire frontier
x=Y[which(Data_2017$type==1),2]/X[which(Data_2017$type==1)]
y=Y[which(Data_2017$type==1),1]/X[which(Data_2017$type==1)]
x<-x[which(eff_org==1)]
y<-y[which(eff_org==1)]

lines(c(0,x[order(x)],max(x[order(x)])), c(0,y[order(x)],0), xlim=range(x), ylim=range(y), pch=16,col="black",lwd=2)

# Colouring the inefficient part of the frontier
y<-y[order(x)][c(5,6)]
x<-x[order(x)][c(5,6)]

lines(c(x[order(x)],max(x[order(x)])), c(y[order(x)],0), xlim=range(x), ylim=range(y), pch=16,col="white",lwd=2,lty=3)


#--------------------------------------------------------------------------------#
####------------- Second stage analysis of inefficiency scores -------------####
#--------------------------------------------------------------------------------#

#----------------------------#
####---- Tobit models ----####
#----------------------------#

####---- Conventional tobit ----####
t=0 # Type defined for conventional farms

Data_con_2017<-Data_2017[which(Data_2017$type==t),]
Data_con_2017$beta_con<-beta_con
Data_con_2017$Shadow_con<-Shadow_con
Data_con_2017<-Data_con_2017[which(Data_con_2017$Shadow_con<0),]

tobit_con<-tobit(Data_con_2017$beta_con~Data_con_2017$All_cattle[which(Data_con_2017$type==t)]
                 + I(Data_con_2017$milk[which(Data_con_2017$type==t)]/(Data_con_2017$milk[which(Data_con_2017$type==t)]+Data_con_2017$oo[which(Data_con_2017$type==t)])*100)
                 + I((Data_con_2017$fix[which(Data_con_2017$type==t)]/Data_con_2017$input[which(Data_con_2017$type==t)])*100)
                 + I((Data_con_2017$X5110[which(Data_con_2017$type==t)]/Data_con_2017$All_cattle[which(Data_con_2017$type==t)]*100))
                 + as.factor(Data_con_2017$X6404[which(Data_con_2017$type==t)])
                 + I(Data_con_2017$X7825[which(Data_con_2017$type==t)]/1000)
                 + I(Data_con_2017$X7826[which(Data_con_2017$type==t)]/1000)
                 + I(Data_con_2017$X7841[which(Data_con_2017$type==t)]/1000)
                 + I(Data_con_2017$Andel_3[which(Data_con_2017$type==t)]*100)
                 + I(Data_con_2017$deep_litter_share[which(Data_con_2017$type==t)]*100)
                 + I(Data_con_2017$Milk_kv[which(Data_con_2017$type==t)] / Data_con_2017$X5110[which(Data_con_2017$type==t)]*100)
                 ,left=0,right=1)
summary(tobit_con)

logLik(tobit_con)
sigma_con=tobit_con$scale
yhat_con=fitted(tobit_con)
Squarred_cor_con <- cor(yhat_con,Data_con_2017$beta_con)^2
Squarred_cor_con

####---- organic second stage ----####
t=1 # Type defined for conventional farms

Data_org_2017<-Data_2017[which(Data_2017$type==t),]
Data_org_2017$beta_org<-beta_org
Data_org_2017$Shadow_org<-Shadow_org
Data_org_2017<-Data_org_2017[which(Data_org_2017$Shadow_org<0),]

tobit_org<-tobit(Data_org_2017$beta_org~Data_org_2017$All_cattle[which(Data_org_2017$type==t)]
                 + I(Data_org_2017$milk[which(Data_org_2017$type==t)]/(Data_org_2017$milk[which(Data_org_2017$type==t)]+Data_org_2017$oo[which(Data_org_2017$type==t)])*100)
                 + I((Data_org_2017$fix[which(Data_org_2017$type==t)]/Data_org_2017$input[which(Data_org_2017$type==t)])*100)
                 + I((Data_org_2017$X5110[which(Data_org_2017$type==t)]/Data_org_2017$All_cattle[which(Data_org_2017$type==t)]*100))
                 + as.factor(Data_org_2017$X6404[which(Data_org_2017$type==t)])
                 + I(Data_org_2017$X7825[which(Data_org_2017$type==t)]/1000)
                 + I(Data_org_2017$X7826[which(Data_org_2017$type==t)]/1000)
                 + I(Data_org_2017$X7841[which(Data_org_2017$type==t)]/1000)
                 + I(Data_org_2017$Andel_3[which(Data_org_2017$type==t)]*100)
                 + I(Data_org_2017$deep_litter_share[which(Data_org_2017$type==t)]*100)
                 + I(Data_org_2017$Milk_kv[which(Data_org_2017$type==t)] / Data_org_2017$X5110[which(Data_org_2017$type==t)]*100)
                 + Data_org_2017$Year_org[which(Data_org_2017$type==t)]
                 ,left=0,right=1)
summary(tobit_org)

sigma_org=tobit_org$scale
yhat_org=fitted(tobit_org)
Squarred_cor_org <- cor(yhat_org,Data_org_2017$beta_org)^2
Squarred_cor_org


#--------------------------------------------#
####---- Average partial effects (APE) ----####
#--------------------------------------------#

#---- APE for conventional----#

# APE for Continous variables
t=0 # Defining conventional farms (type=0)

tobit_data<-cbind(
  Data_con_2017$All_cattle[which(Data_con_2017$type==t)]
  , I(Data_con_2017$milk[which(Data_con_2017$type==t)]/(Data_con_2017$milk[which(Data_con_2017$type==t)]+Data_con_2017$oo[which(Data_con_2017$type==t)]*100))
  , I((Data_con_2017$fix[which(Data_con_2017$type==t)]/Data_con_2017$input[which(Data_con_2017$type==t)])*100)
  , I((Data_con_2017$X5110[which(Data_con_2017$type==t)]/Data_con_2017$All_cattle[which(Data_con_2017$type==t)]*100))
  , as.factor(Data_con_2017$X6404[which(Data_con_2017$type==t)])
  , I(Data_con_2017$X7825[which(Data_con_2017$type==t)]/1000)
  , I(Data_con_2017$X7826[which(Data_con_2017$type==t)]/1000)
  , I(Data_con_2017$X7841[which(Data_con_2017$type==t)]/1000)
  , I(Data_con_2017$Andel_3[which(Data_con_2017$type==t)]*100)
  , I(Data_con_2017$deep_litter_share[which(Data_con_2017$type==t)]*100)
```

```
  , I(Data_con_2017$Milk_kv[which(Data_con_2017$type==t)] / Data_con_2017$X5110[which(Data_con_2017$type==t)]*100))

APE_tobit<-matrix(0,nrow=1,ncol=dim(tobit_data)[2])

for (i in 1:dim(tobit_data)[2]){

  b_xj<- coef(tobit_con)[i+1]
  xb=predict(tobit_con)
  sigma=tobit_con$scale
  invmills <- dnorm(xb/sigma)/pnorm(xb/sigma)
  PE_Ey_ygt0_xj <- b_xj*(1-invmills*(xb/sigma+invmills))
  APE_tobit[1,i]<-mean(PE_Ey_ygt0_xj)

}

# APE for Discrete variables
t=0
xb_c = predict(tobit_con)- tobit_con$coef[6]*(Data_con_2017$X6404[which(Data_con_2017$type==t)])+ tobit_con$coef[6]*(0)
xb_c1 = predict(tobit_con)- tobit_con$coef[6]*(Data_con_2017$X6404[which(Data_con_2017$type==t)])+ tobit_con$coef[6]*(1)

sigma=tobit_con$scale

Eyy_c1<- pnorm(xb_c1/sigma)*xb_c1+sigma*dnorm(xb_c1/sigma)
Eyy_c<- pnorm(xb_c/sigma)*xb_c+sigma*dnorm(xb_c/sigma)
PE <- Eyy_c1-Eyy_c
APE_disc<- mean(PE)

APE_tobit[1,5]<-APE_disc
APE_tobit_con<-APE_tobit

#---- APE for corganic----#
# APE for continous variables
t=1 # Defining organic farms (type=1)

tobit_data<-cbind(Data_org_2017$All_cattle[which(Data_org_2017$type==t)]
                  , I(Data_org_2017$milk[which(Data_org_2017$type==t)]/((Data_org_2017$milk[which(Data_org_2017$type==t)]+Data_org_2017$oo[which(Data_org_2017$type==t)])*100))
                  , I((Data_org_2017$fix[which(Data_org_2017$type==t)]/Data_org_2017$input[which(Data_org_2017$type==t)])*100)
                  , I((Data_org_2017$X5110[which(Data_org_2017$type==t)]/Data_org_2017$All_cattle[which(Data_org_2017$type==t)]*100))
                  , as.factor(Data_org_2017$X6404[which(Data_org_2017$type==t)])
                  , I(Data_org_2017$X7825[which(Data_org_2017$type==t)]/1000)
                  , I(Data_org_2017$X7826[which(Data_org_2017$type==t)]/1000)
                  , I(Data_org_2017$X7841[which(Data_org_2017$type==t)]/1000)
                  , I(Data_org_2017$Andel_3[which(Data_org_2017$type==t)]*100)
                  , I(Data_org_2017$deep_litter_share[which(Data_org_2017$type==t)]*100)
                  , I(Data_org_2017$Milk_kv[which(Data_org_2017$type==t)] / Data_org_2017$X5110[which(Data_org_2017$type==t)]*100)
                  , Data_org_2017$Year_org[which(Data_org_2017$type==t)] )

APE_tobit<-matrix(0,nrow=1,ncol=dim(tobit_data)[2])

for (i in 1:dim(tobit_data)[2]){

  b_xj<- coef(tobit_org)[i+1]
  xb=predict(tobit_org)
  sigma=tobit_org$scale
  invmills <- dnorm(xb/sigma)/pnorm(xb/sigma)
  PE_Ey_ygt0_xj <- b_xj*(1-invmills*(xb/sigma+invmills))
  APE_tobit[1,i]<-mean(PE_Ey_ygt0_xj)

}

# Discrete variables

#APE
xb_c = predict(tobit_org)- tobit_org$coef[6]*(Data_org_2017$X6404[which(Data_org_2017$type==t)])+ tobit_org$coef[6]*(0)
xb_c1 = predict(tobit_org)- tobit_org$coef[6]*(Data_org_2017$X6404[which(Data_org_2017$type==t)])+ tobit_org$coef[6]*(1)
sigma=tobit_org$scale

Eyy_c1<- pnorm(xb_c1/sigma)*xb_c1+sigma*dnorm(xb_c1/sigma)
Eyy_c<- pnorm(xb_c/sigma)*xb_c+sigma*dnorm(xb_c/sigma)
PE <- Eyy_c1-Eyy_c
APE_disc<- mean(PE)

APE_tobit[1,5]<-APE_disc

APE_tobit_org<-APE_tobit


####---- Stargazer - tobit with efficiencies ----####

# # Run when estimating the GHG model with g(0,w)
# tobit_con_GHG<-tobit_con
# tobit_org_GHG<-tobit_org
# Sq_con_GHG<-Squarred_cor_con
# Sq_org_GHG<-Squarred_cor_org
# APE_con_GHG<-APE_tobit_con
# APE_org_GHG<-APE_tobit_org

# # Run when estimating the Revenue model with g(v,0)
# tobit_con_REV<-tobit_con
# tobit_org_REV<-tobit_org
# Sq_con_REV<-Squarred_cor_con
# Sq_org_REV<-Squarred_cor_org
# APE_con_REV<-APE_tobit_con
# APE_org_REV<-APE_tobit_org

names_org<- c("Number of cattle (100)",
              "Milk (DKK)/Total output (DKK)",
              "Fixed costs/Total costs (pct.)",
              "Dairy cattle / all cattle (pct.)",
              "Ownership (other than private)",
              "Cost to consulting - production (1.000 DKK)",
              "Cost to consulting - cattle (1.000DKK)",
              "Cost to consulting - economic (1.000 DKK)",
              "Share of jersey cattle (pct.)",
              "Share of cattle having a deep litter housing system (pct.)",
              "Milk production per dairy cattle (liter/cow)",
              "Year since converted to organic (year)")

names_con<- c("Number of cattle (100)",
              "Milk (DKK)/Total output (DKK)",
              "Fixed costs/Total costs (pct.)",
              "Dairy cattle / all cattle (pct.)",
              "Ownership (other than private)",
              "Cost to consulting - production (1.000 DKK)",
              "Cost to consulting - cattle (1.000DKK)",
              "Cost to consulting - economic (1.000 DKK)",
              "Share of jersey cattle (pct.)",
              "Share of cattle having a deep litter housing system (pct.)",
              "Milk production per dairy cattle (liter/cow)")


####---- APE tabels ----####

## Conventional
APE_table_con<-data.frame(cbind(names_con,t(APE_con_GHG),t(APE_con_REV)))
APE_table_org<-data.frame(cbind(names_org,t(APE_org_GHG),t(APE_org_REV)))

write.csv(APE_table_con,file = "APE_table_con.csv")
```

```r
write.csv(APE_table_org,file = "APE_table_org.csv")

Sq_con<-c(Sq_con_GHG,Sq_con_REV)
Sq_org<-c(Sq_org_GHG,Sq_org_REV)

# Conventional
stargazer(tobit_con_GHG,tobit_con_REV, align=TRUE,type="html",out="Tobit_models_CON_ny1.html",no.space=TRUE, single.row=TRUE,
          digits = 3,dep.var.labels="Inefficiency score",report=('vc*p'),covariate.labels=names_con,column.labels = c("GHG","Mix","Rev"),add.lines=c("Squarred correlation",Sq_con),decimal.mark="," ,digit.separator = ".")

# Organic
stargazer(tobit_org_GHG,tobit_org_REV, align=TRUE,type="html",out="Tobit_models_ORG_ny1.html",no.space=TRUE, single.row=TRUE,
          digits = 3,dep.var.labels="Inefficiency score",report=('vc*p'),covariate.labels=names_org,column.labels = c("GHG","Mix","Rev"),add.lines=c("Squarred correlation",Sq_org),decimal.mark="," ,digit.separator = ".")


#---------------------------------------------------------#
####---- Potentials and average opportunity costs ----####
#---------------------------------------------------------#

####---- Defining direction ----####
# all three models are ran at the same time when defining potentials and average opportunity costs

## GHG model
g_GHG <- cbind(0.00000000000000000000000000000000000000000001,0.00000000000000000000000000000000000000000001, Y[,2]) # Only reducing GHG
g_GHG_con<-g_GHG[which(Data_2017$type==0),]
g_GHG_org<-g_GHG[which(Data_2017$type==1),]

# Benchmarking
sf_solved_GHG_con<-dm.sf.new(xdata = X_con,ydata = Y_con,rts="crs",g=g_GHG_con,wd=wd)
sf_solved_GHG_org<-dm.sf.new(xdata = X_org,ydata = Y_org,rts="crs",g=g_GHG_org,wd=wd)

## Revenue model
g_REV <- cbind(0.00000000000000000000000000000000000000000001, Y[,1],0.00000000000000000000000000000000000000000001) # Only reducing REV
g_REV_con<-g_REV[which(Data_2017$type==0),]
g_REV_org<-g_REV[which(Data_2017$type==1),]

# Benchmarking
sf_solved_REV_con<-dm.sf.new(xdata = X_con,ydata = Y_con,rts="crs",g=g_REV_con,wd=wd)
sf_solved_REV_org<-dm.sf.new(xdata = X_org,ydata = Y_org,rts="crs",g=g_REV_org,wd=wd)

#-------------------------#
####--- Potentials ----####
#-------------------------#

## GHG model
# Removing observations with negative shadow prices
Shadow_GHG_con<-round(sf_solved_GHG_con$p[,2]/sf_solved_GHG_con$p[,1],4)
Shadow_GHG_org<-round(sf_solved_GHG_org$p[,2]/sf_solved_GHG_org$p[,1],4)

Only_GHG_totpot_con<-sum((sf_solved_GHG_con$eff*Y_con[,2])[which(Shadow_GHG_con<0)])
Only_GHG_totpot_org<-sum((sf_solved_GHG_org$eff*Y_org[,2])[which(Shadow_GHG_org<0)])
Only_GHG_totpot_tot<-Only_GHG_totpot_con+Only_GHG_totpot_org

##  REV model
# Removing observations with negative shadow prices
Shadow_REV_con<-round(sf_solved_REV_con$p[,2]/sf_solved_REV_con$p[,1],4)
Shadow_REV_org<-round(sf_solved_REV_org$p[,2]/sf_solved_REV_org$p[,1],4)

Neg_REV_con<-which(Y[,1] %in% Y_con[which(!Shadow_REV_con<0),1])
Neg_REV_org<-which(Y[,1] %in% Y_con[which(!Shadow_REV_org<0),1])

Neg_REV_tot<-c(Neg_REV_con,Neg_REV_org)

Only_REV_totpot_con<-sum((sf_solved_REV_con$eff*Y_con[,1])[which(Shadow_REV_con<0)])
Only_REV_totpot_org<-sum((sf_solved_REV_org$eff*Y_org[,1])[which(Shadow_REV_org<0)])
Only_REV_totpot_tot<-Only_REV_totpot_con+Only_REV_totpot_org

####---- Potentials in volumes ----####
Potential_matrix<-matrix(0,nrow = 2,ncol=6)
colnames(Potential_matrix)<-c("GHG_con","GHG_org","GHG_tot","REV_con","REV_org","REV_tot")
rownames(Potential_matrix)<-c("Only GHG","Only Rev")

#Potential_matrix[1,]<-c(RAD_GHG_totpot_con,RAD_GHG_totpot_org,RAD_GHG_totpot_tot,RAD_REV_totpot_con/1000,RAD_REV_totpot_org/1000,RAD_REV_totpot_tot/1000)
Potential_matrix[1,1:3]<-c(Only_GHG_totpot_con,Only_GHG_totpot_org,Only_GHG_totpot_tot)
Potential_matrix[2,4:6]<-c(Only_REV_totpot_con/1000,Only_REV_totpot_org/1000,Only_REV_totpot_tot/1000)

Potential_matrix
write.csv(Potential_matrix,file = "Totpot.csv")

####---- Potentials in pct. ----####
Potential_matrix_pct<-matrix(0,nrow = 2,ncol=6)
colnames(Potential_matrix_pct)<-c("GHG_con","GHG_org","GHG_tot","REV_con","REV_org","REV_tot")
rownames(Potential_matrix_pct)<-c("Only GHG","Only Rev")

# Potential_matrix_pct[1,]<-c(mean(sf_solved_RAD_con$eff[which(Shadow_RAD_con<0)]),mean(sf_solved_RAD_org$eff[which(Shadow_RAD_org<0)]),RAD_GHG_totpot_tot/sum(Y[-Neg_RAG_tot,2]),
#                              mean(sf_solved_RAD_con$eff[which(Shadow_RAD_con<0)]),mean(sf_solved_RAD_org$eff[which(Shadow_RAD_org<0)]),RAD_REV_totpot_tot/sum(Y[-Neg_RAG_tot,1]))
Potential_matrix_pct[1,1:3]<-c(mean(sf_solved_GHG_con$eff[which(Shadow_GHG_con<0)]),mean(sf_solved_GHG_org$eff[which(Shadow_GHG_org<0)]),Only_GHG_totpot_tot/sum(Y[,2]))
Potential_matrix_pct[2,4:6]<-c(mean(sf_solved_REV_con$eff[which(Shadow_REV_con<0)]),mean(sf_solved_REV_org$eff[which(Shadow_REV_org<0)]),Only_REV_totpot_tot/sum(Y[-Neg_REV_tot,1]))

Potential_matrix_pct

write.csv(Potential_matrix_pct,file = "Totpot_pct.csv")

#-----------------------------------------#
####---- Average opportunity costs ----####
#-----------------------------------------#

####---- Conventional average opportunity costs ----####
Rev_pot_con<-sf_solved_REV_con$eff*Y_con[,1]
Rev_pot_con<-Rev_pot_con[-c(which(sf_solved_REV_con$eff==0),which(!Shadow_REV_con<0),which(sf_solved_GHG_con$eff==0))]
which(Rev_pot_con==0)

GHG_pot_con<-sf_solved_GHG_con$eff*Y_con[,2]
GHG_pot_con<-GHG_pot_con[-c(which(sf_solved_REV_con$eff==0),which(!Shadow_REV_con<0),which(sf_solved_GHG_con$eff==0))]
which(GHG_pot_con==0)
Opp_cost_con<-Rev_pot_con/GHG_pot_con
mean(Opp_cost_con)

####---- Histogram of cenventional opportunity costs ----####
par(mfrow=c(1,2))
hist(Opp_cost_con,xlab="Opportunity cost conventional farms",ylab="",breaks = 40,family="Times New Roman",col="#ddd3c8",border="white",cex.lab=2,cex.axis=1.7,main="")
abline(v = mean(Opp_cost_con), col="black", lwd=2, lty=1)

####---- Organic average opportunity costs ----####
Rev_pot_org<-sf_solved_REV_org$eff*Y_org[,1]
Rev_pot_org<-Rev_pot_org[-c(which(sf_solved_REV_org$eff==0),which(!Shadow_REV_org<0),which(sf_solved_GHG_org$eff==0))]
which(Rev_pot_org==0)

GHG_pot_org<-sf_solved_GHG_org$eff*Y_org[,2]
GHG_pot_org<-GHG_pot_org[-c(which(sf_solved_REV_org$eff==0),which(!Shadow_REV_org<0),which(sf_solved_GHG_org$eff==0))]
which(GHG_pot_org==0)
Opp_cost_org<-Rev_pot_org/GHG_pot_org
mean(Opp_cost_org)
```

```
####---- Histogram of organic opportunity costs ----####
hist(Opp_cost_org,xlab="Opportunity cost organic farms",ylab="",breaks = 30,family="Times New Roman",col="#8b9086",border="white",cex.lab=2,cex.axis=1.7,main="")
abline(v = mean(Opp_cost_org), col="black", lwd=2, lty=1)

#-------------------------------------------------------------------------------------#
####-------------- Second stage analysis of Average opportunity costs --------------####
#-------------------------------------------------------------------------------------#

####---- OLS with average opportunity costs for conventional farms ----####
t=0 # Defining conventional farms

Data_con_2017<-Data_2017[which(Data_2017$type==t),]
Data_con_2017<-Data_con_2017[-c(which(sf_solved_REV_con$eff==0),which(!Shadow_REV_con<0),which(sf_solved_GHG_con$eff==0)),]
Data_con_2017$Opp_cost_con<-Opp_cost_con

con_shadow_model<-lm(Opp_cost_con~All_cattle
                     + I((milk/(milk+oo))*100)
                     + I((fix/input)*100)
                     + I((X5110/All_cattle*100))
                     + as.factor(X6404)
                     + I(Andel_3*100)
                     + I(deep_litter_share*100)
                     + I(Milk_kv / X5110*100)
                     ,data = Data_con_2017)
summary(con_shadow_model)

####---- OLS with average opportunity costs for organic farms ----####
t=1 # Defining organic farms

Data_org_2017<-Data_2017[which(Data_2017$type==t),]
Data_org_2017<-Data_org_2017[-c(which(sf_solved_REV_org$eff==0),which(!Shadow_REV_org<0),which(sf_solved_GHG_org$eff==0)),]
Data_org_2017$Opp_cost_org<-Opp_cost_org

org_shadow_model<-lm(Opp_cost_org~All_cattle
                     + I((milk/(milk+oo))*100)
                     + I((fix/input)*100)
                     + I((X5110/All_cattle*100))
                     + as.factor(X6404)
                     + I(Andel_3*100)
                     + I(deep_litter_share*100)
                     + I(Milk_kv / X5110*100)
                     + Year_org
                     ,data = Data_org_2017)
summary(org_shadow_model)

#### Stargazer for second stage of opportunity costs
stargazer(con_shadow_model,org_shadow_model, align=TRUE,type="html",out="shadow_models.html",no.space=TRUE, single.row=TRUE,
          digits = 3,dep.var.labels="Shadow price",report=('vc*p'),column.labels = c("Con","Org"),decimal.mark="," ,digit.separator = ".")

#-------------------------------------------------------------------------#
####---- Illustrating the weak disposability technology and the ddf ----####
#-------------------------------------------------------------------------#

####---- Costumizing fonts ----####
# library(extrafont)
# windowsFonts()
library(extrafont)
# font_import()
# y
# loadfonts(device = "win")

####---- Weak disposable technology ----####
Input<-c(1,1,1,1,1)
Output_1<-c(10.5,15,18,17,12)
Output_2<-c(5,8.5,12.5,15,17.5)

x_test<-Output_2/Input
y_test<-Output_1/Input

par(mfrow=c(1,1))
plot(c(x_test,14),c(y_test,10),col="#46626f",pch=16,ylab="v",xlab="w",family="Times New Roman",cex.lab=1.7,cex.axis=1.5,frame = FALSE,xlim=c(0,20),ylim=c(0,20),cex=2)
lines(c(0,x_test[order(x_test)],max(x_test[order(x_test)])), c(0,y_test[order(x_test)],0), xlim=range(x_test), ylim=range(y_test), pch=16,col="#46626f",lwd=2)
lines(c(0,x_test[order(x_test)][3:5],max(x_test[order(x_test)])), c(max(y_test),y_test[order(x_test)][3:5],0), xlim=range(x_test), ylim=range(y_test),
pch=16,col="#46626f",lwd=2,lty=3)
lines(c(x_test[3:5],x_test[5]),c(y_test[3:5],0), xlim=range(x_test), ylim=range(y_test), pch=16,col="#ddd3c8",lwd=2)

text(c(x_test+0.1,14+0.1),c(y_test+0.55,10+0.55),c("A","B","C","D","E","F"),family="Times New Roman")

####----- the DDF ----####
Input<-c(1,1,1,1,1)
Output_1<-c(10.5,15,18,17,12)
Output_2<-c(5,8.5,12.5,15,17.5)

x_test<-Output_2/Input
y_test<-Output_1/Input

par(mfrow=c(1,1))
plot(c(x_test,14),c(y_test,10),col="#46626f",pch=16,ylab="v",xlab="w",family="Times New Roman",cex.lab=1.7,cex.axis=1.5,frame = FALSE,xlim=c(0,20),ylim=c(0,20),cex=2)
lines(c(0,x_test[order(x_test)],max(x_test[order(x_test)])), c(0,y_test[order(x_test)],0), xlim=range(x_test), ylim=range(y_test), pch=16,col="#46626f",lwd=2)

#g(v,0)
text(14.3,18.5,expression(paste("g(",v^"k´",",0)")),family="Times New Roman",cex=1.5)
lines(c(14,14),c(10,17.3), xlim=range(x_test), ylim=range(y_test), pch=16,col="#46626f",lwd=2,lty=4)

#g(0,w)
text(3.8,10,expression(paste("g(0,",w^"k´",")")),family="Times New Roman",cex = 1.5)
lines(c(14,4.8),c(10,10), xlim=range(x_test), ylim=range(y_test), pch=16,col="#46626f",lwd=2,lty=4)

#g(v,w)
text(9.35,16.8,expression(paste("g(",v^"k´",w^"k´",")")),family="Times New Roman",cex=1.5)
lines(c(14,10),c(10,16.1), xlim=range(x_test), ylim=range(y_test), pch=16,col="#46626f",lwd=2,lty=4)

# The inffecient part of the frontier
lines(c(x_test[3:5],x_test[5]),c(y_test[3:5],0), xlim=range(x_test), ylim=range(y_test), pch=16,col="#ddd3c8",lwd=2)

# DMUs
text(15,9.8,expression(DMU^"k´"),family="Times New Roman",cex=1.5)
text(c(x_test+0.2,14+0.2),c(y_test+0.6,10+0.6),c("A","B","C","D","E",""),family="Times New Roman")

# Standard output oriented distance function
lines(c(0,17.35),c(0,12.35),col="#46626f",lwd=2,lty=3)
```