

Henriques, Sofia; McLaughlin, Eoin; Sharp, Paul; Tsoukli, Xanthi; Veddel, Christian

Working Paper

Opening the Black Box of the Danish Dairy Cooperatives: A Productivity Analysis

EHES Working Paper, No. 203

Provided in Cooperation with:

European Historical Economics Society (EHES)

Suggested Citation: Henriques, Sofia; McLaughlin, Eoin; Sharp, Paul; Tsoukli, Xanthi; Veddel, Christian (2020) : Opening the Black Box of the Danish Dairy Cooperatives: A Productivity Analysis, EHES Working Paper, No. 203, European Historical Economics Society (EHES), s.l.

This Version is available at:

<https://hdl.handle.net/10419/247133>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

EHES Working Paper | No. 203 | December 2020

Opening the Black Box of the Danish Dairy
Cooperatives: A Productivity Analysis

Sofia Henriques, Lund University

Eoin McLaughlin, University College Cork

Paul Sharp, University of Southern Denmark,
CAGE, CEPR

Xanthi Tsoukli, University of Southern Denmark

Christian Vedel, University of Southern Denmark

Opening the Black Box of the Danish Dairy Cooperatives: A Productivity Analysis*

Sofia Henriques, Lund University

Eoin McLaughlin, University College Cork

Paul Sharp¹, University of Southern Denmark, CAGE, CEPR

Xanthi Tsoukli, University of Southern Denmark

Christian Vedel, University of Southern Denmark

Abstract

The rapid spread of the Danish dairy cooperatives from the 1880s until the First World War is often portrayed as a uniform wave which swept the country. We investigate this using exceptionally detailed micro-level panel data taken from the Operational Statistics of Creameries, which were published from 1898 until after the Second World War. Our database comprises 1419 creameries over the period 1898- 1945 and no less than 131 variables. We document the data, and use a simple fixed effects setup to demonstrate considerable heterogeneity in the productivity of the individual creameries both over time and across space. We conclude by suggesting reasons for this, including scale of production, accessibility of fuel, religious institutions, and more.

JEL Codes: N53, N54, Q13

Keywords: Creameries, Denmark, productivity

¹ Corresponding Author: Paul Sharp (pauls@sam.sdu.dk)

* This project was made possible through generous funding by Independent Research Fund Denmark (DFR – 6109-00123). Moreover, the digitization required many research assistant hours. Two of the authors of this paper started as RAs, but we would also like to thank Peter Baunsgaard, Peter Gerner Ejersbo, Nick Ford, Jakob Baltzer Henriksen, Maja Uhre Pedersen, Andreas Slot Ravnholt and Emil Nørmark Sørensen for extremely conscientious research assistance. Finally, various iterations of this project have been presented at the Cork FRESH Meeting, the 4th HEDG Mini-Workshop for Young Scholars, the 2nd SDU Workshop on Economic and Business History, the 3rd Annual Meeting of the Danish Society for Economic and Social History, two DGPE workshops, and several times at the PhD seminar series of the Department of Business and Economics at the University of Southern Denmark. We would like to thank every participant at these events for insightful questions and comments.

Notice

The material presented in the EHES Working Paper Series is property of the author(s) and should be quoted as such. The views expressed in this Paper are those of the author(s) and do not necessarily represent the views of the EHES or its members

1 Introduction

The rapid spread of the Danish dairy cooperatives from the 1880s until the First World War, based on the use of a new technology, the steam-powered automatic cream separator, a centrifuge, was a defining moment in the development of Denmark, leading to its rapid convergence with the leading economies of the time, and an unusually balanced growth path between town and country (Bjørn 1982, Henriksen 1993, Khaustova and Sharp 2015). Denmark successfully outcompeted traditional suppliers of the important British market for butter, such as Ireland, and was later to become an important provider of knowhow to other countries, such as the United States (Boberg-Fazlic and Sharp 2019) and Russia (Korchmina and Sharp 2020). A sizeable body of research has attempted to explain this, but common to these explanations is an impression that the cooperatives rolled over Denmark in something of a uniform wave, which is often compared to relative failure in other countries. In the present work we document a large microlevel database of Danish creameries from the late-nineteenth century until the Second World War, and use a simple fixed effects setup to demonstrate that there was also heterogeneity between regions of Denmark. Although we leave the explanation of this to future work, we conclude by suggesting some potential hypotheses.

Many reasons have been given for the success of the Danish cooperative creameries. Henriksen (1999) demonstrates that cooperation was particularly well-suited to dairying and that given the technology of the time they were ideally suited to overcoming the problems of potential lock-in and asymmetric information, although they were not necessarily technologically “savvy”, and often dragged their feet in the implementation of new technologies (Henriksen and Hviid 2005). Denmark was, however, quick to embrace “winter dairying”, which allowed producers to enjoy higher prices at times of the year when more traditional operators were unable to produce (Henriksen and O’Rourke 2005). The homogeneous culture of Denmark, in contrast to the divisions of Ireland, have also been suggested as a reason why cooperation found special favor (O’Rourke 2006, 2007), although Henriksen et al (2012) and Henriksen et al (2015) demonstrate that social cohesion was not enough, and explain that the legal system and in particular the ability to enforce contracts was also important, and McLaughlin and Sharp (2019) explain that a lack of sizeable proprietary competitors also explained Denmark’s relative success compared to Ireland. Henriksen et al (2011) demonstrate that the productivity of the cooperatives in terms of butter production owed mostly to their rapid adoption of the centrifuge rather than their organizational form, and Lampe and Sharp (2015) show that farmers were able to increase milk yields through the introduction of multiple innovations in for example breeding and feeding. Henriques and Sharp (2016) explain that Denmark was also fortunate in having a particular geography with a long coastline which allowed coal, a vital input, easy and cheap access to the entire country, although this followed centuries of search for coal in a country where it was practically nonexistent (Ranestad and Sharp 2020). Finally, Lampe and Sharp (2018) and Boberg-Fazlic et al (2020) have demonstrated the role of traditional landed elites who

introduced new practices to Denmark from the eighteenth century, and through continuous innovation laid the basis for the rapid spread of the cooperatives in the final decades of the nineteenth century.

An important innovation of the aforementioned elites was the early adoption of sophisticated agricultural accounting, probably unique in a world perspective at the time, which allowed farmers to make rational decisions regarding both how and what to produce (Lampe and Sharp 2017, 2019). An important part of this was the dissemination of knowledge through educational establishments, scientific journals, extension services, and other publications, which allowed for the rapid diffusion of best practice. Although the collection and standardization of accounting material for landed estates had a longer history, the first attempt to do this for the cooperatives was under M.C. Pedersen in 1884, just two years after the first was established in 1882, at the request of the then chair of the Association of Agricultural Societies of Jutland’s dairy committee, Frederik Friis, who hoped to be able to demonstrate their inferiority compared to privately owned concerns. In fact, Pedersen’s report, which covered just seven cooperatives creameries in western Jutland, ended up demonstrating the opposite and thus providing an important boost to the cooperative movement. The dairy consultant Bernhard Bøggild carried on his work in a number of publications between 1897 and 1892. When creamery associations, representing the cooperatives at a regional level, were established, they took responsibility for this, and various reports were published covering increasing numbers of creameries for different parts of the country between 1891 and 1897 (MDS 1897/1899).¹ In 1897 it was decided to request that all cooperatives submit accounts for publication by a central organization, and it is these “Operational Statistics of Creameries” (*Mejeridriftsstatistik*, MDS) which we have hand-collected from the original published volumes and allow us to construct detailed microlevel panel data over many decades. Our database comprises 1419 creameries over the period 1898-1945 and no less than 131 variables. We are not aware of any other such detailed microlevel database for the major industry of any country covering such a long period, and in such detail. MDS continued to be published until the early 1970s, but increased rationalization of the industry and a lack of comparability to earlier volumes means that we chose to end in 1945.

The remainder of this paper proceeds as follows. In the following section, we explain the background to the collection of the statistics, and the challenges which were faced, and in Section 3 we provide an overview of the data. Section 4 describes our empirical strategy, and Section 5 presents our results. Section 6 concludes with some suggestions for potential avenues for future research using the database.

¹To avoid confusion, we refer to the published volumes of MDS throughout by MDS followed by two years. The first is the year of the data covered in that volume, the second is the year of publication.

2 Background to the MDS

The first volume of MDS (1897/1899) provides a helpful summary of the battle to establish the Operational Statistics. In the winter of 1897, the Creamery Associations of Jutland, Funen and Zealand, which represented the cooperatives in their respective regions, applied for government funding of up to 4000 kroner from the Ministry of Agriculture to collect statistics from their members. They received a reply on September 24, 1897 addressed to the chairman of the United Creamery Associations of Jutland (*samvirkende jydske Mejeriforeninger*). The ministry explained that they had exchanged letters with the Danish Royal Agricultural Society, an organization founded in the eighteenth century and largely representing traditional landed elites, and had ultimately concluded that they could not proceed without first receiving more details of exactly how the information was to be collected and processed, how much it would cost, and whether it was clear that a large proportion of creameries would be willing to share their accounts. Moreover, they requested that further planning should be in collaboration with the Royal Agricultural Society, which had declared itself willing. The Creamery Associations wrote to the Royal Agricultural Society on October 6, 1897, and held their first meeting with them on October 26 of the same year. There was general agreement about the importance of the project, but disagreement regarding how exactly it should be organized. Ultimately, however, they agreed on the following:

1. That it should be led by a committee of three members, one from the Royal Agricultural Society, one from the Creamery Associations, and the Ministry of Agriculture's dairy consultant.
2. That the annual workplan should be determined at a meeting of delegates, where each Creamery Association participated with one representative, and other provincial agricultural organizations and the Danish Association of Dairymen (*Dansk Mejeristforening*) should also be represented. The idea was that this would facilitate fruitful collaboration.
3. The costs were calculated as at least 6000 kroner the first year, and the requested support was increased to this amount.

One member subsequently decided to take issue with the first point, and this led to continued negotiations, which were additionally delayed by the poor health of the president of the Royal Agricultural Society, Jørgen Carl la Cour. Under the new president, Count Gustav Wedell Wedellsborg, point 1 was changed so that the committee should consist of four members: the president of the Royal Agricultural Society, its dairy consultant, and two representatives of the Creamery Associations, while the other points were left substantially unchanged. The Ministry agreed to support this initiative with 4000 kroner for the budget year 1898-99, and parliament accepted this. This was communicated in a letter from the Ministry dated April 15, 1898 to the United Creamery Associations of Jutland, stating that they would leave the responsibility to the Royal Agricultural Society and the Creamery Associations to perform the task to the extent that the amount offered could cover.

On May 16, 1898 the committee as constituted met for the first time, and it was decided that responsibility for organizing the statistics should mostly rest with the Creamery Associations, so that three committee members would now come from them, and just one from the Royal Agricultural Society, and that point 2 should be removed. A new proposal was sent to the ministry proposing that the 4000 kroner would be administered by those four members. Half was to be distributed in amounts of 5 kroner per creamery which was a member of the associations, under the condition that each association should contribute the same amount, that they submitted accounts which could be used for the Operational Statistics, that a report was only printed for an association if at least 15 usable accounts were submitted, and that only two copies of the statistics were to be sent to each creamery in each association. In order for the accounts to be usable, creameries were required to use a certain standardized form, covering 365 days and with an accounting year ending somewhere between October 1, 1897 and January 1, 1898. The other half of the money would be at the disposal of the committee to cover expert advice and printing costs, with 1000 kroner to each activity. The Ministry accepted this revised plan on September 27, 1898 (MDS 1897/1899).

Although it was originally planned that the statistics would be compiled separately for each Association, the committee eventually agreed that it made more sense to collect the reports in one volume, which was subsequently published each year with a short introduction. This was distributed to each Creamery Association with 10 copies per creamery, and one was sent to creameries outside the associations, to parliament members, to chairs of agricultural societies, etc. (MDS 1897/1899). By 1901, the organization was running an increasing deficit and asked the government for an extra 1500 kroner (MDS 1901/1902), but ultimately what happened was that the MDS were combined with the publication of butter price statistics, which were previously published elsewhere, with government support of 9000 kroner. Production was turned over to a Committee for Creamery Statistics (*Udvalg for Mejeri-Statistik*), consisting of one representative from the Royal Agricultural Society, two from the United Danish Agricultural Associations (*De samvirkende danske Landboforeninger*), three from the United Danish Dairy Associations (*De samvirkende danske Mejeriforeninger*), and one from the Association of Danish Dairymen (*Dansk Mejeristforening*) (MDS 1902/1903).

Besides the obvious value of the massive amount of data collected, the introduction to each volume provides a wealth of important information and analysis of the published statistics, touching on issues such as fuel prices, tuberculosis, and war, as well as the more mundane aspects of how to run a creamery based on the results of their statistical analysis. The editors frequently emphasized the importance of submitting accounts, and complained about low compliance rates in the early years, although this improved from around a third in the first years of the twentieth century, to about half by 1910, and two-thirds by the late 1920s, at which point it is stated that “it can be considered completely responsible to use the calculated averages as an expression of the general situation of dairying” (MDS 1929/1930). Although the first submissions of accounts from the individual creameries were patchy and many could not be used, and were thus not published, already in the volume for 1899 improvement is noted, which was greatly

facilitated by the publication of standardized accounting books.

From the beginning, the editors encouraged more creameries to submit their accounts, explaining that the comparison would facilitate improvements for the individual creamery, and for the situation of dairying in the country as a whole. One editor provided a particularly pertinent motivation:

“It can easily be understood that replying to the given questions gives a very reliable picture of the individual creamery before and now, not to mention the historical and general statistical interest that the collection of this information has at a time when it is still possible to collect authentic data from the beginning of the cooperatives. The present generation will, quite naturally, not attach great value to this information, since the development is occurring before their eyes; but future descendants will quite surely value it highly.” (MDS 1900/1901, p.3)

A later editor explained that “Accounts are the compass of the creamery. Without them one manages blindly, and if one is to have one’s accounts accounting-wise in order, it is important that they are not misleading, but point in the direction of the greatest possible return under the given conditions” (MDS 1910/1911, p. XV). Occasionally, editors speculated about the reasons why some creameries did not submit their accounts, for example because they were embarrassed by bad results, because their accounting system was incompatible with that of the MDS, or that they were simply happy to freeride on the efforts of others (MDS 1902/1903), something which was even once described as morally wrong (MDS 1928/1929). There is no doubt that the entire sector was put under a great deal of pressure to provide accurate accounts, and despite the weaknesses in the data discussed below, we believe that our database can be considered to provide a good snapshot of the condition of dairying in Denmark over the period covered.

3 Overview of the database

Here we provide an overview of the variables we have collected from MDS. It should be noted that, partly due to time constraints, we did not collect every piece of published information, and chose to exclude some for reasons of lack of comparability between volumes, or our own subjective lack of interest. The data we decided to collect from MDS can be roughly divided into four categories:

1. Basic characteristics
2. Input/Output
3. Technology and energy sources
4. Financial statistics

Appendix A contains tables with a description of all the variables presented by category as well as the range of years for which the variable is available (Tables A.1-A.4). Furthermore, appendix B contains summary statistics for all of these variables (Tables B.1-B.4).

As noted above, the data collected is not without its flaws, in particular in the early years, with the most obvious issue being that not every cooperative submitted accounts. We can only speculate, as did the editors of MDS, as to the reasons for this, and whether there might be certain characteristics common to those who did not send their accounts. For example, although it might be that “worse” or less efficient creameries were less likely to have kept good accounts which would have made it possible for them to participate, it might also be the case that the most modern creameries simply had less interest in the project, perhaps because they received useful information from elsewhere, or they were content to freeride, which would certainly have been rational at the level of the creamery. Whatever the case, our analysis below demonstrates the heterogeneity we sought to uncover, which might only have been greater if we had access to the full sample of creameries.

In their introductions to each volume of MDS, the editors describe the most important issues that the accounts submitted had. Some were of course simply not compatible with the need to provide meaningful data for comparison between creameries. For example, the first accounts, which only cover Funen and Jutland, totaled 156 of which only 101 could actually be used, either because they remained, which were thus not published, did not cover the relevant period, or for other reasons. Another issue which was raised almost every year was that creameries failed to harmonize their accounting years, although the actual dates covered are recorded in our database. The accounts for cheese production are mentioned as being particularly incomplete for the early years, reflecting the lack of importance attached to this minor part of Danish dairying (although see Henriksen et al, 2012). Making the accounts comparable was also a considerable task, although became easier as standardized accounting forms became more widely used. In the first volume, some Associations chose to anonymize their data, and we thus did not include them (MDS 1897/1899). In the MDS for 1899 (1900), very few accounts were received from Maribo, and these were thus merged with Sorø and Præstø, which also included by mistake those for Odden (Holbæk). As the accounting forms became more standardized, some creameries stopped participating for some years, since their accounts took some work to make compatible, although this issue was only temporary (MDS 1909/1910). Pricing the biproducts of butter production, i.e. buttermilk and skim milk, as well as whey from cheese production, also proved difficult, and the system for this was changed with the MDS for 1916 (1917), meaning that the prices before and after this year are not directly comparable. Finally, the small Association in Bornholm did not send accounts for 1926 and 1927 due to changes in the accounting year. The variables not digitized are within the following approximate categories:

1. Classical accounts (incomes, expenses, etc);
2. Cheese statistics;
3. Milk production and milk fat distributed over the months of the year.

A key issue with the construction of large-scale hand-entered databases such as this one is the validation of the data. We have however used several techniques to ensure the greatest possible accuracy:

1. Judgement based: From previous work, we have knowledge of realistic values of e.g. the milk to butter ratio, which was the main productivity measure recorded (see also below). Thus, we were able to verify or correct unlikely values. Furthermore, we have visually inspected all observations in plots and checked all unlikely entries.
2. Statistical methods: A method centered on outliers in individual variables and a regression based technique.

All of these approaches are presented in detail in appendix C, and might be of value to others considering similar exercises.

4 Empirical strategy

Beyond documenting the database, our aim here is to determine whether the productivity of the individual creameries differed to a significant extent between regions, as well as to rule out trivial explanations for this. As noted above, a common efficiency measure of the time was the milk/butter ratio (MB ratio) (see e.g. Henriksen et al, 2011), that is the weight of milk necessary to produce one weight-unit of butter. Thus, in order to document heterogeneity in productivity we model this ratio non-parametrically using regional dummies and annual fixed effects, i.e.

$$MB_ratio_{it} = \sum_j \beta_j 1[Region_consistent_{it} = j] + \sum_t \gamma_t 1[Year_{it} = t] + \varepsilon_{it} \quad (1)$$

The set that j iterates over contains all regions and the set that t iterates over contains all years. $1[x]$ is an indicator function returning 1 if x is true and 0 otherwise. The reference year is chosen as 1898 (the first year) and Silkeborg is chosen as the reference region, since it is the one with an average MB_ratio closest to the overall average. If all elements of β vary significantly from zero, this is evidence that variation within every single year can be explained by regional dummies i.e. there exists regional variation in productivity.

Principally, we can think of efficiency of butter production in terms of milk input being affected by two factors:

1. The technology used in processing the milk into butter, i.e. the ability to extract as much butterfat from the milk as possible; and
2. The quality of the milk, i.e. the fat or cream content of the milk.

Our main specification does not however capture possible heterogeneity induced from the differing quality of milk, which might principally be determined by variation in the breed of cows, with varying geographic

accessibility². Thus, if we are to argue that any heterogeneity observed can be found in the production itself, we need to rule out this source of variation. Another determinant of productivity is the scale of production. Appendix D documents how some regions reported creameries of much larger size than others. Taken together, this motivates the following alternative semi-parametric specification:

$$MB_ratio_{it} = \sum_j \beta_j 1[Region_consistent_{it} = j] + \sum_t \gamma_t 1[Year_{it} = t] + \delta_1 \log(Cows_now_{it}) + \delta_2 Milk_fat_{it} + \varepsilon_{it} \quad (2)$$

We only have information on milk fat content from 1929. For this reason we show results both with and without milk fat. We refer to the the specification with milk fat as 2A and the specification without as 2B.

Finally, we might suspect that the regional heterogeneity changes over time, and this motivates the following flexible specification:

$$MB_ratio_{it} = \sum_{j,t} \beta_{j,t} 1[Region_consistent_{it} = j \wedge Year_{it} = t] + \delta_1 \log(Cows_now_{it}) + \varepsilon_{it} \quad (3)$$

A problem in using MB ratio as an outcome variable is that some of the heterogeneity will be caused by randomness in the production process, making the ratio an unfit measure of efficiency of the individual creamery. This could be addressed by decomposing the random and inefficiency elements of the heterogeneity in a Stochastic Frontier Model approach³. However, we are not interested in obtaining a measure of individual productivity of each creamery, and for this reason the MB ratio suffices for our purpose and simplifies the analysis somewhat. Nevertheless, an analysis based on SFA was also performed and gave similar results. We used the specification of Battese & Coelli (1995), where explanations of the inefficiency are modeled directly in a maximum-likelihood framework. However, we encountered the so-called wrong skew problem, which causes the Hessian and thereby all classical standard errors to be undefined (Kumbhakar, Parmeter, & Zelenyuk, 2020), thus motivating our focus on the non-parametric specification outlined above.

²See Lampe & Sharp (2015) for an analysis of cow productivity.

³See e.g. Henriksen et al. (2011) and Lampe and Sharp (2015).

Table 1: Regression table for specifcaiton 1 and 2

	Specification 1	Specification 2A	Specification 2B
(Intercept)	26.463 (0.075)***	41.198 (1.120)***	27.539 (0.267)***
Region_consistentBornholm	-0.164 (0.016)***	-0.073 (0.039)	-0.108 (0.009)***
Region_consistentFyn	-0.304 (0.005)***	-0.054 (0.055)	-0.316 (0.014)***
Region_consistentHobro	0.263 (0.009)***	0.060 (0.020)*	0.255 (0.004)***
Region_consistentHolbæk	-0.325 (0.016)***	-0.103 (0.038)*	-0.362 (0.025)***
Region_consistentKøbenhavn og Frederiksborg	-0.247 (0.013)***	0.032 (0.020)	-0.206 (0.012)***
Region_consistentMaribo	-0.286 (0.004)***	-0.004 (0.037)	-0.230 (0.012)***
Region_consistentMidtjydsk	-0.287 (0.008)***	-0.013 (0.023)	-0.284 (0.014)***
Region_consistentRanders	0.065 (0.010)***	-0.052 (0.005)***	0.065 (0.009)***
Region_consistentRibe	0.323 (0.002)***	0.111 (0.021)***	0.297 (0.003)***
Region_consistentRingkøbing	0.699 (0.012)***	0.214 (0.035)***	0.685 (0.012)***
Region_consistentSorø, Præstø og Møn	-0.340 (0.004)***	-0.079 (0.035)	-0.345 (0.006)***
Region_consistentSønderjydsk	0.681 (0.023)***	0.188 (0.033)***	0.677 (0.025)***
Region_consistentThisted	0.335 (0.006)***	-0.080 (0.045)	0.445 (0.012)***
Region_consistentVejle	-0.219 (0.002)***	-0.102 (0.017)***	-0.266 (0.003)***
Region_consistentVendsyssel	0.288 (0.023)***	-0.016 (0.007)	0.283 (0.026)***
Region_consistentViborg	0.325 (0.003)***	0.083 (0.031)*	0.334 (0.006)***
Region_consistentAalborg	0.757 (0.010)***	0.070 (0.043)	0.746 (0.013)***
Region_consistentAarhus	-0.351 (0.002)***	-0.027 (0.025)	-0.351 (0.004)***
log(Cows_now)		-0.076 (0.017)***	-0.163 (0.036)***
Milk_fat		-4.659 (0.310)***	
Regional F	5609979717***	78701263***	1651790***
Year fixed effects	Yes	Yes	Yes
Years	1898-1945	1929-1945	1898-1945
R ²	0.766	0.830	0.763
Num. obs.	32951	8525	27517

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. 'Regional F' is the F-statistic of the hypothesis that all regional parameters equal zero.

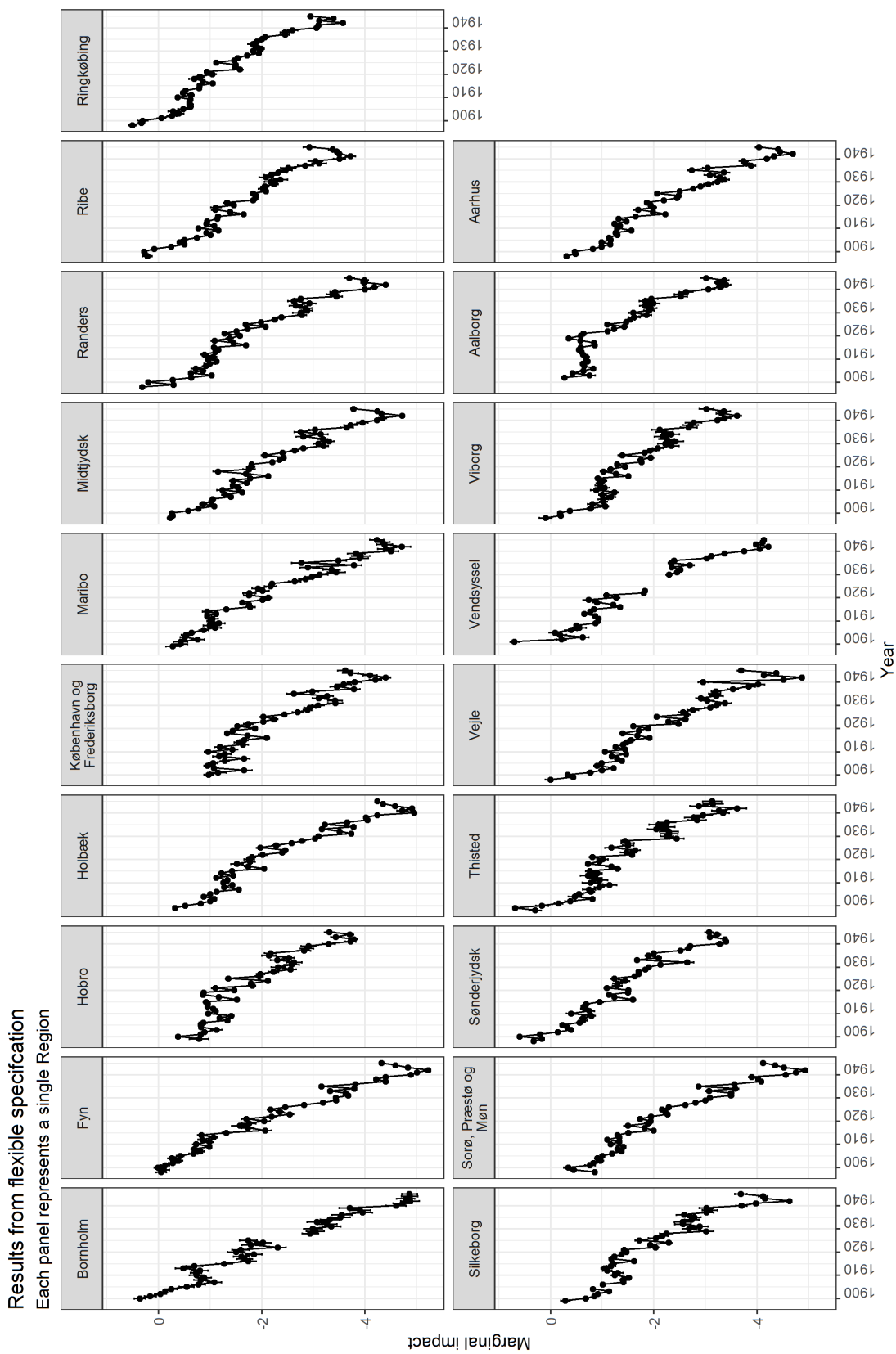


Figure 1: This figure displays all the parameters of specification 3. Each of these should be interpreted as differences from Silkeborg in 1898 at fixed creamery size measured by $\log(\text{cows_now})$.

5 Results

Table 1 presents our regression results from the first two specifications.⁴ The reference region is chosen to match the mean MB ratio for that period, meaning that deviations can be roughly compared to the average. Note that there is considerable regional variation in productivity, as revealed by the very strong F-statistic on the regional dummy parameters. This does not qualitatively change when taking milk quality and production scale into account. It should be noted that although there is regional variation in creamery size, the parameters on the regional dummies do not change much between specifications 1 and 2B. This is evidence that the regional variation is not only determined by regional heterogeneity in creamery size, and that some other factor must be of importance. The parameters do however change a lot when controlling for milk quality.

Because of the relatively large amount of data it is not hard to provide precise estimates, but these also show that our results are economically significant and consistent with at least one feasible mechanism. Thus, what is generally observed is that MB ratios are larger in regions which are in the geographic periphery of Denmark. In Ringkøbing (the worst performer) around 1.05 kg of additional milk is needed for every kg of butter when compared to Aarhus (the best performer). Moreover, 0.24 kg of this difference can be attributed to factors other than milk quality, at least for the years 1929-1945 where we have data on this.⁵ Aarhus is the largest city in Jutland, while Ringkøbing is in the rural western part of Jutland.

Specification 3 contains 890 individual parameters. For this reason, it is simply not feasible to show a regression table for every one of them. However, a figure for each estimate can communicate the information in a far more efficient manner, and Figure 1 contains all the parameter estimates. This shows changes in MB ratio, which are not attributable to the scale of production each year, compared to Silkeborg in 1898, and each point represents a parameter estimate. We have added error bars representing a 95% confidence interval to each estimate, but the error is so small that it is hardly noticeable. Overall, it can be noted that productivity increases across all regions over time, and that this development is fairly consistent. However, the different vertical positions of each of the curves demonstrates how it was not at all the same across the different regions. Again, a clear core-periphery pattern emerges.

As discussed in Section 2, the data does not describe the entire set of creameries in Denmark at the time. Of particular concern is that missing creameries might be so due to some process which is correlated with regions. The most likely root of problems like this could be that less well-performing creameries might have underreported or that some creameries might even have reported fake positively skewed data and that this issue would vary across regions. The typical MB ratio was well-known and might have

⁴Regression tables are generated with Leifeld (2013).

⁵One might be concerned that the changing range of years generates this changed difference. Appendix E contains a robustness check with no qualitative difference in the conclusion, with a difference between Aarhus and Ringkøbing of 1.21 kg, when not controlling for Milk fat in 1929-1945.

provided an incentive for well-performing creameries to submit their impressive reports. On the other hand, underperforming creameries might have reported results which were better than in reality. If this is the case, we expect it to be the case in generally underperforming regions, since the potential signaling gain is greater, and this would thus generate a smaller difference between regions than was actually the case. Consequently, our estimates of regional heterogeneity might be considered a lower bound.

Finally, we have controlled directly for creamery size, but production scale is not a trivial matter. Specifically there might be external economies of scale. Individual creameries were dependent on a sufficient supply of milk in order to ensure the efficient use of the machinery, given the technology of the time. At the other end of the spectrum, if supplies became too large to handle, they might have needed time to invest in an additional centrifuge, for example.

6 Discussion and conclusion

We have presented a new microlevel panel database of cooperative creameries in Denmark for the period 1898-1945. As a first use of this, and in contrast to the implicit assumption often given in the literature that the creameries were a relatively homogeneous institution, we demonstrate considerable heterogeneity between the regions of Denmark in terms of the productivity as measured by the milk/butter ratio. Our econometric results should of course not be given a direct causal interpretation. However, the value of this finding is that it offers an insight into this previously unseen heterogeneity, which might provide inspiration for future research. For now, we note that the productivity of the modern dairies did not spread uniformly, that the regional heterogeneity we observe seems to be robust and attributable to other factors than the quality of the milk input and the production scale. Finally, we observe a clear core-periphery-pattern.

Why might this be? One possibility is institutional. Haue (1978) documents how the new-pietist ‘Inner Mission’ movement and its associated pressure to keep creameries closed on Sundays caused a loss of productivity. Since they were stronger in certain regions, this might be one possibility. Another possibility is geography, and access to fuel in particular. During the First World War and well into the 1920s, coal supplies were limited. For this period, access to coal might have differed according to location, and moreover alternatives, principally peat and firewood, were also mostly available in certain locations. Geography might also have played a role in terms of endowments in infrastructure. Since the main export market was the UK, there might have been an advantage for creameries closest to the coast. Finally, geographical spill-over could also be considered in this context. To some degree the pattern might be caused by agglomeration and external economies of scale in urban regions. Also, scale might have been a driver of productivity, although greater productivity might also have encouraged expansion, thus driving a spurious correlation. We leave all these potential determinants, and others, to future work, which is now made possible by opening the “black box” of the Danish creameries through new data.

References

- Battese, G. E., & Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*, 20 (2), 325–332. <https://doi.org/10.1007/BF01205442>
- Bjørn, C. (1982). *Dansk mejeribrug 1882-1914*. Aarhus: De Danske Mejeriers Fællesorganisation.
- Boberg-Fazlic, N., & Sharp, P. R. (2019). *Immigrant Communities and Knowledge Spillovers: Danish-Americans and the Development of the Dairy Industry in the United States*. CEPR
- N Boberg-Fazlic, PS Jensen, M Lampe, PR Sharp, C Skovsgaard ‘Getting to Denmark’: The Role of Elites for Development Discussion Papers on Business and Economics, University of Southern Denmark 5, 2020
- Haue, H. (1978). Mejerikrigen - En undersøgelse af de søndagshvilende mejeriers udvikling i Danmark. *Fortid Og Nutid*, 1.(17), 359–390. Retrieved from <https://tidsskrift.dk/fortidognutid/article/view/73600>
- Henriksen, I. (1993). The transformation of Danish agriculture. In K. G. Persson (ed.), *The Economic Development of Denmark and Norway since 1870*. Aldershot: Edward Elgar.
- Henriksen, I. (1999). Avoiding lock-in: cooperative creameries in Denmark, 1882-1903. *European Review of Economic History* 3 (1), pp. 57–78.
- Henriksen, I. and Hviid, M. (2005). Diffusion of new technology and complementary best practice: a case study. *European Review of Economic History* 9 (3), pp. 365–97.
- Henriksen, I., Hviid, M., & Sharp, P. (2012). Law and peace: contracts and the success of the Danish dairy cooperatives. *The Journal of Economic History*, 197-224.
- Henriksen, I., Lampe, M., & Sharp, P. (2011). The role of technology and institutions for growth: Danish creameries in the late nineteenth century. *European Review of Economic History*, 15(3), 475-493.
- Henriksen, I., Lampe, M., & Sharp, P. (2012). The strange birth of liberal Denmark: Danish trade protection and the growth of the dairy industry since the mid-nineteenth century 1. *The Economic History Review*, 65(2), 770-788.
- Henriksen, I., McLaughlin, E., & Sharp, P. (2015). Contracts and cooperation: the relative failure of the Irish dairy industry in the late nineteenth century reconsidered. *European Review of Economic History*, 19(4), 412-431.
- Henriksen, I., & O’Rourke, K. H. (2005). Incentives, technology and the shift to year-round dairying in late nineteenth-century Denmark 1. *The Economic History Review*, 58(3), 520-554.
- Henriques, S. T., & Sharp, P. (2016). The Danish agricultural revolution in an energy perspective: a case of development with few domestic energy sources. *The Economic History Review*, 69(3), 844-869.

Khaustova, E., & Sharp, P. (2015). A Note on Danish Living Standards through Historical Wage Series, 1731-1913. *Journal of European Economic History*, 44(3), 143-72.

Korchmina, E. and P. Sharp (2020). Denmark and Russia: What can we learn from the historical comparison of two great Arctic agricultural empires? *Russian Journal of Economics* (forthcoming)

Kumbhakar, S. C., Parmeter, C. F., & Zelenyuk, V. (2020). *Stochastic Frontier Analysis: Foundations and Advances I* (pp. 1–39). https://doi.org/10.1007/978-981-10-3450-3_9-1

Lampe, M., & Sharp, P. (2015). Just add milk: a productivity analysis of the revolutionary changes in nineteenth-century Danish dairying. *The Economic History Review*, 68(4), 1132-1153.

Lampe, M., & Sharp, P. (2017). A quest for useful knowledge: the early development of agricultural accounting in Denmark and Northern Germany. *Accounting History Review*, 27(1), 73-99.

Lampe, M., & Sharp, P. (2018). *A land of milk and butter: how elites created the modern Danish dairy industry*. University of Chicago Press.

Lampe, M., & Sharp, P. (2019). Accounting for the wealth of Denmark: A case study of Smithian growth using the emergence of modern accounting in Danish dairying. *The European Journal of the History of Economic Thought*, 26(4), 659-697.

Leifeld, P. (2013). Conversion of Statistical Model Output in R to LaTeX and HTML Tables. *Journal of Statistical Software*, 55 (8), 1—24. Retrieved from <http://www.jstatsoft.org/v55/i08/>

McLaughlin, E., & Sharp, P. (2019). Competition between organisational forms in Danish and Irish dairying around the turn of the twentieth century. *Business History*, 1-28.

O'Rourke, K. H. (2006). Late 19th century Denmark in an Irish mirror: land tenure, homogeneity and the roots of Danish success. *The state of Denmark: small states, corporatism and the varieties of capitalism*. McGill-Queen's University Press, Montreal, 159-196.

O'Rourke, K. H. (2007). Culture, conflict and cooperation: Irish dairying before the Great War. *The Economic Journal*, 117(523), 1357-1379.

Ranestad, K and Sharp, P. R. (2020). Success through failure? Four Centuries of Searching for Danish Coal. EHES Working Paper No. 183.

Source: Danmarks Mejeri-Drifts-Statistik, 1897/9-1945/6. The first year refers to the data covered by the data in that volume, and the second year refers to the publication date. All volumes are held in the collection of the National Library of Denmark.

Appendix:

Opening the Black Box of the Danish Dairy Cooperatives: A Productivity Analysis

Sofia Henriques, Lund University

Eoin McLaughlin, University College Cork

Paul Sharp, University of Southern Denmark, CAGE, CEPR

Xanthi Tsoukli, University of Southern Denmark

Christian Vedel, University of Southern Denmark

Contents

A Description of data	i
B Summary tables	vii
C Error correction	x
D Creamery sizes	xiii
E Does the range of years make a qualitative difference?	xv

A Description of data

The following table contain a description of all the variables in the database.

Table A.1: Data description: Basic characteristics

Variable name	Unit	Short description	Availability
Year	Years	Year of observation	1898-1945
Trend	Years	Convience variable that has the value 1 for the first year, 2 for the second, etc.	1898-1945
Creamery_name	Characters	Creamery name	1898-1945
Region_merged	Characters	Dairy association but merged to ensure the same region throughout the period	1898-1945
Dairy_association	Characters	Dairy association	1898-1945
Region_consistent	Characters	Region_merged but where a single creamery is only a member of one region throughout the period	1898-1945
Creamery_ID	None	Units of observation	1898-1945
Region_ID	None	ID for Region_consistent	1898-1945
Creamery_no	None	Creamery no. in the original source - useful for lookups	1898-1945
Established_year	Years	Year of establishment	1898-1945
Years_since_est	Count	Number of years since establishment	1898-1945
Production_days	Count	Number of days in the accounting year	1898-1903
Production_year_start	Date	First day of the accounting year	1898-1945
Production_year_end	Date	Last day of the accounting year	1898-1945
Production_year_string	Dates	Raw data entry of production year	1898-1945
Renovation_years	Characters	List of the years that the creamery was renovated	1898-1928
Renovation_year_dumm	Dummy	Dummy for every year the creamery was renovated	1898-1928
Board_members	Count	Number of board members	1905-1909
Shareholders_start	Count	Number of shareholders at the beginning	1900-1909
Shareholders_now	Count	Number of shareholders now	1898-1945

Table A.1: Data description: Basic characteristics (*continued*)

Variable name	Unit	Short description	Availability
Shareholders_large	Count	Number of shareholders with: 1900-1909: More than 30 cows, 1910-1928: More than 20 cows	1900-1928
Shareholders_small	Count	Number of shareholders with less than 4 cows. Less than 3 for years from 1904.	1900-1928
Collection_routes	Count	Number of routes for milk collection from farmers to the creamery	1898-1928
Route_length	Km	Average length of collection routes	1898-1918
Distance_to_railway	Km	Distance to nearest railway or harbour	1898-1912
BovineTB_programme	Dummy	Does the creamery participate in Bovine Tuberculosis reduction?	1924-1928
Row_ID	None	Individual number for every row of data	1898-1945
Butter_brand_member	Dummy	Is the creamery a member of a butter brand association?	1903-1904
Original_order	None	Used for matching up with other data	1898-1945

Table A.2: Data description: Financial

Variable name	Unit	Short description	Availability
Cost_acquisitions	DKK	Expenses for new buildings or large equipment	1929-1945
Amortization	DKK	Loan repayments	1929-1945
Funds	DKK	Set aside for reserve fund	1933-1945
Cost_construction	DKK	Construction cost	1898-1902
Fire_insurance	DKK	Fire insurance	1900-1945
Debt	DKK	Debt at the end of the year	1900-1945
Dividends	DKK	Total profit paid to shareholders	1901-1922
Interest_accr	DKK	Interest accrued	1901-1907
Supplement_classes	Characters	Is supplement given based on classes? If yes, then which classes?	1927-1928
Supplement_dkk	DKK	Supplements in DKK	1927-1928

Table A.2: Data description: Financial (*continued*)

Variable name	Unit	Short description	Availability
Deduction_classes	Characters	Is deduction taken based on classes? If yes, then which classes?	1923-1928
Deduction_dkk	DKK	Deductions in DKK	1923-1928
Price_setting	Characters	What price setting method is used? Factor,(F; 2 or 3), Cream (Fl.) or Fat Units (F.E. / F.E.H)	1908-1928

Table A.3: Data description: Input/Output

Variable name	Unit	Short description	Availability
Cows_start	Count	Number of cows in the first production year	1900-1909
Cows_now	Count	Number of cows in the creamery now	1898-1945
Cows_per_shareholder	Count	Number of cows per shareholder	1898-1928
Daily_milk_start	kg	Average daily milk in the first year of production	1898-1945
Daily_milk_last	kg	Average daily milk last year	1900-1907
Milk_total_start	kg	Total milk production in the first year of production	1898-1945
Milk_per_shareholder	kg	Milk production per shareholder	1898-1899
Milk_per_cow	kg	Milk production per cow	1898-1928
Milk_fat	%	Average percentage of milk fat	1929-1945
MB_ratio	ratio	Kg. of milk per kg. of butter output	1898-1945
Extraction_pct	%	How many kg. of Milk fat makes 100 kg. of butter?	1942-1945
Price_butter	DKK cents (Øre)	Butter price per kg.	1898-1945
Jersey_cows	Count	Number of Jersey cows that the creamery receives milk from	1910-1926
Milk_by_cream	Dummy	Does the creamery price the milk based on the cream content?	1903-1907
Milk_for_butter	kg	Total amount of whole milk used for butter production	1908-1918
Milk_for_butter_alt	kg	Milk used for butter production based on MB ratios	1898-1945

Table A.3: Data description: Input/Output (*continued*)

Variable name	Unit	Short description	Availability
Milk	kg	Amount of whole milk received at the creamery	1898-1945
Butter	kg	Output of total milk	1898-1945
Loss_in_production	kg	Loss in production as the difference between input and output	1898-1945
Pct_loss	%	Loss in production as a percentage	1898-1945
Waste_milk	kg	Registered waste milk	1898-1899
MB_ratio_cal	ratio	Calculated ratio of Milk to Butter	1898-1945
MBdif	ratio	Difference between MB_ratio reported and calculated (Residuals)	1898-1945
Milk_sold	kg	Milk that was sold	1908-1944
Milk_retail	kg	Milk sold to retail	1919-1945
Supplement_milk	kg	Whole milk to others - supplement milk	1919-1928
Milk_cheese	kg	Whole milk used for cheese making	1908-1945
Milk_total	kg	Total milk output	1898-1945
Cream_out	kg	Cream output	1898-1945
Skimbutter_total1	kg	Total amount of skim and butter milk recieved or from earlier production steps	1908-1918
Skimbutter_total2	kg	Total amount of skim and butter milk send out of the creamery	1898-1945
Skimbutter_shh	kg	Skim and butter milk for shareholders	1908-1945
Skimbutter_retoth	kg	Skim and butter milk for retail and for others	1929-1945
Skimbutter_ret	kg	Skim and butter milk sold in retail	1919-1928
Skimbutter_sup	kg	Skim and butter milk for others - supplement milk	1908-1928
Skimbutter_cheese	kg	Skim and butter milk for cheese production	1908-1945

Table A.3: Data description: Input/Output (*continued*)

Variable name	Unit	Short description	Availability
Skimbutter_cream	kg	Skim and butter milk out in cream	1908-1928
Total_out	kg	Total weight out	1898-1943
Butter_export	kg	Butter sold for export	1908-1945
Butter_shhret	kg	Butter for shareholders and export	1932-1945
Butter_expret	kg	Butter for export and shareholders	1930-1931
Butter_distshh	kg	Butter for distributor and shareholders	1929
Butter_retail	kg	Butter in retail	1929-1931
Butter_distr	kg	Butter for distributor	1919-1928
Butter_shh	kg	Butter for shareholders	1908-1928
Butter_other	kg	Butter for others	1908-1928
Butter_imp	kg	Butter sent out implicitly in cream for butter production. Assuming 1 kg. of butter per 4 kg. of cream.	1908-1928

Table A.4: Data description: Technology/Energy

Variable name	Unit	Short description	Availability
Which_refrigerator	Characters	Refrigiator brand	1900-1923
Refrigerator	Dummy	Does the creamery have a refrigerator?	1900-1923
Icehouse_size	m ³ / m ²	Size of icehouse	1900-1923
Icehouse	Dummy	Does the creamery have an icehouse?	1900-1923
Icehouse_raw	Characters	Raw data entry	1900-1925
Coal	Dummy	Does the creamery use coal as fuel?	1900-1928
Brushwood	Dummy	Does the creamery use brushwood as fuel?	1900-1928
Peat	Dummy	Does the creamery use peat as fuel?	1900-1928
Briquettes	Dummy	Does the creamery use briquettes as fuel?	1900-1928

Table A.4: Data description: Technology/Energy (*continued*)

Variable name	Unit	Short description	Availability
Firewood	Dummy	Does the creamery use firewood as fuel?	1900-1928
Coke	Dummy	Does the creamery use coke as fuel?	1900-1928
Heather	Dummy	Does the creamery use heather as fuel?	1900-1928
Sawdust	Dummy	Does the creamery use sawdust as fuel?	1900-1928
Mix	Dummy	Does the creamery use a mix of fuels?	1900-1928
Browncoal	Dummy	Does the creamery use brown coal as fuel?	1900-1928
Electricity	Dummy	Does the creamery use electricity as fuel?	1900-1928
Straw	Dummy	Does the creamery use straw as fuel?	1900-1928
Coal_scrap	Dummy	Does the creamery use coal scraps as fuels?	1900-1928
Centrifuges_brand	Characters	Brand and count of centrifuges	1913-1927
Centrifuges_alt	Count	Alternative count of centrifuges. Missing information is encoded as missing.	1913-1927
Centrifuges	Count	Count of centrifuges. It is assumed that missing information implies one centrifuge.	1913-1927
Q_smell	Dummy	Quality assesment by smell	1922-1928
Q_taste	Dummy	Quality assesment by taste	1922-1928
Q_reduc	Dummy	Quality assesment by reductase test	1922-1928
Elec	Dummy	Creamery lit by electricity	1903-1909
Candle	Dummy	Creamery lit by candles	1903-1945
Gaspetr	Dummy	Creamery lit by gasoline / petrol	1903-1909
Act	Dummy	Creamery lit by acetylin	1903-1945
Light_other	Dummy	Creamery's lighting source unknown	1903-1909
Lighting_raw	Characters	Raw data entry on lighting type used	1898-1944

Table A.4: Data description: Technology/Energy (*continued*)

Variable name	Unit	Short description	Availability
QAA_which	Characters	Which quality assesment association is the creamery a member of?	1908-1928
QAA	Dummy	Dummy for quality assesment association membership	1908-1928
Butter_water_control	Characters	How often does the creamery check for water content of butter?	1913-1928
Churn_brand	Characters	Brand of combined churn (if the creamery uses one)	1908-1924
Churn	Dummy	Dummy for the use of combined churn	1908-1924
Regenerative	Characters	String describing use of a regenerative device	1910-1928
Milk_cooling_temp	°C and others	If the creamery cools skim milk, then to what temperature? (Raw text)	1910-1928
Milk_cooling	Dummy	Does the creamery cool skim milk?	1910-1945
Whey_past	Dummy	Is the whey pasteurized?	1910-1912
Acidification	Characters	Does the creamery acidify the skimmed milk for suppliers? (Yes, no, partially)	1914-1928
Cheese_past	Dummy	Is cheese made from pasteurised milk?	1909-1912
RenCasein	Characters	Does the creamery produce rennet or acid casein?	1910-1913

B Summary tables

This section contains summary tables of all the variables in the database.

Table B.1: Summary statistics: Basic characteristics

Variable	Type	N	Mean	St.Dev	Min	q25	Median	q75	Max	N_categories	N_positive
Year	Numeric	33485	1924.5	12.64	1898	1914	1925	1935	1945	-	-
Trend	Numeric	33485	27.5	12.64	1	17	28	38	48	-	-
Creamery_name	Categorical	33485	-	-	-	-	-	-	-	1646	-
Region_merged	Categorical	33485	-	-	-	-	-	-	-	19	-

Table B.1: Summary statistics: Basic characteristics (*continued*)

Variable	Type	N	Mean	St.Dev	Min	q25	Median	q75	Max	N_categories	N_positive
Dairy_association	Categorical	33485	-	-	-	-	-	-	-	25	-
Region_consistent	Categorical	33485	-	-	-	-	-	-	-	19	-
Creamery_ID	Numeric	33485	709.65	412.71	1	368	702	1064	1428	-	-
Region_ID	Numeric	33485	10.06	5.49	1	5	10	14	19	-	-
Creamery_no	Categorical	33485	-	-	-	-	-	-	-	131	-
Established_year	Numeric	23975	1890.22	5.26	1879	1887	1888	1893	1908	-	-
Years_since_est	Numeric	23975	31.83	13.8	1	21	31	43	64	-	-
Production_days	Numeric	1877	365.08	3.82	304	364	365	365	426	-	-
Production_year_start	Categorical	31691	-	-	-	-	-	-	-	4108	-
Production_year_end	Categorical	31517	-	-	-	-	-	-	-	4007	-
Production_year_string	Categorical	33485	-	-	-	-	-	-	-	10313	-
Renovation_years	Categorical	19540	-	-	-	-	-	-	-	265	-
Renovation_year_dummy	Dummy	19540	0.04	0.19	0	0	0	0	1	-	700
Board_members	Numeric	2639	7.26	1.85	2	7	7	9	26	-	-
Shareholders_start	Numeric	3971	84.54	47.05	0	52	75	103	308	-	-
Shareholders_now	Numeric	32415	147.77	68.25	3	97	137	187	560	-	-
Shareholders_large	Numeric	15004	3.67	4.32	0	1	2	5	91	-	-
Shareholders_small	Numeric	15844	59.2	38	0	32	51	78	324	-	-
Collection_routes	Numeric	18808	9.88	4.19	0	7	9	12	70	-	-
Route_length	Numeric	6686	4.31	1.82	0.19	3.5	3.77	5.65	20.71432	-	-
Distance_to_railway	Numeric	5103	4.51	4.32	0.03	0.94	3.77	5.65	26.36368	-	-
BovineTB_programme	Dummy	1406	0.29	0.45	0	0	0	1	1	-	402
Row_ID	Numeric	33485	17603.6	10153.68	1	8650	17802	26436	34933	-	-
Butter_brand_member	Dummy	926	1	0.07	0	1	1	1	1	-	922
Original_order	Numeric	33485	17603.6	10153.68	1	8650	17802	26436	34933	-	-

Table B.2: Summary statistics: Financial

Variable	Type	N	Mean	St.Dev	Min	q25	Median	q75	Max	N_categories	N_positive
Cost_acquisitions	Numeric	13423	3737.32	10731.6	-5000	0	0	3488.5	362765	-	-
Amortization	Numeric	12450	4551.76	3982.17	0	2025.5	3799.5	5798.75	87114	-	-
Funds	Numeric	10025	340.09	2145.57	-44000	0	0	400	40967	-	-
Cost_construction	Numeric	956	21589.2	7302.58	1	17000	20567	26000	53178	-	-
Fire_insurance	Numeric	31910	78466.55	62337.8	0	35300	67000	100980	4520000	-	-
Debt	Numeric	31667	26665.3	29651.57	0	9000	18000	33300	455300	-	-
Dividends	Numeric	13317	38656.31	28194.41	0	19577	32486	50014	299780	-	-
Interest_accr	Numeric	2860	411.61	323.24	0	188	335	555	5907	-	-
Supplement_classes	Categorical	313	-	-	-	-	-	-	-	6	-
Supplement_dkk	Numeric	41	1601.51	1697.61	100	422.69	874.28	2417.5	7147.37	-	-
Deduction_classes	Categorical	2453	-	-	-	-	-	-	-	31	-
Deduction_dkk	Numeric	1440	715.8	716.28	0	245.63	534.7	921.8	6771.1	-	-
Price_setting	Categorical	8434	-	-	-	-	-	-	-	22	-

Table B.3: Summary statistics: Input/Output

Variable	Type	N	Mean	St.Dev	Min	q25	Median	q75	Max	N_categories	N_positive
Cows_start	Numeric	65	342.63	169.05	82	200	350	444	800	-	-
Cows_now	Numeric	27967	984.54	451.67	13	650	905	1233	8830	-	-
Cows_per_shareholder	Numeric	16749	6.59	3.83	1.2	5.4	6.3	7.3	266.7	-	-
Daily_milk_start	Numeric	17099	78450.03	2.82e+05	380	2000	2836	3937	2025731	-	-

Table B.3: Summary statistics: Input/Output (*continued*)

Variable	Type	N	Mean	St.Dev	Min	q25	Median	q75	Max	N_categories	N_positive
Daily_milk_last	Numeric	3474	8834.79	8105.83	522.5	4474.62	6400	9155.38	58000	-	-
Milk_total_start	Numeric	9516	1.13e+06	7.46e+05	64000	750000	1e+06	1375000	9996064	-	-
Milk_per_shareholder	Numeric	381	13708.55	4613.37	3452	10279.5	13473	16086.5	37111.5	-	-
Milk_per_cow	Numeric	16686	2540.98	470.55	780	2236	2573	2864	4759	-	-
Milk_fat	Numeric	10641	3.75	0.17	2.56	3.63	3.75	3.86	4.73	-	-
MB_ratio	Numeric	32948	24.3	1.35	19.1	23.3	24.5	25.3	29.6	-	-
Extraction_pct	Numeric	3029	85.91	1.57	80	84.8	85.8	86.9	91.6	-	-
Price_butter	Numeric	32546	312.79	127.26	0	204.2	267.7	441.28	737.7	-	-
Jersey_cows	Numeric	11606	2.4	16.42	0	0	0	0	630	-	-
Milk_by_cream	Dummy	2348	0.91	0.29	0	1	1	1	1	-	2131
Milk_for_butter	Numeric	7278	2.31e+06	1.07e+06	144394	1.51e+06	2.17e+06	2.99e+06	6734227	-	-
Milk_for_butter_alt	Numeric	32948	2.55e+06	1.27e+06	75375	1.62e+06	2332785	3.26e+06	9.59e+06	-	-
Milk	Numeric	33485	2.68e+06	1.37e+06	137960.5	1702508	2452942	3421012	48167115	-	-
Butter	Numeric	33485	1.05e+05	54409.1	102	65917	95603	134236	414947	-	-
Loss_in_production	Numeric	32838	-753.74	59199.92	-4.47e+05	-2.32e+04	41.5	21003.12	5175955	-	-
Pct_loss	Numeric	32240	0.01	1.81	-23	-0.9	0	0.9	22.2	-	-
Waste_milk	Numeric	183	78171.4	1.74e+05	0	10240	35664	74621.5	1953669	-	-
MB_ratio_cal	Numeric	33485	28.12	194.09	2.65	24.31	25.22	26.1	3.31e+04	-	-
MBdif	Numeric	32948	1.49	6.53	-23.88	0.08	0.31	1.13	4.3e+02	-	-
Milk_sold	Numeric	18182	68461.36	1.68e+05	0	5865.25	18803.5	54512.5	5603141	-	-
Milk_retail	Numeric	8246	41231.23	1.14e+05	13	5484	13137.5	31427	2840943	-	-
Supplement_milk	Numeric	1114	2.05e+05	2.59e+05	40	33806	90340	2.97e+05	2106382	-	-
Milk_cheese	Numeric	12447	1.26e+05	2.39e+05	0	12287	37392.5	131057	2532831	-	-
Milk_total	Numeric	24246	1.32e+05	2.49e+05	-874.5	9491	39695	127371.5	2840943	-	-
Cream_out	Numeric	16322	8526.47	24746.46	0	1144	2718	6386.75	462463	-	-
Skimbutter_total1	Numeric	7277	2.22e+06	1.03e+06	81953.5	1450717	2073748	2875187	6471865	-	-
Skimbutter_total2	Numeric	33005	2.47e+06	1.23e+06	0	1569670	2263310	3.16e+06	9313864	-	-
Skimbutter_shh	Numeric	29320	2.31e+06	1.22e+06	0.5	1.42e+06	2087702	2.97e+06	8918036	-	-
Skimbutter_retoth	Numeric	13484	85881.72	1.54e+05	13	13524.25	32114.5	76224.25	2033750	-	-
Skimbutter_ret	Numeric	7724	43505.55	54907.13	5.47	14171.75	29360	54000	935348	-	-
Skimbutter_sup	Numeric	8205	96940.37	1.6e+05	1.5	17507.5	41459	96900	1806916	-	-
Skimbutter_cheese	Numeric	13033	2.93e+05	3.1e+05	0	66830	185045	414600	2547755	-	-
Skimbutter_cream	Numeric	13714	3294.14	13990.22	0	497	1090	2553.75	428279.5	-	-
Total_out	Numeric	11913	2.37e+06	1.15e+06	0	1508413	2.18e+06	3.02e+06	9200065	-	-
Butter_export	Numeric	18429	96861.76	56635.89	0	57366	87262.5	126456	563241	-	-
Butter_shhret	Numeric	11351	16111.32	18903.19	11.16	4805	10051	20663	206376	-	-
Butter_expret	Numeric	1682	1.37e+05	62319.33	2399	91139.25	127546	1.75e+05	393148	-	-
Butter_distshh	Numeric	770	1.29e+05	64445.29	12921	84036.75	117896.5	163271	898559	-	-
Butter_retail	Numeric	2412	3566.03	9871.89	15	525	1140	2666.25	159610	-	-
Butter_distr	Numeric	8043	88838.56	48913.6	44	52184.5	80568	116297.5	355424	-	-
Butter_shh	Numeric	14939	3967.68	5779.05	8	1466	2832.5	5050	213105.5	-	-
Butter_other	Numeric	1745	3683.07	7025.88	26	597	1438	3755	116523	-	-
Butter_imp	Numeric	13769	1039.59	2919.32	0	171	375	885	75398	-	-

Table B.4: Summary statistics: Technology/Energy

Variable	Type	N	Mean	St.Dev	Min	q25	Median	q75	Max	N_categories	N_positive
Which_refrigerator	Categorical	7150	-	-	-	-	-	-	-	66	-
Refrigerator	Dummy	7150	0.47	0.5	0	0	0	1	1	-	3393
Icehouse_size	Numeric	5948	207.49	174.46	0	112.48	154.5	247.2	2966.4	-	-
Icehouse	Dummy	9741	0.61	0.49	0	0	1	1	1	-	5948

Table B.4: Summary statistics: Technology/Energy (*continued*)

Variable	Type	N	Mean	St.Dev	Min	q25	Median	q75	Max	N_categories	N_positive
Icehouse_raw	Categorical	13171	-	-	-	-	-	-	-	453	-
Coal	Dummy	18183	0.83	0.37	0	1	1	1	1	-	15133
Brushwood	Dummy	18183	0.05	0.21	0	0	0	0	1	-	854
Peat	Dummy	18183	0.05	0.21	0	0	0	0	1	-	854
Briquettes	Dummy	18183	0.01	0.1	0	0	0	0	1	-	176
Firewood	Dummy	18183	0.06	0.24	0	0	0	0	1	-	1137
Coke	Dummy	18183	0	0.02	0	0	0	0	1	-	8
Heather	Dummy	18183	0	0.01	0	0	0	0	1	-	2
Sawdust	Dummy	18183	0	0.01	0	0	0	0	1	-	1
Mix	Dummy	18183	0	0.02	0	0	0	0	1	-	5
Browncoal	Dummy	18183	0	0.04	0	0	0	0	1	-	33
Electricity	Dummy	18183	0	0.03	0	0	0	0	1	-	15
Straw	Dummy	18183	0	0.04	0	0	0	0	1	-	23
Coal_scraps	Dummy	18183	0	0.01	0	0	0	0	1	-	2
Centrifuges_brand	Categorical	9124	-	-	-	-	-	-	-	127	-
Centrifuges_alt	Numeric	9122	1.8	0.6	1	1	2	2	4	-	-
Centrifuges	Numeric	11737	1.62	0.62	1	1	2	2	4	-	-
Q_smell	Dummy	3298	0.38	0.48	0	0	0	1	1	-	1241
Q_taste	Dummy	3298	0.38	0.48	0	0	0	1	1	-	1238
Q_reduc	Dummy	3298	0.49	0.5	0	0	0	1	1	-	1629
Elec	Dummy	3550	0.2	0.4	0	0	0	0	1	-	721
Candle	Dummy	4310	0	0.03	0	0	0	0	1	-	4
Gaspetr	Dummy	3550	0.76	0.42	0	1	1	1	1	-	2711
Act	Dummy	4310	0.02	0.15	0	0	0	0	1	-	105
Light_other	Dummy	3550	0	0.05	0	0	0	0	1	-	9
Lighting_raw	Categorical	32725	-	-	-	-	-	-	-	30	-
QAA_which	Categorical	5626	-	-	-	-	-	-	-	32	-
QAA	Dummy	5626	0.42	0.49	0	0	0	1	1	-	2346
Butter_water_control	Categorical	9215	-	-	-	-	-	-	-	49	-
Churn_brand	Categorical	8829	-	-	-	-	-	-	-	229	-
Churn	Dummy	8834	0.82	0.38	0	1	1	1	1	-	7241
Regenerative	Categorical	7905	-	-	-	-	-	-	-	119	-
Milk_cooling_temp	Categorical	4928	-	-	-	-	-	-	-	253	-
Milk_cooling	Dummy	5679	0.51	0.5	0	0	1	1	1	-	2896
Whey_past	Dummy	559	0.16	0.36	0	0	0	0	1	-	87
Acidification	Categorical	2622	-	-	-	-	-	-	-	4	-
Cheese_past	Dummy	1035	0.1	0.3	0	0	0	0	1	-	105
RenCasein	Categorical	770	-	-	-	-	-	-	-	6	-

C Error correction

Digitizing 33485 observations of 131 variables ($131 \times 33485 = 4386535$ individual entries) by hand is almost guaranteed to involve some errors. Even a very low error rate - say 2 - would cause around 90000 errors. Any use of this data should consider entry mistakes as a source of error. However, we have done several things to mitigate this potential problem. In total we have used four different procedures - two statistical and two based on domain knowledge.

In the first statistical approach we transformed each variable subtracting the mean, at the creamery level. Let x be a variable of interest then

$$\tilde{x}_{it} = x_{it} - \bar{x}_i, \tag{C.1}$$

where i represents the creamery and t the year. Every \tilde{x}_{it} more than 5 standard deviations away from the mean was flagged as a potential error. Intuitively such an observation is very different from what is typical for that particular creamery.

The second statistical approach is regression based. We calculated so-called dif betas for each variable in the following year fixed effects regression

$$\log(\text{Butter}_{it}) = \alpha_t + \beta_0 \log(\text{Milk}_{it}) + \beta_1 x_{it} + \varepsilon_{it} \tag{C.2}$$

This can be thought of as a rough estimate of a Cobb-Douglas butter production function with different TFP each year, α_t , at constant levels of x . Dif beta is a measure for each it of how much our estimate of β_1 would change if observation x_{it} was left out. Intuitively, a large value of dif beta corresponds to observation it being very influential for the conclusions that would be drawn about the size of β_1 . We might fear that a single observation is so influential that it might alter the opinion of economic or statistical significance in applications of the data. If this observation turns out to be wrong, this would undermine the validity of an analysis using this variable.

We ran the regression for all variables replacing x by the specific variable in each iteration. This gives a dif beta measure for each observation and variable pair. Only a handful were so large that they would create the risk of false inference - larger than 2 White standard errors. These were checked and corrected where appropriate. Furthermore, we flagged any dif betas larger than 4 standard deviations of the dif beta measure for each variable. These were also checked and corrected when wrong.

Moreover, we applied two correction procedures based on judgement. We know that MB ratios are typically in a range between 20 and 30. We know that a shareholder would not have several million cows. And we know that a creamery cannot have a fraction of cows associated with it. Using such obvious but domain knowledge facts, we have corrected anything that simply did not seem correct.

Last, we have inspected every single observation in context using a scatter plot with lines connecting each creamery to itself over time. An example of such a plot can be found in figure C.1. We have meticulously checked 2000 such plots - one for each region/variable pair.

In the early phases of the project we mistakenly digitized 1942 twice and applied all the correction procedures independently twice to this. This apparent misfortune allowed us to make an informal test of the accuracy of the data. Given that mistakes are rare, random and independent, it would be unlikely

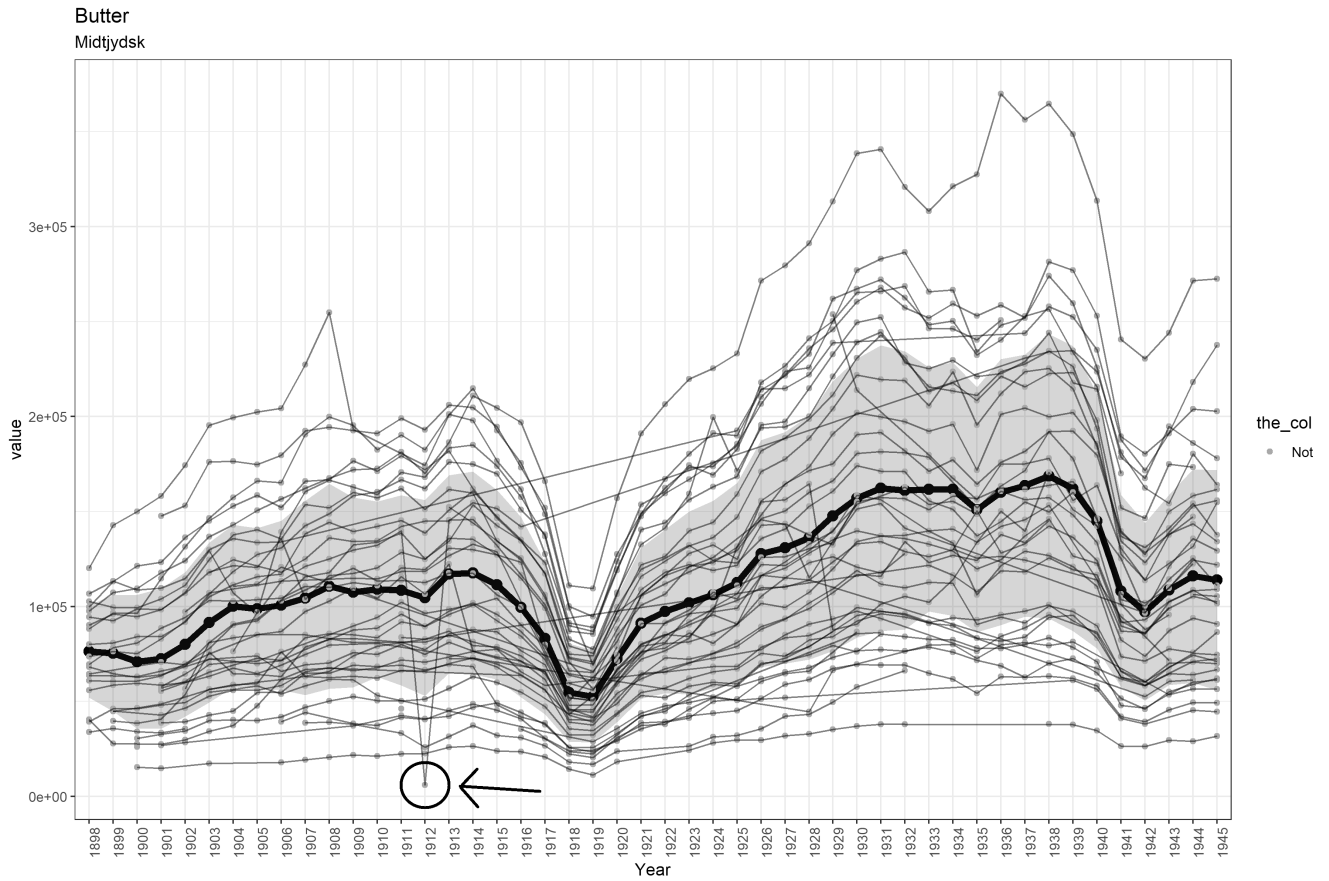


Figure C.1: In this plot Butter production is plotted against time with every observation of the same creamery connected by lines. Deviations from the typical range for a creamery might suggest a mistyped observation. An example of a suspicious observation is encircled. Any such potentially wrong observations were checked and corrected if they were wrong.

that the same mistake is made twice. Differences between the two versions of 1942 are thus an indication of the error rate. One problem, of course, is that errors might not be independent. Specifically, something unclear in the source material might cause the same mistake in both versions of 1942. However, this is probably rare and thus does not invalidate this approach as an informal test of accuracy. Using this approach we find a 1.57% error rate, when taking the risk of *independently* making the same mistake twice into account.¹

D Creamery sizes

In some regions the creameries are much larger than in other regions. For this reason we might suspect that regional differences in scale productivity is the only reason that we see regional variation in productivity. This is documented by the following figure D.1

¹If a is the true (independent) error rate, then the risk of making the same mistake twice is a^2 . Then we would observe a rate of x mistakes, where $x = a - a^2$. We use $a \approx x + x^2$ as an approximation for the true error rate, which is valid as long as a is close to zero.

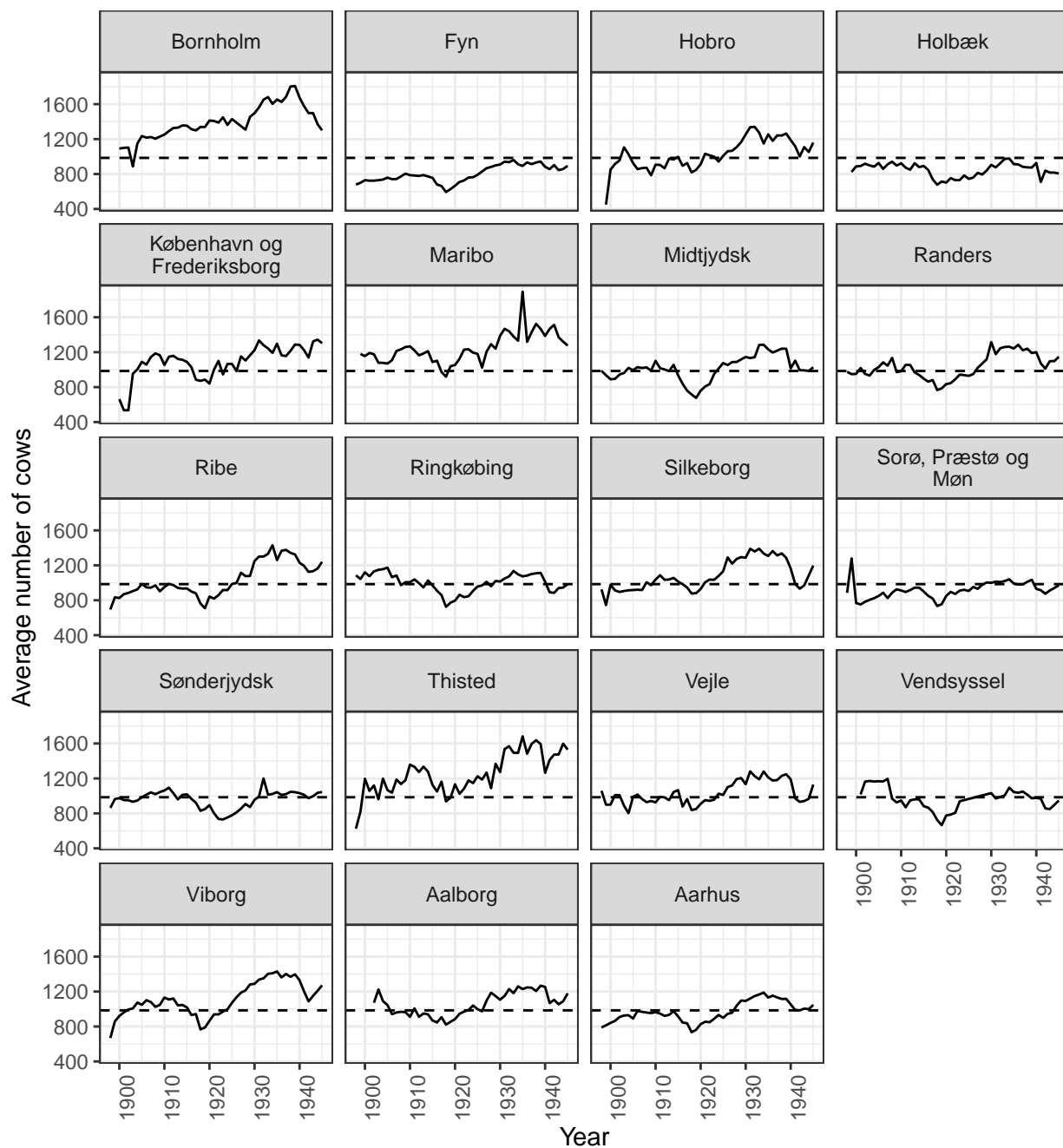


Figure D.1: Average size of creameries in each region measured as average number of cows. The dotted line is the average for all entries in MDS.

E Does the range of years make a qualitative difference?

Table E.1 reports regression specification 2B but only on the years 1929-1945. It should be noted that the qualitative conclusion of regional heterogeneity and a core-periphery pattern holds.

	Model 1
(Intercept)	25.159204*** (0.111741)
Region_consistentBornholm	-0.613893*** (0.065154)
Region_consistentFyn	-0.888437*** (0.051265)
Region_consistentHobro	0.350562*** (0.032868)
Region_consistentHolbæk	-0.692980*** (0.068300)
Region_consistentKøbenhavn og Frederiksborg	-0.263931** (0.056143)
Region_consistentMaribo	-0.527202*** (0.054954)
Region_consistentMidtjydsk	-0.378354*** (0.039505)
Region_consistentRanders	-0.092053 (0.042550)
Region_consistentRibe	0.420588*** (0.067805)
Region_consistentRingkøbing	0.760114*** (0.041364)
Region_consistentSorø, Præstø og Møn	-0.606102*** (0.052010)
Region_consistentSønderjydsk	0.706837*** (0.066369)
Region_consistentThisted	0.540613*** (0.049101)
Region_consistentVejle	-0.342395** (0.081228)

	Model 1
Region_consistentVendsyssel	0.113145 (0.061896)
Region_consistentViborg	0.527558*** (0.053373)
Region_consistentAalborg	0.776153*** (0.045391)
Region_consistentAarhus	-0.447817*** (0.041937)
log(Cows_now)	-0.192557*** (0.016420)
Regional F	57091
Regional F p-value	0
Year fixed effects	Yes
R ²	0.605346
Adj. R ²	0.604083
Num. obs.	10972

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table E.1: Results from specification 2A on 1929-1945

EHES Working Paper Series

Recent EHES Working Papers

2020

- EHES 202 On the Origins of the Demographic Transition. Rethinking the European Marriage Pattern
Faustine Perrin
- EHES 201 Did taller people live longer? Influence of height on life span in rural Spain, 1835-2019,
Francisco J. Marco-Gracia, Javier Puche
- EHES 200 Regional market integration and the emergence of a Scottish national grain market,
Daniel Cassidy, Nick Hanley
- EHES 199 The Pre-kautilyan Period: a sustainable model through ancient economic ideas and practices,
Satish Deodhar, Sriram Balasubramanian
- EHES 198 Accounting for Growth in Spain, 1850-2019,
Leandro Prados de la Escosura, Joan R. Rosés
- EHES 197 Capital in Spain, 1850-2019,
Leandro Prados de la Escosura
- EHES 196 Lordships, state capacity and beyond: literacy rates in mid-nineteenth-century Valencia,
Francisco J. Beltrán Tapia, Alicia Gómez-Tello, Julio Martínez-Galarraga, Daniel A. Tirado-Fabregat
- EHES 195 Builders' Working Time in Eighteenth Century Madrid,
Mario García-Zúñiga
- EHES 194 Is there a Refugee Gap? Evidence from Over a Century of Danish Naturalizations,
Nina Boberg-Fazlic, Paul Sharp
- EHES 193 Epidemics, inequality and poverty in preindustrial and early industrial times,
Guido Alfani

All papers may be downloaded free of charge from: www.ehes.org

The European Historical Economics Society is concerned with advancing education in European economic history through study of European economies and economic history. The society is registered with the Charity Commissioners of England and Wales number: **1052680**