Using Machine Learning for Supply and Demand Predictions in the German Milk Market

Dissertation to obtain the doctoral degree of
Agricultural Sciences (Dr. sc. agr.)

Faculty of Agricultural Sciences
University of Hohenheim
Institute of Agricultural Policy and Markets

submitted by
Dominik Baaken
from Neuss
2023

Date of Oral Examination: December 13, 2023

Dean: Prof. Dr. Ralf Vögele
Examination Chair: Prof. Dr. Thilo Streck
Supervisor and First Reviewer: Prof. Dr. Sebastian Hess
Second Reviewer: Prof. Dr. Jens-Peter Loy
Additional Examiner: Prof. Dr. Christine Wieck

Acknowledgments

This thesis would not have been possible without the support of numerous people.

First and foremost, my deepest appreciation goes to my supervisor, Prof. Dr. Sebastian Hess. I would like to thank you for trusting my research abilities and allowing me the academic freedom to develop and implement my own ideas. Your valuable mentorship and constructive support, whether in Kiel, in a fully remote setting during the pandemic, or in Hohenheim, have been essential to my professional and personal development.

I would also like to acknowledge the dedicated efforts of the examining committee in reading and reviewing this thesis.

Furthermore, I would like to thank all my former and current colleagues from the Department of Dairy and Food Industry Economics at the University of Kiel and the Department of Agricultural Markets at the University of Hohenheim, not only for the academic exchange, but also for the much-needed amount of socialising during and after work.

A special thanks also goes to my family, especially my parents, for their continuous and invaluable support throughout this journey. Likewise, I thank all my friends for their understanding and patience at times when my commitments have required occasional absences from celebrations. Last but not least, I would like to thank my girlfriend, Annika, for supporting and believing in me during this doctoral project. This has been a great boost to my energy and motivation.

Summary

The German milk market is driven by various ongoing trends on both the domestic supply and the international demand side. This results in increasingly volatile prices, as well as increasing production costs, and both risks continue to induce dairy farms going out of business. Particularly since the abolition of the milk quota in March 2015, market participants have therefore expressed a desire for more reliable forecasting tools at the regional level in order to be able to make strategic and operational decisions with greater planning certainty. However, such forecasting models at the farm or regional level do not currently exist or are not publicly available. This dissertation fills this research gap by developing a forecasting model for predicting regional milk production in Lower Saxony.

The first of four research chapters, Chapter 3, compares five different Machine Learning (ML) models and a traditional linear regression (OLS) model based on time trends, direct and indirect weather influences, and price events. There are slight differences in forecast accuracy between the ML methods for a 12-month forecast period. The advantage of the ML methods over OLS regression is particularly evident in the classification of farms into average and fast-growing farms. Nonlinear developments induced by investments can be predicted more accurately using ML methods instead of econometric modelling. Furthermore, differences in the efficiency of the methods are apparent: while comparable estimation approaches achieve similar accuracies, the training speed of the models varies considerably.

ML methods promise high forecasting accuracy, but causal inferences are difficult to derive. In particular, the comparison of several estimation approaches makes the presentation of causal relations difficult as the models offer different or no interpretation possibilities. For this reason, Chapter 4 presents the relationship between seasonal weather conditions and seasonal milk production. This chapter incorporates the influences of direct and indirect weather conditions as well as time and price trends into the model. A Fixed Effects (FE) estimator is used to model quarterly milk production for a panel dataset from Lower Saxony. The results mainly illustrate the influence of farm decisions on milk production, which is stronger than the influence of weather conditions. Contrary to expectations, the influence of weather conditions during the growing season cannot be significantly demonstrated. Instead, there is a positive effect of warmer and drier weather in almost all quarters except autumn.

Chapters 3 and 4 address the supply side of the German milk market, while chapters 5 and 6 deal with the demand for raw milk from vending machines.

The current challenges in the German milk market require a high degree of flexibility from farmers. More and more farmers are therefore seeking alternative sales channels to generate additional income beyond supplying milk to dairy companies. For example, direct marketing of raw milk through on-farm vending machines has been an option for a number of years; this allows farmers to generate potential additional income with relatively little effort and risk. Over the years, however, it has become evident that many farmers close their raw milk vending machines after a short time. This raises the question of for which farms and under which conditions on-farm vending machines would be an economically viable investment.

In Chapter 5, a forecasting model is developed based on a nationwide survey of farmers and several location factors that predicts the raw milk sales of the farms surveyed. A methodological approach to ML, the Xtreme Gradient Boosting (XGB) algorithm, achieves sufficiently accurate values to qualify the model as a practical tool. Farm managers who are still undecided could enter their own values into the model and thus secure their investment decision. The influence of the variables on the prediction is investigated using SHapley Additive exPlanation (SHAP) values. The results reveal that sales of raw milk from vending machines are influenced less by individual marketing measures than by various location factors such as population density, proximity to a city, and location along a road with commuter traffic. It can be concluded that there is additional sales potential if farmers would be allowed to place the vending machine in an optimal location away from the farm, thus reaching the target group more effectively and reducing inefficient marketing activities.

Chapter 6 is based on a consumer survey in Germany. Seemingly unrelated regression (SUR) is used to model willingness to pay (WTP) and frequency of purchase. The results suggest that in this form of marketing, consumers especially value a 'fair' price for the producer and are less price-sensitive. On average, customers' WTP is higher than the current milk price and varies between consumer groups. Consumers with a closer connection to milk production are willing to pay more for raw milk but purchase it less frequently. It also appears that as consumers get older, they are more likely to buy raw milk but are less willing to pay for it. These and other consumer characteristics can help farmers tailor their marketing activities to the needs of their customers and increase the efficiency of their additional sales channels.

Overall, this dissertation demonstrates the potential applications and limitations of ML methods for considering supply and demand in the German milk market. The forecasting models can serve as a potential tool for farmers to better weight strategic and operational decisions, thus contributing to more efficient agriculture.

Zusammenfassung

Gestiegene Anforderungen auf der Produktions- und Nachfrageseite des deutschen Milchmarktes erschweren es landwirtschaftlichen Betrieben zunehmend, sich in einem herausfordernden Umfeld zu behaupten. Volatile Preise und langanhaltende Niedrigpreisphasen sowie gestiegene Produktionskosten führen zu einer fortschreitenden Konsolidierung der Zahl der Milchviehbetriebe. Insbesondere seit dem Wegfall der Milchquote im März 2015 wird von einigen Marktteilnehmern der Wunsch nach Prognoseinstrumenten auf regionaler Ebene geäußert, um strategische und operative Entscheidungen mit größerer Planungssicherheit treffen zu können. Prognosemodelle auf betrieblicher oder regionaler Ebene sind jedoch derzeit nicht verfügbar bzw. nicht öffentlich zugänglich. Diese Forschungslücke wird in der vorliegenden Dissertation durch die Entwicklung eines Prognosemodells zur Vorhersage der regionalen Milchproduktion in Niedersachsen geschlossen.

In dem ersten von vier Forschungskapiteln, Kapitel 3, werden fünf verschiedene Modelle des Maschinellen Lernens (ML) und ein herkömmliches lineares Regressionsmodell (OLS) auf Basis von zeitlichen Entwicklungen, direkten und indirekten Wettereinflüssen sowie Preisereignissen miteinander verglichen. Zwischen den jeweiligen ML-Methoden zeigen sich für einen Prognosezeitraum von 12 Monaten leichte Unterschiede in der Prognosegenauigkeit. Der Vorteil der ML-Methoden gegenüber der OLS-Regression zeigt sich insbesondere bei der Unterteilung der Betriebe in durchschnittlich wachsende und stark wachsende Betriebe. Nicht-lineare Entwicklungen, die durch Investitionen induziert werden, können mit ML-Methoden wesentlich genauer vorhergesagt werden als auf Basis des OLS Schätzers. Darüber hinaus wird die unterschiedliche Effizienz der Methoden deutlich: während vergleichbare Schätzansätze ähnliche Genauigkeiten erreichen, variieren die Trainingsgeschwindigkeiten der Modelle erheblich.

ML-Methoden versprechen eine hohe Prognosegenauigkeit Aussagen über kausale Zusammenhänge sind jedoch nur schwer ableitbar. Insbesondere der Vergleich mehrerer Schätzansätze erschwert die Darstellung, da die Modelle jeweils unterschiedliche bzw. keine Interpretationsmöglichkeiten bieten. Aus diesem Grund wird in Kapitel 4 explizit der Zusammenhang zwischen saisonalen Wetterbedingungen und saisonaler Milchproduktion dargestellt. Auch in diesem Kapitel werden die Einflüsse der direkten und indirekten Wetterbedingungen sowie der zeitlichen und preislichen Trends in das Modell integriert. Mit Hilfe eines Fixed-Effects (FE)-Schätzers wird die Milchproduktion pro Quartal für einen Panel-Datensatz aus Niedersachsen modelliert. Die Ergebnisse verdeutlichen vor allem den Einfluss be-

trieblicher Entscheidungen auf die Milchproduktion, der im Vergleich zum Einfluss von Wetterbedingungen stärker ausfällt. Entgegen den Erwartungen kann ein Einfluss der Witterung in der Vegetationsperiode nicht signifikant nachgewiesen werden. Stattdessen zeigt sich ein positiver Effekt wärmerer und trockenerer Witterung in fast allen Quartalen mit Ausnahme des Herbstes.

Kapitel 3 und 4 befassen sich mit der Angebotsseite des deutschen Milchmarktes, Kapitel 5 und 6 hingegen mit der Nachfrage an Rohmilchautomaten.

Die aktuellen Herausforderungen des deutschen Milchmarktes verlangen von den Landwirten ein hohes Maß an Flexibilität, um in diesem herausfordernden Umfeld bestehen zu können. Immer mehr Landwirte suchen nach alternativen Absatzmöglichkeiten, um neben der Milchlieferung an die Molkerei ein zusätzliches Einkommen zu generieren. Seit einigen Jahren bietet sich hierfür die Direktvermarktung von Rohmilch über Automaten auf dem Hof an. Auf diese Weise können Landwirte mit relativ geringem Aufwand und überschaubarem wirtschaftlichen Risiko ein potenzielles Zusatzeinkommen erwirtschaften. Im Laufe der Jahre hat sich jedoch gezeigt, dass viele Betriebsleiter den Rohmilchautomaten nach kurzer Zeit wieder schließen. Es stellt sich daher die Forschungsfrage, für welche Milchviehbetriebe und unter welchen Bedingungen eine solche Investition wirtschaftlich sinnvoll erscheint.

Dazu wird in Kapitel 5 mit Hilfe einer deutschlandweiten Befragung von Landwirten und einer Vielzahl von Standortfaktoren ein Prognosemodell entwickelt, das den Rohmilchabsatz der betrachteten Betriebe vorhersagt. Ein Methodenansatz des ML, der Xtreme Gradient Boosting (XGB)-Algorithmus, erreicht ausreichende Genauigkeitswerte, sodass das Modell für den praktischen Einsatz geeignet ist. Noch unentschlossene Betriebsleiter könnten ihre eigenen Werte in das Modell einpflegen und so ihre Investitionsentscheidung absichern. Der Einfluss der Variablen auf das Prognoseergebnis wird mit Hilfe von SHapley Additive exPlanation (SHAP)-Werten dargestellt. Die Ergebnisse zeigen, dass der Rohmilchabsatz an Automaten weniger von einzelnen Marketingmaßnahmen als maßgeblich von verschiedenen Standortfaktoren wie z.B. Bevölkerungsdichte, Stadtnähe und Lage an einer Straße mit Pendlerverkehr beeinflusst wird. Daraus lässt sich ableiten, dass zusätzliches Absatzpotenzial besteht, wenn es Landwirten rechtlich ermöglicht würde, ihren Automaten abseits vom Betrieb an einem optimalen Standort aufzustellen und somit die Zielgruppe besser zu erreichen. Bis dahin eher ineffiziente Marketingmaßnahmen in Hofnähe könnten stattdessen reduziert werden.

Kapitel 6 basiert auf einer deutschlandweiten Verbraucherbefragung. Mit Hilfe einer seemingly unrelated regression (SUR) wird die Zahlungsbereitschaft und Kauf-

häufigkeit in einem gemeinsamen Modell dargestellt. Es zeigt sich, dass die Konsumenten bei dieser Vermarktungsform insbesondere den „fairen“ Preis für den Erzeuger zahlen möchten und weniger preissensibel sind. Die Zahlungsbereitschaft der Kunden liegt im Durchschnitt über dem aktuellen Milchpreis und variiert je nach Konsumentengruppe. Verbraucher, die einen engeren Bezug zur Milcherzeugung haben, sind zwar bereit, mehr für Rohmilch zu bezahlen, kaufen aber seltener dort ein. Es zeigt sich auch, dass mit zunehmendem Alter die Verbraucher zwar häufiger Rohmilch kaufen, aber weniger bereit sind, dafür zu bezahlen. Diese und weitere Charakterisierungen der Konsumenten können den Landwirten helfen, ihre Marketingaktivitäten auf die Bedürfnisse der Kunden auszurichten und die Effizienz ihres zusätzlichen Absatzkanals zu erhöhen.

Insgesamt können im Rahmen dieser Dissertation die Einsatzmöglichkeiten und Limitationen von ML-Methoden zur Betrachtung von Angebot und Nachfrage auf dem deutschen Milchmarkt aufgezeigt werden. Die Prognosemodelle können Landwirten als potentielles Hilfsmittel dienen, um strategische und operative Entscheidungen besser abwägen zu können und somit zu einer effizienteren Landwirtschaft beitragen.

Contents

List of Figures

List of Tables

Abbreviations

BLUE Best Linear Unbiased Estimator

BLE Bundesanstalt für Ernährung und Landwirtschaft

BMEL Bundesministerium für Ernährung und Landwirtschaft

CAP Common Agricultural Policy

CART Classification and Regression Tree

Destatis Statistisches Bundesamt

DP dependence plot

DWD Deutscher Wetterdienst

ECM energy-corrected milk yield

EN Elastic Net

FE Fixed Effects

FGLS Feasible Generalised Least Squares

GBM Gradient Boosting Machine

GHG greenhouse gas

GLS Generalised Least Squares

IPCC Intergovernmental Panel on Climate Change

LKV Landeskontrollverband Niedersachsen e.V.

LSDV Least Squares Dummy Variables

LSN Landesamt für Statistik Niedersachsen

Max. maximum value

Min. minimum value

ML Machine Learning

NN Neural Network

OLS Ordinary Least Squares

PR purchase reason

R^2 coefficient of determination

ReLu Rectified Linear Units

RMSE root mean squared error

RSS residual sum of squares

SD standard deviation

SHAP SHapley Additive exPlanation

SOR Stimulus-Organism-Response

SPEI Standardised Precipitation-Evapotranspiration Index

SUR seemingly unrelated regression

THI Temperature-Humidity-Index

WTP willingness to pay

XGB Xtreme Gradient Boosting

Chapter 1

Introduction

1.1 Motivation

Supply and demand are the drivers of a well-functioning market (Mankiw et al., 2008). Analysis of these drivers has traditionally been an important scientific issue. Adam Smith formulated the basic idea of markets in his book *The Wealth of Nations*, stating that individual self-interest and competition form an 'invisible hand' that leads to an efficient allocation of resources (A. Smith, 1776). This and other economic theories assume the existence of a perfect competitive market, which entails certain assumptions. These include the existence of numerous small firms and customers, the rational behaviour of all market participants as *homo economicus*, the perfect information of all market participants, the homogeneity of the traded goods, and the possibility for all market participants to react quickly to changes (Baumol & Blinder, 2010).

In practice, however, there is usually an imperfect market characterised by various deviations from the ideal assumptions. These can vary in severity from industry to industry. Agricultural markets, such as the German milk market, have one particularly notable characteristic: the production of goods depends on a variety of biological processes (Burgess et al., 2023). While industrial production can be fine-tuned down to the smallest detail at short notice, operational decisions on farms are much slower, so it is only possible to react to market events with a considerable time lag (Hoehl & Hess, 2022). In addition, the required input factors are not always available at the same quality and quantity due to vegetation processes (Köster, 2010). Moreover, although complete information is usually available to market participants at any given time, the complexity of the interrelationships in agricultural production makes it difficult to predict future developments (Hansen, 2002).

Imperfect markets lead to market distortions, which manifest themselves in different ways. The highly volatile milk prices of recent years illustrate the imperfection of the German milk market. The market has been in turmoil for a number of years, and this is particularly evident in milk prices. Within one year, the Kiel commodity value of milk ex farm (*Kieler Rohstoffwert Milch*), which is considered a leading indicator for price developments on the German milk market, fell from 66.6 cents/kg (May 2022) to 36.4 cents/kg milk (May 2023), while the average commodity value in 2021 was 39.1 cents/kg (Institut für Ernährungswirtschaft, 2023). In recent years, the milk price has also regularly been subject to considerable fluctuations. Periods of low prices have occurred repeatedly. In some cases, the milk price has even fallen below the 30 cents/kg threshold (e.g., 2009–2010, 2015–2016; BMEL (2023a)).

Previous policy interventions to regulate the market, such as the introduction of a milk quota system in the European Union, have not been able to prevent these

fluctuations and have been considered inefficient (Frick & Sauer, 2018). However, there is considerable policy interest in reducing market volatility to achieve welfare gains, use resources more efficiently, and thus promote more sustainable milk production. Short- and medium-term forecasting tools that provide farm managers with information on their own and expected regional milk volumes are one possible tool to achieve these goals (Petropoulos et al., 2022). Such tools will enable them to make strategic and operational decisions based on solid forecasts, resulting in economically and environmentally sustainable production decisions not only for the individual farm but also for the whole milk market. In addition, forecasting models could be used to support investment decisions when farmers are considering establishing additional financial pillars. By predicting potential demand, uncertainty about the economic success of such an investment could be reduced, leading to economically and environmentally efficient investment decisions.

The challenge to developing forecasting tools consists of modelling the market while considering all relevant factors. For many years, supply and demand have been incorporated into the theory of an existing economic construct in order to draw conclusions about causal relationships within the market (Altemeier, 1981; Bouamra-Mechemache et al., 2008; Kraatz, 2012, e.g.). Typically, econometric estimates are used to express these relationships in terms of elasticities to identify market drivers (Gujarati & Porter, 2009). However, forecasting tools are less interested in the causal relationships and effects of different parameters and instead aim to predict future developments based on historical data (Athey & Imbens, 2019).

In recent years, Machine Learning (ML) has evolved along with high-performance computing to provide new opportunities for detecting, quantifying, and understanding data-driven farm processes (Liakos et al., 2018). ML methods offer a way of ensuring practice-relevant forecasting based on large datasets. Supervised ML is capable of making accurate predictions and may be preferable to traditional econometric approaches, even with small datasets (Breiman, 2001b). Rather than building and applying a model based on existing theory, ML reads patterns from existing data structures that can be used to infer future developments (Mullainathan & Spiess, 2017). The focus of ML methods is therefore on predictive accuracy, whereas econometric approaches tend to focus on the statistical properties of parameters (Storm et al., 2020).

Unlike traditional economic and social science research, the ML models do not provide direct recommendations for policy action. Instead, the development of the ML models creates practical applications. The integration of the models as a forecasting tool for farmers or dairies, for example, offers market participants direct

and practical added value. Farmers will be able to input their own data into the respective tools to generate a farm-specific forecast and draw conclusions for their strategic and operational farm decisions. While policy recommendations need to be translated into action through a lengthy process before they have an impact on the farmer or the market, macroeconomic impact can be achieved in a direct way if the tools developed are sufficiently applied.

In addition to the development of forecasting tools, the factors influencing the market are also of interest. A comprehensive understanding of these factors enables stakeholders to make informed decisions and manage their operations efficiently. Farmers can benefit from this understanding by identifying strategies to mitigate the adverse effects of weather conditions, planning for seasonal variations, and improving overall milk production. Some types of ML models, such as deep learning models, are known as 'black boxes' that are highly recursive and therefore difficult to interpret (Rudin, 2019). However, there are also explainable ML techniques that provide tools for understanding the influence of the parameters on prediction results (Lundberg et al., 2020). Some methods, such as decision trees, offer interpretation features, such as measures of variable importance, to interpret prediction models (Hastie et al., 2009). Another method for interpreting ML models is the calculation of SHapley Additive exPlanation (SHAP) values, which originates from cooperative game theory. SHAP values provide the importance of individual features by quantifying the contribution of each feature to the predictive output of the model (Lundberg & Lee, 2017). While other feature importance methods focus on the main effects of features, SHAP values additionally consider interactions between variables. This feature allows for the identification of complex relationships, providing a more comprehensive understanding of how features collectively contribute to model output. Furthermore, by aggregating individual SHAP values across multiple instances, it is possible to gain insight into the overall behaviour and performance of the model. However, it is important to note that SHAP values do not provide causal statements and cannot be used to evaluate statistical parameters (Heskes et al., 2020). Instead, they provide an overview of the importance of the variables in relation to the prediction outcome. Explainable ML can therefore contribute to understanding the drivers within a market, but it cannot replace the use of econometric models. If causal relationships are to be represented in a specific way by elasticities, then the use of econometric models is indispensable.

1.2 Challenges of the German Milk Market

German milk production was the fifth largest in the world in terms of volume in 2021 (FAOSTAT, 2023). With a production volume of 32.5 million metric tonnes of cow's milk, Germany led the ranking of the largest milk producers in the EU-27 in 2021, ahead of France (24.8 million tonnes) and Poland (14.9 million tonnes; Eurostat (2023)). In terms of revenue, the dairy sector is the strongest sector of the agricultural and food industry in Germany (BMEL, 2023c). However, structural change has also been affecting the dairy sector for several years. In particular, the abolition of the European milk quota system on 31 March 2015 has meant that dairy farms can expand their production arbitrarily. This has accelerated the already-prevailing structural change in the dairy sector. While the total milk volume increased by approximately 14.8% between 2000 and 2021, the number of dairy farms more than halved over this period, dropping from around 138,500 to around 54,800 (BLE, 2022). Accordingly, the average size of farms has grown significantly during this period. Whereas in 2000 a dairy farm in Germany held an average of about 33 cows, in 2021 there were 70 cows per farm (BLE, 2022). Large regional differences can also be observed: in the southern federal states, farms tend to have a smaller structure, with 44 cows per farm (Bavaria) or 58 cows per farm (Baden-Württemberg); in the northern federal states, farms hold an average of just over 100 animals per farm; and in the eastern, 'new' federal states, farms are the largest, with up to 241 animals per farm in Mecklenburg-Vorpommern (BLE, 2022). These regional differences illustrate the heterogeneity of the dairy sector in the federal territory of Germany and complicate the nationwide introduction of environmental policy measures.

There are several reasons for the progress of structural change. The volatility of milk prices is a contributing factor. Figure 1.1 shows the development of the average milk price paid ex farm in Germany in the period from 2007–2022. It should be noted that price fluctuations during this period range from 22.2 cents/kg (07/2009) to 61.6 cents/kg (12/2022; BMEL (2023a)). Low-price phases emerged repeatedly, confronting farmers with enormous financial challenges. It therefore seems even more astonishing that milk production did not decline during these periods but rather remained stable to a certain extent, so that these phases endured until the milk price stabilised (see Figure 1.1). Typical fluctuations in milk volume were evident throughout the year, with the annual peak typically in spring and the lowest yield often in autumn. A steady increase of 19.1% in annual milk production per farm from 2007 to 2020 can be identified, while there was a slight decrease of 1.9% from 2020 to 2022 (Eurostat, 2023).

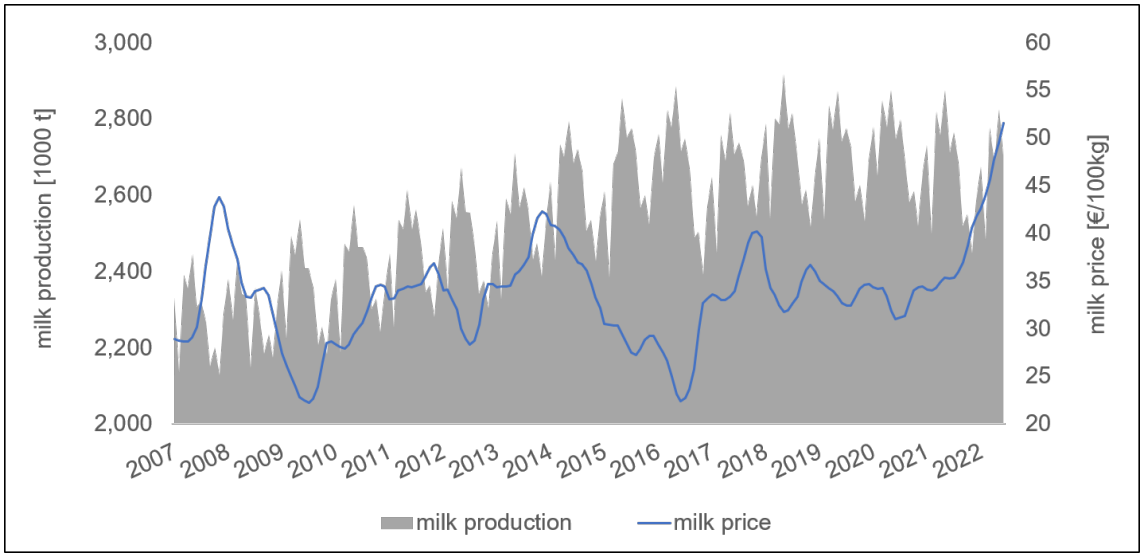

Figure 1.1: Development of milk price and production in Germany, 2007–2022

Source: Own illustration based on BMEL (2023a); Eurostat (2023)

The development of milk production is conditioned by its dependence on weather conditions (Gisbert-Queral et al., 2021). Of all the economic sectors, the agricultural sector is most sensitive and vulnerable to weather events and climate change (IPCC, 2014). The production of livestock is particularly directly affected by climate change as livestock performance and welfare respond to higher temperatures and the increasing frequency of extreme weather events (Hill & Wall, 2015). High temperatures in the form of heat waves are identified as the main risk to dairy production (Key & Sneeringer, 2014) and also decrease dairy farmers' revenues (Finger et al., 2018). Heat stress is related not only to temperatures but also to relative humidity, and lactating cows are more likely to experience it on hot and humid summer days (Qi et al., 2015). These weather conditions can reduce performance, decrease animal comfort, increase mortality rates, and force losses (Mader et al., 2006; Megersa et al., 2014). The negative impact on milk production is directly due to the weather, but there are also indirect effects caused by changes in the availability of fodder and pasture as well as the distribution of pests and parasites (Gauly et al., 2013). Due to changing weather conditions, dairy farmers must manage reduced water availability and thus changes in forage quantity and quality (Megersa et al., 2014).

While output prices are highly volatile, input prices have increased significantly, especially since the outbreak of the Russian invasion of Ukraine in early 2022. This is compounded by other macroeconomic developments such as a changed interest rate environment and frequent supply shortages in the Asian region due to pandemic restrictions (Brunetti et al., 2022). As a result, the purchase prices of agricultural

inputs increased by 38.3% between 2012 and 2022 (Destatis, 2023b). In particular, operating resources relevant to milk production, such as electricity (+45.5%), fertilisers (+97.6%), and animal feed (+45.7%), have become significantly more expensive in recent years (Destatis, 2023b). The land market environment also appears to be tense. The average rental price for agricultural land increased by about 42.3% between 2010 and 2020, with rental prices for arable land rising by 59.6% and those for grassland rising by 43.1% (Destatis, 2023e). In addition, the average purchase values for agricultural land have risen considerably over the same period, almost doubling (Destatis, 2023a). The reasons for this development are manifold and can vary regionally. The main influencing factors include livestock density, the increase in biogas plants, and the price level of agricultural products (Langenberg et al., 2017).

In addition to financial challenges, dairy farmers are facing increasing requirements for sustainable production. Evidence is growing that farms seeking to maintain competitive milk production must implement and document sustainability requirements on an ongoing and comprehensive basis. However, the collection of indicators for environmental and animal welfare as well as social and economic aspects remains challenging due to the limited availability of appropriate and cost-effective measurement approaches (Calker et al., 2005). Nevertheless, consumers are increasingly demanding such evidence in the form of labels (Mehlhose et al., 2019). For instance, consumers are willing to pay a premium for milk produced by cows that are mainly pasture-based and not continuously confined indoors (Weinrich et al., 2014). While the steadily declining number of farms has reduced the amount of contact between consumers and producers, thus leading some consumers to feel disconnected from agricultural production (La Trobe & Acott, 2000), there is a desire among many consumers for short value chains (Jarzębowski et al., 2020). Several German consumers prefer locally produced food, transparent and traceable food supply chains, and animal welfare certifications (BMEL, 2018; Feldmann & Hamm, 2015), and these desires are partly accompanied by a higher willingness to pay (WTP) for these products (Emberger-Klein et al., 2016). The demand for organically produced milk has also increased in recent years. The share of organic milk increased from 0.96% (2001) to 3.96% (2021) of the total milk supply (BLE, 2022). Simultaneously, the price difference between conventionally and organically produced milk has steadily narrowed. While at the beginning of 2021 there was still a difference of 15.9 cents/kg, at the end of 2022 the difference is only 3.1 cents/kg (BMEL, 2023a, 2023b).

Cow milk demand is declining in Germany. Per capita milk consumption fell by

4.4% compared to previous year to 47.8 kg/person in 2021, the lowest level since 1991 (BLE, 2023b). One reason for this could also be the increasing preference for milk substitutes. The sales volume of milk substitutes increased by approximately 25% annually since 2018, reaching 251 million litres in 2020 (Nielsen, 2020). The turnover of milk substitutes has thus also grown significantly in recent years, but at around €470 million in 2020 it remains far below the revenue of cow milk, which reached €16 billion (Statista Consumer Market Insights, 2023; BMEL, 2023c). However, according to Statista Consumer Market Insights (2023), sales of milk substitutes are expected to grow steadily in the coming years, reaching €1.2 billion by 2027.

The environmental policy of combating climate change is another major challenge for the German milk market. The agricultural sector was responsible for about 11% of greenhouse gas (GHG) emissions in the European Union in 2019 (European Environmental Agency, 2021). In Germany, the share was slightly lower at 7.4% in 2021, but livestock farming in particular accounts for a large share of GHG emissions. For instance, about 56.4% of GHG emissions from the agricultural sector are methane emissions resulting from digestion processes, the storage and spreading of livestock fertiliser, and the storage processes of fermentation residues from renewable raw materials (Umweltbundesamt, 2023). Thus, dairy farming was responsible for about 51.5% of methane and nitrous oxide emissions from agriculture in 2021, accounting for about three quarters of total greenhouse gas emissions from the agricultural sector (Fuß et al., 2023).

The European Green Deal is intended to lead to a climate-neutral Europe by 2050 and to protect the environment. Agriculture is to become more sustainable and food systems more resilient (European Commission, 2020b). The Green Deal is implemented in the agricultural sector through the Farm-to-Fork strategy and calls for a 20% reduction in the use of mineral fertilisers, a 50% reduction in the use of pesticides, and a 50% reduction in the nitrogen balance surplus. The share of landscape elements with high diversity is to be increased by at least 10% and the share of organic farming by at least 25% in less than a decade (by 2030; European Commission (2020a)).

Realising the objectives of the Green Deal is strongly linked to the Common Agricultural Policy (CAP) 2023–27. Implementation of the new CAP will be at the national level, with national strategic plans (European Commission, 2020a). The revised CAP is based on a 'green architecture' that requires all beneficiaries of the CAP to comply with mandatory rules ('conditionality') and statutory management requirements that apply to all farmers regardless of whether they receive CAP support. In addition, farmers receiving CAP support are expected to respect good agri-

cultural and environmental conditions. Voluntary 'eco schemes' are also introduced to incentivise climate- and environment-friendly practises, which are mandatory for national strategic plans but optional for farmers (European Commission, 2022). The challenges for farmers and the need for adaptation are accordingly significant. In a study commissioned by the Grain Club, Henning et al. (2021) simulated impacts of these adjusted regulations on the German agricultural sector using the CAPRI model and demonstrated a significant decline in production as a result. High adjustment costs could arise, leading to high losses, especially in the dairy industry, thus reinforcing structural change (Henning et al., 2021).

Meanwhile, the restructuring of livestock farming has been discussed in Germany for some time, primarily to ensure better animal welfare in German livestock farming. The Federal Ministry of Food and Agriculture, Bundesministerium für Ernährung und Landwirtschaft (BMEL), convened a network of experts, called the 'Borchert Kommission', in 2019 to address the restructuring of livestock farming in Germany. The Commission has submitted proposals to the Federal Ministry, including plans for financing, labelling, and investments (Borchert Kommission, 2020). By June 2023, no decisions had been taken on these proposals, but farmers will also have to face these challenges soon.

1.3 Objectives and Research Questions

Stakeholders in the German milk market face a number of challenges, as outlined above. To survive in such a challenging environment, efficient use of resources and innovative farming are required. The aim of this thesis is to develop forecasting tools that support resource efficiency in the German milk market. The thesis deals with two different aspects of the German milk market: supply and demand.

Firstly, milk production in the region of Lower Saxony is examined. While in previous studies milk production was mostly predicted on the basis of the lactation curve of individual cows (Adediran et al., 2012), this chapter focuses on the farm and regional level. The question is whether past developments can be used to predict future milk production. ML will be used to develop prediction models that can be used in practice. In this context, the following research questions are addressed:

- How can regional milk production be predicted as accurately as possible using farm, price, and weather data?
- What is the best method for predicting regional milk production?

In addition to an accurate forecast of milk supply, the factors influencing milk production are also of interest. To better understand the impact of different weather,

price, and trend developments during the year, direct and indirect influences on seasonal milk production are analysed. Several studies have investigated annual milk production and the influence of weather conditions in this regard (Perez-Mendez et al., 2019; Key & Sneeringer, 2014). However, the particular focus of this work is to examine intra-seasonal variations. In addition, although the direct influence of weather on dairy performance has been extensively studied (Gisbert-Queral et al., 2021; Qi et al., 2015), there is still a lack of information on the causal relationship between weather and dairy production as an indirect consequence of variations in feed quantity and quality. Therefore, the following research questions are considered:

- What seasonal weather conditions affect seasonal milk production?
- What influence do other factors have on seasonal milk production?

The second aspect to consider is the sale of raw milk from vending machines. As outlined above, dairy farms need to adapt if they are to continue to survive in this competitive market. Adaptation can take different forms. For example, some farms have chosen to reduce unit costs by increasing the size of the farm structure, taking advantage of economies of scale, and thereby gaining a financial advantage (Hallam, 1991). Others diversify their farms and are able to offset losses in the dairy sector during periods of low prices by adding other sectors (Schöpe, 2011). The establishment of a new financial pillar in the form of direct marketing is another possibility for diversification (Rutz et al., 2016).

For several years, the sale of raw milk through raw milk vending machines has been an attractive option for the dairy farms. The number of open raw milk filling stations is constantly changing as many farms set up a raw milk vending machine, but some close it down after a short time. Farmers complain about the time needed to operate the vending machine and the comparatively low income they receive. This raises the question of which farms and locations would be profitable for such an investment. The development of a forecasting model to support farm managers in their investment decisions is therefore the objective of this research approach. An accurate forecasting model can contribute to an economical use of resources by enabling farms to install vending machines in promising locations and to refrain from investing in less promising locations. With regard to the influencing factors, it is also of interest to know which factors have a particularly positive or negative effect on milk sales in order to be able to make recommendations for action. This leads to the following research questions:

- How can sales of raw milk from vending machines be predicted as accurately as possible?

- What factors influence sales of raw milk from vending machines?

Understanding the target group also contributes to the efficient use of resources. The needs, desires, and specific characteristics of the respective consumer groups can be better assessed, and the offer design and marketing measures can be adapted accordingly. This aspect is addressed in Chapter 6, which focuses on the factors influencing frequency of consumption and WTP at raw milk vending machines. This chapter aims to answer the following research questions:

- What is the customer's WTP and what are the factors that influence their WTP?
- What are the motivations for regularly buying raw milk from vending machines?

1.4 Outline

Following the introduction presented above, this dissertation is structured as follows.

Chapter 2 presents the econometric and ML methods used to forecast and analyse the factors influencing supply and demand in the German milk market.

Chapters 3–6 address the previously identified research questions. Figure 1.2 shows the schematic structure of the corresponding research chapters. The research issues range from supply to demand and from forecasting to influencing factors.

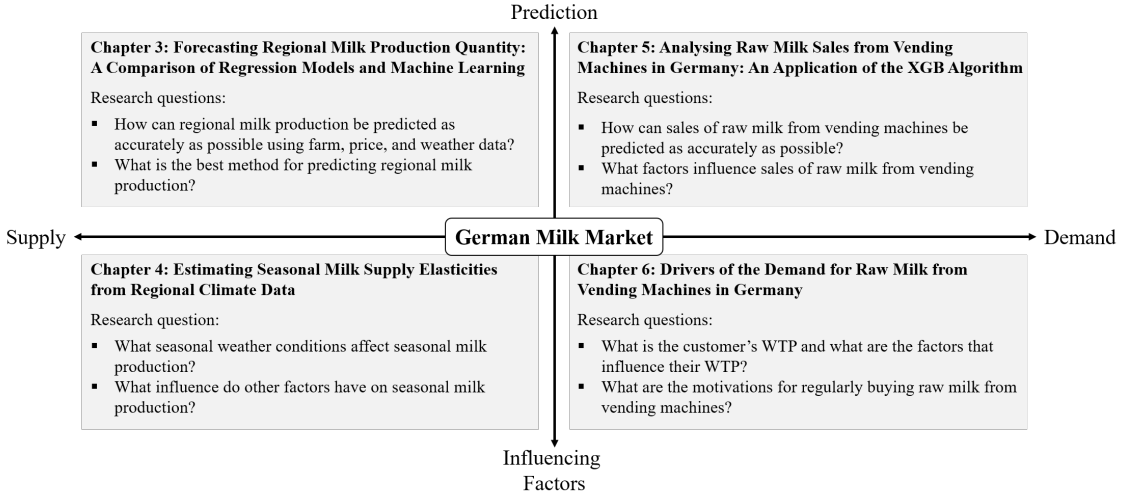

Figure 1.2: Schematic structure for answering the research questions

In Chapter 3, different ML models and a linear regression model are developed to predict milk production in the region of Lower Saxony. The results are compared in terms of prediction accuracy and practicality.

Chapter 4 examines the relationships presented for milk production in Lower Saxony, in particular for correlations between seasonal weather conditions and seasonal milk production. In addition, other factors influencing milk production during the year are analysed.

Chapter 5 predicts the sales of raw milk from vending machines and then considers the influence of the parameters used.

Chapter 6 presents the factors influencing customers' frequency of purchase and willingness to pay at raw milk vending machines in order to characterise the different consumers.

Chapter 7 provides a comprehensive overview of the main findings of this dissertation and offers a general discussion of the applied data and methodologies. Additionally, this chapter provides conclusions, policy implications, and recommendations for future research.

Chapter 2

Methods

The aim of this thesis is to model supply and demand in the German milk market. Thus, causal relationships within the market are illustrated, and predictions for future developments are made based on historical patterns. As in most empirical economic studies, the influence of a change in one variable on the development of another variable is of interest (Wooldridge, 2010). In most cases, economic models describe deterministic relationships between dependent and independent variables by specifying a functional form and representing a simplified version of reality. A linear form of representation allows for a simple form of representation and interpretation of the relationships under consideration (Greene, 2002). The Ordinary Least Squares (OLS) estimator has some attractive statistical properties that have made it one of the most powerful and popular methods of regression analysis (Gujarati & Porter, 2009). It therefore serves as the starting point for all modelling in the different chapters of this thesis.

In certain applications, a modification of the OLS estimator is necessary due to insufficient fulfilment of the assumptions, which affects its efficiency as an estimator. Therefore, a Fixed Effects (FE) estimator for a panel dataset is also considered in Section 2.1.2. In another special case, where the error terms of two dependent variables are correlated, the criteria for OLS as the Best Linear Unbiased Estimator (BLUE) are not met. Thus, the method of seemingly unrelated regression (SUR) is described in Section 2.1.3.

In addition to econometric modelling, advances in high-performance computing in recent years have opened up new possibilities in the field of data-driven farm processes. ML has emerged as a powerful tool to detect, quantity, and understand these processes (Liakos et al., 2018). ML offers great potential for expanding the toolbox of applied economists (Storm et al., 2020). While the typical goal of ML algorithms is to make predictions, economists have additionally discovered the role of precise predictions in policy evaluation and causal analysis (Athey, 2018). Whereas there are a variety of ML types, the ML approaches used in this thesis (regularised regression, tree-based methods, neural networks, and stacking) are introduced in Section 2.2.

2.1 Econometric Approaches

The starting point of econometric analysis is the definition of certain conditions of the question to be studied. Thus, the first step is to define precise, deterministic relationships between different variables in a linear regression model, the most useful tool in the econometrician's kit (Greene, 2002). The generic form of a multiple linear

regression model is described as follows:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \varepsilon_i \;\; or \;\; y_i = x_i'\beta + \varepsilon_i \tag{1}$$

where y_i is the dependent variable and $x_{i1},\ x_{i2},\ \ldots,\ x_{ik}$ are independent, observable random scalars, ε_i is the unobservable random disturbance or error term, and $\beta_0,\ \beta_1,\ \ldots,\ \beta_k$ are the unknown population parameters (constants) that should be estimated (Wooldridge, 2010). The underlying theory specifies the dependent and independent variables in the model. The term ε_i is a random disturbance, so called because the overall stable relationship is 'disturbed'. The influence of each variable cannot always be perfectly captured in a model; measurement errors in particular make representation difficult (Greene, 2002). The analysis considers all possible samples of size N that could have been drawn from the same population, where y_i and ε_i (and often x_i) are treated as random variables. Each observation represents one realisation of these random variables (Verbeek, 2012). In order to make sense of the model, assumptions must be made, since for any value of β, a set of ε_i can always be defined such that Equation 1 holds exactly for each observation. A common assumption is that the expected value of ε_i, given all the explanatory variables in x_i, is zero, denoted by $E\{\varepsilon_i|x_i\} = 0$. This assumption is referred to as the exogeneity of the x-variables, since x_i is uncorrelated with ε_i. Under this assumption, it can be concluded that:

$$E\{y_i|x_i\} = x_i'\beta \tag{2}$$

which means that the regression line $x_i'\beta$ describes the conditional expectation of y_i for the values of x_i. The coefficients β_k measure how the expected value of y_i changes when the value of x_{ik} is changed, holding the other elements of x_i constant (=ceteris paribus condition). The error term ε_i therefore has not only a statistical but also an economic interpretation. However, $E\{\varepsilon_i|x_i\} = 0$ may not always be justified, as in many cases the unobserved error term ε_i and the observed variable x_i are related, so that an interpretation of the regression coefficients as a measure of causal effects must be made with caution (Verbeek, 2012).

2.1.1 Ordinary Least Squares Estimator

The process of obtaining an approximate value for β as a given sample (y_i, x_i) with $i = 1, \ldots, N$ is commonly referred to as an estimator. The results obtained from a given sample is called an estimate, which is a vector of numbers. However, the estimator itself is a vector of random variables, as it can vary with different

samples. The OLS estimator is the most widely used estimator in econometrics and the following description of its functioning is based on Verbeek (2012). The OLS procedure involves finding values for the unknown parameters that minimise the residual sum of squares (RSS) $\sum \varepsilon_i^2$. The OLS estimator for β is expressed as:

$$b = \left(\sum_{i=1}^{N} x_i x_i' \right)^{-1} \sum_{i=1}^{N} x_i y_i \tag{3}$$

By checking the second order conditions, it can be observed that b corresponds to the solution of the RSS minimisation problem. The resulting linear combination of x_i is then given by:

$$\widehat{y_i} = x_i' b \tag{4}$$

which is the best linear approximation of y from $x_1, \ldots, x_k$ and a constant. The term 'best' refers to the fact that the sum of squared differences (approximation errors) is minimal for the least squares solution b.

Assuming the existence of an underlying 'true' model as defined in Equation 1, the vector of random variables b represents a sampling scheme. The objective is to estimate the true unknown parameter vector β using b as an approximation. While a single estimate is obtained from a given sample, the quality of the estimate is evaluated based on the properties of the underlying estimator. Due to its dependence on the randomly sampled population, b follows a sampling distribution.

In order to apply the OLS estimator, several assumptions have to be met, including the Gauss-Markov assumptions, given by:

$$E\left\{\varepsilon_i\right\} = 0; \quad i = 1, \ldots, N \tag{5}$$

$$\{\varepsilon_i, \ldots, \varepsilon_N\} \ \textit{and} \ \{x_i, \ldots, x_N\} \ \textit{are independent} \tag{6}$$

$$V\left\{\varepsilon_i\right\} = \sigma^2; \quad i = 1, \ldots, N \tag{7}$$

$$cov\left\{\varepsilon_i, \varepsilon_j\right\} = 0, ; \quad i, j = 1, \ldots, N; \quad i \neq \ j \tag{8}$$

Assumption 5 assumes that the expected value of the error term is zero, indicating that the regression line is, on average, accurate. Assumption 6 suggests that the error terms and the explanatory variables are unrelated and uncorrelated. Assumption 7 states that all error terms have the same variance, known as homoscedasticity, while assumption 8 requires zero correlation between different error terms, ruling out any form of autocorrelation. Combined, assumptions 5, 7 and 8 imply that the error terms are uncorrelated samples drawn from a distribution with zero mean and

constant variance σ^2. Under the assumptions of 5–8, the OLS estimator b for β has several desirable properties. Most notably, it is unbiased, meaning that in repeated sampling, the estimator is expected to be on average equal to the true value of β. This is expressed as $E\{b\} = \beta$. Overall, the OLS estimator b is considered to be the BLUE for β.

2.1.2 Fixed Effects Estimator

In some cases of the analysis of the German milk market, the data include observations on several entitlements over time, which allows the analysis of changes and trends over time (Verbeek, 2012). Panel data regression differs from regular time series or cross-sectional regression in that its explanatory variables vary in two dimensions, per individual and over time (Gujarati & Porter, 2009). The generic linear regression model refers to:

$$y_{it} = \alpha + X'_{it}\beta + \varepsilon_{it} \tag{9}$$

where $i = 1, \ldots, N$ denotes individuals and thus the cross-sectional dimension. The subscript $t = 1, \ldots, T$ indicates the time series dimension. α is a scalar, β is $K \times 1$ and X_{it} is the itth observation on K explanatory variables. In most of the panel data applications, a one-way error component model is used for the disturbances, where it holds:

$$\varepsilon_{it} = \mu_i + \nu_{it} \tag{10}$$

where μ_i is the unobservable individual-specific effect and ν_{it} is the residual disturbance (Baltagi, 2008).

One possible estimator for calculating the unknown parameter β in a panel regression is the FE or within-groups estimator. The operation of this estimator is described below and is based on Baltagi (2008). The term 'fixed' refers to the fact that the term μ_i does not vary over time, making it a group-specific constant. The remaining disturbances vary over time and individuals and are stochastic with ν_{it} independent and identically distributed $IID(0, \sigma_\nu^2)$. The X_{it} are assumed independent for the ν_{it} for all i and t. It must be assumed that the OLS assumption $E\{x_{it}\varepsilon_{it}\} = 0$ is violated as it seems unrealistic that the observations are independent. This means that $E\{x_{it}\mu_i\} \neq 0$ but it still holds $E\{x_{it}\nu_{it}\} = 0$. In vector form, the panel regression model can also be written as:

$$y = \alpha\iota_{NT} + X\beta + Z_\mu\mu + \nu \tag{11}$$

The matrix y has dimensions $NT \times 1$, while X has dimensions $NT \times K$. Z is formed by $Z = [\iota_{NT}, X]$, where ι_{NT} is a vector of ones of dimension NT. $Z_\mu = I_N \otimes \iota_T$ where I_N is an identity matrix of dimension N, ι_T is a vector of ones of dimension T. The symbol $\otimes$ stands for the Kronecker product. Z_μ acts as a selector matrix of ones and zeros and serves as a matrix of individual dummies that can be included in the regression to estimate the fixed parameters μ. OLS is applied to Equation 11 to estimate α, β and μ. Note that Z has dimensions $NT \times (K+1)$ and Z_μ, the matrix of individual dummies, has dimensions $NT \times N$. If N is large, Equation 11 will contain a substantial number of individual dummies, resulting in a large matrix to be inverted by OLS with dimensions of $(N+K)$. However, since α and β are the parameters of interest, an alternative approach is to use the Least Squares Dummy Variables (LSDV) estimator. This involves multiplying the model by Q and running OLS on the resulting transformed model:

$$Qy = QX\beta + Q\nu \tag{12}$$

This approach takes advantage of the property that $QZ_\mu = Q\iota_{NT} = 0$, since $PZ_\mu = Z_\mu$ with $P = Z_\mu {Z'_\mu}^{-1} Z'_\mu$ (the projection matrix on Z_μ), meaning that the Q-matrix eliminates the individual effects. This allows a regression of $\widetilde{y} = Qy$, with each element $(y_{it} - \bar{y}_{i.})$ on $\widetilde{X} = QX$, with each element $(x_{it,k} - \bar{x}_{it,k})$ for the kth regressor, where $k = 1, \ldots, K$. This approach involves the inversion of a $(K \times K)$ matrix instead of $(N+K)\times(N+K)$ as in Equation 11. The resulting OLS estimator is shown in Equation 13, with its variance, shown in Equation 14:

$$\widetilde{\beta} = (X'Q\ X)^{-1} X'Qy \tag{13}$$

$$var\left(\widetilde{\beta}\right) = \sigma_\nu^2 \left(X'Q\ X\right)^{-1} = \sigma_\nu^2 \left(\widetilde{X}'\widetilde{X}\right)^{-1} \tag{14}$$

Applying Generalised Least Squares (GLS) to Equation 12, using the generalised inverse, will also give $\widetilde{\beta}$.

The FE estimator faces challenges such as reduced degrees of freedom, multicollinearity among regressors, and the inability to estimate time-invariant variables. It relies on individual dummies, which can exacerbate multicollinearity, and eliminates the effects of time-invariant variables through Q- or mean-transformation. This results in perfect multicollinearity, leading to potential errors in regression packages. However, if the true model follows the individual dummies approach, the LSDV estimator is the BLUE, assuming ν_{it} is the standard classical disturbance with mean 0 and variance-covariance matrix $\sigma_\nu^2 I_{NT}$. While the FE estimator is consistent

as $T \to \infty$, when T is fixed and $N \to \infty$, only the FE estimator of β is consistent. The FE estimators of the individual effects $(\alpha + \mu_i)$ are not consistent, due to the increasing number of parameters as N increases.

2.1.3 Seemingly Unrelated Regression

If two dependent variables are used as response variables and thus have similar explanatory variables, it must be assumed that the two error terms correlate. In such a case, estimating the regression parameters equation by equation using OLS is unlikely to yield efficient estimates (Peremans & Van Aelst, 2018). Therefore, Zellner (1962) developed SUR models. SUR models account for the underlying covariance structure of the error terms in the equations. A SUR model consists of G equations, where each nth equation is defined as follows:

$$y_n = X_n \beta_n + u_n; \quad n = 1, 2, \ldots, G \tag{15}$$

where y_n is a vector of the observations in the dependent variable, X_n is a $K \times n$ matrix of the K exogenous variables, β_n is the estimated coefficient vector and u_n is a vector of the disturbance terms of the ith equation. The vectors can be noted among themselves as in a stacked system, so that β_n forms a K-dimensional vector of unknown parameters to be estimated by the model. It holds $K = \sum_{n=1}^{G} K_n$. It is assumed that there is no correlation of the error terms across the observations, so that $E[u_{ni} u_{mj}] = 0 \; \forall \; i \neq j$ with equation indices n and m and observation indices i and j holds.

It is explicitly assumed that the SUR model blocks are connected with each other via the assumption of contemporaneous correlation: $E[u_{ni} u_{mj}] = \sigma_{nm}$. This results in the following covariance matrix for all error terms: $E[uu'] = \Omega = \Sigma \otimes I_N$, where $\Sigma = [\sigma_{nm}]$ is the (contemporaneous) disturbance covariance matrix, $\otimes$ is the Kronecker product, I_N is an identity matrix of dimension N, and N is the number of observations in each equation.

Each G equation could be estimated individually by OLS, ignoring the correlation between the different error terms, resulting in inefficient estimators. As a modification of OLS, the GLS estimation can deal with this type of correlation. Therefore, the equation for the GLS estimator is as follows:

$$\hat{\beta}_{GLS} = \left(X' \left(\Sigma^{-1} \otimes I_N \right) X \right)^{-1} X' (\Sigma^{-1} \otimes I_N) y \tag{16}$$

Since in most cases the covariance Σ is unknown, the Feasible Generalised Least Squares (FGLS) estimator is used. This estimator calculates the elements of Σ by

$\hat{\sigma}_{nm} = \hat{u}_n'\hat{u}_m/N$, where $\hat{u}_m$ is the residual vector of the mth block obtained from OLS and then replaces Σ in GLS by the resulting estimator $\hat{\Sigma}$. This process can be repeated several times until the estimation coefficient is convergent.

2.2 Machine Learning Approaches

ML, a subset of artificial intelligence, is becoming increasingly accessible to applied economists due to advances in data availability and computing power (Storm et al., 2020). With its ability to use statistical methods to learn from data, ML provides a powerful toolbox for economists in a wide range of fields, including agriculture, the environment and development (Coble et al., 2018). While ML methods focus primarily on forecasting accuracy and errors, economists are recognising the importance of improved forecasting in policy evaluation and causal analysis (Athey, 2018). However, there is a distinction between ML and traditional econometric methods, with econometricians often focusing on deriving statistical properties of estimators for hypothesis testing (Mullainathan & Spiess, 2017). The ML literature typically focuses on algorithm development and prediction, including tasks such as classification and regression (Baylis et al., 2021). As a very simple example, Athey and Imbens (2019) model the conditional distribution of an outcome Y_i given a vector-valued regressor X_i as follows:

$$Y_i|X_i \sim N(\alpha + \beta^T X_i, \sigma^2) \tag{17}$$

In econometric approaches, the focus would be on estimating α and β using the OLS. In ML settings, the goal is to predict the outcome (Y_{N+1}) for new units N+1 based on their regressor values for that new unit (X_{N+1}). In a linear regression setting this would result in:

$$\hat{Y}_{N+1} = \hat{\alpha} + \hat{\beta}^T X_{N+1} \tag{18}$$

with the loss associated squared error $\left(Y_{N+1} - \hat{Y}_{N+1}\right)^2$. The goal is to find estimators $(\hat{\alpha},\ \hat{\beta})$ that have good properties associated with the loss function.

ML terminology is slightly different from econometric notation: The data used to estimate the parameters is commonly referred to as the training sample, as the model is trained rather than estimated. The predictors, covariates or regressors are called features, and the regression parameters can also be called weights. Prediction tasks are divided into supervised learning problems, where both the predictors/features X_i and the outcome Y_i are observed, and unsupervised learning problems, where

only the X_i are observed and the goal is to cluster them or estimate their joint distribution. Problems involving unordered discrete responses are typically called classification problems. (Athey & Imbens, 2019)

In ML, the ultimate goal is to obtain accurate predictions when the model is applied to previously unseen test data (James et al., 2017). To achieve this, it is important to evaluate the performance of the model on test data, rather than relying solely on training data. Out-of-sample comparisons, rather than in-sample goodness-of-fit measures, are used to ensure an unbiased assessment of model fit (Athey & Imbens, 2019). This can be done by dividing the available data into training, validation and test sets (Hastie et al., 2009). The training set is used for model estimation, while the validation set, also known as the development set, is used to monitor the model's out-of-sample prediction error and select the best performing model (Goodfellow et al., 2016).

Overfitting is a well-known challenge in ML, and researchers aim to strike a balance between model flexibility and overfitting (Athey & Imbens, 2019). ML models can be very flexible and can fit very specific relationships in the data, including non-linear patterns (Baylis et al., 2021). However, this flexibility also increases the risk of overfitting, where the model may learn spurious patterns that do not generalise well to unseen data. To mitigate overfitting, regularisation techniques are commonly used in ML to control the complexity of a model, with higher complexity increasing the risk of overfitting, while lower complexity can lead to underfitting (Goodfellow et al., 2016).

To select the optimal model, it is necessary to tune certain hyperparameters that affect the regularisation. Typically, the approach to tuning hyperparameters is to choose the levels that give the best results in the objective function of the validation set. In cases where the objective function aims to minimise the out-of-sample prediction error, the hyperparameters that give the lowest prediction error on the validation set are chosen (Yang & Shami, 2020). While the train/validation/test split approach is straightforward for datasets with many observations, k-fold cross validation is a common variant used for smaller datasets (Baylis et al., 2021). In this approach, the data is randomly divided into equally sized subsamples or 'folds' (Mullainathan & Spiess, 2017). The model is then trained successively on k folds, with one fold held back for evaluation. This process is repeated for different regularisation parameters, and the parameter with the best estimated mean performance is selected (Baylis et al., 2021). This helps to find the right level of model complexity that balances between overfitting and underfitting.

2.2.1 Regularised Regression

Regularisation is based on the addition of a penalty term to a linear regression that shrinks the coefficients towards zero. This is why it is also known as 'shrinkage'. Consequently, coefficients with low explanatory power are 'penalised' and thus forced out of the model (James et al., 2017). This procedure is particularly effective for datasets with a large number of explanatory variables, as only coefficients that differ from zero and contribute substantial predictive power are included in the model (Storm et al., 2020). The best-known regularisation methods are LASSO (first presented by Tibshirani (1996)) and Ridge Regression (developed by Hoerl and Kennard (1970)). Their optimisation functions are represented by minimising the RSS in Equation 19 and an additional penalty term, as in equations 20 (LASSO) and 21 (Ridge), according to James et al. (2017). Here, the coefficients β_j are minimised:

$$RSS = \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 \tag{19}$$

$$\min_{\beta_j} \left(RSS + \lambda \sum_{j=1}^{p} |\beta_j| \right) \tag{20}$$

$$\min_{\beta_j} \left(RSS + \lambda \sum_{j=1}^{p} \beta_j^2 \right) \tag{21}$$

Both regularisation methods use a tuning parameter λ, which can be determined by the data using cross-validation. The influence of the regularisation increases with increasing λ (James et al., 2017). The methods differ in the configuration of their penalty terms, also called L_1- (LASSO) and L_2-term (Ridge). The Elastic Net (EN) penalty term is shown in Equation 22 and forms a combination of both penalty terms with $\alpha \, \epsilon \, [0, 1]$ (Zou & Hastie, 2005).

$$(1 - \alpha) \, \lambda \sum_{j} |\beta_j| + \alpha \lambda \sum_{j} \beta_j^2 \tag{22}$$

The EN uses automatic variable selection via the L_1-part, which improves the prediction by inducing a grouping effect during variable selection so that highly correlated features are reduced (Ogutu et al., 2012).

2.2.2 Tree-based Methods

As a more powerful (multidimensional) alternative to classical non-parametric statistics, tree-based methods can be considered (Athey & Imbens, 2019). Decision trees

are used to divide a dataset into maximally homogeneous subgroups based on existing features (Storm et al., 2020). Thus, a so-called Classification and Regression Tree (CART) generates binary splits in each feature's contribution to the loss function (Breiman et al., 1984). This is done, as described in James et al. (2017), by partitioning the prediction space into J distinct and non-overlapping regions, $R_1, \ldots, R_J$. For each observation that falls into R_j, the same prediction is made, which is simply the mean of the response values for the training observations in R_j. The goal is to find boxes $R_1, \ldots, R_J$ that minimise the RSS given by:

$$\sum_{j=1}^{J} \sum_{i \epsilon R_j} \left(y_i - \hat{y}_{R_j}\right)^2 \tag{23}$$

where $\hat{y}_{R_j}$ is the mean response for the training observations within the jth box. Since it is computationally infeasible to consider every possible partition of the feature space into J boxes, the minimisation problem is approached algorithmically by recursive binary splitting. For each partition, a value s is chosen, such that two new sub-boxes are generated:

$$R_1(j,s) = \{X | X_j < s\} \;\; and \;\; R_2(j,s) = \{X | X_j \geq \; s\} \tag{24}$$

with the minimising function for j and s:

$$\sum_{i: x_i \epsilon R_1(j,s)} \left(y_i - \hat{y}_{R_1}\right)^2 + \sum_{i: x_i \epsilon R_2(j,s)} \left(y_i - \hat{y}_{R_2}\right)^2 \tag{25}$$

where $\hat{y}_{R_k}$ is the mean response for the training observations in $R_k(j,s)$. This process is applied recursively until there is a pre-specified depth or number of end-leaves.

Decision trees are almost always grown using regularisation by a complexity parameter α, since large trees are likely to overfit. One approach is to add a penalty term to the RSS, linearly scaled by the number of subsamples (leaves), with α chosen by cross-validation. A deep tree is first estimated and then pruned to a shallower tree using cross-validation to determine the optimal tree depth. This sequential process of growing and pruning the tree helps to avoid missing splits that may be based on subtle interactions (Athey & Imbens, 2019).

The random forest algorithm combines a collection of decision trees $\{h(x, \Theta_k)\}$, where x are inputs and Θ_k are independent, identically distributed random matrices. These were created by splitting the dataset randomly into subgroups of observations and features, with each subgroup k differing only slightly. For each subgroup, a

decision tree $h(x, \Theta_k)$ is created, from which the random forest predictor is formed by taking the average over k of the trees (Breiman, 2001a). This working principle is illustrated step-by-step in Figure 2.1.

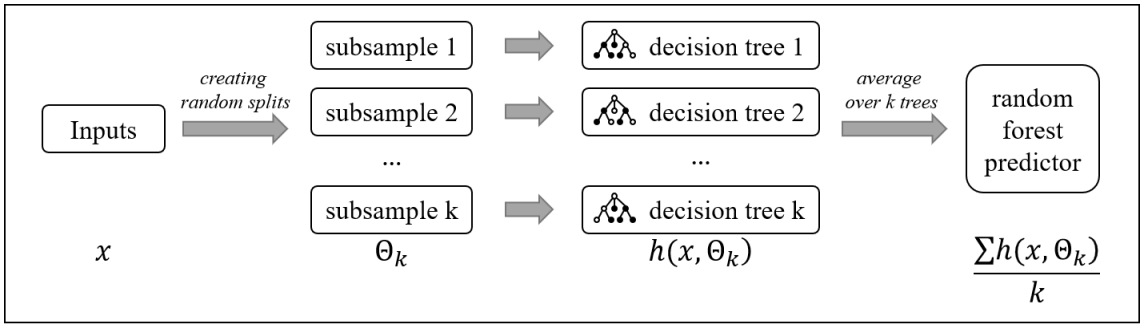

Figure 2.1: Schematic structure of a random forest

Source: Own illustration based on Breiman (2001a)

Another way to optimize the predictive power of decision trees is the method of boosting. This method was first presented by Schapire (1990), who ran repeatedly so-called 'weak learner' over various distributed subsamples and then combined these on a single predictor. The weak learners are assigned in each round $t = 1, \ldots, T$ to a uniform distribution D_t. In each round t, the algorithm maintains a set of weights based on accuracy. After evaluating the errors in each round, the weights are adjusted so that the more incorrect values are weighted more heavily to make up for the disadvantage over the other values. The weak learner in the following round then combines the prediction with the residuals of the previous round. This process is performed by t rounds, resulting in a weighted majority vote with the weight of each weak learner as a function of its accuracy (Freund & Schapire, 1997).

In this thesis, two boosting algorithms are applied: Xtreme Gradient Boosting (XGB), introduced by T. Chen and Guestrin (2016), and Gradient Boosting Machine (GBM), proposed by Friedman (2001). Both XGB and GBM use gradient boosting and optimise for accuracy by updating weights based on a gradient loss function. XGB can be thought of as an optimised and enhanced implementation of GBM that addresses some of the limitations of traditional GBM and provides additional features such as built-in support for regularisation, parallelization, missing value handling, categorical feature handling, and performance optimisations (T. Chen & Guestrin, 2016).

The structure of these algorithms requires the specification of several hyperparameters. The 'number of trees' hyperparameter controls the number of boosting rounds or iterations. Increasing the number of trees allows the model to learn

more complex relationships in the data, but may also increase the risk of overfitting (Schapire, 1999). The 'learning rate' (η) determines the step size at each boosting iteration. A smaller learning rate makes the model more conservative, requiring more boosting rounds to achieve the same level of complexity. This parameter helps to control the trade-off between learning speed and model accuracy (Friedman, 2001). The 'max depth' hyperparameter limits the depth of each decision tree in the boosting process. Deeper trees can capture more complex interactions in the data, but can also lead to overfitting (Xia et al., 2017). The 'min child weight' hyperparameter specifies the minimum sum of instance weights required in a child node for further splitting, which prevents overfitting. A higher value makes the algorithm more conservative, reducing the complexity of the model, while a lower value allows more splits, potentially capturing finer details but increasing the risk of overfitting (T. Chen & Guestrin, 2016). The 'subsample' and 'colsample by tree' hyperparameters introduce randomness into the algorithm. 'Subsample' specifies the fraction of observations to be used for each tree, while 'colsample by tree' specifies the fraction of features (columns) to be randomly sampled for each tree. These hyperparameters add stochasticity to the model and reduce the correlation between individual trees, which can improve generalisation, prevent overfitting, and improve computation speed (Friedman, 2002).

In order to achieve an optimal performance and to controll the model complexity, the hyperparameters can be optimised using different approaches (Yang & Shami, 2020). A commonly and in this thesis used method is grid search, where a predefined set of variables is tested in different combinations to find the best solution based on a minimum root mean squared error (RMSE) (Injadat et al., 2020).

2.2.3 Neural Networks

Typical Neural Networks (NNs) or multilayer perceptrons are structured as shown in Figure 2.2. They consist of several connected layers, each with multiple neurons. Their goal is to approximate a function f that learns from input values x to define an output vector y. NNs are basically a mapping of $y = f(x; w)$, where the weights w are continuously updated until they generate the best approximation of the function (Goodfellow et al., 2016). The weights and neurons in the centre of the network cannot be observed directly, so they are called the hidden layer (Hastie et al., 2009).

The strength of the connection between neurons is represented by the weights, which are multiplied by the input values and added to a constant ϑ that acts as the initial value for the optimisation (Athey & Imbens, 2019). The sum of these products is fed to the hidden layer via the activation function σ. Various types of

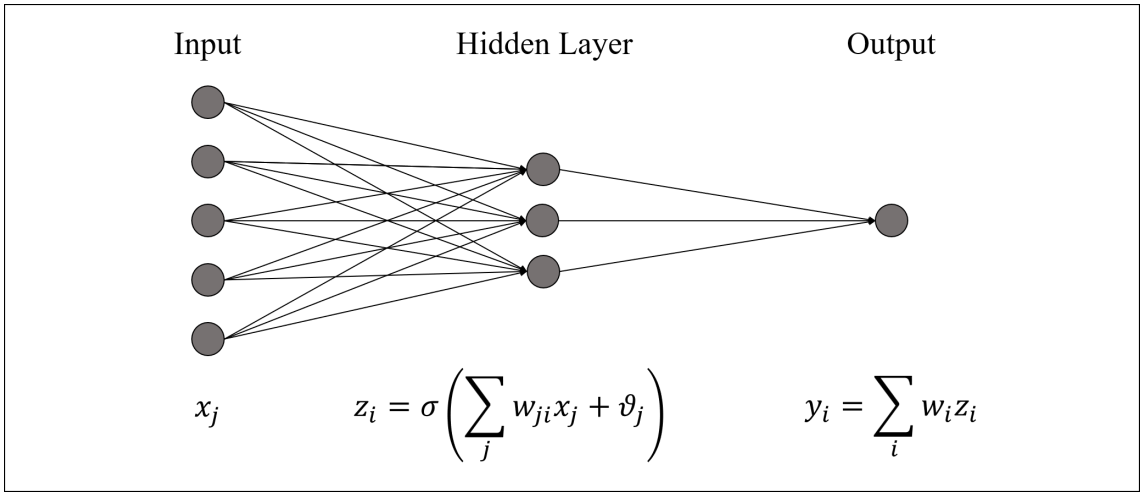

Figure 2.2: Schematic structure of a neural network

Source: Own illustration based on Goodfellow et al. (2016)

activation functions are available; for regression, the Rectified Linear Units (ReLu) form is considered particularly suitable here (M. Nielsen, 2015) and is defined by:

$$\sigma(z) = max\{0,\ z\} \tag{26}$$

In the specific case of Figure 2.2, the activated values are then fed into the output layer, although this serves only as a simplification. NNs are called nets because of their interconnection of different functions. In this case, three hidden layers are used so that a chain structure in the form of $f(x) = f^{(3)}(f^{(2)}\left(f^{(1)}(x)\right))$ occurs (Goodfellow et al., 2016). In a fully connected, also called dense, NN, each layer k has a structure given by $f^{(k)}(x) = h^{(k)} = g^{(k)}\left(W^{(k)T}h^{(k-1)} + b^{(k)}\right)$ with $h^{(0)} = x_j$, where $W^{(k)}$ is a matrix of unknown weights w_{ji}, and $b^{(k)}$ is a column vector of basic factors ϑ_j (Baylis et al., 2021).

The weights are updated and optimised by a loss function, similar to the optimisation of a linear regression (Goodfellow et al., 2016). In the NN context, this is called backpropagation and can be easily derived using the chain rule for differentiating the minimisation function of the RSS (Hastie et al., 2009). A global minimiser of $W^{(k)}$ is typically not desired, as it is likely to result in an overfitted solution. Instead, regularisation is required, which can be achieved directly through a penalty term or indirectly through early stopping (Hastie et al., 2009).

Again, this approach allows for different hyperparameters to be specified, so fine-tuning via e.g. grid-search is required (Athey & Imbens, 2019). Since all NNs are desigend with an input and output layer, the complexity of a NN is primarily determined by the number of hidden layers and the number of neurons in each

layer (Goodfellow et al., 2016). These hyperparameters are set and adjusted to achieve sufficient capacity to effectively model the objective functions. Deep NNs with more layers and neurons can learn complex representations, but may also be prone to overfitting. Shallow networks with fewer layers are simpler but may have limited representational capacity (Koutsoukas et al., 2017). The learning rate (η) determines the step size used to update the network weights during training, controlling the speed of convergence and the stability of the learning process (Athey & Imbens, 2019). While a higher learning rate can lead to faster convergence but also instability, a lower learning rate works the other way around and can lead to getting stuck on suboptimal solutions (Yang & Shami, 2020). The hyperparameter 'epochs' determines the number of times the entire dataset is passed through the NN during training (Bengio, 2012). Too few epochs can lead to underfitting, while too many can lead to overfitting (Sinha et al., 2010). The optimal number of epochs is typically determined through experimentation and monitoring the performance on a validation dataset. Techniques such as early stopping can be used to prevent overfitting by stopping training when validation performance plateaus or declines (Hastie et al., 2009). 'Mini batch size' refers to the number of training examples processed together before updating the network's weights (M. Nielsen, 2015). Larger batch sizes may provide more stable gradient estimates but require more memory (S. Smith et al., 2017). Smaller batch sizes introduce more stochasticity and can help the network escape local minima, but they may result in noisy updates (Yang & Shami, 2020).

2.2.4 Stacking

The stacking method, also known as stacked generalisation, combines several algorithms via a so-called 'super-learner'. This principle was first introduced by Wolpert (1992), who proposed the idea of improving the prediction performance by combining several individual models, which is also referred to as a refined version of cross-validation that improves on the winner-takes-all principle of cross-validation. Van der Laan et al. (2007) developed this concept and the following explanation of this learning technique is based on this literature.

In a regression analysis, the objective is to estimate the regression function $\psi_0(X)$, which represents the conditional expectation $E_0(Y|X)$ of the output variable $Y \in \mathcal{Y}$ given the input variables $X \in \mathcal{X}$. This is achieved by minimising the expected squared error loss, represented as $E_0 L(O|\psi)$, where $L(O|\psi)$ denotes the squared difference between the observed output value Y and the predicted value $\psi(X)$ with the observations $O_i = (X_i,\ Y_i) \sim\ P_0,\ i = 1, \ldots, n$.

The super learner algorithm uses V-fold cross-validation to evaluate and select the best candidate estimators. The candidate estimators, denoted as $\hat{\Psi}_j,\ j = 1, \ldots, J$, represent mappings from the empirical probability distribution P_n into the parameter space Ψ, which consists of functions of X. For each validation sample $v \in 1, \ldots, V$, the candidate estimators $\hat{\Psi}_j$ are applied to the corresponding training sample, denoted as $P_{nT(v)}$. The resulting estimates ψ_{njv}, are the realised values of the jth estimator when trained on the training sample $P_{nT(v)}$. The new dataset consists of pairs $(Y_i,\ Z_i)$, where Z_i is a vector consisting of the J predicted values obtained by applying the J estimators trained on the training sample $P_{nT(v(i))}$ to the input variable X_i where $v(i)$ is the validation sample to which it belongs. The set of possible outcomes for Z_i is denoted by $\mathcal{Z}$.

The stacking algorithm requires an additional user supplied prediction algorithm $\widetilde{\Psi}$ which estimates the regression $E(Y|Z)$ based on the dataset (Y_i, Z_i). For simplicity, this dataset is represented as $P_{n,Y,Z}$, so that Ψ is a mapping from $P_{n,Y,Z}$ to $\widetilde{\Psi}(P_{n,Y,Z})$, where it represents a function from $\mathcal{Z}$ to $\mathcal{Y}$. This prediction algorithm Ψ aims to minimise the cross-validated risk, which is defined as the average squared difference between the observed output values and the predicted values over the training samples, while considering penalisation or cross-validation techniques to prevent overfitting.

By applying the cross-validated risk minimiser $\widetilde{\Psi}$ to the dataset $P_{n,Y,Z} = (Y_i, Z_i) : i = 1, \ldots, n$, a mapping $\hat{\Psi}^*$ is obtained from the original data $P_n \equiv (Y_i, X_i) : i = 1, \ldots, n$ to the predictor

$$\widetilde{\Psi}\left(\left\{Y_i, Z_i = \left(\widetilde{\Psi}_i\left(P_{nT(\nu_i)}\right)(X_i) : j = 1, ..., J\right) : i = 1, ..., n\right\}\right) \tag{27}$$

The resulting predictor $\psi_n^* = \hat{\Psi}^*(P_n)$ represents the actual predictor obtained when the estimator/learner $\hat{\Psi}^*$ is applied to the original sample P_n. It is important to note that ψ_n^* belongs to the function space $\Psi^* \equiv f : \mathcal{Z} \longrightarrow \mathcal{Y}$, which represents functions mapping Z onto the outcome set $\mathcal{Y}$. Therefore, the super learner $\hat{\Psi}(P_n)(X)$ for a given value X based on the data P_n is defined as the evaluation of the predictor $\psi_n^* = \hat{\Psi}^*(P_n)$ at the J predicted values $\hat{\Psi}_j(P_n)(X)$ produced by the J candidate estimators for X.

Chapter 3

Forecasting Regional Milk Production Quantity: A Comparison of Regression Models and Machine Learning

3.1 Introduction

Germany is the fourth largest milk-producing country in the world in terms of volume (FAOSTAT, 2023). In the last 10 years, the volume of milk produced in the country has increased by about 11.7% (BLE, 2022). Milk production plays a significant role in the economic sector of agriculture, with the dairy sector accounting for about 18.9% of the value of goods produced (BMEL, 2020). In the period from 2007–2022 the milk price has been very volatile, with a fluctuation range of 22.2 cents/kg (07/2009) to 61.6 cents/kg (12/2022; BMEL (2023a)). Since the end of the milk quota in Europe in 2015, it has become much easier for farmers to expand their production capacities. One consequence of this is a lack of information among market participants about the volume of milk that will be produced in the months ahead.

Some dairies try to estimate the milk volume on the basis of voluntary self-reporting and self-assessment by farmers. However, it is not sufficient to simply ask farmers about their production intentions in order to build up a forecast index for predicting milk supply volumes because not all production-relevant influences can be assessed equally correctly by all farms at any given time (Larcher et al., 2015). Moreover, from an economic point of view, strategic responses are to be expected in the case of self-reporting. The challenge in forecasting milk quantities lies in the dependence on slow, biological processes (Hoehl & Hess, 2022). These complicate farmers' ability to fine-tune and adjust their production volumes directly.

For many years, econometric models have been estimated based on the lactation curve of individual dairy cows and thus for an entire farm (Adediran et al., 2012). In addition, recently there are various models for estimating individual cow performance in terms of milk quality and quantity using ML (Murphy et al., 2014; Nguyen et al., 2020; Soyeurt et al., 2020; Bovo et al., 2021). However, predictive models for a larger number of farms or even whole regions do not yet exist or are not publicly available. In this context, ML methods could offer a way of ensuring practise-relevant forecasting based on large datasets. Instead of building and applying a model based on existing theory, ML reads patterns from existing data structures from which inferences about future developments can be drawn (Mullainathan & Spiess, 2017). The focus of ML methods is therefore especially on predictive accuracy while econometric approaches rather concentrate on the statistical properties of parameters (Storm et al., 2020).

This chapter applies different forecasting approaches of ML to predict milk quantities and compares them, using a conventional regression model as a reference. It identifies which type of ML algorithms can be applied to predict regional milk pro-

duction and how they differ from each other using a combination of trend, weather and price variables. The relationships of the used variables are presented in the next section. The data structure and preparation are described in Section 3.3, followed by an introduction to the models and methods applied in Section 3.4. Section 3.5 presents and discusses the results of the algorithms before conclusions are drawn in the last section.

3.2 Theoretical Background

While economists usually describe aggregate dairy farm output in terms of a production technology with milk and meat output as a function of capital, labour and land inputs, the discipline of animal sciences has traditionally followed a different path: Production on dairy farms is particularly determined by the lactation curve of individual dairy cows. This can be illustrated as a mathematical formula to represent the course of milk production over time. Since Wood (1967) first established the gamma function, a variety of models have been developed to represent biological processes as accurately as possible. The course of the lactation curve is mainly based on the genetic material of the animals under consideration (Brügemann et al., 2011). Thus, a similar curve per animal can be observed over several lactation phases. Deviations from the original curve can be caused by various factors such as grazing conditions (Baudracco et al., 2012), seasonality (Wood, 1969), feeding conditions (Kolver & Muller, 1998), diseases (Collard et al., 2000) and other disturbances (Tekerli et al., 2000) such as heat. While farmers and their herd management can control most factors, feed conditions depend on weather as an additional component: Several studies show the influence of weather on pasture and forage quality (Parsons et al., 2019; Schlenker & Roberts, 2009). Therefore, weather can be expected to influence milk production in the short and medium term. In addition, livestock conditions are also related to weather. A combination of temperature, relative humidity, solar radiation, wind speed and precipitation can cause heat stress in animals, leading to reduced milk production in terms of quantity and quality (Mader et al., 2006). High temperatures can also cause increased mortality in dairy herds (Vitali et al., 2015).

Farmers' herd management is modelled in different ways in the literature. Dual production functions model the quantity of milk produced by farms as a relationship between output quantities and input and output prices. This is based on indirect cost or profit functions that are designed to maximize profits. For example, Blayney and Mittelhammer (1990) use a profit function of a generalised Box-Cox form to calculate price elasticities of the quantity of produced milk. Thus, it is possible to simulate the amount by which the milk supply changes in relation to price changes. Bouamra-

Mechemache et al. (2008) also use a dual approach and determine short and medium-term price elasticities for the quantity of milk produced using a normalized quadratic profit function. Both models illustrate the relationship between milk production and input and output prices. Farmers act as a profit maximiser in this relationship, making management decisions to adjust their milk production at given prices.

The amount of milk produced in a region is the sum of the amount of milk produced by the farms located there. In the last 20 years, the number of farms in Germany has fallen by more than half, but the quantity of milk produced has increased by 14.4% in the same period (BLE, 2022). This is caused by ongoing structural change, with smaller and less profitable farms ceasing production due to various factors such as economics of size and scale (Hallam, 1991) as well as natural conditions and agricultural prices in specific regions (Neuenfeldt et al., 2019). In contrast, larger and highly professionalized dairy farms can significantly increase their production through increased remounting and expansion of their facilities.

Figure 3.1 shows the relations of the determinants of milk production outlined above. For the forecasting models presented in this chapter, the influencing factors are grouped into three main components: weather, time, and price variables. It is assumed that milk production in a region k is a function depending on the production function on farm i with cows j kept and their lactation curve. This results in a function depending on herd performance, herd management, market prices, and other factors. This simplified representation allows complex on-farm interactions and macroeconomic developments to be illustrated by so called time variables as a trend in milk production.

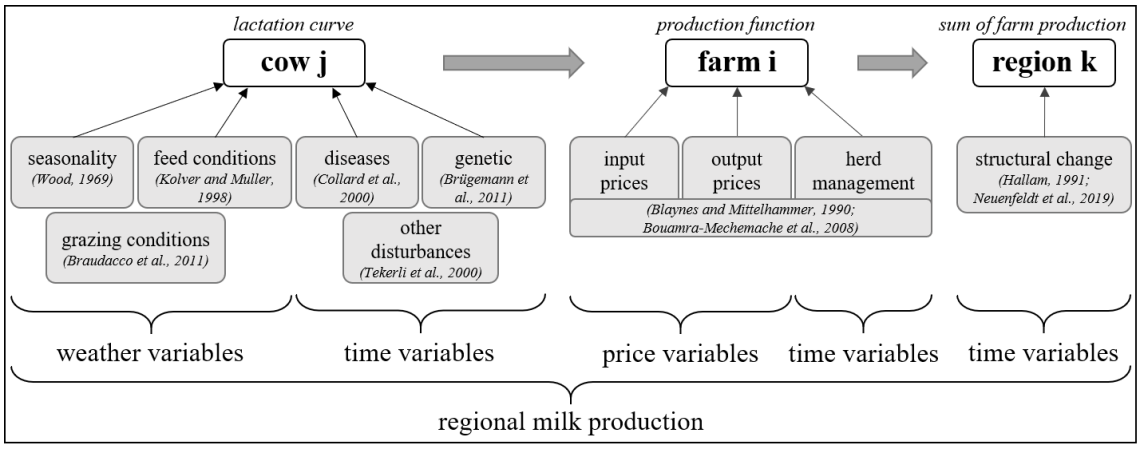

Figure 3.1: Schematic structure of milk production

An exact definition of the variable groups used with the specific explanation of individual variables is given in the next section.

3.3 Data

The dataset used refers to monthly observations in the period 09/2010 to 09/2019, and thus falls roughly within the timeframe when the abolition of the milk quota had already been decided and prepared for by an annual increase in quota volumes ('soft landing'). Variables of individual farm milk production as well as weather and price data are included in the models.

The data on milk production are provided by the Lower Saxony State Inspection Association, Landeskontrollverband Niedersachsen e.V. (LKV). The association compiles a monthly milk performance test for all member farms in Lower Saxony (a region in Northern Germany) and provides this data for 291 randomly selected dairy farms. Farmers are allowed to drop one value per year, so the dataset contains 11 data points per farm per year. The missing monthly value is replaced by a linear imputation procedure, so that a balanced panel dataset can be used. This generates data points for about 30,000 dairy cows in Lower Saxony that are aggregated by an average value for the respective farms. Thus, a single monthly value is obtained by calculating the mean milk production of the cows milked on each farm. This value is multiplied by 30.5 days and the number of milked animals per farm and month, corresponding to the average amount of milk produced per farm and month. The energy-corrected milk yield (ECM) is considered, calculated according to Sjaunja (1990). The number of animals kept per farm varies from fewer than 10 to more than 300 animals in the applied dataset. The mean is about 100 cows per farm. The size class of farms with over 200 dairy cows corresponds to the distribution in Lower Saxony with about 8.5% share of the total number of farms. The number of smaller farms with less than 50 dairy cows (17.7%) is underrepresented compared to the distribution in Lower Saxony (34.5%) (LSN, 2020). Accordingly, the majority (73.7%) of observed farms are in the 50–199 dairy cow size class. However, this can be helpful in terms of forecasting as structural change is ongoing and the share of the size classes is continuously shifting towards larger farms (BLE, 2020).

Figure 3.2 shows the development of milk volume over the observed period for the dataset of $N = 291$ farms. The overlapping areas represent the respective quantiles with different colours according to the legend. The mean is presented as a black line. Significant seasonal variations in monthly milk production can be observed throughout the year. Typically, the annual maximum occurs in May, while the annual minimum is generally observed in September. Furthermore, there is an overall continuous increase of 35.6% in average monthly milk production per farm from 64,445 ECM (2011) to about 87,398 ECM (2018). The dataset shows substantial differences in milk production between farms, confirming the heterogeneity of farm

size mentioned above. Larger farms produced a disproportionate amount of milk compared to the other farms. This is confirmed by the mean being higher than the median, which falls within the middle of the 40-60% quantile.

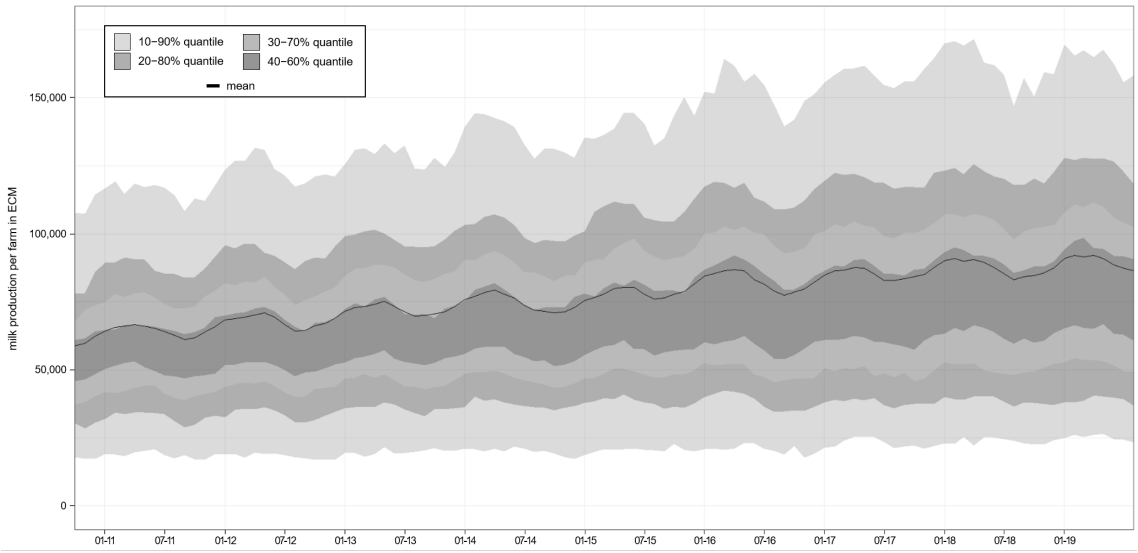

Figure 3.2: Monthly milk production of the $N = 291$ farms during observed period

Note: The quantiles are represented as overlapping areas.

As shown above, it must be assumed that milk production in a region is largely determined by short- and long-term climatic influences (Hoehl & Hess, 2022). This particularly affects the quality of basic feed and the physiology of the animals. It would therefore appear necessary to include regional weather and climate data in the forecast. For weather variables, the 'COSMO Regional Reanalysis' database of Germany's National Meteorological Service, Deutscher Wetterdienst (DWD), is used. In these numerical weather models, observations and analyses are combined so that a large number of parameters are available for a grid of about 6 km x 6 km resolution.

Milk prices are provided by the Federal Ministry of Agriculture and Food, BMEL, and refer to conventionally produced cow's milk in Lower Saxony at actual fat and protein content. For feed prices, the index for purchase prices of agricultural enterprises determined by the Federal Statistical Office is used.

All features applied are listed with minimum value (Min.), Mean, maximum value (Max.) and standard deviation (SD) in Table 3.1.

Variable description	**Min.**	**Mean**	**Max.**
Monthly milk production per farm [ECM]	1746.65	77141.11	414150.63
Monthly milk production per farm [kg]	1409.10	76196.64	401700.25
Monthly milk production per cow per farm [ECM]	150.51	860.09	1330.24
Cloud cover [%]	47.86	77.86	96.89
Sunshine duration [h]	11.91	152.92	356.26
Precipitation [mm]	0.35	51.66	268.94
Relative humidity [%]	54.09	80.14	93.61
Maximum temperature [°C]	1.75	21.10	39.84
Mean temperature [°C]	-3.14	13.41	27.66
Minimum temperature [°C]	-18.17	1.26	13.69
Soil temperature [°C]	-3.29	10.01	23.96
Zonal wind speed [km/h]	-21.31	3.78	22.19
Temperature-Humidity-Index [Index]	27.99	55.97	76.12
Milk price [cents/l]	20.57	32.76	42.40
Feed price [Index, 2015=100]	90.70	101.74	122.40

Table 3.1: Descriptive statistics of the applied features

3.4 Methods

This section presents the different models used to predict milk production in the region of Lower Saxony. The basic model is a multiple linear regression based on the OLS estimator. Compared with this basic model, a number of ML models are set up to compare the difference to the econometric approach and between the different ML algorithms for the prediction of milk production.

For the analysis, the dataset is randomly divided into a training set and a test set. This allows the models and algorithms to learn the patterns of the data in the training set (80% of the data). In a second step, the prediction precision of the algorithms can be evaluated by applying them to the test set, which accounts for 20% of the data. The prediction is to be determined for a medium-term period of 12 months. Time series data are known to contain temporal dependencies that should be considered when randomly splitting the data to avoid leakage and overfitting (Lones, 2021). However, the model structure allows these temporal dependencies to be included in the model even with a random split. This model structure is described in more detail in the following sections.

All analyses were performed using R version 4.0.2 (R Core Team, 2021) and RStudio version 1.3.1056 (RStudio Team, 2020) for statistical computing on a laptop equipped with an Intel Core i7-8565U processor clocked at 1.8 GHz and 16 GB RAM.

3.4.1 Basic Model

The basic model assumes that the milk production of farm i can be explained by an interaction of economic and climatic conditions in month t and previous months (lags) n. For the prediction period of 12 months, it is assumed $n = 1, \ldots, 12$ and previous months are considered by adding further lags as $m = 0, \ldots, 6$. Economic effects can be divided into price effects for inputs and outputs and farm-specific management factors. The influence of weather on milk yield can be divided into direct and indirect effects. Direct weather effects typically include heat stress for dairy cows and impairment of their metabolism (Hill & Wall, 2017). Indirect weather effects result from quality and quantity differences in feeding (Harder et al., 2019).

The basic model can be described as follows:

$$y_{i,t-n} = T_{t-(n+m)} + W_{i,t-(n+m)} + P_{t-(n+m)} + \mu_i + \varepsilon_{i,t-n} \tag{28}$$

where $y_{i,t-n}$ = average monthly quantity of milk [kg ECM] produced by farm i at time (month) t lagged by n months, $T_{t-(n+m)}$ = time variables, $W_{i,t-(n+m)}$ = weather variables, $P_{t-(n+m)}$ = price variables, μ_i = dummy for farm i and $\varepsilon_{i,t-n}$ = error term. The time variables are divided into date, year, and factor variables for each month. The weather variables section contains ten different weather aspects at time t lagged by n+m months, which can be assigned to the coordinates of farm i. These are precipitation [mm], minimum temperature [°C], maximum temperature [°C], mean temperature [°C], soil temperature [°C], relative humidity [%], sunshine duration [h], cloud cover [%], zonal wind speed [km/h] and Temperature-Humidity-Index (THI) (calculated according to (Mader et al., 2006)). The monthly milk prices for Lower Saxony [cents/l] and a feed price index at time $t - (n + m)$ are used as price variables.

3.4.2 Machine Learning Models

The model specifications presented in the previous section are applied to the dataset by means of different estimation methods, comparing conventional regressions (OLS) with 'supervised learning' ML methods. As there is a number of supervised ML approaches, the analysis here is limited to three different model types: regularised regression, tree-based methods, and a neural network approach. To find an appropriate level of model complexity and thus to avoid an overfitted model, some parameters of the model are validated using cross-validation. This involves dividing the dataset into k parts, each with the same number of observations. On each occasion, the out-of-sample prediction error is iteratively predicted using all but one of

the k parts. By averaging the out-of-sample prediction error over the k estimators, the expected value of the out-of-sample prediction error can be estimated (Baylis et al., 2021).

3.4.2.1 Regularised Regression

Regularisation is a technique used in linear regression to reduce the impact of coefficients with low explanatory power by adding a penalty term. This penalty term encourages the coefficients to shrink towards zero, effectively excluding them from the model (James et al., 2017). Regularisation is particularly useful when dealing with datasets that have a large number of explanatory variables. It ensures that only coefficients that contribute significantly to the predictive power and are different from zero remain in the model. Two well-known regularisation methods are LASSO (introduced by Tibshirani (1996)) and ridge regression (proposed by Hoerl and Kennard (1970)). These methods involve minimising the RSS, as shown in Equation 29, along with an additional penalty term for LASSO (see Equation 30) and for Ridge (Equation 31), as described by James et al. (2017). The objective is to minimise the coefficients β_j.

$$RSS = \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 \tag{29}$$

$$\min_{\beta_j} \left(RSS + \lambda \sum_{j=1}^{p} |\beta_j| \right) \tag{30}$$

$$\min_{\beta_j} \left(RSS + \lambda \sum_{j=1}^{p} \beta_j^2 \right) \tag{31}$$

Both regularisation methods, LASSO and Ridge Regression, rely on a tuning parameter λ, which can be determined from the data through cross-validation. The strength of the regularisation effect increases as λ increases (James et al., 2017). These methods differ in the configuration of their penalty terms, known as the L_1 term (LASSO) and the L_2 term (Ridge). Another approach, known as EN, combines both penalty terms and is represented by Equation 32. The parameter $\alpha \ \epsilon \ [0, 1]$ determines the combination ratio between the L_1 and L_2 penalties (Zou & Hastie, 2005).

$$(1 - \alpha)\, \lambda \sum_{j} |\beta_j| + \alpha \lambda \sum_{j} \beta_j^2 \tag{32}$$

The EN employs automatic variable selection through its L_1 penalty term, which

improves prediction performance by grouping highly correlated features and reducing their influence (Ogutu et al., 2012). In this particular case, an α value of 0.5 is assumed and the tuning parameter λ is validated using 10-fold cross-validation.

The 'glmnet' library (Friedman et al., 2010) was used to perform regularised regression analysis.

3.4.2.2 Tree-based Methods

Decision trees are used to partition a dataset into highly homogeneous subgroups based on available characteristics (Storm et al., 2020). This process, known as CART, creates binary splits that contribute to the loss function (Breiman et al., 1984). The random forest algorithm combines multiple decision trees, denoted as $h(x, \Theta_k), k = 1, \ldots, K$ where x represents inputs and Θ_k represents independent, identically distributed random matrices. These decision trees are generated by randomly dividing the dataset into subgroups of observations and features, with each subgroup k having slight variations. For each subset, a decision tree $h(x, \Theta_k)$ is constructed, and the random forest predictor is obtained by averaging the predictions of all the trees (Breiman, 2001a). Figure 3.3 provides a step-by-step illustration of this procedure.

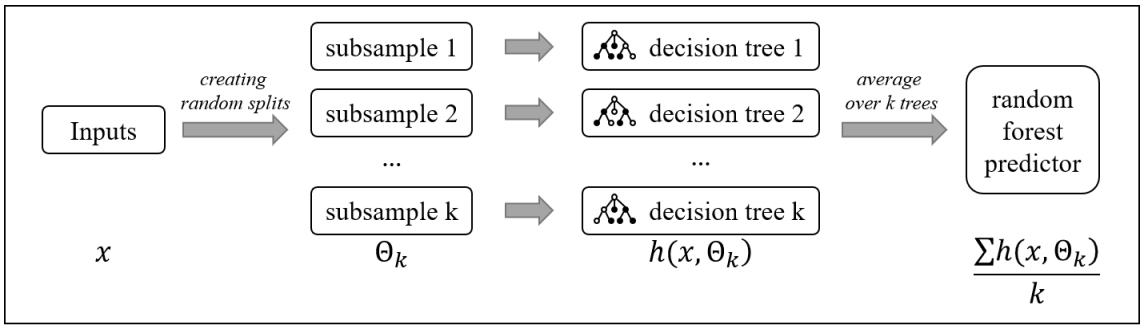

Figure 3.3: Schematic structure of a random forest

Source: Own illustration based on Breiman (2001a)

An alternative approach to improving the predictive performance of decision trees is boosting. This technique was first introduced by Schapire (1990), who iteratively applied weak learners to different distributed subsamples and then aggregated them into a single predictor. Weak learners are assigned a uniform distribution D_t in each round $t = 1, ..., T$. The algorithm maintains a set of weights based on accuracy during each round. By evaluating the errors, the weights are adjusted to give more weight to incorrect values to compensate for their disadvantage compared to other

values. The weak learner in the next round combines its prediction with the residuals from the previous round. This process continues for t rounds, resulting in a weighted majority vote where the weight of each weak learner is determined by its accuracy (Freund & Schapire, 1997).

Two boosting algorithms are used to predict milk production in this case: XGB and GBM. These algorithms optimise weights by updating them based on a gradient loss function. Boosting algorithms have hyperparameters that need to be specified. To achieve the most accurate results, different approaches can be used to optimise the hyperparameters of ML algorithms (Yang & Shami, 2020). One commonly used method is grid search, where a predefined set of hyperparameters is tested in different combinations to find the optimal combination based on minimising the RMSE (Injadat et al., 2020).

The hyperparameters used for fine tuning are shown in the appendix. Finally, the following hyperparameters were used for the XGB algorithm: number of trees = 2000; learning rate (η) = 0.1; maximum depth of a tree (max depth) = 7; minimum sum of weights of all observations required a child (min child weight) = 1; fraction of observations to be randomly sampled for each tree (subsample) = 0.8; fraction of features to be randomly subsampled (colsample by tree) = 0.8; 10-fold cross-validation, and the following hyperparameters for the GBM algorithm: number of trees = 2000; η = 0.01; max depth = 7; min child weight = 5; subsample = 0.8; colsample by tree = 1; 10-fold cross-validation.

The XGB was done using the 'xgboost' library (T. Chen et al., 2020), while the GBM was done using the 'h2o' library (LeDell et al., 2021).

3.4.2.3 Neural Networks

NNs, also known as multilayer perceptrons, have a typical structure as shown in Figure 3.4. These networks consist of interconnected layers, each consisting of many neurons. Their objective is to approximate a function f, which learns from input values x and produces an output vector y. In essence, NNs establish a mapping $y = f(x; w)$, where the weights w are continuously adjusted to achieve the optimal approximation of the function (Goodfellow et al., 2016). The weights and neurons in the innermost part of the network are not directly observable and are therefore referred to as the hidden layer (Hastie et al., 2009).

The interconnection between neurons in a NN is determined by the weights, which are multiplied by the input values and combined with a constant term ϑ, which serves as a starting point for optimisation (Athey & Imbens, 2019). The resulting products are summed and passed through the activation function σ, as

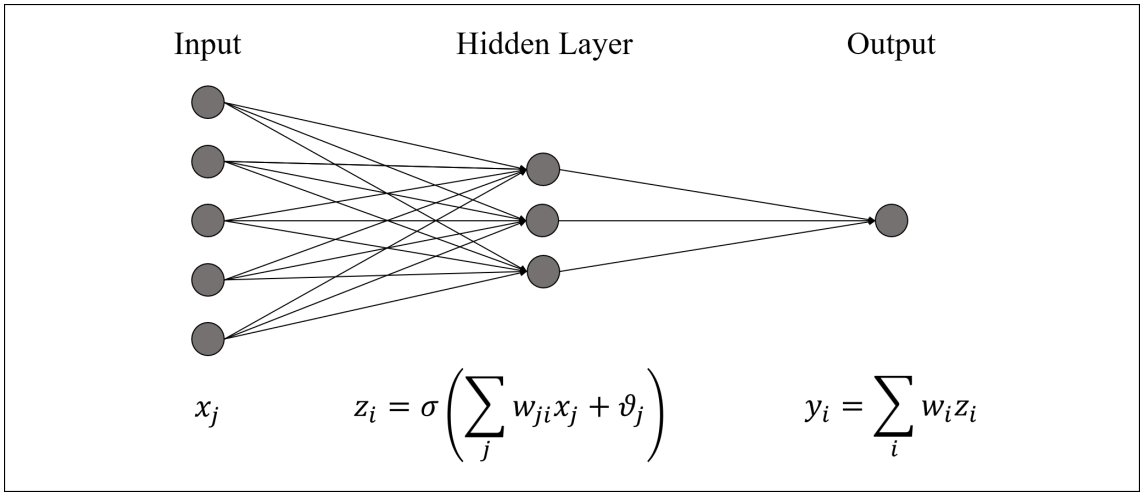

Figure 3.4: Schematic structure of a neural network

Source: Own illustration based on Goodfellow et al. (2016)

commonly used in a regression case of the ReLu type (see Equation 33 and Figure 3.4).

$$\sigma(z) = max\{0,\ z\} \tag{33}$$

In the specific case shown in Figure 3.4, the activated values are then transmitted to the output layer. However, this simplification is for illustrative purposes only. NNs are so called because of their composition of interconnected functions. In this particular case, three hidden layers are used, resulting in a cascading structure described by $f(x) = f^{(3)}(f^{(2)}\left(f^{(1)}(x)\right))$. The weights are updated and optimised by a loss function, similar to the process used in linear regression (Goodfellow et al., 2016).

Again, various hyperparameters can be specified, so that fine-tuning is necessary. The set of possible hyperparameters is shown in the appendix. The following hyperparameters are chosen for the application of the NN for milk quantity forecasting: three hidden layers with 200, 50 and 50 neurons; optimization with backpropagation; 100 epochs; stopping criterion = RMSE; minimum batch size = 32; 10-fold cross-validation.

The NN was built using the 'h2o' library (LeDell et al., 2021).

3.4.2.4 Stacking

The stacking method, as the name suggests, combines multiple algorithms through a 'super learner' to improve prediction performance by exploiting the strengths of different models. While the choice of learning algorithm is flexible, the use of the super learner requires each of the preceding models to be trained with the same

V-fold cross-validation. The resulting predictions from the base learner L are then stacked into a $V \times L$ matrix, which serves as the input data for the super learner (Van der Laan et al., 2007). Consequently, the stacking method is often seen as an extended version of cross-validation that improves upon the winner-takes-all principle (Wolpert, 1992). In this particular case, the stacking model includes the EN, GBM and NN models, with the XGB model excluded due to its similarity to GBM and the limited added value it would provide.

The algorithms were combined by the stacking method using the 'h2o' library (LeDell et al., 2021).

3.5 Results and Discussion

To evaluate the different algorithms and models, the trained algorithms are applied to the test set. Through this approach, different values for the relative and absolute measure of accuracy of the models can be observed. Therefore, the coefficient of determination (R^2) is compared to measure the relative error (Figure 3.5) and the RMSE for an absolute error value (Figure 3.6). The adjusted R^2 does not need to be calculated in this instance because the explanatory power of specific variables is not primarily of interest; instead the focus is on the overall forecast accuracy, without considering degrees of freedom.

Figure 3.5 compares the accuracies of the different algorithms. In addition, the differences with regard to the R^2 are represented for different variations of the target variable. The R^2 values for the monthly milk quantity per farm corrected for fat and protein content (ECM), as well as without correction and as average milk quantity per cow per farm are shown in different grey shades. Considering the initial model for the monthly milk quantity per farm in ECM (coloured in dark grey), it is noticeable that for all models a R^2 above 0.93 is achieved and thus the accuracy of all models is relatively high. While the results for the test data are shown in full colour, the bars for the results of the training data are not filled in and are just behind the filled bars. Since the respective values do not show large differences between training and test results, it can be assumed that none of the models suffers from overfitting.

By comparing the differences between the various target variables in Figure 3.5 on the basis of the shading, it is noticeable that there is almost no difference in the prediction accuracy for the monthly milk quantity with and without fat and protein correction. However, a reduction of the R^2 can be observed for the monthly average milk quantity per farm per cow for all applied models. It seems that the models can predict the development of the number of animals better than the differences at individual cow level. Additionally, the difference between the R^2 of the non-linear

ML algorithms for training and test data are larger for the prediction of the milk production per cow per farm. In contrast to the other models, the models here are therefore slightly overfitted.

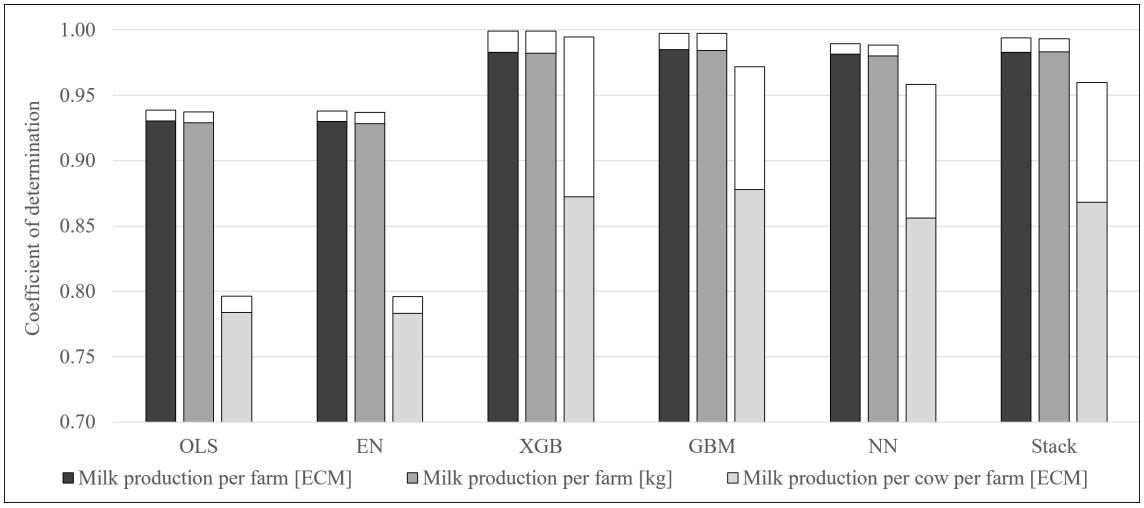

Figure 3.5: Coefficient of determination for all applied algorithms in training and test set

Note: Filled bars refer to R^2 for test data, unfilled bars behind refer to R^2 for training data.

The RMSE is illustrated as an absolute value in Figure 3.6 to be able to estimate the deviations from the monthly milk yield per farm. The values indicate the standard deviation of the residuals and are mainly below 15,000 for linear models and below 6,000 for non-linear approaches. In relation, an average farm in the dataset produces about 79,698 ECM milk per month. The differences between the respective models correspond to the ratios of the R^2 (see Figure 3.5), while in this case a lower RMSE indicates a more precise model. The filled columns correspond to the RMSE of the training dataset and the unfilled columns correspond to the results of the test dataset.

A systematic difference in the size of the R^2 and RMSE can also be observed between the algorithms. While the models XGB, GBM, NN and the stacking algorithm all have fairly similar accuracies in each case, a significant deviation from the linear approaches such as the OLS and the EN can be recognized. One possible explanation for the differences in the results for R^2 and RMSE in linear and non-linear approaches could be the occurrence of two different farm groups. Figure 3.7 shows the quantity of milk produced over time by ten selected farms representing these two kinds of farm groups. Farm group *a* is defined as those farms which increase the quantity of milk over proportionately in the observed period as a result

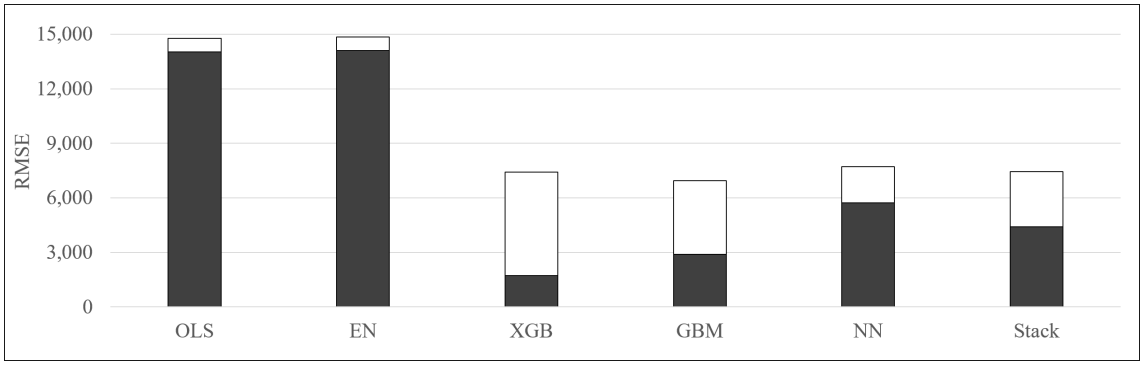

Figure 3.6: RMSE for all applied algorithms for milk production per farm in ECM

Note: Filled bars refer to R^2 for training data, unfilled bars behind refer to R^2 for test data.

of massive investments (see Figure 3.7, farms F-J). In farm group b, in contrast, an average change in milk quantity with regular seasonal fluctuations can be identified (see Figure 3.7, farms A-E).

Figure 3.7: Milk production over time for ten picked farms

This development can be seen, on the one hand, by the average annual growth rate and, on the other hand, by abrupt changes in the monthly milk volume. For this reason, an indicator function is formed that distinguishes between these two groups, which allows a separate calculation of R^2 for each group. That function is defined as follows:

$$I_{GR}(x,z) = \begin{cases} a & if\ x_i > GR_{mean} \wedge z_i > GR_{max} \\ b & if\ x_i \leq GR_{mean} \wedge z_i \leq GR_{max} \end{cases} \tag{34}$$

where x_i is the average annual growth rate of milk production for farm i and

$GR_{mean} = Q_{0.75}(x)$. z_i corresponds to the maximum annual growth rate of milk production for farm i and $GR_{max} = Q_{0.75}(x)$. This division based on the 0.75 quantiles assigns each farm a subgroup a or b according to the indicator function.

The majority of farms (83.6%) can be assigned to farm group b – moderately growing farms. The average annual change in milk volume in this group is about 2.2%, while in farm group a – strongly growing farms – this rate is significantly higher at about 9.5%. Group a accounts for 16.4% of farms showing substantial higher maximum annual growth rates ranging from 18.9% to 71.5% during the observation period. This suggests that these farms are growing not only through on-farm remounting, but through larger investments based on purchased animals. In group b, by contrast, maximum annual growth rates range from 1.5% to 36.7%.

Figure 3.8 depicts the R^2 for the respective algorithms for group a and b. The columns grouped side by side show that the R^2 of the group of fast-growing farms (group a) is lower in the linear approaches (OLS, EN). In particular, compared with the non-linear approaches, the R^2 in these respective groups is significantly lower, with a maximum value of 0.88 (OLS for group a). For the group of moderately growing farms the linear approaches achieve considerably higher R^2s than for the entire dataset. These differences between group a and b are not evident to the same extent for the non-linear estimation approaches. Although the R^2s of group a are slightly lower in all cases except for the XGB, this can also be attributed to the smaller database as group a only accounts for 16.4% of the data. These differences could be caused by investments as a driving factor of growth in group a. Milk production in this group does not increase linearly, but partly in step changes due to investments in new barn constructions and/or expansion. This development can be explained more effectively by a non-linear algorithm.

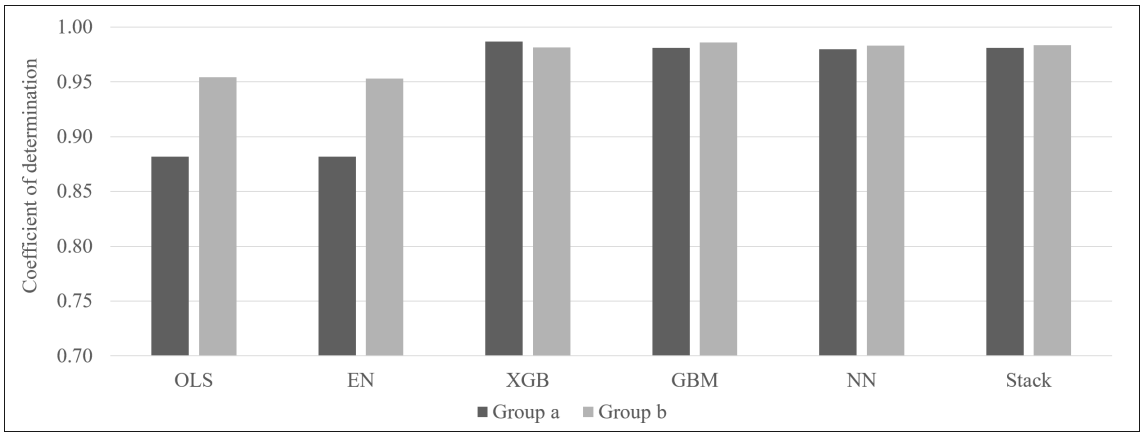

Figure 3.8: Coefficient of determination in the comparison of farm groups a and b

The accuracy not only differs between the two groups of farms, but also between

the predictions for individual farms. In order to illustrate these differences, Figure 3.9 shows the respective R^2 per farm in a violin plot. Again, the difference between the linear and non-linear estimation approaches becomes apparent. While the linear estimators also predict some farms very poorly (with $R^2 < 0.25$), the non-linear estimators show these lower values less often. On average (indicated by the red dot in each violin plot) there is also a significant difference between the two estimation approaches. However, Figure 3.9 also shows differences between the non-linear algorithms themselves. In particular, there is a difference in the proportion of relatively small R^2, with the GBM algorithm showing these values most rarely. The R^2 for XGB, NN, and stacking algorithm are again relatively similar. There is also a small advantage for GBM over the other non-linear algorithms when averaged over all farms.

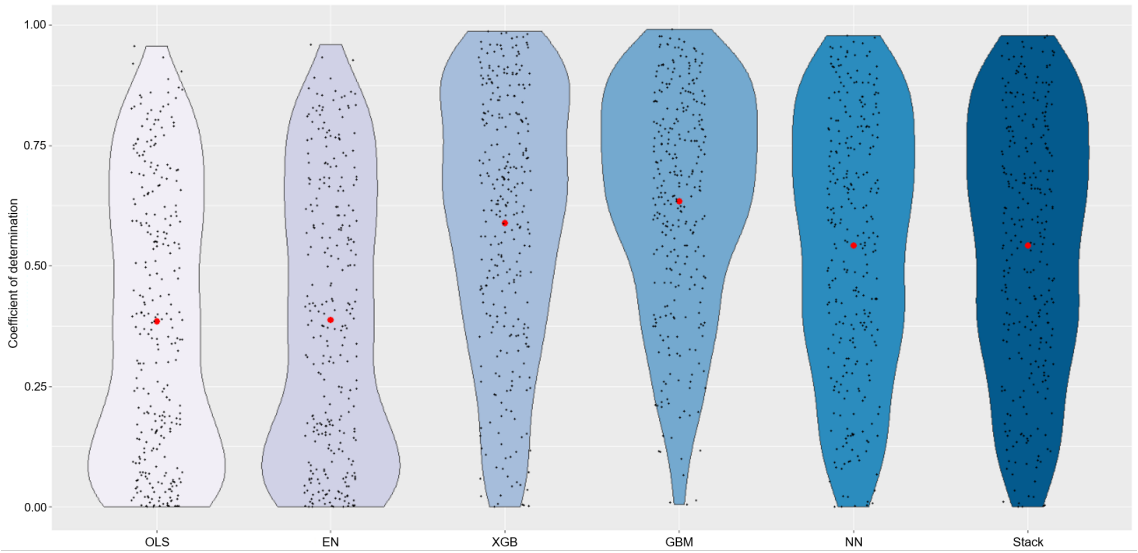

Figure 3.9: Coefficient of determination on individual farm level

Note: Black dots represent R^2 of individual farms, red dots the average of all farms.

The aim of this work is to compare forecasting models for regional milk production. Therefore, in addition to examining the results at farm level, it is appropriate to consider the regional level. Summing the predicted values of milk production for each county allows a comparison of the R^2 at county level. Figure 3.10 shows the respective R^2 per county for the different models. It can be seen that for some counties the results are quite similar for all algorithms. It is also notable that the R^2 for OLS and EN are mostly only slightly different, and that the other ML algorithms also have very comparable R^2 among themselves. There are only a few exceptions where clear differences can be seen. For example, in the districts of Göttingen, Helmstedt, Osterode am Harz and Uelzen, the monthly milk quantity can

only be predicted to a limited extent by all models with significantly lower R^2 than in other districts. This also highlights the differences between linear and non-linear estimation approaches, as there are larger differences in R^2 between these two types. This may be explained by the comparatively smaller number of farms in these counties in the dataset, which may affect the accuracy of the models. However, it can be concluded that the non-linear ML algorithms seem to be able to predict more accurately with less data.

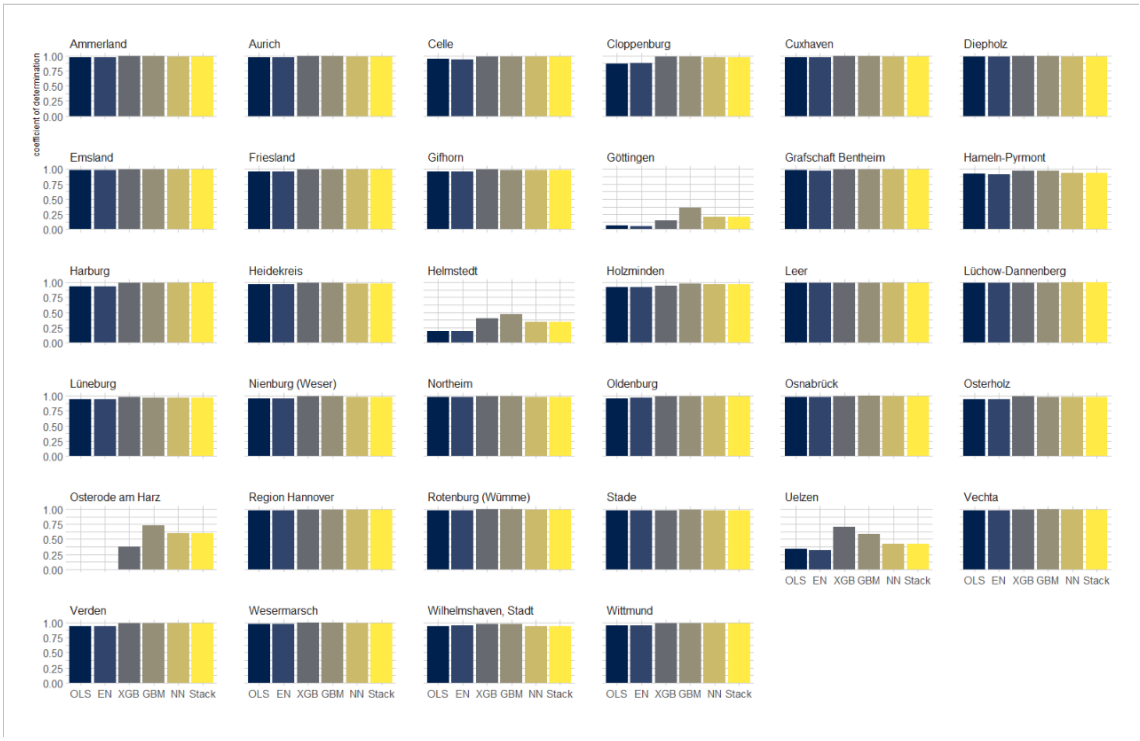

Figure 3.10: Coefficient of determination on county level

Another approach to compare the models focuses on the time required to train the algorithms. This is of particular interest for a possible future extension of the dataset to a larger number of farms or to a larger region, e.g. the whole of Germany. The differences between the non-linear approaches are considerable. The XGB algorithm is well ahead of the GBM, NN and stacking method with an average training time of about 7.1 minutes. The other approaches require about 3.8 (GBM) to 4.7 (NN) times longer training times. Stacking also takes substantially longer than the other methods, at around 84.9 min, because it adds the training times of the combined methods and its own execution time.

The use of the stacking method was originally intended to combine the advantages of the respective algorithms. For example, some algorithms tend to underestimate and others tend to overestimate the predicted values. However, such added

value from combining the previously used algorithms cannot be identified here as the R^2 of the stacking algorithm is not higher than the values of the previously combined algorithms themselves (see Figure 3.5).

Meanwhile, the training times of the linear models are much faster at about 0.2 minutes (OLS) and 0.4 minutes (EN).

3.6 Conclusions

The comparison of different ML methods with each other and with OLS regression provides valuable insights into their suitability for forecasting regional milk production and thus for assisting managers in the dairy industry and farmers in their operational and strategic decisions.

Four non-linear ML approaches (namely XGB, GBM, NN, stacking) are evaluated alongside two linear approaches (OLS, EN). While the R^2 for EN is almost the same as the OLS results, the non-linear ML algorithms demonstrate significantly superior accuracy in predicting monthly milk quantities per farm. This discrepancy is particularly pronounced when farms with high and average annual growth rates are considered separately. The R^2s of the non-linear algorithms are significantly higher than the results of the traditional OLS approach, especially for farms with above-average growth. It is evident that OLS predictions are often imprecise when predicting milk production for several farms at individual farm level, whereas ML methods tend to offer more reliable predictions. The R^2 for the non-linear ML algorithms are comparatively higher, regardless of the level of observation.

Among the ML algorithms, the GBM algorithm proves to be the most accurate both at the individual farm level and at the county level, as well as for the entire dataset. However, the XGB algorithm turns out to be the most efficient algorithm, considering its significantly shorter training time. While the R^2 of the non-linear algorithms differ only slightly, XGB requires only about a quarter of the time to train the models compared to the other algorithms. This parameter is particularly relevant as it is reasonable to expand the dataset with additional farms to improve accuracy, even in counties with a limited number of farm.

Furthermore, it is plausible to extend the model beyond the region of Lower Saxony in order to be able to predict milk production not only in Northern Germany, but in the whole of Germany. A temporal extension is also conceivable. While currently only short- and medium-term milk production is predicted, further research is needed to predict long term trends.

3.7 Appendix

	Option 1	**Option 2**	**Option 3**	**Option 4**
Learning rate (η)	0.01	0.05	0.1	0.3
Maximum depth of a tree	1	3	5	7
Minimum weights[a]	1	3	5	7
Subsample[b]	0.65	0.8	1	-
Colsample[c]	0.8	0.9	1	-

Table 3.2: Grid search parameters for fine-tuning the XGB and GBM algorithm

Note: [a]minimum sum of weights of all observations required in a child, [b]fraction of observations to be randomly samples for each tree, [c]fraction of columns to be subsampled

	Option 1	**Option 2**	**Option 3**
Neurons hidden layer I	200	100	50
Neurons hidden layer II	200	100	50
Neurons hidden layer III	200	100	50
Epochs[a]	100	50	25
Mini batch size[b]	34	32	16

Table 3.3: Grid search parameters for fine-tuning the NN algorithm

Note: [a]number of times to iterate the dataset, [b]value for the mini-batch size; smaller values lead to a better fit, while larger values can speed up and generalise better

Chapter 4

Estimating Seasonal Milk Supply Elasticities from Regional Climate Data

4.1 Introduction

Agricultural production is greatly determined by weather (Gisbert-Queral et al., 2021). Both average weather conditions and extreme events contribute to the annual volatility of crop yields, representing a significant threat to global crop production (Lobell et al., 2013). Among all economic sectors, agriculture is recognised as the most sensitive and vulnerable to the impacts of climate change (IPCC, 2014). Climate change has a direct impact on livestock production, as the performance and well-being of livestock are affected by rising temperatures and the increased occurrence of extreme weather events (Hill & Wall, 2015). In particular, heat waves have been identified as a major risk factor in dairy production, leading to reduced income for dairy farmers (Key & Sneeringer, 2014; Finger et al., 2018). Heat stress, which is determined not only by high temperatures but also by relative humidity, is particularly experienced by lactating cows during hot and humid summer days (Qi et al., 2015). Such weather conditions can lead to reduced performance, impaired animal comfort, increased mortality and forced culling (Mader et al., 2006; Megersa et al., 2014). These adverse effects can be directly attributed to weather conditions, while indirect effects result from changes in feed and pasture availability as well as changes in the distribution of pests and parasites (Gauly et al., 2013). As weather conditions continue to change, dairy farmers face reduced water availability, which in turn affects feed quantity and quality (Megersa et al., 2014).

While the direct effect of weather on dairy performance has been thoroughly investigated (Gisbert-Queral et al., 2021; Perez-Mendez et al., 2019; Qi et al., 2015, e.g.), there is a lack of information about the causal relationship between weather and dairy production as an indirect effect of varying forage quantity and quality. This chapter investigates the influence of seasonal weather on seasonal milk production in Northern Germany. Germany's milk production is the fifth largest in the world (FAOSTAT, 2023) and since 2000 has increased by about 16.7% (BLE, 2022). In the federal state of Lower Saxony in Northern Germany in particular, dairy production has increased by around 40% in the last 20 years and contributes up to a quarter of Germany's milk production (BLE, 2022).

The examination of farm data allows a more accurate depiction of management decisions and farmers' adaptation to different weather conditions. As milk production follows the biological processes of lactation curves (Wood, 1967), farmers can influence milk production by taking management decisions that have longer time lags (Hoehl & Hess, 2022). While various studies investigate yearly milk production and the impact of the weather (Perez-Mendez et al., 2019; Key & Sneeringer, 2014, e.g.), intra-seasonal variations are of particular interest here.

To assess this, a panel dataset with monthly farm-level data from Lower Saxony was used for analysing the influence of seasonal weather on seasonal milk production.

The chapter is organised as follows: Section 4.2 gives an overview of the causal relationship between weather, forage quality and quantity, and milk production, based on the available literature. In Section 4.3 the dataset is explained, while the applied methods are shown in Section 4.4. Finally, the results are discussed in Section 4.5 and the conclusions follow in the last section.

4.2 Conceptual Framework and Hypotheses

The milk yield of cows is determined by multiple factors, including reproduction, genetics, environment, management and nutrition (Bach et al., 2008). Nutrition in particular has been studied extensively in the scientific literature (Gaines, 1927; Thomas et al., 2001; Buza et al., 2014, e.g.). According to this literature, the metabolism of dairy cows aims to maintain the energy balance in order to provide energy for performance components such as milk production or lactation (Wangsness & Muller, 1981).

The most important factors influencing feed intake are firstly the cows' individual energy requirements and secondly the quality of feed, particularly with regard to the digestibility and degradation rate of forage (Van Soest, 2018). The regulation of feed intake depends not only on the feed intake itself, but also on the filling of the rumen, which is determined by the microbial fermentation and passage of the feed (Forbes, 1996).

In this context, various feed-intake models have been estimated in order to measure the influence of the most important parameters on milk yield. Jensen et al. (2015) compared five prediction models for the feed intake of dairy cows and found that live weight, milk yield and net energy lactation (NEL) content in forage are the most important parameters. The impact of forage quality has also been demonstrated in a review by Wangsness and Muller (1981), who mainly emphasise the quality of feed and the adjustment of forage according to the lactation curve. Furthermore, the addition of feed additives offers further potential for positively influencing milk yield. This is also evident in a meta-analysis by Hristov et al. (2005), who compared the influence of specific nutrients on milk yield and milk protein yield using 563 different diets. The comparison of these diets confirmed the relationship between feed intake and milk production, as well as the influence of different additives. It is therefore evident how important the organisation of feed management is for dairy farms. For example, the difference between maize silage and ryegrass

silage is that feed intake, milk yield, and milk protein content are significantly higher when feeding maize silage (Baldinger et al., 2014). Likewise, feeding legume silage has a significantly positive effect compared with feeding grass silage in terms of feed intake, milk production and milk ingredients (Steinshamn, 2010). In a comparison of an indoor feeding system with a grazing system that allows for a high daily intake of fresh grass in Ireland, Kennedy et al. (2005) found no difference in the quantity of produced milk over a seven-week trial period in early spring, but slight advantages of the grazing system over the indoor system in terms of milk protein content. These studies show that both feed quality and feed management are important in milk production.

The quality of the respective types of forage is mainly determined by weather conditions since this has an impact on both the quantity and quality of the crop yield (Lesk et al., 2016). Primarily weather conditions during the growth phase affect the yields of different crops. For instance, Schlenker and Roberts (2009) estimated the relationship between yields of corn, soybeans and cotton based on daily weather data during the growing season. Based on their flexible nonlinear model, they showed that yields increase up to a temperature threshold of 29 °C (corn) or 30 °C (soy). Above this boundary, negative effects can occur that are estimated to be very harmful. This has also been confirmed by Eck et al. (2020), who studied the effect of temperature and precipitation during the growing season on crop yields in the south-eastern United States. They discovered that high temperatures, particularly in July and August, can lead to yield losses. The impact of precipitation during the growing season is heavily dependent on crop type and local geography. For corn and soy, they estimated significant positive correlations. Another study by Zscheischler et al. (2017) modelled the weather-yield relationship using a linear model based on bivariate return periods of specific climate conditions. They showed that yields increase along a gradient from hot and dry to cold and wet weather. Especially in summer, most crops are sensitive to warm and hot conditions. These relationships can also be observed with grassland quality, as grassland productivity is strongly correlated with annual rainfall, as well as with summed annual temperature (Smit et al., 2008).

The causal chain of milk production with the effect of weather conditions on animals and on grassland and forage quality formed the basis for the analysis of the causal relationship between seasonal weather conditions and milk production during the year. The conceptual causal framework of seasonal milk production is thus presented in Figure 4.1.

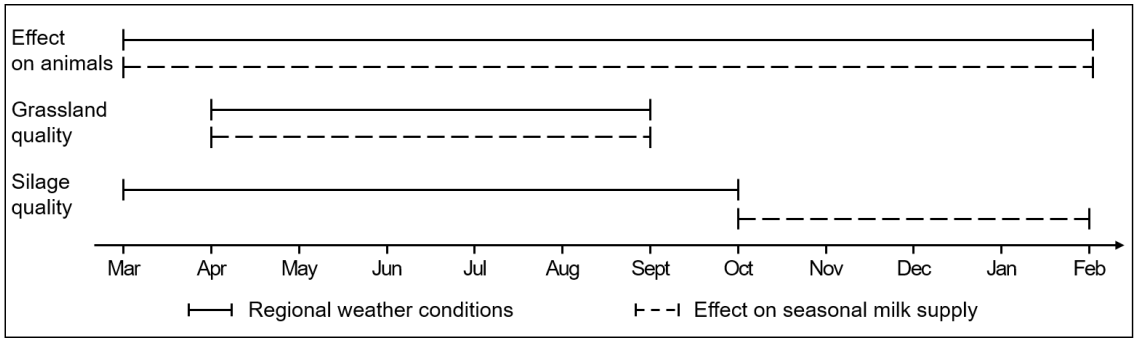

Figure 4.1: Conceptual causal structure of seasonal milk production

It is expected that the weather conditions during the vegetation season affect milk production through the quality of fresh grass and of silage. However, the effect of silage quality on production will only become evident after a time lag of several months because silage requires several weeks for the fermentation process (Weinberg & Muck, 1996). In addition, silage from the first cut is not typically used before the end of the vegetation period in autumn (Cherney & Picasso Risso, 2020).

In order to investigate the impact of weather parameters on milk production, a multiple linear regression model was used with milk production as the dependent variable. Milk volume is a function of various economic decisions and climatic conditions. To illustrate this correlation, operational decisions were represented by using the milk price $p_{t,q}$, and a trend variable δ_t, as well as the number of cows $q_{it,q}$. Milk quantity is particularly conditioned by individual farm performance; hence a fixed-effects model was used that corrected the regression for the individual part of each farm. Additionally, variables representing the quality characteristics of milk $V_{it,q}$ were included, illustrating urea content and somatic cell count.

This resulted in the general model of milk quantities $y_{it,q}$ produced by farm i in quarter q as a sum of 3 following months starting in March, for $i = 1, \ldots, N$ farms in the sample and $q = Mar - May, Jun - Aug, Sep - Nov, Dec - Feb$; a subscript t indicated the use of a multi-year sample:

$$y_{it,q} = f(c_{it,q-k},\ r_{it,q-k},\ q_{it,q},\ p_{t,q},\ V_{it,q},\ \delta_t) \tag{35}$$

Direct weather effects are represented by the THI $c_{it,q}$ calculated according to Mader et al. (2006). Indirect weather effects can be attributed to the variable Standardised Precipitation-Evapotranspiration Index (SPEI) $r_{it,q-k}$ based on the Penman-Monteith equation (Allen et al., 1998) as an indicator of the climatic water balance regarding 3 preceding months. The SPEI variable entered the model in the form of time lags $k = \{1, 2, 3\}$.

4.3 Data

To model the causal weather-milk relationship, a dataset with monthly milk data and both daily and monthly weather data for the period 01/2012 to 12/2018 was provided by the Lower Saxony State Inspection Association, LKV, and Germany's National Meteorological Service, DWD. The observations thus fell roughly within the time period when the abolition of the milk quota had already been agreed by policy makers and was prepared for through an annual increase in quota volumes ('soft landing'). In addition to data for milk production and weather conditions, trend and price variables were included into the model to represent management decisions based on these factors.

The data of milk production quantities are collected by the LKV on a monthly basis as part of regular monthly on-farm control of milk quantity and quality control. Data from a random sample of 291 milk producers from Lower Saxony were used for this analysis. As milk on farms is tested in 11 out of 12 months, one monthly data value per farm and year is missing. This missing value was imputed by a linear procedure, resulting in a balanced panel with the dimension of $N \times T$ where $N = 291$ farms and $T = 84$ months. This provided the average amount of milk produced per day for each individual farm. It was then multiplied by 30.5 to give the average monthly milk volume. Finally, in order to reflect the respective fat and protein content within the estimation, the ECM, according to Sjaunja (1990), was used as the dependent variable.

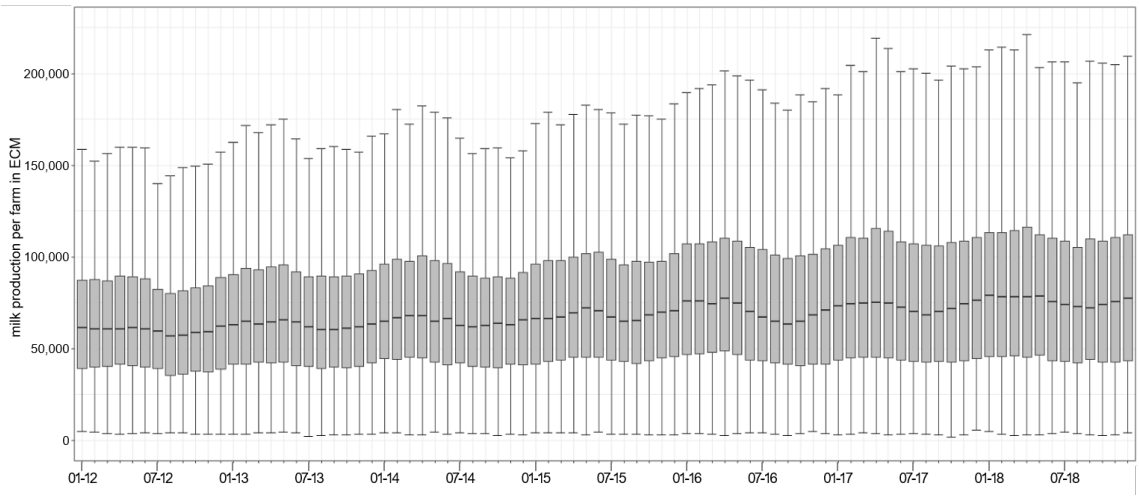

Figure 4.2: Boxplots of the monthly milk production of the $N = 291$ farms during observed period

Figure 4.2 shows the computation of the dependent variable in the dataset during the observed period. Typical fluctuations in milk volume are evident throughout the

year, with the annual peak typically in May and the lowest yield often in September. A steady increase of 28.7% in average milk production per farm from 67,910 ECM (2012) to about 87,398 ECM (2018) can be identified. The boxplots illustrate the different sizes of farms, which continued to increase over the observation period.

The DWD provides data on various weather variables with a spatial resolution of 6 km x 6 km. These data originate from the so-called 'COSMO regional reanalysis'. This is a combined numerical model of observations and model-based analysis that maps regional weather and climatic conditions onto a spatial grid, so that the data can be assigned to the coordinates of the observed farms. The THI is calculated using daily temperature and relative humidity values and then summed up for the respective quarter. The SPEI calculation uses monthly values of the minimum temperature, maximum temperature, wind speed, cloud coverage and precipitation. The monthly SPEI values are then summed up on a quarterly level.

Both indices are shown in Figure 4.3 during the observed period on a monthly basis. It can be seen that the THI follows wavelike movements during the year, with the maximum usually reached between July and September and the minimum in the period from January to March. In contrast, the values for the SPEI were more volatile during the observed period. No definite seasonal course can be identified.

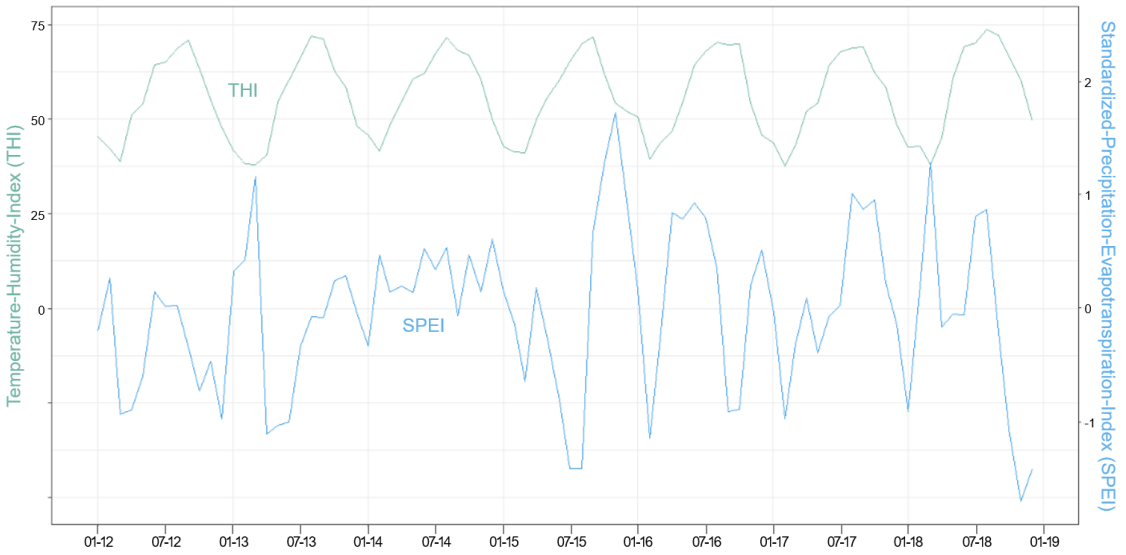

Figure 4.3: Weather conditions during the observed period

Besides weather conditions, milk production depends on various management decisions. These are mostly based on the milk price paid and were therefore included in the model. Farm gate raw milk prices were provided by the German Federal Ministry of Agriculture and Food, BMEL, and correspond to the price of conventionally produced cow's milk in Lower Saxony at actual fat and protein content.

4.4 Methodology

Based on the data described in the previous section, the model from Equation 35 was estimated as a FE panel regression, according to the following regression equation:

$$log(y_{it,q}) = c_{it,q} + r_{it,q-k} + log(q_{it,q}) + log(p_{t,q}) + log(V_{it,q}) + \delta_t + \alpha_i + \varepsilon_{it,q} \quad (36)$$

where $y_{it,q}$ = average milk production [ECM] per farm i in quarter q and year t; $c_{it,q}$ = average THI in quarter q, year t and farm i; $r_{it,q-k}$ = average SPEI in quarter $q-k$ with $k = 1, 2, 3$ and year t; $q_{it,q}$ = average number of cows kept; $p_{t,q}$ = average milk price in quarter q and year t; $V_{it,q}$ = vector of quality characteristics including urea content and cell number; δ_t = trend variable for year t; α_i = fixed effects for individual farm i and $\varepsilon_{it,m}$ = error term. For each quarter q, a separate FE model was estimated, so that each quarter could be modelled individually based on the variables.

In contrast to the milk and weather variables, the milk price variable varied only in time and not for individual farms. As the trend variable, a variable δ_t for the respective year t was used in order to be able to account for macroeconomic changes in addition to individual development of herd sizes and milk yield. By adding farm-level fixed effects (α_i), the estimation could be controlled for farm-specific and time-invariant factors such as productivity (Verbeek, 2012). The function was expressed in a log-log and log-linear transformation to present the regression coefficients of the continuous variables as elasticities (Gujarati & Porter, 2009).

For the data presented here, the FE estimator showed significant differences compared with the random-effects estimator under the usual Hausman test, therefore only the FE model was considered to be unbiased (Hausman, 1978). The Breusch-Pagan test for each model indicated the presence of inconsistent error terms. Therefore, to avoid heteroscedasticity, the standard errors were clustered according to White (1980) on an individual farm level using the library 'sandwich' for panel data (Zeileis, 2006). The FE models were estimated using the 'plm' library (Croissant & Millo, 2008).

4.5 Results and Discussion

The four separately estimated models for the respective quarters have similar R^2 as shown in Figure 4.4. The adjusted R^2 is highest in spring (March, April, May) with 0.7549. The autumn quarter (September, October, November) can be explained worst by the model and has the lowest adjusted R^2 of 0.6941. Due to the seasonal

course of the lactation curve, especially many animals stand dry in autumn. A varying milk production of the available animals therefore has a relatively stronger effect on the milk quantity of the entire farm at this time of year and is therefore less well explained by the model.

Figure 4.4 shows the regression results for each quarter in each column. The rows show the estimated coefficients of the corresponding dependent variables. The estimated coefficients are presented first in each box and in the column behind the corresponding p-values. With a significance level of $\alpha = 0.95$, the Null-Hypothesis of an estimated coefficient being equal to zero was rejected for all p-values < 0.05. These coefficients are shown in bold in the Figure 4.4.

According to the theoretical framework as shown in Figure 4.1, one would have expected the estimated coefficients of the trend variable to be significantly positive in all quarters, as there is a clear growth trend over the years. However, contrary to expectations, this variable is only significantly positive in the autumn. It can therefore be concluded that the increase in milk production is not consistently due to an increase in milk yield per animal, but is positively influenced in particular by the number of animals kept.

The estimated coefficients for the variable 'number of cows' are significantly positive in all quarters. This shows that the farm management decisions have a significant impact on the milk yield of the whole farm. A comparison of the coefficients shows that the greatest influence on milk production per farm can be observed in the winter and spring quarters. In these months, according to the seasonal lactation curve, a particularly large number of animals are recently lactating, so that the milk quantity in this period is relatively higher than in the other quarters. The total milk yield can therefore be influenced most by the number of cows kept in these quarters. Conversely, the influence of this variable is smallest in the autumn quarter, as there are relatively fewer animals lactating, in line with seasonal milk production as described by Timlin et al. (2021) before. Thus, a one per cent change in the number of livestock leads to a change in milk production of less than one per cent (0.986). For the remaining quarters, on the other hand, the rate of change or elasticity is just over one per cent.

Autumn is also the only quarter in which the milk price has a significant effect on milk yield. The positive estimated coefficient indicates that farmers react to the market price in this quarter and make operational decisions on feed management in such a way that animals produce more milk in the autumn and thus a higher milk price leads to a higher milk quantity. In particular, autumn is the lowest feed cost period of the year due to a fresh fodder harvest and relatively low feed prices

compared to the rest of the year. It therefore makes sense for farmers to adjust feed management to increase milk yield during this period, when variable costs are lower.

y=log(monthly milk production)	Q1/Spring coef	p	Q2/Summer coef	p	Q3/Autumn coef	p	Q4/Winter coef	p
Trend variable	-0.001	0.346	-0.001	0.428	**0.004**	**0.015**	0.000	0.771
log(number of cows)	**1.013**	**0.000**	**1.004**	**0.000**	**0.986**	**0.000**	**1.017**	**0.000**
log(urea content)	0.012	0.429	**-0.033**	**0.021**	**-0.040**	**0.012**	-0.010	0.540
log(somatic cell count)	**-0.022**	**0.000**	**-0.024**	**0.000**	-0.011	0.134	**-0.015**	**0.028**
log(milk price)	-0.012	0.348	0.013	0.160	**0.046**	**0.000**	0.002	0.880
SPEI Q1	0.001	0.511	0.001	0.298	0.001	0.417	0.000	0.825
SPEI Q2	0.000	0.990	0.000	0.616	0.002	0.061	0.000	0.915
SPEI Q3	0.000	0.937	0.002	0.065	**-0.002**	**0.028**	**-0.004**	**0.000**
SPEI Q4	0.000	0.883	0.001	0.227	0.001	0.253	**0.002**	**0.025**
THI	**0.037**	**0.005**	**0.079**	**0.049**	0.001	0.984	**0.037**	**0.004**
adj. R^2	0.7549		0.7389		0.6941		0.7456	

Figure 4.4: Estimated coefficients of the regression models

Note: Colours correspond to seasons (spring/Q1/Mar–May = green, summer/Q2/Jun–Aug = yellow, autumn/Q3/Sep–Nov = orange, winter/Q4/Dec–Feb = blue; p-values$<$0.05 and their corresponding coefficients are printed in bold with a grey background)

The urea content gives an indication of the protein supply to the microbes in the rumen or cow. A high urea content indicates an excess of degraded feed protein and a lack of energy for microbial protein formation in the rumen (Gustafsson & Carlsson, 1993). These high levels typically occur in late summer and autumn on grass-based farms, as the grass of later cuts is often high in protein but low in energy. This effect also has a negative impact on milk yield and can also be represented in this model. Thus, a significant negative coefficient was estimated for the summer and autumn quarters, representing exactly this correlation, namely that high urea contents in summer and autumn lead to lower milk yield during this time.

Another component considered is the somatic cell count of the milk. The somatic cell count of the milk can be used to assess the health of the udder. If the somatic cell count is increased, a reduced milk yield can be expected (Hagnestam-Nielsen et al., 2009). This correlation can be confirmed in the results (see Figure 4.4). A significant negative coefficient was estimated in almost all quarters except autumn. Accordingly, an increase in cell count leads to a reduced milk yield. A comparison of the coefficients shows that they are highest in the warmer quarters (spring and summer). This can be explained by an increased THI at this time, which creates optimal conditions for pathogen growth. This increases the infection pressure and

therefore the risk of clinical udder infections and increased somatic cell counts.

The focus of this analysis was on the correlations between direct and indirect weather effects on seasonal milk production. With regard to direct weather effects, a positive significant coefficient was estimated for the variable THI in all quarters except autumn. This shows that milk yield can be increased by a drier and warmer climate. This seems to contradict previous findings by e.g. Mader et al. (2006) who found a negative effect of heat stress on dairy cows. However, the difference with this chapter is that here average values are presented rather than extreme values. It can therefore be seen that there is a positive influence with increasing THI. However, it is not possible to show the boundary of this positive influence by using a linear regression model. According to the seasonal weather conditions, it is also clear that the positive influence of the THI is highest in summer. This can be explained by the fact that, on average, the THI is highest in this quarter of the year and can therefore have the greatest influence. In the autumn quarter, on the other hand, there is no effect from the THI. It can therefore be assumed that other effects overlap the effects of the THI in this quarter.

While the direct weather influences only have an effect in the respective quarters, the SPEI is considered as an indirect influence with $0-3$ quarters, i.e. with $0-9$ months lag. For this reason, individual variables have been included in the model for each quarter. It is particularly striking that the SPEI has no significant effect in the spring and summer quarters. Neither in the immediate months nor in the lagged months does the drought index have an effect on the milk production of the farms. Only the summer quarter shows a slightly significant positive coefficient (with $p = 0.065$), indicating that a higher SPEI has a positive effect on milk production. It should be noted that the SPEI is on a scale from $< -2.33 = \textit{very dry}$ to $> 2.33 = \textit{very wet}$. A higher SPEI therefore corresponds to wetter weather. Especially the SPEI in autumn seems to have a significant influence on the seasonal milk production. A significant negative coefficient was estimated for milk production in autumn itself as well as in winter. According to the estimated model, drier weather in autumn has a positive effect on milk production. This correlation could be due to the fact that grassland quality is less relevant in these months but on farms with maize silage, for example, this is harvested and fed fresh. Schlenker and Roberts (2009) have already shown that a drier climate has a positive effect on maize yields. As a result, a positive effect on milk yield can also be inferred. Another positively significant coefficient was estimated for the influence of winter SPEI on winter milk production. Here it can be seen that the wetter the winter, the more milk is produced.

The lack of representation of the effect of growing season weather on milk production over the course of the year in the model may have several reasons. One could be that mainly extreme droughts affect milk production negatively (Roche et al., 2009). In this case, however, only average values are included in the model and the influence of extreme weather conditions is not considered. Furthermore, farms may be able to adapt to changing weather conditions and thus avoid negative effects. Potential strategies to address the adverse effects of weather variability on forage and dairy production may include a variety of measures such as changing feeding practices, implementing climate-adapted pasture management techniques or using other mitigation strategies (Hopkins & Del Prado, 2007).

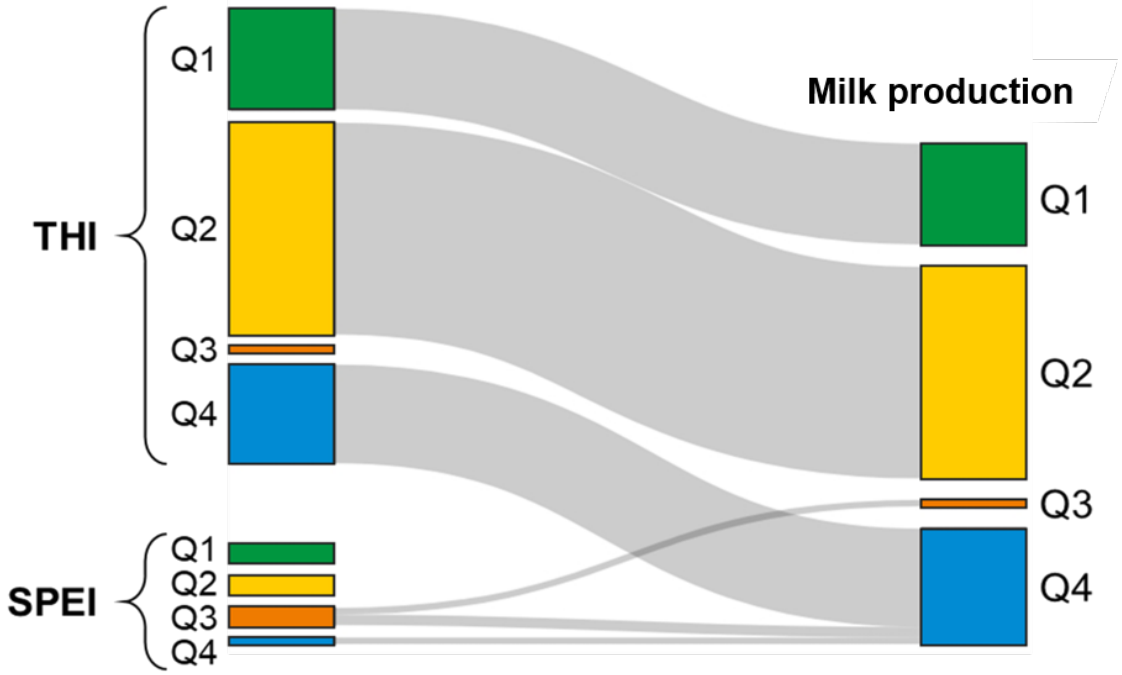

Figure 4.5: Sankey diagram for significant weather effects ($\alpha = 0.95$)

A comparison of the size of the coefficients for each variable shows a clear difference between the groups of variables. The estimated coefficients for the number of animals are significantly higher than all the other coefficients. It is therefore clear that the milk production of a farm can be controlled in particular by the number of animals. The other variables have less influence. A comparison of the direct and indirect influence of the weather also shows a clear difference between the two groups. To illustrate this and to show the seasonal influences over the course of the year, the weather influences are shown graphically in a Sankey diagram in Figure 4.5. The Sankey diagram shows in each case the significant effects of weather conditions in quarter $q-k$ on milk production in quarter q when considering a significance level of

$\alpha = 0.95$. The different sizes of the bars demonstrate the corresponding sizes of the estimated coefficients. It can be noted that the THI has both a stronger and more frequent significant influence on milk production than the SPEI. In addition, it is again shown graphically that the spring and summer quarters have a less significant influence on milk production than originally assumed. Instead, weather conditions in autumn and winter have a significant effect on milk production.

The coefficients shown here can be interpreted as supply elasticities. For example, if the number of animals in spring changes by 1%, the quantity of milk produced changes by 1.013%. To illustrate these effects, an example farm with all the averages of the farms considered is presented here. This results in a farm with an average of about 101 cows and an average milk yield of about 8,850 litres per animal per year. The resulting average seasonal milk production is shown over the quarters in Figure 4.6. The effect of an increase of one unit in the THI over the whole year is entered as a proxy for the other independent variables considered. For the spring quarter, this results in an increase in milk volume of about 1,032 litres, which would correspond to an additional turnover of €516 at a milk price of, for example, 50 cents. Over an average year, the same conditions would result in an additional turnover of around €2,035. Another example is the effect of an 1% change in the number of somatic cells. If this were to increase by 1% over the course of the year, this would result in a loss of approximately €777 for the same example farm.

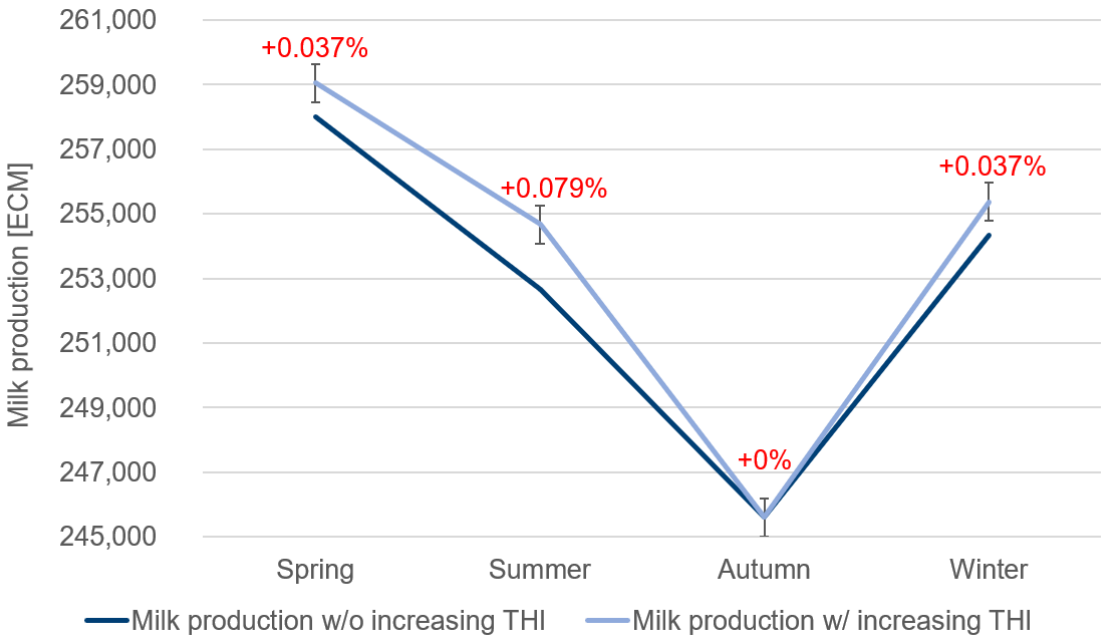

Figure 4.6: Effect of an one unit increase in THI

Note: Sample farm is characterised by 101 cows and 8,850 litres of milk yield.

4.6 Conclusions

Seasonal milk production in Lower Saxony is influenced by several factors. While the research question of this chapter was primarily aimed at the influence of seasonal weather conditions, it was found that the number of animals on a farm has the greatest influence on the milk production of the farms. In particular, the indirect influence of weather conditions through varying feed quality and quantity throughout the year cannot be clearly confirmed by the results. It can be concluded that dairy farms with successful feed management also cope well with difficult weather conditions. The advantage of these farms over the other farms is constant regardless of the seasonal weather.

The direct effect, represented here by the THI, can be shown to be significant in almost all quarters except autumn. At the same time, this effect illustrates the limits of examining the relationship using a linear regression. It can be assumed that above a certain limit of the THI, heat stress occurs for the animals and accordingly a negative influence of the variable THI on milk yield would be expected. Therefore, future studies could additionally investigate the research questions using non-linear approaches as well as extreme weather events and their effect on seasonal milk production in particular. This is also of current relevance in view of the advancing climate change.

Economic factors such as the milk price paid appear to have only a strategic and rarely an operational influence, as milk yield is only positively affected by the milk price in the autumn quarter. Farm managers should also ensure that udder health and the protein-energy ratio of the feed rations are under control at all times of the year, otherwise there will be a negative impact on farm profits.

Overall, understanding the complex interactions of different factors influencing seasonal milk production is crucial for farm management and further research is needed to account for non-linear relationships and the effects of extreme weather events.

Chapter 5

Analysing Raw Milk Sales from Vending Machines in Germany: An Application of the XGB Algorithm

5.1 Introduction

The last 20 years have seen a considerable decline in the number of agricultural farms in Germany (Destatis, 2023d, 2023c). This structural change is particularly evident in the dairy production sector, where there has been a 58% fall in the number of dairy farms between 2001 and 2020 (Destatis, 2023g). There are many reasons for this ongoing structural change (Schaper et al., 2011). In particular, a lack of financial competitiveness is leading to farms being forced out of business (Bragg & Dalton, 2004). While there has been a significant increase in milk price since the end of 2021, there have been repeated low-price phases over the last 10 years (BMEL, 2023a). Moreover, the average price of 33.7 ct per litre (in the period 2011-2021) is only a fraction of the price charged by food retailers, which could be an indication of retailers' oligopsony power in the German dairy value chain (Grau & Hockmann, 2018).

Particularly since the abolition of the milk quota in 2015, a growing number of farmers have been trying to build up a second financial pillar by selling part of their milk production directly to consumers through so-called raw milk filling stations. This specific form of marketing is strictly regulated by Germany's Raw Milk Quality Regulation (*Rohmilchgüteverordnung*) as it involves the sale of unprocessed raw milk, which may cause health issues unless consumers heat it once above 100 degrees Celsius. Nevertheless, this distribution channel gives farmers a low-threshold entry into direct marketing for comparatively low investment costs, and could thus represent a possible second financial pillar for dairy farms.

Raw milk filling stations could attract considerable interest from consumers. The steadily decreasing number of farms has reduced the amount of contact between consumers and producers, leading some consumers to feel disconnected from agricultural production (La Trobe & Acott, 2000). However, there is a desire among many consumers for short value chains (Jarzębowski et al., 2020), as evidenced by several German consumers' preference for locally produced food, transparent and traceable food supply chains, and certified animal welfare (BMEL, 2018). This demand is partly accompanied by a greater WTP a premium for such products (Emberger-Klein et al., 2016).

Direct marketing of raw milk as a second financial pillar could therefore represent an option for some farms. However, the type of farms for which the commissioning of a raw milk vending machine appears economically reasonable is unclear. Some vending machine operators complain about the high maintenance involved in return for relatively low sales volumes. This raises the research question as to how to predict sales of raw milk from vending machines. Farmers who are considering acquiring a

raw milk vending machine could use this information to evaluate their investment decision more effectively. Additionally, the influence of various marketing measures and location factors on the prediction result is of interest so that the drivers of this market can be better understood.

For this analysis, a cross-sectional survey about milk vending machines was conducted throughout Germany and supervised ML in combination with SHAP values were used to consider the importance of individual factors in the prediction of raw milk sales (Lundberg & Lee, 2017). The use of this methodology offers the advantage of accurate prediction and applies explainable ML to highlight the origin of the results, making it more favourable than traditional econometric approaches in this case. In order to demonstrate this advantage, an OLS estimator is applied as a benchmark model.

The chapter is organised as follows: Section 5.2 describes the conceptual framework underlying the modelling of raw milk sales from vending machines, Section 5.3 explains the data collection procedure and Section 5.4 the methodology. Section 5.5 presents the results, which are then discussed in terms of limitations of this modelling approach. Finally, conclusions and recommendations are provided for farmers, policy makers and food safety authorities.

5.2 Conceptual Framework

Influential factors describing the success of agricultural direct marketing have been presented in numerous studies in the literature (Pottebaum, 1989; Kuhnert & Wirthgen, 1997; Wirthgen & Maurer, 2000; Knuck, 2020, e.g.). However, it is unclear whether these factors are also relevant for the sale of raw milk from farms through milk vending machines. The sales volume of raw milk filling stations is influenced by various supply and demand factors, which are shown schematically in an extended Stimulus-Organism-Response (SOR) model (see Figure 5.1).

The basis for modelling the consumer's purchase decision is a SOR model, which was developed by Woodworth (1929) as a modification of the classical stimulus-response theory of Pavlov (1927). It consists of three components – stimulus, organism and response – and attempts to explain consumer behaviour by considering environmental influences (Woodworth, 1929). This model, which has been used to analyse consumer decisions and behaviour (Hempel & Hamm, 2016; Sultan et al., 2021, e.g.), has not yet been applied to the raw milk market.

The theoretical structure of the SOR model remains the same for various products, although the design of the individual phases varies depending on the product and the consumer (Kotler et al., 2007). In this case, Figure 5.1 shows the connec-

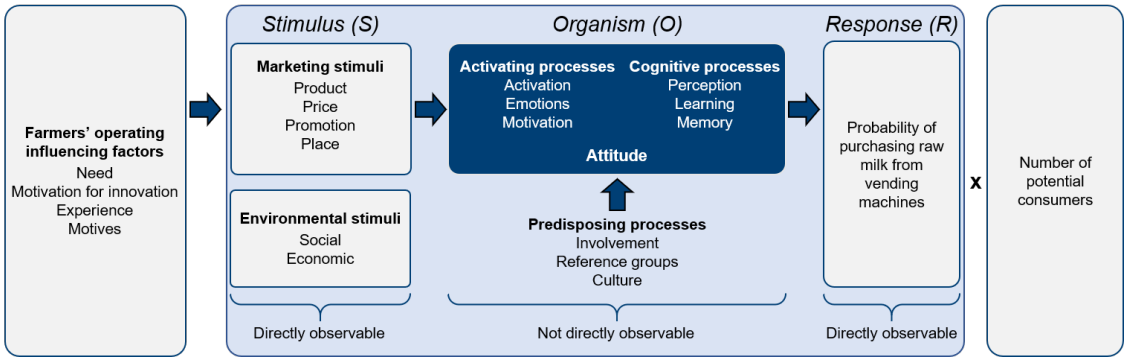

Figure 5.1: Extended SOR model for raw milk sales from vending machines

Source: Own illustration based on Kotler et al. (2007)

tions for the purchase of raw milk at vending machines (R_{ik}), which is influenced by various stimuli. A stimulus is defined as an outside force that surrounds and influences an individual in its psychological operation (Young & Young, 2016). A distinction is made between controlled (=marketing) (S_{Mik}) and uncontrolled (=environmental) stimuli (S_{Eik}) (Kroeber-Riel & Weinberg, 2019). The relationships of the SOR model can be described as follows in Equation 37:

$$R_{ik} = g(S_{Mk},\ S_{Eik},\ O_{ik}) \tag{37}$$

for consumer i with $i = 1, \ldots, N$ at vending machine k with $k = 1, \ldots, K$. The processes of the individual organism (O_{ik}) are not directly observable, as shown in Figure 5.1, since they are active and cognitive processes as well as attitudes, which vary individually per consumer and thus are difficult to measure (Kotler et al., 2007).

Marketing stimuli result from operational decisions, where operators of raw milk filling stations attempt to attract consumers' attention (Kroeber-Riel & Weinberg, 2019). A marketing mix of product, price, promotion, and place helps generate potential buying interest (Meffert et al., 2018). For the sale of raw milk through vending machines, the farmer can only differentiate the product by its production method, as only the raw milk product is considered here. Therefore, the impact of organic production on the marketing success of raw milk was investigated. A success factor study by Mellin et al. (2006) shows that a high level of customer satisfaction is determined not only by the quality of the products offered, but above all by the atmosphere of the sales premises. When implementing this form of marketing, operators take different approaches, starting with the choice of vending machine manufacturer. Despite offering similar functions, the models of various manufactur-

ers differ significantly in price. In addition, producers can offer other products to make the main product, raw milk, more attractive. Wille et al. (2018) revealed that expanding the product range, for example by including eggs, cheese or sausages in the product range, could be a success factor in direct agricultural marketing. A wider range of products can appeal to a larger group of potential customers and contribute to customer satisfaction as they can meet other needs besides the purchase of raw milk without additional effort. Individual products are sold either through trust-based cash registers or from farm shops or other vending machines. Vending machines offer the advantage that refrigerated products, such as processed dairy products or meat products, can also be sold. Another benefit of the vending machines is that they can be open 24/7. However, as there have been thefts from vending machines, some farmers have abandoned all-day opening.

Pricing for the sale of raw milk is limited to price per litre. Price differentiation for larger purchase quantities is not technically possible with current vending machine models.

Various studies have emphasised the role of marketing measures in direct agricultural marketing (Pottebaum, 1989; Wirthgen & Maurer, 2000). For advertising to be successful, it has to attract attention (Pepels, 1994). Individual advertising campaigns, such as flyers, newspaper advertisements, and online advertising, differ significantly in their reach and efficiency. Especially in areas where the clientele are more likely to be commuters, tourists or random shoppers, it is important to attract attention with outdoor advertising in the form of street signs (Lopez-Pumarejo & Bassell, 2009). Social media applications are also increasingly influencing consumers' purchasing decisions (Horstmann & Schulze, 2011). Some agricultural direct marketers use the internet to publicise their farms and products to a wider audience via their website or Facebook page. By continuously providing photographs and up-to-date information, they attempt to gain benefits indirectly through social identification and perceived value (S. Chen & Lin, 2019). Thus, when investigating raw milk sales, various marketing measures were considered.

The installation of a raw milk vending machine is governed by a complex set of regulations that are subject to monitoring and enforcement by district veterinarians. In Germany, these district veterinarians are in charge of food safety issues at farm level. Hygiene concerns are the main reason for authorities not permitting these machines to be positioned a long way from the barns, with distance used as a proxy for food safety (§ 17 (4) Tier-LMHV, German Animal Food Hygiene Regulation). Consequently, the farmer is unable to influence the location where raw milk is sold. Therefore, it cannot be considered a factor through which farmers can introduce

varying stimuli, but it is still an influencing factor for raw milk sales and was considered for prediction purposes. Müller-Hagedorn et al. (2012) describe the central role of accessibility in direct agricultural marketing. Farms that are located near a busy road have a large number of potential customers in their vicinity every day due to high traffic volumes. In particular, commuter locations have an increased likelihood of consumers calling in to purchase raw milk on their way to or from work. In addition to regular shoppers, irregular consumers may also stop at the raw milk vending machine to make a purchase while on holiday or a weekend trip. Being located near a city can have a positive effect on the likelihood of purchases from raw milk vending machines as urban consumers could be more willing to buy fresh products directly from the producer.

However, due to direct competition with food retailing, substitution effects could occur that have a negative impact on vending machine sales. Food retailers have also recognised the trend for regional products and are increasingly offering processed and conveniently packaged milk from local farms because it corresponds with consumer demand (Feldmann & Hamm, 2015). Furthermore, retailers can offer consumers a full range of food items, which a dairy farm usually cannot. Thus, the distance to the nearest supermarket is also considered. Direct competition can also arise from other vending machines in the vicinity of vending machine operators. Alternatively, up to a certain density of vending machines, there can also be a positive effect if the perception of this type of direct marketing is improved by an increased number of vending machines in the consumers' immediate area.

Consumers are also influenced by environmental stimuli (S_{Eik}), which are mainly socio-economic. In particular, the social experiential environment that arises from direct contact with fellow human beings in a region is considered an environmental stimulus (Kroeber-Riel & Weinberg, 2019). To represent these drivers, various proxies were used, e.g. per capita income, which provides information about the prevailing purchasing power and consumption habits. The variables 'rurality', 'proportion employed in the agricultural sector' and 'number of dairy farms' stand for the consumers' relationship with agriculture. While contact with agricultural producers tends to be less frequent in urban regions, contact points between farmers and consumers are more likely in more rural areas. In particular, the relationship with the producer of raw milk could be stronger in a region with a higher density of dairy cattle than in a region with fewer dairy farms. Other environmental influences are the average age of the population and the proportion of university graduates in order to take any educational influence into account.

While the location stimuli are predetermined by the location of the farm and

thus the location of the vending machine, marketing stimuli can be influenced by the farmer. Therefore, the SOR model according to Kotler et al. (2007) is extended to include the farmer's influencing factors as an additional component (see Figure 5.1). This results in a mix of need (F_{Nk}), motivation (F_{Mk}), experience (F_{Ek}), and motives (F_{Rk}) that influence the farmer to undertake different activities. For instance, the need for farmers to establish a second financial pillar varies depending on their current economic situation. This is defined here by the number of cows, average milk yield, the milk price paid, and the lease fee customary in the region. Moreover, a distinction can be made according to the dairy farms' degree of specialisation, thus the share of milk production out of total farm revenue was also examined.

The age of the farm manager can also have an impact on marketing stimuli, as younger farmers tend to be more able to develop entrepreneurial skills and more open to new developments in farm activities (De Wolf et al., 2007). In addition, experience using the vending machine may build up over time, which can have a positive effect on marketing stimuli, as measured by the year in which the raw milk vending machine was commissioned. The motives for commissioning a raw milk filling station can be financial, as shown above, but other reasons can also be relevant, such as maintaining the image of agricultural production and public relations work for the farm. A distinction is therefore made between purely financial motives, purely image-related motives, and the indication of both. The relationship between marketing stimuli and farm structures can therefore be represented by the following equation:

$$S_{Mk} = h(F_{Nk},\ F_{Mk},\ F_{Ek},\ F_{Rk}) \tag{38}$$

As a result of the SOR model, the response refers to the purchase decision of an individual consumer. Depending on the circumstances of the farm, marketing and environmental stimuli, and the individual organism, there is a certain probability that the purchase decision for raw milk will be positive for the consumer. The location of the raw milk vending machine gives the farmer a potential addressable market based on population density (C_k). The classical SOR model can therefore be extended by this additional factor. Accordingly, the sales (M_k) of a farm k are determined by the following equation:

$$M_k = f(g(h(F_{Nk},\ F_{Mk},\ F_{Ek},F_{Rk}),\ S_{Eik},\ O_{ik}), C_k) \tag{39}$$

5.3 Data Collection

The empirical analysis used operational and personal data for German farmers who market raw milk through vending machines. The operational data were obtained in an anonymous survey. After a pre-test, the survey commenced in April 2020. According to the district veterinary offices, 850 milk filling stations were in operation in Germany at the time of the survey. Approximately 72% of farmers with a raw milk vending machine could be contacted by telephone, email or post. The addresses of the remaining vending machine operators could not be obtained for data protection reasons. From this number of contacted farms, 154 farmers expressed interest in completing the survey, thus a random distribution of observations within Germany can be assumed. The distribution of these farms is shown in Figure 5.2, which also illustrates the sales volume of raw milk in 2019 by the size of the blue circles. The remaining locations of raw milk filling stations for which the address was available are shown as red dots. This reveals a fairly close correspondence between the distribution of milk filling stations and the distribution of survey locations. The interviews took place in the form of personal telephone interviews (37%) as well as by sending interview forms via email and post (63%).

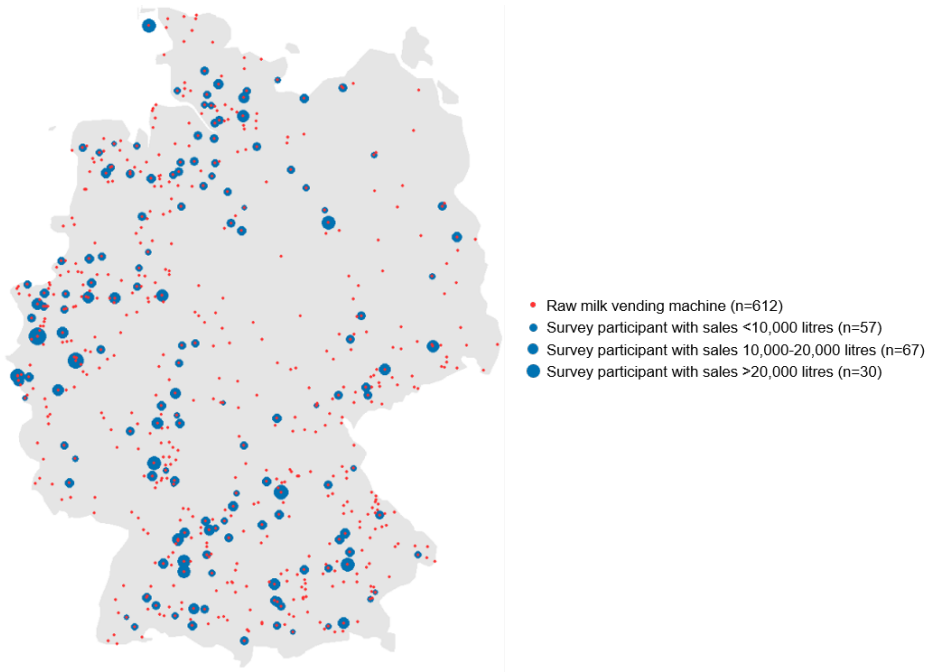

Figure 5.2: Raw milk vending machine locations in Germany, 2019

Note: For data privacy reasons, not all addresses could be obtained, resulting in only 72% of all raw milk vending machine locations being shown.

In order to present the socio-economic circumstances in the areas in which the raw milk vending machines are located, data from the statistical offices of the federal states and the federal government, as well as from the Thünen-Landatlas (2023), were used. As raw milk is an untreated product, its physiological properties make it perishable and it should be refrigerated at all times. The catchment areas of its consumers are smaller than those of other products. Therefore, it can be assumed that the factors that significantly influence marketing success operate on a small regional scale. Socio-demographic characteristics were therefore surveyed at least at municipality level and sometimes, depending on data availability, at county level.

5.4 Methodology

The conceptual framework presents the relations within the extended SOR model for the prediction of raw milk sales. Figure 5.1 shows that raw milk sales are a function of operational factors, marketing and environmental stimuli, individual organisms, and the addressable market (see Equation 39). For the prediction, these factors were used as independent variables/features, and raw milk sales in 2019 for farm k was the dependent variable/target variable M_k. As a benchmark model, these relations are modelled using a multiple linear regression based on the OLS estimator, resulting in the following equation:

$$M_k = F_{Nk} + F_{Mk} + F_{Ek} + F_{Rk} + S_{Eik} + O_{ik} + C_k + \varepsilon_{ik} \tag{40}$$

Supervised ML was then applied as a methodology. It is capable of making accurate predictions and can be preferable to traditional econometric approaches, even with small datasets (Breiman, 2001b). In recent years, improvements in high-performance computing have created new ways of analysing data-driven agricultural processes; ML has played a key role in detecting, quantifying, and improving the understanding of these processes (Liakos et al., 2018). As shown by a review by Sharma et al. (2020), the whole agricultural supply chain can benefit from ML techniques in terms of data-driven decision-making by practitioners, researchers and policymakers. For applied economists, ML also offers great potential for expanding their toolbox (Storm et al., 2020). While the typical goal of ML algorithms is to make predictions, economists have discovered the role of precise predictions in policy evaluation and causal analysis (Athey, 2018). Nevertheless, certain ML models, e.g. deep learning models, are commonly referred to as 'black boxes' due to their complex and recursive nature, which presents challenges in terms of interpretability (Rudin, 2019). In contrast, some methods, such as decision trees, provide interpretation

features, such as measures of variable importance, to interpret prediction models (Hastie et al., 2009).

In this chapter, the ML algorithm XGB, developed by T. Chen and Guestrin (2016), was used to predict raw milk sales through vending machines in Germany. The basic principle of this algorithm was originally presented by Friedman (2001) and is a specific type in the large family of classification and decision trees. The first decision tree, the so-called CART, was demonstrated by Breiman et al. (1984) and creates binary splits to divide the dataset into maximally homogeneous subgroups based on its contribution to the loss function. As a combination of multiple decision trees, the random forest algorithm creates a random split into subgroups of variables and observations, and takes an average across all generated subsamples and their decision trees (Breiman, 2001a). Similarly, gradient-boosted trees work by training the model sequentially tree by tree and then combining these on a predictor (Schapire, 1990). The algorithm functions as an additive model, which builds the next tree by repeatedly fitting trees to the residuals of the previous tree (Efron & Hastie, 2021). In this iteration process, the gradient-boosting algorithm re-estimates the predictor weights, so that poorly predicted observations are given increasing weights over time (Athey & Imbens, 2019). The weights are adjusted using a gradient descent algorithm to find the direction and constantly update the model parameters. The XGB algorithm developed by T. Chen and Guestrin (2016) uses this learning technique, adding second partial derivatives to efficiently minimise the loss function. For regularisation, a strict penalty structure is added to avoid overfitting.

Regarding the structure of the XGB, a set of hyperparameters had to be tuned to achieve state-of-the-art results. Various approaches can be used to optimise ML algorithms in order to adjust the hyperparameters (Yang & Shami, 2020). One frequently used method is grid search, in which a user-specific set of values is established, whose all possible combinations are tested so that the optimal set can be selected (Injadat et al., 2020). For the fine-tuning of the XGB algorithm to predict raw milk sales, the grid-search parameters shown in Table 5.1 were used.

It is common to split a dataset into a training, validation and test dataset to find the appropriate level of model complexity and prevent overfitting (Hastie et al., 2009). However, for smaller datasets, a k-fold cross validation can alternatively be applied. This involves splitting the dataset into k parts, each with the same number of observations. On each occasion, all data are used except for one of the k parts, which is then used to iteratively predict the out-of-sample prediction error. The expected value of the out-of-sample prediction error can be estimated by averaging the out-of-sample prediction error across the k estimators (Baylis et al., 2021). In the

	Option 1	**Option 2**	**Option 3**	**Option 4**
Learning rate (η)	0.01	0.05	0.1	**0.3**
Maximum depth of a tree	1	**3**	5	7
Minimum weights[a]	**1**	3	5	7
Subsample[b]	**0.65**	0.8	1	-
Colsample[c]	**0.8**	0.9	1	-

Table 5.1: Grid search parameters for fine-tuning the XGB algorithm

Note: [a]minimum sum of weights of all observations required in a child, [b]fraction of observations to be randomly samples for each tree, [c]fraction of columns to be subsampled

present case, a three-fold cross-validation was used to fine-tune the hyperparameters with number of trees = 5,000 and an early stopping criterion for 10 rounds. The optimal combination of hyperparameter values received from the grid search was based on minimised RMSE and is marked in bold in Table 5.1. Finally, a random split was applied of 70% of the data for training and 30% as test data.

All variables that were not dummy variables were logarithmised so that the scales could be approximated. In this way, it was possible to standardise the variables fairly, but still present and interpret the effects based on the variable values. Some variables showed correlations with each other, which in these cases could be explained by considering both the category and the specific response option. For instance, a variable was used for online advertising and advertising via Facebook or a website. In order for predictions to be as accurate as possible, these correlated variables are not necessarily problematic, but when interpreting the variable importance, these correlations should be considered.

For the interpretation of the XGB algorithm, SHAP values were used. This method is described in Lundberg and Lee (2017) and allows to describe the individual contributions of features to the predictions made by the model. The idea was initially based on game theoretically optimal Shapley values, which provide the marginal contribution φ_n of each feature n that represents the effect on the model output (Shapley, 1953). For a subset $S \subseteq F$ (where F is the set of all features), two models are trained to compute the difference between them: one model $\left(g_{S\cup\{n\}}\left(x_{S\cup\{n\}}\right)\right)$ with feature n and the other $g_S\left(x_S\right)$ without that feature, where $x_{S\bigcup\{n\}}$ and x_S are input features. The Shapley values were calculated as follows:

$$\varphi_n = \sum_{S\subseteq F\setminus\{n\}} \frac{|S|!\left(|F|-|S|-1\right)!}{|F|!}\left(g_{S\bigcup\{n\}}\left(x_{S\bigcup\{n\}}\right) - g_S(x_S)\right) \tag{41}$$

SHAP values combine the Shapley value explanation based on an additive feature attribution method, as shown in the following linear function for binary features h:

$$g(x) = h(x') = \varphi_0 + \sum_{n=1}^{M} \varphi_n {x'}_n \tag{42}$$

where $x' \in \{0,1\}^M$ and equals 1 when a feature is observed, otherwise equals 0, and M is the number of input features (Lundberg & Lee, 2017). It is proposed that the SHAP summary plot that displays the magnitude, prevalence and direction of a feature's effect is used (Lundberg et al., 2020). This provides interpretable insights into the model's decision-making process, helping to understand the impact of each feature on the model's predictions. To illustrate the heterogeneous data structure as well as reflect the differences in SHAP values between farms, additional SHAP dependence plots (DPs) were used. These plots show how a feature's value x_{ik} impacts the prediction of every sample in the data. Here, the differences and advantages compared to the results of a linear regression model become apparent, since regression models can only represent average relationships but are unable to reflect the heterogeneity of individual variables on the model output.

To test the sensitivity of model accuracy and variable importance, robustness tests were applied by excluding the ten most important and least important variables (calculated as SHAP values) from the model. This procedure allows the stability of the model to be tested with respect to omitted variable bias due to unobserved confounders. While for the prediction without the most important features the accuracy decreased significantly, for the prediction without the least important features only a slightly reduced R^2 could be measured. The order of the variables in terms of importance for the prediction also remained fairly constant, thus the influence of the most important variables could be confirmed by this check.

All analyses were performed using R version 4.0.2 (R Core Team, 2021) and RStudio version 1.3.1056 (RStudio Team, 2020) for statistical computing with the library 'xgboost' version 1.6.0.1 for modelling (T. Chen et al., 2020) and the library 'SHAPforxgboost' version 0.1.1 (Liu & Just, 2020) for calculating and presenting SHAP values.

5.5 Results

5.5.1 Descriptive Results

Among the survey respondents, the first raw milk filling station opened in 1999. Over the years, 153 further vending machines have been continuously added by

the respondents, with the majority (approximately 70%) installing their vending machines in 2015, 2016 and 2017 shortly after the abolition of the milk quota on 31 March 2015. The sales volumes at the vending machine are relatively heterogeneous and range between 2,200 and 70,000 litres per year. On average, approximately 14,767 litres are sold at a vending machine annually, with 80% of the respondents having annual sales below 20,000 litres. The sales figures correspond with revenue figures in most cases (approximately 78%), as the typical price for 1 litre of raw milk is €1.00. Higher prices are charged in particular by farms that produce their milk organically. For this segment, the average price is about €1.18 per litre, whereas the average price for conventional farms is about €0.99 per litre. The share of organic milk producers (15% of respondents) is significantly higher than the proportion of organic dairy farms among all German dairy farms in 2020 (approximately 9%, (Destatis, 2023g, 2023f)). Figures on the share of farms with organic production and milk vending machines in Germany are not yet available.

Several of the farms surveyed operate other farm branches in addition to milk production. Thus, 30% of the farmers surveyed reported that less than 50% of their revenue comes from milk production. However, about 35% of the respondents earn more than 90% of their turnover from the sale of milk. In most cases, the revenue that can be derived from the additional sale of raw milk is fairly small in relation to the total revenue from milk production. About 80% of the farmers indicated that the share is less than 10%, and for 60% of the respondents the share of raw milk sales out of total turnover is less than 5%.

Farm size varies from 14 to 1,500 cows, with the average of about 136 cows significantly above the Germany-wide average of 68 cows in 2020 (Destatis, 2023g). The annual milk yield of about 9,040 litres/cow is also slightly above the German average of 8,246 in 2020 (BLE, 2023a).

The reasons for commissioning raw milk vending machines were not just financial. While the majority of farms surveyed were satisfied with the relationship with their dairy, most were fairly dissatisfied with the milk price it paid them. On a Likert scale from 1=very dissatisfied to 5=very satisfied, the survey resulted in a mean of 3.86 for satisfaction with the dairy (SD=1.03) and a mean of 2.51 (SD=1.14) for satisfaction with the milk price. When asked specifically about their motives for introducing the milk vending machines, 25% of the farmers indicated that they were only financial. About a quarter of the farmers interviewed said they had hopes of improving the image of agricultural production, conducting public relations work and/or increasing contact with consumers.

Variable description	Min.	Mean	Max.	SD
Number of cows	14.00	136.36	1500.00	170.32
Average annual milk yield per cow [l]	3,500	9,039.68	13,000	1542.50
Average milk price in the state (2011-2021) [€/100 kg]	32.33	34.15	35.77	1.33
Lease fee in the county, 2010 [€/ha]	69.00	241.43	612.00	103.92
Share of milk production in total farm revenue [%]	0.20	0.71	1.00	0.22
Age of the farmer	27.00	45.31	66.00	9.64
Year of commissioning	1999	2,014.78	2019	3.08
Financial motive [dummy]	0.00	0.2483	1.00	0.4335
Image building motive [dummy]	0.00	0.2483	1.00	0.4335
Both motives [dummy]	0.00	0.4621	1.00	0.5003
Organic production [dummy]	0.00	0.1523	1.00	0.3605
Vending machine model[a]	0.00	0.4430	2.00	0.6083
Additional product range [dummy]	0.00	0.8224	1.00	0.3835
Additional product sale via trust box [dummy]	0.00	0.3092	1.00	0.4637
Additional product sale via farm store [dummy]	0.00	0.0789	1.00	0.2705
Additional product sale via vending machines [dummy]	0.00	0.5066	1.00	0.5016
Raw milk selling price [€/l]	0.60	1.0230	1.50	0.1263
Opening hours: 24/7 [dummy]	0.00	0.7895	1.00	0.4090
Online advertising[b]	0.00	1.4172	2.00	0.8355
Listed in Google Maps [dummy]	0.00	0.7632	1.00	0.4266
Newspaper advertising [dummy]	0.00	0.2649	1.00	0.4427
Sign advertising [dummy]	0.00	0.8609	1.00	0.3472
Flyer advertising [dummy]	0.00	0.4094	1.00	0.4934
Facebook page [dummy]	0.00	0.6291	1.00	0.4846
Website [dummy]	0.00	0.3311	1.00	0.4722
Location near (federal) highway [dummy]	0.00	0.4605	1.00	0.5001
Location with commuter traffic [dummy]	0.00	0.4768	1.00	0.5011
Location near a town or city [dummy]	0.00	0.3245	1.00	0.4697
Accessibility of a supermarket in the municipality, 2017 [min driving time]	2.46	5.8132	15.40	2.3403
Location in tourist area [dummy]	0.00	0.2848	1.00	0.4528
Location in recreational area [dummy]	0.00	0.2914	1.00	0.4559
Raw milk vending machines in the county, 2019 [machines per 100 km²]	0.04	0.5627	3.16	0.5383
Annual income per capita in the municipality, 2016 [€000]	11.70	19.5816	45.55	4.3858
Rurality in the municipality, 2016[c]	-3.29	-0.0825	1.13	1.0432
Share of employees in the agricultural sector in the county, 2019 [%]	0.0009	0.0142	0.0491	0.0118
Number of dairy farms in the county, 2016	4.00	273.82	1400.00	242.73
Average age in the county, 2019	41.00	44.95	49.70	1.67
Proportion of university graduates in the county, 2020 [%]	0.0642	0.1220	0.2926	0.0432
Population density in the municipality, 2019 [population/km²]	10.00	346.64	3,535.00	496.31
Annual raw milk sales, 2019 [l]	2,200	14,767.63	70,000	10825.43

Table 5.2: Descriptive statistics of the applied features

Note: [a]vending machine model [0=Risto, 1=Brunimat, 2=Millymat], [b]online advertising [1=only offline, 2=only online, 3=both online and offline], [c]rurality calculation based on Thünen-Landatlas (2023)

The total investment costs are on average around €19,475 and only deviate strongly upwards in a few exceptional cases. The majority of the farms (approximately 74%) had to spend between €10,000 and €30,000 to commission the vending machine. The monthly costs are about €292 on average, and comprise costs for electricity, material, advertising, and labour. A large proportion of the farms (about 72%) make a profit from the raw milk vending machine. For a quarter of the respondents, the business of selling raw milk is a plus-minus-zero business, and about 3% of the farms were making a monetary loss at the time they were surveyed.

The survey was conducted a few months after the outbreak of the Covid-19 pandemic. The impact was clearly noticeable on sales from raw milk filling stations. About 87% of the farmers surveyed reported an increase in their sales as a result of the pandemic. However, other effects during the pandemic could not be considered due to the survey design.

All variables applied are listed with Min., Mean, Max. and SD in Table 5.2.

5.5.2 Results

As a first step, both the OLS regression and the XGB algorithm were trained on raw milk sales in 2019 as a target variable and then applied to the test dataset.

From the entire dataset of $N = 154$, the model used $N = 120$ observations since the XGB algorithm can only deal with complete data points and all farms with missing values were excluded. Filling in missing values by interpolation did not increase the R^2 in that case. This may be attributed to the heterogeneous data structure since this complicates interpolation.

During the development of the model, different numbers of trees for the XGB algorithm were tested and the optimal number was found to be 25. This number of trees resulted in a maximum R^2 of 0.4721 for the test dataset and a R^2 of 0.9421 for the training set. It was evident that the model was slightly overfitted, as the R^2s of the training and test datasets were relatively widely separated, which should be kept in mind when interpreting the SHAP values. In comparison, the OLS resulted in a R^2 of 0.6597 for the training set and -0.0689 for the test dataset. This shows the advantage of this ML algorithm over traditional linear regression in that case. Although the OLS can also explain a large proportion of the raw milk sales in the training dataset by the model structure, the model does not apply and generalise these learned structures well to new data, which is reflected in a negative R^2. The following analysis therefore focuses on the results of the XGB model. The estimated regression results for the OLS benchmark model are presented in the appendix.

SHAP values were calculated in order to examine the influence of various vari-

ables on predicting raw milk sales specifically in this dataset. To ensure that these correlations are not limited to the presented dataset and cannot be extrapolated to the population, some robustness tests were carried out, including removing the most and least important features, applying cross-validation, and using a train/test split. Moreover, at about 19%, the data basis already represents a considerable part of the milk vending machine structure in Germany. However, it should still be borne in mind that the heterogeneous structure presented here could deviate in reality.

In Figure 5.3, all the variables involved are sorted by their average influence for the prediction of raw milk sales from vending machines. They are in descending order, thus the most important variable is put first. Variable importance was measured in SHAP values, which show the contribution of each variable to the outcome of the model (raw milk sales). While in the left column all average SHAP values per variable are listed as numeric values, every individual SHAP value per observation is shown as a dot distributed along the x-axis in the middle of the figure. This allows not only average relationships in the dataset to be represented, but also illustrates the heterogeneity of farms. For instance, the most relevant variable 'population density' shows a large dispersion of SHAP values. This means that there is a wide variety of influence on raw milk sales from population density. For some farms, the impact is rather positive and for some farms it is rather negative, but on average it has more impact on the prediction than other factors. In contrast, the variable 'location in a recreational area' has rather close and small SHAP values, which are associated with a lower effect on raw milk sales on average.

The feature of population density has the greatest influence on the predicting result and thus on raw milk sales. The ten most important variables are all features that are either related to the location (e.g. accessibility of supermarket, rurality) or operational orientation (e.g. share of milk production sales on total sales, age of farm manager). In contrast, individual marketing measures, both for advertising (flyers, signs etc.) and product design (opening hours, vending machine model etc.), are rather less relevant for the prediction of raw milk at vending machines. This demonstrates that a farm's raw milk sales are not defined by individual measures, but by a mixture of multiple measures. The variation between farms due to these measures does not form a sufficiently strong pattern for the XGB algorithm to recognise in order for conclusions to be drawn about raw milk sales. This may also reflect the heterogeneous structure of the farms. The only exceptions are the features 'additional product sales via vending machines' and 'listed in Google Maps', which are classified as more important than other marketing measures in terms of average SHAP values.

In addition to the average and the distribution of SHAP values along the x-axis, Figure 5.3 also presents the value of the variables themselves: the colour ranges from yellow (low value) to purple (high value). This contrast can best be recognised by the dummy variables, where all values equal to zero are yellow and all values equal to one are purple.

For the two marketing measures mentioned, Figure 5.3 illustrates an association between the purple dots and positive SHAP values, while the yellow dots are negatively associated. Based on this, it can be concluded that the presence of vending machines on Google Maps and the additional sale of other products through vending machines has a positive impact on raw milk sales. In particular, vending machines can serve as a channel for offering refrigerated processed dairy products that meet the preferences of raw milk consumers, thereby contributing to increased sales of raw milk.

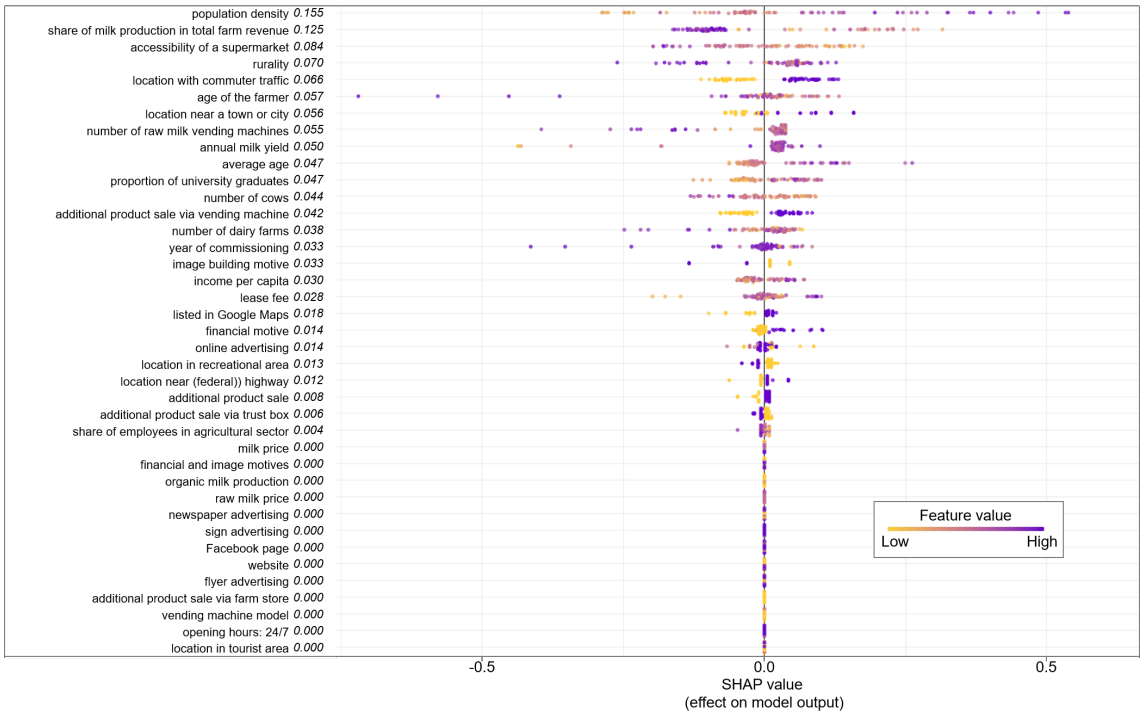

Figure 5.3: SHAP summary

Note: High average SHAP values (left column) indicate a high effect on average on the model output; individual SHAP values per farm are distributed along the x-axis; feature values correspond with the colour from low (yellow) to high value (purple).

However, location plays a major role in the sales of raw milk. This is reflected in the importance of population density, various marketing stimuli regarding the place, and some environmental stimuli features (see Figure 5.3). Based on the different shades of the SHAP values in Figure 5.3, a positive, almost linear relationship

between density and influence on the prediction of raw milk can be shown for population density. In regions with a greater population density, higher sales can thus be predicted. Greater population densities occur mainly in areas near a city and less in rural areas. Therefore, a positive correlation between variable value and SHAP value is also shown for the feature 'city proximity'. According to Figure 5.3, a location with commuter traffic also has a positive effect on raw milk sales, as the probability of purchase increases due to drivers regularly passing the vending machine. In contrast, the location along a (federal) highway plays a less important role. A location in a recreational or tourist area is also ranked as less important for raw milk sales. Consequently, it can be stated that raw milk sales are driven less by occasional purchases and more by regular purchases on the way to or from work.

Surprisingly, the price of raw milk is rather unimportant for predicting raw milk sales. This may be attributed to the homogeneity of the milk price across farms. Most farmers charge a price of €1.00, so consumers are less likely to be influenced by the price than by other factors.

The dependence of variable and SHAP values can also be represented in a DP. Figure 5.4 depicts the dependencies for rurality, accessibility of a supermarket, and vending machine density as examples of the several important location features. While a rather linear relationship to the SHAP values can be identified for the accessibility of a supermarket, for the other two variables the advantage of the XGB algorithm becomes apparent as it can also represent non-linear relationships. For rurality, this means that a negative correlation between the variable value and the SHAP value can only be recognised from a relatively high value of rurality. Rurality is calculated by different variables of land use, whereby effects on raw milk sales can be confounded.

In contrast, the variable 'accessibility of a supermarket' is defined by one single variable, and reveals a negative, almost linear relationship to the SHAP values. This shows that an increase in distance to the supermarket has a decrease in the influence on raw milk sales. Thus, it can be concluded that in regions that have a large number of supermarkets and are thus more urban, the willingness to buy raw milk is higher. A greater distance to the supermarket does not operate here as a positive marketing stimulus, instead the proximity to urban areas can be represented based on this variable. This confirms the statements made earlier about the variables 'population density' and 'proximity to a city'. The variable 'vending machine density' illustrates the competitive situation with other raw milk vending machines in the vicinity. According to Figure 5.4, there seems to be a 'sweet spot' for vending machine operators where there are barely enough vending machines in

the vicinity for consumers to be aware of the offer but they are not yet outcompeted by other vending machines.

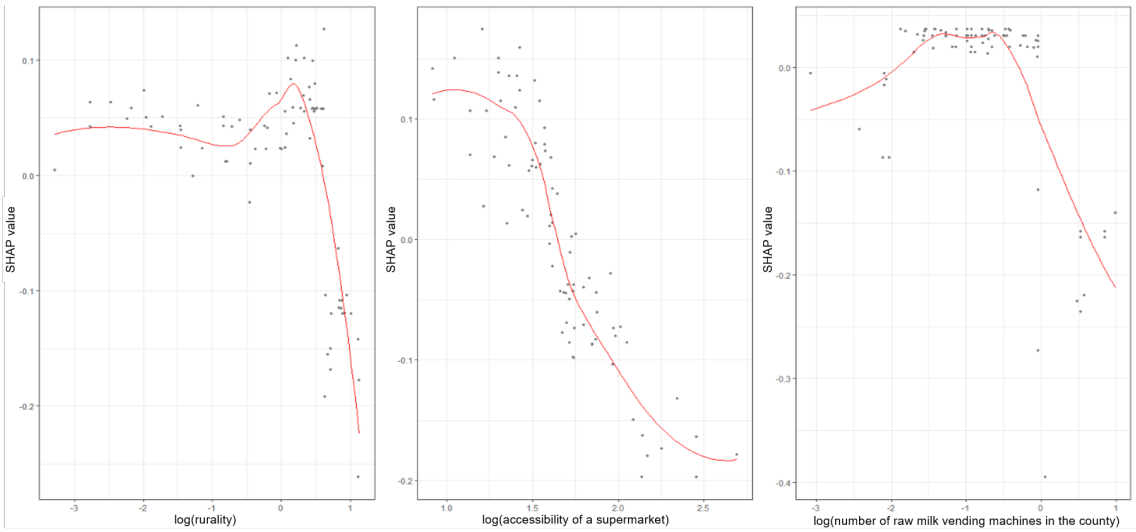

Figure 5.4: Dependence plots of location stimuli

Note: The y-axis shows the corresponding SHAP value, the x-axis shows the logarithmised value of the corresponding variable.

As well as the marketing stimuli regarding place, environmental stimuli also influence raw milk sales. These are less important on average than the marketing features mentioned above (see Figure 5.3) and show less significant correlations. For example, Figure 5.5 shows that there is a positive correlation between average age in a region and SHAP values, but this becomes less evident with increasing age. The pattern is similar for regional education. Although a positive effect on the prediction of raw milk sales can be roughly identified, the distribution of SHAP values is very noisy. For per capita income, the SHAP values are also very widely distributed in relation to the variable value. Together with the fact that the SHAP values are also rather less relevant on average, it can therefore be concluded that income has a very heterogeneous effect on sales and no clear and significant relationship can be shown.

Apart from location factors, farm-specific factors can also make an important contribution to an accurate prediction. The algorithm learns which farm structures generate which raw milk sales based on the patterns of strategic decisions. Several farm-specific factors emerge as important predictors, three of which are shown in the DP in Figure 5.6. The influence of the farm's degree of specialisation (measured as the share of milk production out of total production) is relatively heterogeneous. However, it can clearly be observed that highly specialised farms with >65% share of revenue have a negative effect on raw milk sales from vending machines. Thus, it can be concluded that these farms are sufficiently efficient due to their specialisation and

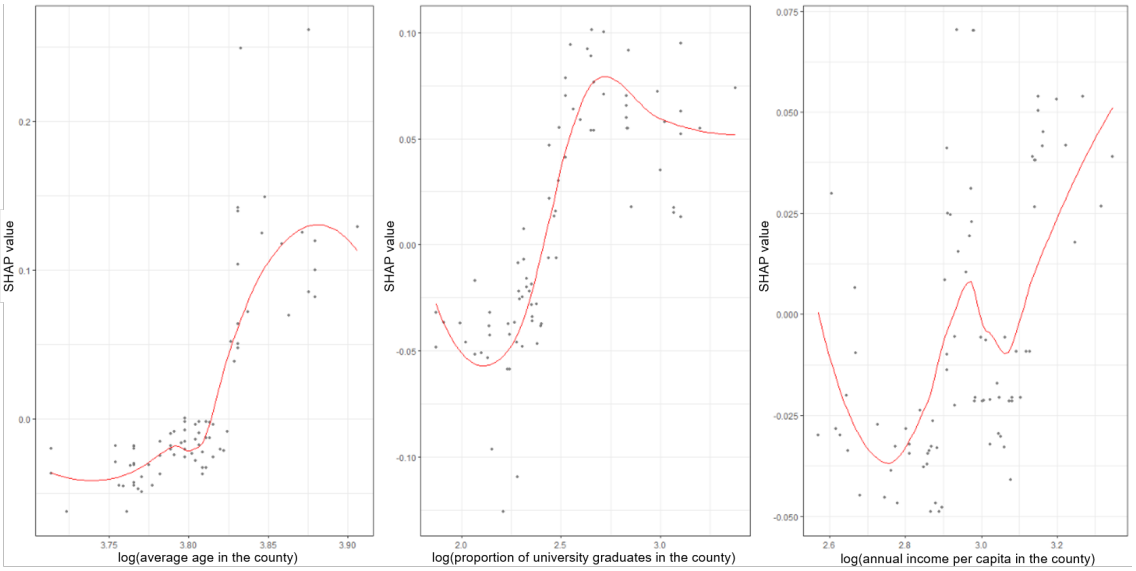

Figure 5.5: Dependence plots of environmental stimuli

Note: The y-axis shows the corresponding SHAP value, the x-axis shows the logarithmised value of the corresponding variable.

do not necessarily need to develop a second financial pillar. Another interpretation could be that more diversified farms have a higher level of expertise in developing and establishing new business segments.

Regarding the feature 'age of the farmer', an initial positive effect on raw milk sales can be observed with increasing age. This can be attributed to the accumulation of experience in farm management, which leads to improved performance. However, this positive effect plateaus and eventually turns negative, indicating that older farm managers have a negative impact on raw milk sales. This suggests that as farm managers age, they may find it difficult to operate a vending machine economically due to their reluctance to adopt new and innovative approaches. As a result, they may struggle to reach as many consumers as their younger counterparts.

The SHAP values of the farms depending on milk yield are divided into two groups: firstly, farms with an average milk yield of 9,337 litres, and secondly, farms with a well below-average milk yield of less than 7,000 litres (average milk yield = 5,333 litres), which have clearly lower SHAP values. Possibly, farms with significantly lower milk yields perform worse in terms of both milk yield and raw milk sales and are just unsuccessful farmers. Alternatively, these farms may not have the primary objective of achieving high milk yields and high raw milk sales, but have other motives, and therefore both produce and sell below average.

The second hypothesis could be supported by examining the motives of farmers to establish a raw milk vending machine. The variables 'financial motive' and 'image

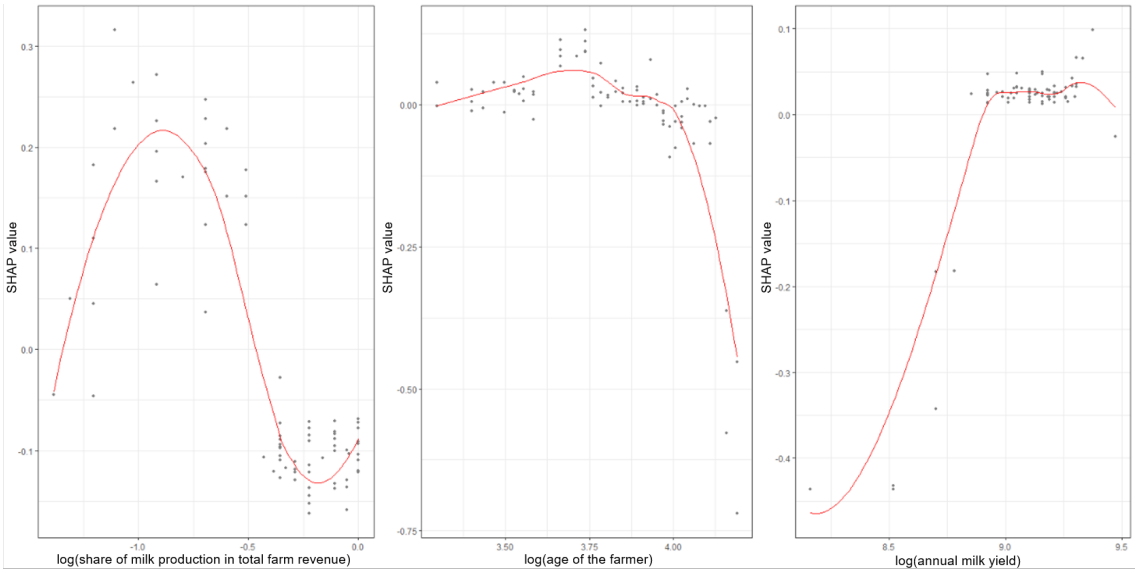

Figure 5.6: Dependence plots of operational farm stimuli

Note: The y-axis shows the corresponding SHAP value, the x-axis shows the logarithmised value of the corresponding variable.

building motive' show that farm managers are pursuing different objectives when they install a vending machine. If the primary motive is financial, farmers also focus on increasing raw milk sales, so that a positive effect of this variable on raw milk sales can be observed. In contrast, if the main focus is on improving the farm's image or doing public relations work, there is a negative effect on the SHAP values of raw milk sales (see Figure 5.3). Farm managers who are not primarily interested in additional financial income are also satisfied with lower sales. If the farm manager mentions both motives, this variable has no effect on the prediction of raw milk sales.

5.6 Discussion

The results show that the XGB algorithm provides significantly better accuracy than OLS regression when predicting new values. In addition, the SHAP values can provide insight into the importance of different factors in the prediction. However, there are some limitations of the methods that need to be considered.

The difference in R^2 for the training and test datasets shows that overfitting is present. This can lead to a bias in the SHAP values as these are calculated on the basis of the training data. Variables may be considered more important than they are in reality, or correlations may be over-interpreted by such a complex algorithm.

The model structure presented here uses a variety of variables. For the results of OLS regression, this means that some variables may be correlated and significant

effects may overlap. In contrast, the XGB algorithm is better able to deal with a large number of variables by regularising less important variables. In addition, the SHAP representation allows the individual influence of each variable on the prediction result to be shown, so that it cannot be assumed that bias has been introduced by correlating variables.

However, it is likely that there are other factors influencing raw milk sales that are not included in this model. These omitted variables and additional unobserved heterogeneity should be taken into account when interpreting the results and could be addressed in future research. For example, the factors considered relate mainly to the characteristics of the farms and their potential customers. Less attention has been paid to existing customers and their specific socio-demographic factors and individual needs. This focus is justified by the fact that predicting raw milk sales was the main focus of the analysis, so the data would not be available for a farm considering an investment decision. Future research could focus on this aspect. For example, by interviewing current consumers, the causal relationships between influencing factors and raw milk sales could be analysed in a different way.

For the current data collection, it should also be noted that there may have been a bias in the selection of farms. Although a random number of farms responded to the survey request, this method of reaching respondents may have resulted in more motivated farms and farms with better performing businesses responding. Self-reporting of raw milk sales is also subject to a degree of uncertainty as farmers had the opportunity to inflate their sales volumes.

5.7 Conclusions

As the trained XGB algorithm provides reasonable accuracies for both the training and test datasets, sales of raw milk can also be predicted for data outside the observed farms using the variable structure presented. Reluctant farm managers could therefore include a prediction of possible sales volumes in their investment decision using the algorithm shown here.

An examination of the influencing factors and the heterogeneity of the farms through the calculation of SHAP values showed that individual marketing measures tend to have less influence on raw milk sales. Therefore, it is not possible to recommend a single, ideal measure for farms; rather they should choose a mix of various measures that suit their individual farm.

Instead, location has the greatest impact on a farm's raw milk sales. However, the location of the vending machine is tied to the farm, as the sale of the unprocessed raw milk is tied to the location of the farm under § 17 (4) Tier-LMHV (German

Animal Food Hygiene Regulation). Farmers are therefore unable to influence this factor, which crucially determines the sale of raw milk. Consideration should be given to whether alternative approaches for food safety authorities could be devised to ensure the hygiene of vending machines and no longer link this to the distance to the farm. Thus, this new type of direct marketing could be established in closer geographic proximity to consumers, and offer an additional financial income source for even more farmers. However, farmers should be aware that this would massively change the sales ambience of raw milk vending machines, thus losing one of their unique selling propositions. As this is still hypothetical, further research is needed into the potential for off-farm sales of raw milk.

5.8 Appendix

The benchmark model was a multiple linear regression, as shown in Equation 40, based on the OLS estimator. All independent and dependent variables were logarithmised, except for the dummy variables and the variables calculated in a special way, which are indicated below by a note. The model was applied to the training data, resulting in the following table.

Independent variables	**Coefficients**	**p**
Intercept	246.803	0.669
Number of cows	0.080	0.549
Average annual milk yield per cow	-0.080	0.888
Average milk price in the state (2011-2021)	1.760	0.492
Lease fee in the county, 2010	0.194	0.485
Share of milk production in total farm revenue	-0.110	0.649
Age of the farmer	-0.097	0.778
Year of commissioning	-32.928	0.665
Financial motive [dummy]	0.236	0.557
Image building motive [dummy]	-0.006	0.989
Both motives [dummy]	0.171	0.623
Organic production [dummy]	-0.061	0.876
Vending machine model[a]	-0.105	0.586
Additional product range [dummy]	0.009	0.979
Additional product sale via trust box [dummy]	0.223	0.450
Additional product sale via farm store [dummy]	0.534	0.122
Additional product sale via vending machines [dummy]	0.259	0.376
Raw milk selling price	-0.791	0.437
Opening hours: 24/7 [dummy]	0.228	0.441
Online advertising[b]	-0.055	0.758
Listed in Google Maps [dummy]	-0.067	0.738
Newspaper advertising [dummy]	0.024	0.886
Sign advertising [dummy]	0.381	0.247
Flyer advertising [dummy]	0.118	0.528
Facebook page [dummy]	0.193	0.473
Website [dummy]	0.021	0.928
Location near (federal) highway [dummy]	0.071	0.672
Location with commuter traffic [dummy]	0.310	0.060
Location near a town or city [dummy]	0.181	0.468
Accessibility of a supermarket in the municipality, 2017	-0.550	0.035
Location in tourist area [dummy]	-0.096	0.659
Location in recreational area [dummy]	-0.159	0.352
Raw milk vending machines in the county, 2019	-0.299	0.063
Annual income per capita in the municipality, 2016	-0.378	0.503
Rurality in the municipality, 2016[c]	-0.019	0.907
Share of employees in the agricultural sector in the county, 2019	-0.018	0.898
Number of dairy farms in the county, 2016	-0.039	0.737
Average age in the county, 2019	1.414	0.636
Proportion of university graduates in the county, 2020	0.256	0.487
Population density in the municipality, 2019	0.281	0.086
Degrees of freedom	41	
R^2	0.6537	
Adjusted R^2	0.3243	

Table 5.3: Regression results for the OLS model

Note: [a]vending machine model [0=Risto, 1=Brunimat, 2=Millymat], [b]online advertising [1=only offline, 2=only online, 3=both online and offline], [c]rurality calculation based on Thünen-Landatlas (2023)

Chapter 6

Drivers of the Demand for Raw Milk from Vending Machines in Germany

6.1 Introduction

In Germany, the number of farms almost halved between 2001 and 2020 (Destatis, 2023d, 2023c). The number of dairy farms in the country has also been declining for many years, decreasing by over 58% in the same period (Destatis, 2023g). This trend has led to a growing disconnect between consumers and agriculture, as fewer people have direct experience of farming, and the distance between the place of production and place of consumption has extended and become increasingly non-transparent (La Trobe & Acott, 2000). At the same time, there is growing demand for more direct connections with farmers as many German consumers prefer locally produced food, transparent and traceable food supply chains, and certified animal welfare (BMEL, 2018). These desires are partly accompanied by a greater WTP for these products (Emberger-Klein et al., 2016).

One way for consumers to purchase from farmers directly is through the use of raw milk vending machines. These allow consumers to purchase unprocessed raw milk directly via self-service from the farm without the need for intermediaries. Raw milk vending machines not only provide an easy opportunity for consumers to get in touch with milk production, but also offer an additional source of income for farmers with comparatively low investment costs.

Despite the potential benefits of raw milk vending machines, little is known about which consumers purchase raw milk from these machines and why, and what factors influence their WTP for raw milk.

This chapter aims to answer these questions by investigating purchase frequency and WTP for raw milk from vending machines in Germany. In addition to socio-economic characteristics, motives and influencing factors will be analysed. By acquiring a better understanding of consumers and influencing factors, this chapter aims to provide insights that may be especially useful for farmers who are considering the use of raw milk vending machines and adapting their marketing activities.

6.2 Theoretical Background

The SOR model, originally proposed by Woodworth (1929) as an adaption of the classic stimulus-response theory of Pavlov (1927), provides a framework for understanding consumer behaviour. This model, illustrated in Figure 6.1, has three components: stimulus, organism, and response. The basic idea is to understand which environmental stimuli influence consumers' perceptions so that they make certain purchase decisions (Woodworth, 1929). Although it has been widely used to analyse consumer behaviour (Hempel & Hamm, 2016; Sultan et al., 2021, e.g.), this

particular formulation of the SOR model has not yet been applied in the context of the raw milk market. While an extended version of the SOR model was presented in Chapter 5, the current chapter focuses on the original version.

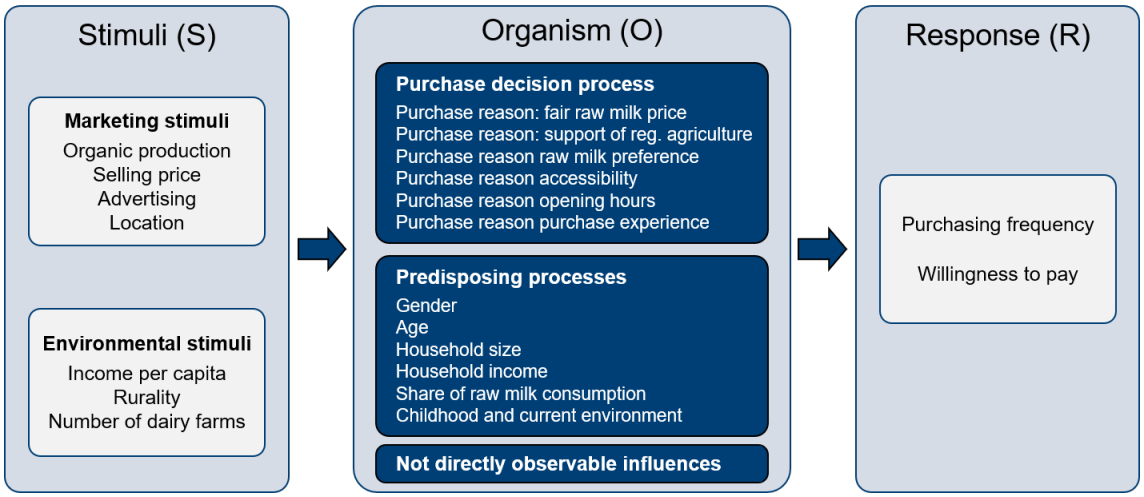

Figure 6.1: SOR model for consumption of raw milk from vending machines

Source: Own illustration based on Kotler et al. (2007)

The underlying theoretical framework of the SOR model remains consistent across different products, although the specific stages and their design may vary depending on the product and the consumer involved (Kotler et al., 2007). Figure 6.1 illustrates the different factors influencing consumption and WTP for raw milk from vending machines. In this context, stimuli are defined as external forces that encompass and influence an individual's psychological processes (Young & Young, 2016). These stimuli can be further categorised into controlled stimuli, derived from marketing efforts, and uncontrolled stimuli, derived from the environment (Kroeber-Riel & Weinberg, 2019).

Operational decisions can create marketing stimuli to gain consumer attention (Kotler et al., 2007). In the case of a raw milk filling station, operators use different strategies to attract consumers. A marketing mix including product, price, promotion, and place is used to generate potential purchase interest (Meffert et al., 2018). In the context of selling raw milk through vending machines, the farmer's ability to differentiate the product is limited to the production method, as only the raw milk product is under consideration in this scenario. Thus, the organic production of milk is included in the analysis as an influencing factor. Pricing for raw milk is rather inflexible too. A survey of farmers ($n = 154$, described in Chapter 5) reveals that a price of €1.00 is prevalent at raw milk vending machines. Nevertheless, there are also farms with lower or higher product prices, so this factor is considered to affect

consumers. Advertising around raw milk vending machines varies from placing signs around the farm to regularly maintaining social media accounts. For consideration in the SOR model, the focus is only on the operation of advertising in general: offline, online or a combination of both. The location for selling raw milk is confined to the farm due to legal requirements and is consequently predetermined, therefore it cannot be considered a factor through which farmers can introduce varying stimuli.

However, the stimuli of the location are still included in the model since they are generated by the environment. The stimuli that the environment can provide are mainly socio-economic in structure. Since raw milk is an untreated product, its physiological characteristics make it perishable and it should always be kept refrigerated. Consumer catchment areas are smaller than those of other agricultural products, therefore socio-economic characteristics are considered at municipality level. In particular, the social experiential environment that arises from direct contact with fellow human beings in a region are considered as an environmental stimulus (Kroeber-Riel & Weinberg, 2019). To represent these drivers, various proxies are used, e.g. per capita income, which provides information about the prevailing purchasing power and its consumption habits. The variables 'rurality' and 'number of dairy farms' stand for the consumers' relationship with agriculture. While contact with the agricultural producer tends to be less frequent in urban regions, in more rural areas contact points between farmers and consumers are more likely. In particular, the relationship with the producer of raw milk could be stronger in a region with a higher density of dairy cattle than in a region with fewer dairy farms.

The presented stimuli affect an organism whose purchase decision processes remain hidden in most models (Kotler et al., 2007). However, through the design of the consumer survey, reasons for individual purchase decisions can also be taken into consideration for the organism's response in this chapter. Additional individual influencing factors include consumers' socio-economic characteristics (such as gender, age, household size, and household income), as well as the share of raw milk regularly consumed out of total milk consumption. The purchase process is also influenced by the relationship with agricultural production, which is represented in this chapter by the consumers' environment in which they grew up or currently live. In addition to the factors that can be observed here, there are still individual factors that impact consumers' decisions, which can only be represented using an error term.

Consumer response is measured based on two variables in this chapter: the WTP for one litre of raw milk and the purchase frequency of raw milk from vending machines per month.

6.3 Data Collection

The database is mainly a survey of consumers of raw milk at vending machines. Supported by operators of raw milk vending machines, posters featuring information about the survey were placed in the immediate vicinity of vending machines. Consumers could take part in the online survey by scanning a QR code. Consumers at 68 locations across Germany participated in the survey between April and October 2022. In addition, consumers were interviewed in person at 11 raw milk vending machines. Overall, 656 customers participated in the survey, with 543 answering the questionnaire online, and about 17% responded on site. The locations of the consumer survey are illustrated in Figure 6.2.

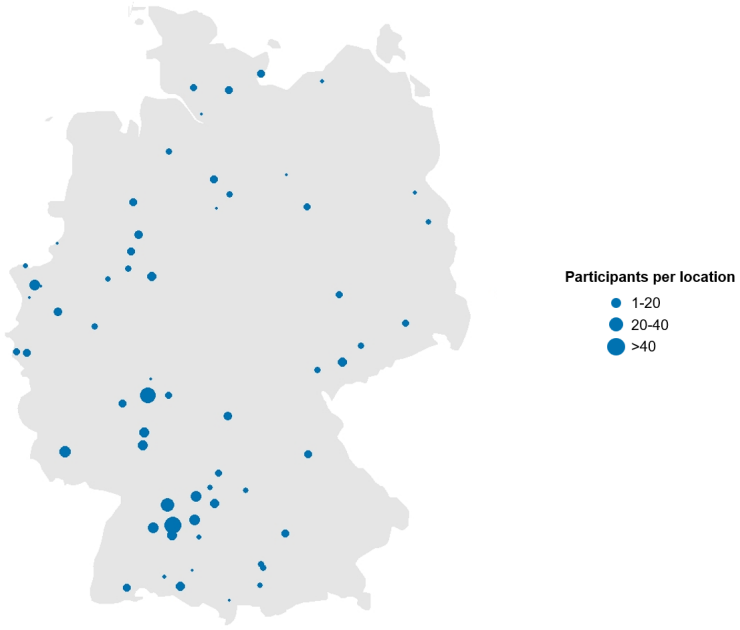

Figure 6.2: Raw milk vending machine locations for the consumer survey (n=68)

Consumers were only surveyed at locations where farmers had previously participated in a separate survey on their experiences with raw milk vending machines (see Chapter 5), to enable these data to be included in the analysis as well. In particular, the sites' equipment and marketing activities were assessed.

In order to present the socio-economic circumstances in the areas where the raw milk vending machines are located, data from the statistical offices of the federal states and the federal government were used. Thus, besides the socio-economic information on the consumers themselves, the market environment of the raw milk vending machines can also be considered as an influencing factor.

6.4 Methodology

The relations presented in the SOR model above can be illustrated in a linear regression model. The response variables 'purchase frequency' and 'willingness to pay' serve as dependent variables. Purchase frequency was measured as an ordinally scaled variable on a scale ranging from daily to sporadic purchases. To include this variable in a linear regression model, numerical values are assigned for the scale, allowing purchase frequency to be reported in terms of the number of monthly purchases (R_{iK}). This specific assignments for the conversion are shown in Table 6.1.

Ordinal:	Daily	Several times a week	Once a week	Several times a month	Once a month	Sporadically
Nominal:	28	12	4	2	1	0.3

Table 6.1: Converting the variable 'purchase frequency' from ordinally to nominally scaled

WTP represents the consumer's intention to buy a product at a maximum acceptable price (Varian & Varian, 1992). For this hypothetical consideration, the consumers were asked to tick in steps of €0.10 what price they would be willing to pay for one litre of raw milk at the vending machine. This results in the second dependent variable: WTP (R_{iW}).

According to Kotler et al. (2007), and as shown in Figure 6.1, the two dependent variables of purchase frequency (R_{iF}) and WTP (R_{iW}) are a function of stimuli (S_i) and individual influencing variables (O_i) for $i = 1, \ldots, N$ individuals of the dataset. Following equations 43 and 44, the vectors of the individual stimuli (S_i) and influencing variables of the organisms (O_i) are divided into several vectors according to different categories (M, E, R, S, I) for the econometric estimation:

$$R_{iF} = f(S_{iM}, S_{iE}, O_{iR}, O_{iS}, O_{iI}, Q_i\ ) \tag{43}$$

$$R_{iW} = f(S_{iM}, S_{iE}, O_{iR}, O_{iS}, O_{iI}, Q_i) \tag{44}$$

Consumer decisions can be influenced by two types of stimuli: marketing stimuli (S_{iM}) and environmental stimuli (S_{iE}). The organism's individual influencing variables are divided into three subgroups: reasons for purchasing raw milk at vending machines (O_{iR}), socio-economic characteristics of the individual consumers (O_{iS}), and their involvement in raw milk consumption (O_{iI}).

Furthermore, the two dependent variables depend on certain specifications (Q_i) of the survey. For instance, some participants were interviewed on site, while others

completed the survey online. This could have led to differences in the answers given to the questions, especially in the assessment of WTP. As the survey was conducted over several months, macroeconomic effects (e.g. inflation) may have had different effects on the participants during the survey period. This factor is therefore also considered.

Given that both dependent variables are used as response variables in the SOR model and thus have similar explanatory variables, it is reasonable to assume a correlation between the two error terms. Consequently, an estimation of the regression parameters equation by equation using OLS is unlikely to provide an efficient estimate (Peremans & Van Aelst, 2018). To address this problem, Zellner (1962) introduced SUR models. SUR models allow for the underlying covariance structure of the error terms across equations. A SUR model is a set of G equations, where each equation n is defined as:

$$y_n = X_n\beta_n + u_n; \quad n = 1, 2, \ldots, G \tag{45}$$

where y_n represents a vector of the observations in the dependent variable, while X_n is a matrix of dimension $K \times n$ consisting of K exogenous variables, β_n is the estimated coefficient vector and u_n is a vector of the disturbance terms of the ith equation. These vectors can be organised in a stacked system, allowing β_n to form a K-dimensional vector of unknown parameters to be estimated by the model. It is important to note that $K = \sum_{n=1}^{G} K_n$. Furthermore, it is assumed that there is no correlation between the error terms across the observations. This assumption implies that $E[u_{ni}u_{mj}] = 0 \; \forall \; i \neq j$, where n and m represent equation indices, and i and j are observation indices.

The assumption of contemporaneous correlation is explicitly stated in the SUR model by linking the blocks of equations. This assumption is expressed by the equation $E[u_{ni}u_{mj}] = \sigma_{nm}$, where $\Sigma = [\sigma_{nm}]$ represents the contemporaneous disturbance covariance matrix. Consequently, the covariance matrix for all error terms can be written as $E[uu'] = \Omega = \Sigma \otimes I_N$. In this context, $\otimes$ denotes the Kronecker product, I_N is an identity matrix with dimension N, and N corresponds to the number of observations in each equation.

Estimating each of the G equations individually using OLS without accounting for the correlation between the error terms can lead to inefficient estimators. To address this issue, the GLS estimator provides a modification of OLS that can handle such correlations. The equation for the GLS estimator is expressed as follows:

$$\hat{\beta}_{GLS} = \left(X'\left(\Sigma^{-1} \otimes I_N\right)X\right)^{-1} X'(\Sigma^{-1} \otimes I_N)y \tag{46}$$

Since the covariance Σ is usually unknown, the FGLS estimator is applied. This estimator calculates the elements of Σ by $\hat{\sigma}_{nm} = \hat{u}'_n \hat{u}_m / N$, where $\hat{u}_m$ is the residual vector of the mth block obtained from OLS. The estimated covariance matrix $\hat{\Sigma}$ is then entered into the GLS estimation. This iterative process can be repeated until the coefficient estimates converge.

The covariates for the respective equations have been selected as shown in Figure 6.1, except for the marketing stimuli 'raw milk price' and 'advertising', which are omitted for the modelling of WTP. There may be an endogeneity problem between current and potential raw milk prices. As currently no suitable exogenous proxies to instrument the variable are existent, the model is estimated without this variable. The variables for advertising are not expected to affect WTP. Furthermore, for efficient estimation by SUR, the covariates need to be at least slightly different otherwise the estimated coefficients would be numerically equivalent to the results of an OLS.

To exclude linear dependence of individual independent variables (multicollinearity), the variance inflation factor of the model is checked according to Backhaus et al. (2015). When using dummy variables per municipality or county, collinearities with the additional location characteristics can be detected. For this reason, dummies are not used to represent location as an influencing factor, but rather its socio-economic characteristics. The Breusch-Pagan test is conducted to determine whether the individual equations within the SUR models are correlated (Breusch & Pagan, 1979). Based on Ramsey's RESET test (Ramsey, 1969), there is no misspecification of the regression equations.

All analyses were performed in R version 4.0.2 (R Core Team, 2021), and Rstudio version 1.3.1056 (RStudio Team, 2020). The SUR model is estimated using the library 'systemfit' (Henningsen & Hamann, 2008).

6.5 Results and Discussion

6.5.1 Descriptive Results

All variables applied are listed with Min., Mean, Max. and SD in Table 6.2 and some are described in more detail below.

The participants in the survey are relatively evenly split between male (44%) and female (56%) (with <1% diverse). The average age of the participants is 46.6 years, which is slightly older than the total population of Germany in 2021 (average age 44.7 years, (Regionalstatistik, 2023)). A comparison of the participants who took part in the survey online or on-site shows that the on-site participants, who have an

Acronym	Category	Variable description	Min.	Mean	Max.	SD
R_{iW}	Response	Willingness to pay [ct]	80	147.49	250	34.82
-	Response	Purchase frequency (ordinal)	1	3.16	7	1.37
R_{iF}	Response	Purchase frequency (nominal)[a]	0.3	6.62	28	5.31
S_{iM}	Stimuli	Organic production [dummy]	0	0.09	1	0.29
S_{iM}	Stimuli	Raw milk price [€]	0.6	1.00	1.4	0.10
S_{iM}	Stimuli	Offline advertising [dummy]	0	0.24	1	0.43
S_{iM}	Stimuli	Online advertising [dummy]	0	0.08	1	0.27
S_{iE}	Stimuli	Annual income per capita [€]	13,050	21,217	38,749	4,671
S_{iE}	Stimuli	Rurality[b]	-3.29	-0.56	1.13	1.05
S_{iE}	Stimuli	Number of dairy farms per county	14	179.22	981	167.06
O_{iR}	Organism	purchase reason (PR) fair milk price[c]	1	6.31	7	0.97
O_{iR}	Organism	PR supporting regional farmers[c]	2	6.51	7	0.84
O_{iR}	Organism	PR raw milk preference[c]	1	5.15	7	1.71
O_{iR}	Organism	PR accessibility[c]	2	5.53	7	1.29
O_{iR}	Organism	PR opening hours[c]	1	5.55	7	1.46
O_{iR}	Organism	PR shopping experience[c]	1	3.76	7	1.66
O_{iS}	Organism	Gender[d]	1	1.56	3	0.51
O_{iS}	Organism	Age	17	46.63	83	12.51
O_{iS}	Organism	Household size[e]	1	3.41	5	0.91
O_{iS}	Organism	Household income[f]	1	4.41	8	1.82
O_{iS}	Organism	Share of raw milk consumption [%]	1	67.67	100	33.01
O_{iS}	Organism	Current environment[g]	1	5.21	7	1.45
O_{iS}	Organism	Childhood environment[g]	1	4.94	7	1.90
Q_i	Survey	On-site survey [dummy]	0	0.13	1	0.34
Q_i	Survey	Month of survey participation	4	5.59	10	1.60

Table 6.2: Descriptive statistics of the applied variables

Note: [a]conversion from ordinal to nominal [see Table 6.1], [b]calculation based on Thünen-Landatlas (2023), [c]measured on 7-point Likert scale [1=very unimportant,...,4=neutral,...,7=very important], [d][1=male, 2=female, 3=diverse], [e][1=single, 2=single with children, 3=with partner, 4=with partner and children, 5=other], [f][1≤ €1,000,...,8 ≥ €7,000], [g][1=very urban,...,4=suburban,...,7= very rural]

average age of 54.4 years, are significantly older than the online participants (average age 45.9 years). This difference may be explained by the fact that internet affinity in Germany decreases with increasing age, especially on mobile devices (Initiative D21, 2022).

The majority of participants (around 66%) currently live in a more rural area, while only around 13% live in a more urban environment. However, around 62% of the respondents grew up in a more rural area, so it can be said that the surveyed consumers tended to move from the city to the countryside.

Respondents mostly live in a household with their partner (approx. 36%) or also with children (approx. 48%). The monthly net income of these households was positioned on a scale of €1,000 increments. On average, the net household income of the respondents is between €4,000 and €5,000 per month. Compared with German's national net household income of €3,813 in 2021 (Destatis, 2023h), the respondents' income is higher.

Raw milk is purchased regularly by most of the respondents. Around 40% indicate that they buy several times a week, and 30% indicate that they buy once a week from the raw milk vending machine. The quantities purchased are mostly small, ranging between one and three litres per purchase for about 77% of the respondents. This can probably be attributed to the fact that it is an untreated product that either has to be consumed promptly or processed further. For further processing, about 30% of the respondents use the raw milk for cooking and about 31% for baking (with multiple choices possible). All the participants purchase the raw milk for their own consumption, with about 80% not boiling the milk first, although almost all (94%) have seen the mandatory notice for prior boiling. The share of raw milk out of total milk consumption is very variable. There is a large group that consumes almost only raw milk (48%), with >90% raw milk as a proportion of total milk consumption. The share of the other participants is distributed fairly homogeneously at between 1% and 90%.

The majority of respondents (approx. 80%) consider the price paid at the raw milk filling station to be appropriate, while 19% still feel the price paid so far is too low. When asked what price the respondents would still be willing to pay (in €0.10 increments) for one litre of raw milk, more people (approx. 55%) are willing to pay a price of up to €1.50 per litre than not. However, only about 23% are willing to pay a price of €1.60 or more, with the proportion becoming even smaller as the price rises.

Figure 6.3 shows the distribution of WTP according to purchase frequency as a violin plot. It reveals that regular consumers are on average less willing to pay

for one litre of raw milk. It is also noticeable that less frequent consumers have a higher range of WTP. In particular, the WTP of the group of consumers who buy raw milk only once a month or sporadically is rather heterogeneous.

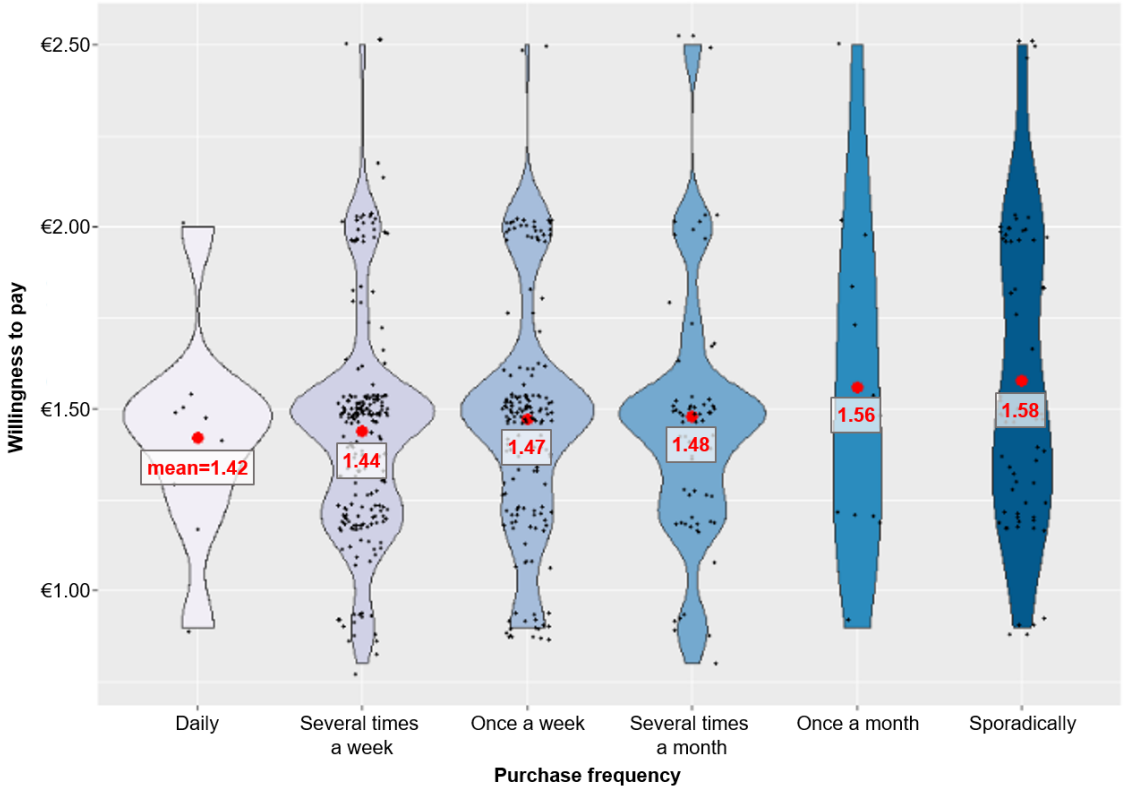

Figure 6.3: Violin plot for WTP per purchase frequency

The reasons for purchasing raw milk are measured on a 7-point Likert scale. The respondents could thus indicate whether the reasons were between very unimportant to very important to them. In particular, categories related to the production of raw milk (support for regional agriculture, fair raw milk price) were rated as most important. The PRs that are more related to the product offering (opening hours, accessibility, shopping experience) are rated as less important on average by the respondents.

6.5.2 Regression Results

Table 6.3 shows the results of the SUR model for WTP and purchase frequency as dependent variables. The McElroy R^2 indicates the explanatory power of the whole model and equals 0.3901. As assumed in the methods section, there is a correlation ($r = -0.1372$) between the residuals of the two models. According to the R^2 calculated separately for the models, the variable structure shown can explain purchase frequency better (*adj.* $R^2 = 0.4777$) than WTP (*adj.* $R^2 = 0.1838$).

The estimated coefficients and the corresponding p-values are listed in Table 6.3. It becomes apparent that several stimuli influence the consumers' WTP. For instance, the customer is more willing to pay for raw milk at the vending machine if the product has been produced organically. In this case, organic production is assigned a higher value than conventional production and acts as a positive stimulus.

The estimated coefficients of the environmental stimuli are negatively significant for WTP (except for rurality). This means that a higher income in the community as well as a higher number of dairy farms lead to a lower WTP. This could indicate that consumers in regions that have less involvement with dairy production are more willing to pay for a product that is produced further away from where they live, even if the local income may be lower. The product seems to be valued more in these areas than in communities where there is automatically more interaction between those producing milk and those consuming it. In contrast, the model estimates a significant positive coefficient for the variable 'rurality'. This shows that the number of dairy farms and rurality (calculated by the Thünen Institute using land use data) are not correlated. The more rural a region, the greater the WTP. One reason for this could be that consumers in rural areas are generally prepared to spend more on food and are therefore more willing to support the producer financially.

If consumers are explicitly asked about their reasons for buying raw milk from vending machines, it becomes apparent that the main reason for a higher WTP is a fair milk price for the producer. This implies that the same consumers do not consider that the milk price charged at other points of sale, such as grocery stores, is fair. They are therefore willing to make a greater effort to buy milk for a fair milk price, so that accessibility of the milk vending machine is unimportant for the WTP. This is also indicated by a significant negative coefficient for the variable 'PR accessibility'.

Socio-economic characteristics only have a partially significant effect on consumers' WTP. Younger consumers have a higher WTP in the dataset considered, while the WTP increases with rising income. A significant correlation between gender, household size, and the environment in which the consumers currently live or grew up in the past cannot be shown. The share of raw milk out of total milk consumption also has no significant influence on WTP.

While most of the respondents participated in the survey online and anonymously, some were interviewed face-to-face on site. Here it can be seen that participants in an anonymous survey appear to be more willing to indicate a lower price for WTP than respondents who gave answers directly to the questioner. While macroeconomic conditions, particularly in terms of inflation rate, have changed steadily

Acronym	Independent variables	WTP (log)		Purchase frequency (log)	
		Coeff.	p	Coeff.	p
-	Intercept	5.994	0.000	-4.817	0.030
S_{iM}	Organic production (dummy)	0.196	0.000	0.153	0.404
S_{iM}	Raw milk price (log)	-	-	-0.638	0.219
S_{iM}	Offline advertising (dummy)	-	-	0.215	0.039
S_{iM}	Online advertising (dummy)	-	-	0.225	0.142
S_{iE}	Income per capita (log)	-0.143	0.003	0.074	0.730
S_{iE}	Rurality	0.023	0.026	-0.061	0.174
S_{iE}	Number of dairy farms (log)	-0.026	0.012	0.123	0.007
O_{iR}	PR fair milk price (log)	0.383	0.000	0.772	0.002
O_{iR}	PR supporting regional farmers (log)	0.009	0.883	-0.287	0.290
O_{iR}	PR raw milk preference (log)	0.003	0.897	-0.008	0.928
O_{iR}	PR accessibility (log)	-0.129	0.000	0.070	0.641
O_{iR}	PR opening hours (log)	0.032	0.262	0.064	0.597
O_{iR}	PR shopping experience (log)	0.015	0.388	-0.069	0.354
O_{iS}	Gender	0.018	0.327	-0.133	0.093
O_{iS}	Age (log)	-0.064	0.040	0.313	0.017
O_{iS}	Household size (log)	-0.010	0.698	0.092	0.382
O_{iS}	Household income (log)	0.062	0.003	-0.140	0.114
O_{iS}	Share of raw milk consumption (log)	-0.006	0.598	0.857	0.000
O_{iS}	Current environment (log)	-0.018	0.524	0.255	0.034
O_{iS}	Childhood environment (log)	0.015	0.391	0.056	0.439
Q_{iS}	On-site survey (dummy)	0.099	0.001	-0.107	0.390
Q_{iS}	Month of survey participation (log)	0.056	0.127	-0.598	0.000
	Degrees of Freedom	520		517	
	R^2	0.2125		0.4989	
	Adjusted R^2	0.1838		0.4777	
	McElroy R^2	0.3901			

Table 6.3: Regression results for the SUR model

Note: Explanatory variables on raw milk price and advertising have been removed from estimation equation on WTP due to potential reverse causality bias.

over the course of the survey, these do not appear to have any influence on customers' WTP for raw milk at vending machines, based on the variable 'month of survey participation'.

For the purchase frequency of raw milk from vending machines, the dummy variable 'organic production' has no significant effect, so consumers do not seem to differentiate between organic and conventional milk in their purchase decision. Additionally, the milk price paid has also no significant impact on the frequency of purchase. This contradicts the economic expectation of decreasing demand for a product with an increased price, and confirms that consumers are willing to pay more for raw milk and that their decision to buy is driven by an appreciation of its production rather than a low price.

For the variable 'advertising', a distinction is made between only offline and only online advertising by using dummy variables. The remaining farmers use both types of advertising to attract consumers, so there were no farmers among the respondents who do not engage in any advertising measures. Table 6.3 presents a significant positive coefficient for offline advertising, which shows that farm managers can address regular customers more effectively via offline advertising, such as signs on the farm or newspaper ads. For online advertising, there is no significant effect on consumer purchase frequency.

In Table 6.3, the statistical effects of the estimated coefficients from income in the region and rurality appear not significant, based on their p values. However, the number of dairy farms has a significant positive influence on consumers' purchase frequency. While this variable has a negative influence on WTP and thus on the appreciation of the product, it appears that in areas where there is a greater relationship with dairy farming, customers are not willing to pay a higher price for raw milk, but consume the product more regularly at the current price.

The main reason for regularly purchasing raw milk is, according to the positively significant estimated coefficient, to give the producer a fair milk price. This also confirms the above statement that consumers are not influenced by price, as their intention to buy is based more on the value of the production and thus on better compensation for the producer. Other PRs, such as support for regional agriculture, a preference for raw milk over pasteurised milk, the accessibility of the vending machine, its opening hours or the shopping experience, have no influence on the purchase frequency of the respondents, according to the model results shown in Table 6.3.

Socio-economic information on participants indicates that male consumers purchase raw milk more frequently than female consumers and that purchase frequency

increases with age. While household size and household income have no influence on purchase frequency, there is a significant positive correlation with the share of raw milk consumption out of total milk consumption. Furthermore, the frequency of raw milk purchases increases the more rural the respondents perceive their current environment to be. This confirms the statement in the context of the variable 'number of dairy farms' that consumers who have a greater connection with agriculture and the producer buy there more often.

The variable 'month of participation' shows a negatively significant coefficient, which can be explained by the fact that consumers who buy more frequently were made aware of the survey earlier or were made aware of it more often over time and therefore were more likely to participate in it.

6.6 Conclusions

According to the results of the SUR model, consumers of raw milk from vending machines would agree in theory to pay more for one litre of raw milk. Operators could therefore consider moderately increasing their prices, as a higher milk price does not negatively influence consumers' purchasing frequency. Indeed, the main reason for buying raw milk from vending machines rather than milk in supermarkets is to pay a fair price to the producer. Future advertising could therefore focus particularly on this aspect. It should be noted that regular purchasers are more likely to be reached through offline advertising. However, farmers should bear in mind that differences were observed between locations and consumer groups. These differences are particularly evident when comparing areas with different relationships with agriculture: consumers who have less personal connections with milk producers are willing to pay more for milk, but do not purchase as frequently at the locations. Farmers should therefore distinguish between buyers from more rural areas and urban consumers who are on average willing to pay more.

Chapter 7

General Discussion and Conclusions

The objective of this thesis is the analysis of the supply and demand of the German milk market. Chapters 3–6 address individual aspects of this issue, with each examining different but related research questions. Chapters 3 and 4 focus on milk production, or supply in the German milk market. Chapters 5 and 6 consider demand in the milk market, specifically the demand for raw milk via vending machines. In both thematic sections, predictions of future supply and demand are made, and the influence of various factors is examined.

In the following sections, the basic results of the supply and demand strands in the German milk market are summarised and placed in the context of existing literature; this is followed by a critical discussion of the applied data and methods. Conclusions are then drawn, and recommendations for potential measures are given. The chapter concludes with an outlook on further research directions.

7.1 Summary of the Results

Milk production can be modelled in various ways and has been the subject of numerous scientific studies since the definition of the lactation curve by Wood (1967) (Kolver & Muller, 1998; Mader et al., 2006; Brügemann et al., 2011, e.g.). Several reproductive, genetic, environmental, management, and nutritional parameters influence milk production and can be included in the analysis (Bach et al., 2008). In the present work, particular attention was paid to the relationship between milk production and weather conditions.

Comparing the respective R^2s for a test dataset, it is consistently shown for all model approaches that the milk quantity per farm can be predicted more accurately than for an individual animal. This differs from previous studies where both econometric approaches (Adediran et al., 2012, e.g.) and ML models (Murphy et al., 2014, e.g.) focused on the exact modelling of the lactation curve and thus on the milk production of the individual cow when predicting milk production. In contrast, Chapter 3 focuses on factors that affect the entire farm.

Farm management decisions can be short-term and operational, such as feed and farm management, or long-term and strategic, such as increasing the size of the farm or herd. Strategic decisions, in particular, have a significant impact on the amount of milk produced on farms. Thus, by dividing the farms in the dataset into farms with average growth and farms with large growth steps, it can be shown that ML methods can predict investment-driven milk production significantly more accurately than OLS models.

The differences between ML methods themselves in terms of prediction accuracy are marginal. However, they differ in terms of their practical use as the methods

require different amounts of computing power and time. The consideration of this parameter is also relevant given that it seems reasonable to extend the dataset with additional farms to increase accuracy, even in counties with few farms. While the GBM algorithm provides the most accurate prediction at all levels of aggregation (from individual farms to the whole dataset), the XGB algorithm can be identified as the most efficient method as it provides very accurate prediction accuracy and requires only a quarter of the computational time of the GBM algorithm. This confirms the advantages that T. Chen and Guestrin (2016) intended to achieve when developing this algorithm.

The comparison of the methods used at different levels of aggregation illustrates the above statements about prediction accuracy. Considering the individual farm, the county, and the whole dataset, the ML methods are more suitable than OLS for making the most accurate prediction for each new test dataset. The results thus confirm the statements of Storm et al. (2020) that ML methods can provide accurate predictions.

However, these methods focus less on the statistical properties of the parameters, thus making it difficult to interpret the results (Storm et al., 2020). While elasticities can be calculated from OLS results to represent marginal effects (Gujarati & Porter, 2009), there are no clear parameters for ML methods that can perfectly represent causal relationships. Although there are approaches such as calculating variable importance and calculating the influence of variables through SHAP values, interpretation is more difficult compared to linear regression models.

In Chapter 4, this application case is addressed by modelling the causal relationships in milk production using FE estimators so that seasonal supply elasticities can be estimated. The results confirm the influence of farm management decisions presented in Chapter 3. In particular, the number of animals has the greatest influence on the quantity of milk produced per farm compared to other factors. This influence can vary during the year. The estimated coefficients show that the quantity of milk per quarter per farm is least influenced by the number of animals kept in the autumn, suggesting that the number of lactating animals is lowest in this quarter. A similar seasonal pattern has been described in previous studies (Timlin et al., 2021, e.g.).

The effects of weather can be divided into direct and indirect effects. The subdivision assumes that the weather has a direct effect on animal metabolism and feed quality on pasture farms (Hill & Wall, 2015) and that the weather has an indirect effect on milk production by influencing feed quality and quantity (Megersa et al.,

2014). To assess the direct effect, the THI is considered over the course of the year. The results reveal significant effects, except in the autumn quarter. Thus, increasing THI has a positive effect on milk production. This seems to contradict the statements of Mader et al. (2006) that heat stress has a negative effect on animals' metabolism. However, it should be noted that Chapter 4 does not consider extreme values but average values. Schlenker and Roberts (2009) indirectly confirmed this positive effect in their study by demonstrating an increasing yield of maize and soybean with increasing temperature in a flexible non-linear model. Only above a certain temperature limit do negative effects on forage quality occur. This positive effect is also discussed in Chapter 4, although the modelling in a linear regression model does not explicitly clarify the negative effect that occurs later.

Similarly, the influence of the water balance index SPEI is significantly lower than the influence of the number of animals kept, as measured by the size of the estimated coefficients. Furthermore, the effect is not consistently significant in all quarters. It seems that a low SPEI in autumn can lead to increased milk production in autumn and winter. This contradicts the findings of previous studies (Qi et al., 2015, e.g.) that hot and dry weather conditions generally have a negative effect on milk production. However, it should again be noted that average relationships were estimated and do not refer to extreme weather events. Contrary to the expectations of the relationships presented in the theoretical framework, spring weather conditions do not seem to have a significant effect on seasonal milk production.

Chapter 4 also presents the relationships between milk production and milk components. It is determined that in the warmer quarters (spring and summer) the increased somatic cell count of the animals has a stronger negative influence on milk production than in quarters with moderate temperatures and correspondingly lower disease pressure on the animals. This confirms previous results showing a negative effect of increasing THI on somatic cell count (Sanker et al., 2013).

In addition, the estimated coefficients for the urea content of the animals suggest that the protein-energy ratio of the diet may become unbalanced in the summer and autumn quarters, with negative effects on milk production. This correlation was also shown by Steinwidder et al. (2008), especially in late summer and autumn on farms rearing animals on grass-based diets.

While Chapters 3 and 4 consider the supply side of the German milk market, Chapters 5 and 6 address the demand side. Both chapters focus on the demand for raw milk from vending machines, which are increasingly used by farmers as an additional source of income. As this way of marketing milk is relatively new

and many vending machines shut down within a few years, farmers are uncertain whether such an investment is profitable for their business. Chapter 5 therefore forecasts possible sales volumes to answer the question of which location factors favour the sale of raw milk through vending machines as well as in which locations and for which farms the investment is worthwhile.

The results demonstrate that the data and model structure presented allow an accurate prediction of raw milk sales through vending machines. Farmers can therefore apply such a model to their own data to support their investment decisions. In addition, individual marketing activities have a rather small effect on the prediction of sales volumes. Marketing activities therefore need to be tailored to the individual business. A clearly significant correlation for different measures, as presented in theoretical studies (Wirthgen & Maurer, 2000, e.g.), cannot be established here. Exceptions are the listing of the vending machine on Google Maps and the offer of additional products via a vending machine, both of which have a positive effect on raw milk sales.

The location of the raw milk vending machine, which is determined by the farm location, plays an important role. In fact, a number of location factors influence raw milk sales. In particular, the number of potential customers, the accessibility of the nearest supermarket, the rurality of the location, and the proximity to a city have a significant impact on raw milk sales. Findings from similar studies, such as Wille et al. (2018), on the influence of location on direct marketing can therefore also be confirmed for the direct marketing of raw milk via vending machines. The fact that location along a road with commuter traffic has a positive influence, whereas location in a recreational and tourist area has a less significant influence on raw milk sales, suggests that consumers of raw milk are more likely to be regular buyers and less likely to be occasional buyers who became aware of the vending machine during an excursion or holiday.

DPs of individual SHAP values illustrate relationships within the heterogeneous data structure. This allows non-linear relationships to be represented graphically, in contrast to OLS regressions. A number of non-linear relationships become apparent when looking at farm-specific structures. The positive influence on raw milk sales initially increases with the age or experience of the farm manager, which is in line with previous literature (De Wolf et al., 2007). This effect changes, and raw milk sales decrease at an age when farm managers may be less open to new developments.

The algorithm can also infer the potential sales volume of raw milk from the milk yield per cow. The model divides farms into herds with a milk yield below or above approximately 7,000 litres per year. Farms with a lower milk yield are likely

to have significantly lower sales, suggesting that these farms are less interested in pure sales figures than in other objectives when opening and operating a raw milk vending machine. This is also confirmed by the explicit question about the motives for opening a vending machine. If farm managers are motivated by mainly financial reasons, this would have a positive impact on raw milk sales. By contrast, if farm managers had purely image-related reasons, there would be a negative influence on sales figures because it appears that the focus is less on pure profit and more on public relations or other image-related motives for the farm.

Whereas Chapter 5 considers the demand for raw milk from vending machines mainly from the farmers' point of view, Chapter 6 examines the consumers' point of view. The results of the SUR modelling support that the main reason for consumers to buy raw milk from vending machines is to support the producer financially. This reason has a significant positive correlation with both WTP and purchase frequency. It can therefore be concluded that the stimuli presented in the literature (Kotler et al., 2007; Kroeber-Riel & Weinberg, 2019) are not congruent with the marketing of this product.

When buying raw milk from vending machines, customers focus less on price characteristics and more on producer support. On average, the WTP is higher than the current milk price, and the current milk price has no significant influence on the frequency of purchase. In addition, one would expect that an increase in the number of dairy farms in the region due to increased competition would have a negative impact on consumer purchase frequency (Mankiw et al., 2008). However, there is a significant positive correlation, suggesting that consumers who have more interactions with milk production are more likely to support it financially.

The analysis of these variables also illustrates that the factors that influence purchase frequency and WTP are, in some cases, fundamentally different. The number of dairy farms in a region has a positive effect on purchase frequency but a negative effect on WTP. Thus, consumers with less connection to milk production still have a higher appreciation of the product. Moreover, as household income increases, so does WTP, as theoretically expected (Kotler et al., 2007). In addition, organic production is valued more by consumers due to its higher WTP. A higher value is placed on the product characteristics of organic production so that consumers' WTP increases, as theorised in Kroeber-Riel and Weinberg (2019).

Further contrasting effects of the purchase frequency and WTP stimuli can be observed when considering the age of the consumer. Purchase frequency increases with age, while WTP decreases. This also explains why there is a significant positive

correlation between purchase frequency and offline advertising. For example, sign or newspaper advertising is on average more likely than online advertising to reach consumers who buy more frequently and who tend to be older. This highlights the need for familiarity with the target group in order to reach them with marketing stimuli. Internet affinity in Germany declines with increasing age (Initiative D21, 2022), so it seems sensible to use offline advertising to reach this target group.

7.2 Applied Data and Methodological Approaches

The models presented, their results, and their conclusions can only be as good as the data and methods used allow. These elements are therefore discussed below.

Separate datasets were compiled for the analysis of supply and demand on the German milk market. The dataset for supply is a panel dataset, which made it possible to observe milk production of the farms under consideration over a longer period of time. The use of panel data allows for individual heterogeneity to be controlled for, thereby reducing the risk of biased estimators (Hsiao, 2022). In addition, panel data offer more variability and less collinearity due to the wealth of information they cover (Klevmarken, 1989). However, the wealth of the data can also be a complicating factor as consistently rich data need to be collected over the period of interest (Baltagi, 2008). For supply modelling, milk production data is collected consistently through a monthly performance test by the LKV, but the farmer has the option of removing one of the 12 measured values per year. This both results in one missing monthly value per year and could also lead to bias as a month is not removed at random; rather, the farmer can selectively delete a qualitatively or quantitatively worse month. In Chapter 3 and 4, the missing monthly value is filled by linear imputation because there is a trade-off between an unbalanced panel and a balanced panel with a limited number of imputed values. Another alternative would be to aggregate the available data at an annual observation level. However, this option was not considered here because of the desire to show seasonal variations and especially correlations. In Chapter 4, data were aggregated by summing daily and monthly values at the quarterly level. Although this type of aggregation can potentially obscure some effects, it serves as a compromise between obtaining reasonably accurate data and developing models that remain interpretable.

It must also be critically discussed that monthly milk yield is not given as the observed milk yield per month but is extrapolated to the whole month by sampling one measurement per month. Daily milk yield can also vary during the month (Timlin et al., 2021). However, as the milk yield check is carried out on every animal on the farm, it can be assumed that the effect of the daily variation of the

total number of animals over the month should be cancelled out.

The applied weather data originate from the regional reanalysis of the DWD. The analyses combine numerical weather models with observations from stations and can thus provide high-resolution weather data on a grid of 6 km x 6 km (Bollmeyer et al., 2015). This allows individual weather data to be assigned to each farm. Compared to using station data, more accurate data can be assigned to the farm. However, it should be noted that this accuracy can be deceptive. For example, although the weather affects the metabolism of the animals and the quality of the grassland at the farm location, additional forage areas may be located further away from the farm location and may experience different weather conditions.

Simplification is also necessary for input and output prices as these are not available for each individual farm. However, the average prices at the federal and state levels are only intended to represent economic trends, so farm-specific price variations would only complicate the model unnecessarily.

The modelling of demand for raw milk from vending machines is mainly based on surveys carried out throughout Germany. Both operators of raw milk vending machines and consumers of raw milk were interviewed. When carrying out a survey, a number of problems can arise that have to be considered in the further analysis.

As it was not possible to survey all farmers with a raw milk vending machine, a random sample was taken that represented a selection of the total population of raw milk vending machines to allow general conclusions to be drawn (Häder & Häder, 2022). A random sample can be assumed for the selection of farmers surveyed given that all operators of raw milk filling stations were contacted and a random sample responded. However, it should be noted that more motivated and particularly better-performing farmers may have reported back using this procedure.

The feedback from consumers was also random, as it was possible to participate in an online survey throughout Germany via a QR code at raw milk vending machines. Nevertheless, it is important to mention that demographic differences in income, education, gender, and age may exist between the total population and web respondents in online surveys (Couper & Coutts, 2006). For this reason, consumers were also interviewed on-site at selected locations. In contrast to the online survey locations, the on-site surveys tended to be spread across the south-west of Germany, so a geographical focus was chosen here for practical reasons. As a standalone database, the on-site survey would therefore have a geographical bias, but in combination with the online survey, it can still be considered an appropriate sampling design.

Furthermore, the model was modified by adding a dummy variable to distinguish between online and on-site surveys. This allows the online effect to be represented by a variable in the model. This is particularly relevant with regard to possible response bias due to socially desirable answers (Esser, 1991). Questions about consumers' WTP in particular could have been biased in the non-anonymous on-site interviews compared to the anonymous online responses. This effect can be represented by the dummy variable.

Another confounding factor in a quantitative survey can arise from spatial or temporal differences in the respective surveys (Schnell et al., 2018). While the spatial differences are intentional for modelling purposes in order to represent the heterogeneous structure of consumers distributed across Germany, the macroeconomic environment has constantly changed over the period of the consumer survey. Of particular note, rising inflation rates may have influenced consumer responses. To deal with this problem, a dummy variable was added to the model to correct for the month of observation.

Another possible source of bias in the results is the modification of the dependent variable 'purchase frequency'. The data originally collected was measured on an ordinal scale but was then converted to a nominal scale during the modelling process. As a result, the exact number of monthly purchases is only an approximation. As purchase frequency is based on consumer reports rather than direct observation, there is a potential for inaccuracy in the reported values. While the use of the nominal value implies an accurate measurement, it should be noted that the ordinal value is also an approximate estimate of purchase frequency. Therefore, the difference in accuracy between ordinal and nominal information is relatively small. However, to apply the SUR modelling technique, it is necessary to use the nominal value.

Additional data on location factors was provided by the statistical offices of the federal states and the Thünen-Landatlas. It should be noted that not all data could be used at the same level of observation. For example, not all variables are available at the municipality level, so some factors were considered at the next higher level of aggregation, namely the county level. A trade-off had to be made between using the same level but aggregating at a higher level or accepting that the levels of aggregation of the respective variables were different. As raw milk is an unprocessed product to be consumed or processed immediately after purchase, the catchment areas for raw milk vending machines are likely to be smaller. Therefore, it is preferable to use lower aggregation levels. In any case, the level of consideration should be taken into account when interpreting the results. There is also a difference in the year of collection of the variables. Due to the different data

sources, it was difficult to find the same survey year for all variables, so for each variable, the closest year to the year of observation of the dependent variable (2019) was chosen. This effect must also be taken into account when considering the results.

The analysis of the German milk market was conducted using ML methods in particular. The advantages of ML methods over econometric models lie primarily in their predictive accuracy (Breiman, 2001b). These differences in the accuracy of the models could be shown in the chapters 3 and 5. ML models are more flexible and are not bound to the strict framework of a linear economic model (Athey & Imbens, 2019). Unlike econometric models, there is no need to make inflexible assumptions about possible relationships (Storm et al., 2020). This is particularly advantageous when considering relationships that depend on biological or social processes, which in most cases are far from linear.

Data-driven approaches therefore allow complex data structures to be processed (Goodfellow et al., 2016). Chapter 3 demonstrated that non-linear trends in milk production in particular can be predicted more accurately using ML methods. In addition, ML methods are more scalable and can be more easily extended with additional data (Athey & Imbens, 2019). However, this is only partly evident for predicting milk production because certain algorithms require extensive training due to their hyperparameter structure. In addition to having a negative impact on training time, hyperparameter structure can also have a significant impact on model overfitting. With poor or no hyperparameter tuning, ML methods face an increased risk that models perform very well on the training dataset but do not generalise well and cannot accurately predict using new data (Yang & Shami, 2020). Less complex structures in econometric models can reduce the risk of overfitting but predict less accurately. This is also evident in Chapter 3 regarding the different R^2s for the models with milk production per animal instead of per farm as the dependent variable. The R^2s for the training and test datasets show a significant difference for the ML methods, whereas the differences for the linear approaches are relatively small. Thus, the ML models are slightly overfitted, but their accuracies are still better than the econometric models.

The more complex structures of ML models also complicate the results interpretation. While in econometric models statistical parameters can be evaluated to show causal relationships, many ML methods produce black box models whose parameters are difficult to interpret (Rudin, 2019). It was therefore quite complex to present the influence of each parameter on milk production in a sophisticated way in Chapter 3. In particular, the comparison between the different ML models is

complicated because not all models offer the same possibilities for interpretation. In this context, the use of a linear regression model, as in Chapter 4, is superior to the ML methods for delineating causal relationships between seasonal factors and seasonal milk production.

By contrast, Chapter 5 deals with only one method, the XGB algorithm, so that the influence of the parameters of the ML method can be interpreted using SHAP values. These values quantify the importance of individual features by measuring their contribution to the model's predictions (Lundberg and Lee, 2017). However, they are not intended to be used to draw causal conclusions but rather provide an overview of the importance of the variables in the prediction outcome. SHAP values assume that the model being explained is reliable and trustworthy. If the underlying model is poorly trained, biased, or suffers from overfitting, the explanations derived from SHAP values may not accurately reflect the true importance of the features. It is essential to ensure that the model itself is robust and reliable before relying on SHAP values for interpretation. For example, in Chapter 5, a clear difference in the R^2 of the training and test datasets shows that overfitting is present. Furthermore, due to the method of data collection, not all influencing factors could be included in the model. Both omitted variables and unobserved heterogeneity could therefore have biased the results of the SHAP values. The variables used may have been attributed a greater influence within the model than they exert in reality. Applying an ML model is therefore a fair compromise in this case because, although some bias may have occurred, the accuracy and representation of non-linear relationships are still advantageous compared to an econometric approach.

Both econometric methods and ML models are highly dependent on data quality. Potential problems are handled differently depending on the approach. ML methods recognise the structure of existing data based on existing patterns and can make predictions for new data based on this (Mullainathan & Spiess, 2017). The composition of the data structure of the training dataset must therefore be representative of the population so that a random new sample of the test data has similar structures and enables high prediction accuracy. This can be a major challenge, as discussed above in this chapter. In econometric approaches, it is possible to apply specific model specifications to specific data structures. For example, in Chapter 4, the panel structure was estimated by FE models, allowing systematic bias due to different farms to be accounted for in the analysis. In addition, Chapter 6 supports that econometric models can also be adjusted for bivariate dependent variables. Specially adapted econometric models exist for a large number of cases, such as time series analysis or endogenous variables.

When evaluating ML and econometric approaches, a clear distinction needs to be made between applications. If the goal is to make the most accurate prediction possible, ML methods are much more accurate, flexible, and efficient. If the focus is instead on causal relationships, econometric models can clarify the relationships more simply and clearly.

7.3 Conclusions and Recommendations

The results confirm that reliable and accurate supply and demand predictions can be achieved for the German milk market. The demonstrated forecasting capability, using the established data and model structure, provides a solid basis for integrating these models into publicly available tools to enable market participants to reap the benefits of predictive analytics in their decision-making processes. This integration would facilitate access to predictive capabilities, promote informed decision-making, and improve overall market efficiency.

An illustrative example of this concept can be found on the website www.mipro.uni-kiel.de, which was developed as part of the third-party project MIPRO ('Development of a regional milk quantity forecasting tool for Lower Saxony') funded by the Ministry of Food, Agriculture and Consumer Protection Lower Saxony. Interested stakeholders, such as farmers or dairies, can use this website by entering their annual milk volume and farm location. Based on this information, the tool generates a personalised forecast of milk quantity for the next 12 months that is specific to each farm. Users can also fine-tune the model's parameters, expressed as percentages, and observe the resulting impact on the monthly milk production trend through visual graphs. This interactive feature provides users with valuable insight into expected milk production levels and the potential influence of various factors, thus assisting in future decision-making and enabling farmers to better balance operational and strategic decisions as well as effectively manage market volatility. Furthermore, if a significant number of farms adjust their production in line with forecast expectations, this can result in benefits for both producers and consumers. These benefits include a reduction in price volatility as equilibrium prices are reached more quickly. In addition, the availability of improved information allows for a more resource-efficient use of inputs in milk production, minimising the production of surplus quantities that may not be immediately needed by the market.

The effectiveness of the additional information provided by the model depends on the regular and meaningful use of the platform by farmers and their ability to incorporate the information gained into their decision-making. Therefore, it would

be beneficial to receive political support to promote the dissemination of this forecasting tool. In addition, to ensure the accuracy of the model in the long term, the integrity of the underlying data structure needs to be maintained. Currently, the model is trained on data up to September 2019, and keeping the data up-to-date is crucial to improving accuracy by incorporating short-term developments. In addition, expanding the dataset both spatially and substantively could increase the impact of the model, but securing the necessary funding for this expansion would require political support.

An analogous application could be envisaged for forecasting raw milk sales. As illustrated in Chapter 5, the sales volume of raw milk through vending machines can be predicted with a remarkable degree of accuracy by using specific farm data and corresponding location data. Consequently, this information could be incorporated into farmers' investment considerations. However, to facilitate widespread adoption, it is essential that the tool is made available to farmers free of charge. This will require financial and therefore political support. By using the information obtained, farmers can make more informed decisions, potentially leading to an increase in the establishment of profitable milk filling stations and a reduction in the number of unsuccessful raw milk vending machines. Such a scenario would promote the visibility of this sales channel, which would subsequently have a positive impact on the entire sector.

The development of these practical applications is particularly promising through the methods of ML. The higher accuracy of the methods favours their use over traditional methods. However, recommendations for action cannot be derived directly from the results of ML models. Methods are needed to interpret the prediction models and their parameters, such as the calculation of SHAP values. Alternatively, additional econometric models are estimated.

By estimating an FE model, the results in Chapter 4 indicate that farmers control their production mainly through operational decisions, while weather conditions have less influence on seasonal milk production. Farmers should therefore focus on effective feed management, maintaining udder health throughout the year, and other management decisions, regardless of weather conditions. These factors have a significant impact on milk production and can contribute to stable farm profitability.

The sales volume of raw milk through vending machines is mainly determined by location and other prevailing factors. However, for hygiene reasons, the location of the vending machine is linked to the location of the farm. The disadvantages of location therefore have to be compensated for by marketing measures, which can be costly and thus reduce profits. One possible consideration would therefore be

to ensure the hygiene of the vending machine in a different way and, at the same time, to decouple the location of the vending machine from the location of the farm. A central and well-connected location could reach a larger group of consumers. Locations with a higher WTP, such as those closer to a city, could also be chosen to increase the farm's profits. However, in the case of a location off the farm, it should be noted that although the location factors are improved, at the same time the buying experience is changed as the unique selling proposition of buying directly from the producer is no longer given.

Operators of raw milk vending machines should consider implementing moderate price adjustments based on consumer preferences and market dynamics. According to the results, consumers are willing to pay higher prices to support local producers. Marketing campaigns should emphasise fair prices for producers to resonate with consumers and build brand loyalty. Differentiated marketing approaches tailored to the specific needs and characteristics of rural and urban consumers can increase the effectiveness of promotional efforts.

A comprehensive understanding of consumer behaviour, based on factors such as their level of agricultural involvement, geographical location, and socio-economic characteristics, is essential for developing targeted marketing strategies for raw milk vending machines. For example, results show that offline advertising channels are more effective in reaching regular purchasers, who tend to have stronger connections to agricultural production. In addition, urban consumers are more likely to pay premium prices for raw milk from vending machines. By leveraging these consumer insights, operators can create compelling marketing messages and campaigns that resonate with their target audiences, resulting in increased customer engagement and improved marketing results.

7.4 Outlook and Future Research

The models' dependence on data quality has already been clearly demonstrated. To ensure accurate forecasting of milk production and demand for raw milk through vending machines in the coming years, future research should especially focus on improving data quality. This can improve existing models and create new research opportunities.

Improved data quality can be achieved by regularly updating the supply forecasting database to reflect short-term market developments. In addition, expanding the dataset to include more farms or whole regions could improve forecast accuracy, although the marginal utility may decrease as the number of farms within a region increases.

Apart from the quantitative expansion of the dataset, the qualitative expansion of the data also promises additional benefits. Distinguishing between the two types of growth has shown that the models can, in principle, predict differently for each group. It is therefore important to be able to make this distinction more precisely so that the algorithm can be better adapted to each growth group. For this purpose, future research should question farm managers about their past operational decisions and especially their willingness to take risks. An overview of farmers' basic attitudes could be a possible variable to improve the differentiation between the two groups and thus improve the prediction. In addition to the distinction between the two growth groups, a comprehensive survey of farmers can provide valuable insights into their management decisions. Milk production has been found to be significantly influenced by these operational decisions. Integrating these parameters into the forecasting model and examining the causal relationships for seasonal milk production offers the potential to contribute to a more efficient use of resources within the German milk market. Just adding information about conventional or organic production, pasture farming, or indoor farming could provide significant added value.

The investigation of the causal relationships of seasonal milk production is currently based on average relationships. Although very high-resolution data is already included into the model through daily weather observations, the data is still summarised as a sum at a monthly or quarterly observation level. Future research should go beyond average observations and examine the impact of extreme weather conditions. As climate change progresses, extreme weather events will become increasingly important. Understanding their impact on milk production can facilitate better risk management and adaptation strategies for dairy farmers. In this respect, the limitations of linear regression models as well as of existing data structures in capturing the complex relationship between milk production and weather conditions should be recognised. Future research should explore additional non-linear models to gain a more accurate understanding of the relationship between weather and milk production.

To improve the raw milk demand database, an increase in the number of participants is recommended to increase the robustness of the results. Furthermore, the geographical focus of on-site surveys can be overcome by conducting a national on-site survey at vending machines, thus eliminating location-specific biases. In addition, complementing the WTP analysis with a choice experiment would allow for the study of consumers' actual behaviour rather than relying solely on their stated choices.

Furthermore, the current interviews were limited to existing customers of raw milk from vending machines, and it would be valuable to include interviews with non-consumers of raw milk. Understanding the motivations and concerns of non-consumers can help identify unrealised market potential and inform strategies that appeal to a wider range of consumers, thereby supporting the growth and viability of the raw milk vending machine market.

Demand prediction for raw milk from vending machines was based on sales data and influencing factors from 2019. Expanding the dataset over several years to create a panel dataset promises to both validate the currently presented factors and improve the accuracy of the forecasting model. Again, the data was only reported by farmers and would be more accurate if measured. In addition, the number of surveyed farmers with raw milk vending machines could be increased, but dairy farms without raw milk vending could also be surveyed. In this way, a two-stage model could be developed in which the first step would be to model the factors influencing farmers to start selling raw milk from vending machines, and the second step would be to model the expected sales level.

In light of the above conclusions, raw milk filling stations should be allowed outside dairy farms to reach a larger and more affluent clientele. The first question is how to ensure and control the hygiene of the vending machines. Secondly, future research should investigate whether the factors influencing the sale of raw milk away from a farm would change given that the change of location would massively alter the shopping experience.

References

Adediran, S., Ratkowsky, D., Donaghy, D., & Malau-Aduli, A. (2012). Comparative evaluation of a new lactation curve model for pasture-based holstein-friesian dairy cows. *Journal of Dairy Science*, *95*(9), 5344–5356.

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration – guidelines for computing crop water requirements – FAO irrigation and drainage paper 56. *FAO, Rome*, *300*(9).

Altemeier, K. (1981). Gleichgewichtssimulationen für den Milchmarkt der BR Deutschland. *German Journal of Agricultural Economics/Agrarwirtschaft*, *30*(9), 257–266.

Athey, S. (2018). The impact of machine learning on economics. In *The economics of artificial intelligence: An agenda* (pp. 507–547). University of Chicago Press.

Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, *11*, 685–725.

Bach, A., Valls, N., Solans, A., & Torrent, T. (2008). Associations between nondietary factors and dairy herd performance. *Journal of Dairy Science*, *91*(8), 3259–3267.

Backhaus, K., Erichson, B., & Weiber, R. (2015). *Fortgeschrittene multivariate Analysemethoden: Eine anwendungsorientierte Einführung* (Vol. 3). Springer.

Baldinger, L., Zollitsch, W., & Knaus, W. (2014). Maize silage and italian ryegrass silage as high-energy forages in organic dairy cow diets: Differences in feed intake, milk yield and quality, and nitrogen efficiency. *Renewable Agriculture and Food Systems*, *29*(4), 378–387.

Baltagi, B. H. (2008). *Econometric analysis of panel data* (Vol. 4). John Wiley & Sons.

Baudracco, J., Lopez-Villalobos, N., Holmes, C., Comeron, E., Macdonald, K., Barry, T., & Friggens, N. C. (2012). e-cow: an animal model that predicts herbage intake, milk yield and live weight change in dairy cows grazing temperate pastures, with and without supplementary feeding. *Animal*, *6*(6), 980–993.

Baumol, W. J., & Blinder, A. S. (2010). *Economics: Principles and policy* (Vol. 11). South-Western Cengage Learning.

Baylis, K., Heckelei, T., & Storm, H. (2021). Machine learning in agricultural economics. In *Handbook of agricultural economics* (Vol. 5, pp. 4551–4612). Elsevier.

Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. In *Neural networks: Tricks of the trade* (Vol. 2, pp. 437–478). Springer.

Blayney, D. P., & Mittelhammer, R. C. (1990). Decomposition of milk supply response into technology and price-induced effects. *American Journal of Agricultural Economics*, *72*(4), 864–872.

BLE. (2020). *Bericht zu Markt- und Versorgungslage mit Milch und Milcherzeugnissen.* Bericht der Bundesanstalt für Ernährung und Landwirtschaft.

BLE. (2022). *Erzeugung von Milch, Milchfett und Milcheiweiß in Deutschland nach Kalenderjahren.* Bericht der Bundesanstalt für Ernährung und Landwirtschaft.

BLE. (2023a). *Milchwirtschaft auf einen Blick in Deutschland nach Kalenderjahren.* Retrieved from `https://www.ble.de/SharedDocs/Downloads/DE/BZL/Daten-Berichte/MilchUndMilcherzeugnisse/JaehrlicheErgebnisse/Deutschland/Dt_Grundlagen/Milchwirtschaft_auf_einen_Blick_406003001_11.html` (Last accessed on 13-04-2023)

BLE. (2023b). *Pressemitteilung: Milchmarkt 2021: Pro-Kopf-Verbrauch von Milch sinkt auf 47,8 Kilogramm.* Retrieved from `https://www.ble.de/SharedDocs/Pressemitteilungen/DE/2022/220502_Milchmarkt_2021.html` (Last accessed on 13-04-2023)

BMEL. (2018). *Deutschland wie es isst.* Der BMEL Ernährungsreport 2018.

BMEL. (2020). *Produktionswert des Bereichs Landwirtschaft, landwirtschaftliche Gesamtrechnung.* Bericht des Bundesministerium für Ernährung und Landwirtschaft.

BMEL. (2023a). *Preise für konventionell erzeugte Kuhmilch.* Statistische Monatsberichte, 2007-2022.

BMEL. (2023b). *Preise für ökologisch/biologisch erzeugte Kuhmilch (MBT-0301440-0000).* Bericht des Bundesministerium für Ernährung und Landwirtschaft.

BMEL. (2023c). *Verkaufserlöse der Landwirtschaft nach Erzeugnissen.* Retrieved from `https://www.bmel-statistik.de/landwirtschaft/landwirtschaftliche-gesamtrechnung/verkaufserloes` (Last accessed on 31-03-2023)

Bollmeyer, C., Keller, J. D., Ohlwein, C., Wahl, S., Crewell, S., Friederichs, P., ... Steinke, S. (2015). Towards a high-resolution regional reanalysis for the European CORDEX domain. *Quarterly Journal of the Royal Meteorological Society*, *141*(686), 1–15.

Borchert Kommission. (2020). *Empfehlungen des Kompetenznetzwerks Nutztierhaltung.* Bericht der Borchert Kommission.

Bouamra-Mechemache, Z., Jongeneel, R., & Réquillart, V. (2008). Impact of a gradual increase in milk quotas on the EU dairy sector. *European Review of Agricultural Economics*, *35*(4), 461–491.

Bovo, M., Agrusti, M., Benni, S., Torreggiani, D., & Tassinari, P. (2021). Random forest modelling of milk yield of dairy cows under heat stress conditions. *Animals*, *11*(5), 1305–1323.

Bragg, L. A., & Dalton, T. J. (2004). Factors affecting the decision to exit dairy farming: a two-stage regression analysis. *Journal of Dairy Science*, *87*(9), 3092–3098.

Breiman, L. (2001a). Random forests. *Machine Learning*, *45*, 5–32.

Breiman, L. (2001b). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), 199–231.

Breiman, L., Friedman, J. H., Stone, C., & Olshen, R. (1984). *Classification and regression trees.* Taylor & Francis.

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society*, 1287–1294.

Brügemann, K., Gernand, E., Von Borstel, U., & König, S. (2011). Genetic analyses of protein yield in dairy cows applying random regression models with time-dependent and temperature x humidity-dependent covariates. *Journal of Dairy Science*, *94*(8), 4129–4139.

Brunetti, A., Wollmershäuser, T., Bandholz, H., Becker, N., Fahimi, Y., Lesch, H., & Schnabl, G. (2022). Inflation auf Rekordniveau – wie kann die Politik die Folgen abfedern? *ifo Schnelldienst*, *75*(7), 3–25.

Burgess, A. J., Masclaux-Daubresse, C., Strittmatter, G., Weber, A. P., Taylor, S. H., Harbinson, J., ... Baekelandt, A. (2023). Improving crop yield potential: Underlying biological processes and future prospects. *Food and Energy Security*, *12*(1).

Buza, M., Holden, L., White, R., & Ishler, V. (2014). Evaluating the effect of ration composition on income over feed cost and milk yield. *Journal of Dairy Science*, *97*(5), 3073–3080.

Calker, K. J. V., Berentsen, P. B., Giesen, G. W., & Huirne, R. B. (2005). Identifying and ranking attributes that determine sustainability in dutch dairy farming. *Agriculture and Human Values*, *22*, 53–63.

Chen, S., & Lin, C.-P. (2019). Understanding the effect of social media marketing activities: The mediation of social identification, perceived value, and satisfaction. *Technological Forecasting and Social Change*, *140*, 22–32.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794).

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., ... Li, Y. (2020). xgboost: Extreme gradient boosting [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=xgboost`

Cherney, J., & Picasso Risso, V. (2020). Systems for temperature humid areas. In *Forages, the science of grassland agriculture* (Vol. 2). John Wiley & Sons.

Coble, K. H., Mishra, A. K., Ferrell, S., & Griffin, T. (2018). Big data in agriculture: A challenge for the future. *Applied Economic Perspectives and Policy*, *40*(1), 79–96.

Collard, B., Boettcher, P., Dekkers, J. M., Petitclerc, D., & Schaeffer, L. (2000). Relationships between energy balance and health traits of dairy cattle in early lactation. *Journal of Dairy Science*, *83*(11), 2683–2690.

Couper, M. P., & Coutts, E. (2006). Online-Befragung. Probleme und Chancen verschiedener Arten von Online-Erhebungen. In *Methoden der Sozialforschung* (pp. 217–243). Springer.

Croissant, Y., & Millo, G. (2008). Panel data econometrics in R: The plm package. *Journal of Statistical Software*, *27*(2), 1–43.

Destatis. (2023a). *Durchschnittlicher Kaufwert für landwirtschaftliche Grundstücke: Deutschland (bis 2020), Größenklassen der Ertragsmesszahlen, Größenklassen der veräußerten Fläche (61521-0100).* Retrieved from `https://www-genesis.destatis.de/genesis/online` (Last accessed on 13-04-2023)

Destatis. (2023b). *Index der Einkaufspreise landwirtschaftlicher Betriebsmittel: Deutschland, Jahre, Landwirtschaftliche Betriebsmittel (61221-0001).* Retrieved from `https://www-genesis.destatis.de/genesis/online` (Last accessed on 13-04-2023)

Destatis. (2023c). *Landwirtschaftliche Betriebe, Fläche: Deutschland, Jahre, Bodennutzungsarten (41141-0001).* Retrieved from `https://www-genesis.destatis.de/genesis/online` (Last accessed on 13-04-2023)

Destatis. (2023d). *Landwirtschaftliche Betriebe, Landwirtsch. genutzte Fläche (Rep. ASE): Deutschland, Jahre (bis 2007), Rechtsform, Größenklassen der landwirtschaftlich genutzten Fläche (41122-0001).* Retrieved from `https://www-genesis.destatis.de/genesis/online` (Last accessed on 13-04-2023)

Destatis. (2023e). *Landwirtschaftliche Betriebe mit gepachteten Flächen, Pachtentgelt: Bundesländer, Jahre, Bodennutzungsart, Rechtsformen, Größenklassen der landw. genutzten Fläche (41141-0110).* Retrieved from `https://www-genesis.destatis.de/genesis/online` (Last accessed on 13-04-2023)

Destatis. (2023f). *Landwirtschaftliche Betriebe mit ökologischem Landbau, Tiere in Betrieben mit ökologischem Landbau, Tiere in ökologischer Haltung: Deutschland, Stichtag, Tierarten (41141-0008).* Retrieved from `https://www-genesis.destatis.de/genesis/online` (Last accessed on 13-04-2023)

Destatis. (2023g). *Landwirtschaftliche Betriebe mit Viehhaltung, Viehbestand: Deutschland, Stichtag, Tierarten (41141-0004).* Retrieved from `https://www-genesis.destatis.de/genesis/online` (Last accessed on 13-04-2023)

Destatis. (2023h). *Laufende Wirtschaftsrechnungen (LWR) – Einkommen, Einnahmen und Ausgaben privater Haushalte nach den monatlichen Haushaltsnettoeinkommen 2021.* Retrieved from `https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Einkommen-Konsum-Lebensbedingungen/Einkommen-Einnahmen-Ausgaben/Tabellen/liste-haushaltsnettoeinkommen.html#115602` (Last accessed on 13-04-2023)

De Wolf, P., McElwee, G., & Schoorlemmer, H. (2007). The European farm entrepreneur: A comparative perspective. *International Journal of Entrepreneurship and Small Business*, *4*(6), 679–692.

Eck, M. A., Murray, A. R., Ward, A. R., & Konrad, C. E. (2020). Influence of growing season temperature and precipitation anomalies on crop yield in the southeastern United States. *Agricultural and Forest Meteorology*, *291*.

Efron, B., & Hastie, T. (2021). *Computer age statistical inference, student edition: algorithms, evidence, and data science* (Vol. 6). Cambridge University Press.

Emberger-Klein, A., Menrad, K., & Heider, D. (2016). Determinants of consumers' willingness-to-pay for fairly-produced, locally grown dairy products. *German Journal of Agricultural Economics*, *65*, 94–111.

Esser, H. (1991). Die Erklärung systematischer Fehler in Interviews. In *Person–Situation–Institution–Kultur* (pp. 59–78). Duncker & Humblot.

European Commission. (2020a). *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. A Farm to Fork Strategy for a fair, healthy and environmentally-friendly food system (EUR-Lex - 52020DC0381).* Communication from the European Commission.

European Commission. (2020b). *Working with parliament and council to make the CAP reform fit for the European Green Deal.* Communication from the European Commission.

European Commission. (2022). *Key reforms in the new cap. A fair deal for farmers.* Retrieved from `https://agriculture.ec.europa.eu/common-agricultural-policy/cap-overview/cap-2023-27/key-reforms-new-cap_en` (Last accessed on 31-03-2023)

European Environmental Agency. (2021). *Annual European Union greenhouse gas inventory 1990–2019 and inventory report 2021. Submission to the UNFCCC secretariat. (EEA/PUBL/2021/066).* Report of the European Environmental Agency.

Eurostat. (2023). *Cows' milk collection and products obtained - monthly data.* Retrieved from `https://ec.europa.eu/eurostat/databrowser/view/apro_mk_colm/default/table?lang=en` (Last accessed on 31-03-2023)

FAOSTAT. (2023). *Statistical database of the food and agriculture organization of the United Nations.* Retrieved from `https://www.fao.org/faostat/en/#data/QCL` (Last accessed on 31-03-2023)

Feldmann, C., & Hamm, U. (2015). Consumers' perceptions and preferences for local food: A review. *Food Quality and Preference*, *40*, 152–164.

Finger, R., Dalhaus, T., Allendorf, J., & Hirsch, S. (2018). Determinants of downside risk exposure of dairy farms. *European Review of Agricultural Economics*, *45*(4), 641–674.

Forbes, J. M. (1996). Integration of regulatory signals controlling forage intake in ruminants. *Journal of Animal Science*, *74*(12), 3029–3035.

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, *55*(1), 119–139.

Frick, F., & Sauer, J. (2018). Deregulation and productivity: Empirical evidence on dairy production. *American Journal of Agricultural Economics*, *100*(1), 354–378.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189–1232.

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, *38*(4), 367–378.

Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22.

Fuß, R., C, V., & C, R. (2023). *Zahlen Fakten – Treibhausgas-Emissionen aus der Landwirtschaft.* Retrieved from `https://www.thuenen.de/de/themenfelder/klima-und-luft/emissionsinventare-buchhaltung-fuer-den-klimaschutz/treibhausgas-emissionen-aus-der-landwirtschaft` (Last accessed on 31-03-2023)

Gaines, W. (1927). The energy basis of measuring milk yields. *Journal of Animal Science*, *1927*(1), 33–36.

Gauly, M., Bollwein, H., Breves, G., Brügemann, K., Dänicke, S., Daş, G., ... Wrenzycki, C. (2013). Future consequences and challenges for dairy cow production

systems arising from climate change in Central Europe – a review. *Animal*, *7*(5), 843–859.

Gisbert-Queral, M., Henningsen, A., Markussen, B., Niles, M. T., Kebreab, E., Rigden, A. J., & Mueller, N. D. (2021). Climate impacts and adaptation in US dairy systems 1981–2018. *Nature Food*, *2*(11), 894–901.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Grau, A., & Hockmann, H. (2018). Market power in the german dairy value chain. *Agribusiness*, *34*(1), 93–111.

Greene, W. (2002). *Econometric analysis* (Vol. 5). Pearson.

Gujarati, D. N., & Porter, D. C. (2009). *Basic econometrics* (Vol. 5). McGraw-Hill/Irwin.

Gustafsson, A. H., & Carlsson, J. (1993). Effects of silage quality, protein evaluation systems and milk urea content on milk yield and reproduction in dairy cows. *Livestock Production Science*, *37*(1-2), 91–105.

Häder, M., & Häder, S. (2022). Stichprobenziehung in der quantitativen Sozialforschung. In *Handbuch Methoden der empirischen Sozialforschung* (pp. 425–441). Springer.

Hagnestam-Nielsen, C., Emanuelson, U., Berglund, B., & Strandberg, E. (2009). Relationship between somatic cell count and milk yield in different stages of lactation. *Journal of Dairy Science*, *92*(7), 3124–3133.

Hallam, A. (1991). Economies of size and scale in agriculture: an interpretive review of empirical measurement. *Applied Economic Perspectives and Policy*, *13*(1), 155–172.

Hansen, J. W. (2002). Realizing the potential benefits of climate prediction to agriculture: Issues, approaches, challenges. *Agricultural Systems*, *74*(3), 309–330.

Harder, I., Stamer, E., Junge, W., & Thaller, G. (2019). Lactation curves and model evaluation for feed intake and energy balance in dairy cows. *Journal of Dairy Science*, *102*(8), 7204–7216.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). Springer.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the Econometric Society*, 1251–1271.

Hempel, C., & Hamm, U. (2016). Local and/or organic: A study on consumer preferences for organic food and food from different origins. *International Journal of Consumer Studies*, *40*(6), 732–741.

Henning, C., Witzke, P., Panknin, L., & M, G. (2021). *Ökonomische und Ökologische Auswirkungen des Green Deals in der Agrarwirtschaft.* Simulationsstudie der Effekte der F2F-Strategie auf Produktion, Handel, Einkommen und Umwelt mit dem CAPRI-Modell.

Henningsen, A., & Hamann, J. D. (2008). systemfit: A package for estimating systems of simultaneous equations in R. *Journal of Statistical Software*, *23*, 1–40.

Heskes, T., Sijben, E., Bucur, I. G., & Claassen, T. (2020). Causal shapley values: Exploiting causal knowledge to explain individual predictions of complex models. *Advances in Neural Information Processing Systems*, *33*, 4778–4789.

Hill, D. L., & Wall, E. (2015). Dairy cattle in a temperate climate: the effects of weather on milk yield and composition depend on management. *Animal*, *9*(1), 138–149.

Hill, D. L., & Wall, E. (2017). Weather influences feed intake and feed efficiency in a temperate climate. *Journal of Dairy Science*, *100*(3), 2240–2257.

Hoehl, S., & Hess, S. (2022). Liquidity, hedging and the survival of north german dairy farms. *European Review of Agricultural Economics*, *49*(1), 208–236.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67.

Hopkins, A., & Del Prado, A. (2007). Implications of climate change for grassland in Europe: impacts, adaptations and mitigation options: a review. *Grass and Forage Science*, *62*(2), 118–126.

Horstmann, F., & Schulze, B. (2011). Landwirtschaftliche Direktvermarktung: Neue Potenziale durch Social Media? In *Qualität und Effizienz durch informationsgestützte Landwirtschaft, Fokus: Moderne Weinwirtschaft* (pp. 97–100). Gesellschaft für Informatik e.V.

Hristov, A., Price, W., & Shafii, B. (2005). A meta-analysis on the relationship between intake of nutrients and body weight with milk volume and milk protein yield in dairy cows. *Journal of Dairy Science*, *88*(8), 2860–2869.

Hsiao, C. (2022). *Analysis of panel data* (Vol. 4). Cambridge University Press.

Initiative D21. (2022). *D21 Digital Index 2021/2022: Jährliches Lagebild zur Digitalen Gesellschaft.* Bericht der Initiative D21.

Injadat, M., Moubayed, A., Nassif, A. B., & Shami, A. (2020). Multi-split optimized bagging ensemble model selection for multi-class educational data mining. *Applied Intelligence*, *50*, 4506–4528.

Institut für Ernährungswirtschaft. (2023). *Kieler Rohstoffwert Milch.* Retrieved from `https://www.ife-kiel.de/kieler-rohstoffwert-milch/` (Last accessed on 13-06-2023)

IPCC. (2014). *Climate change 2014: impacts, adaptation, and vulnerability. Part A: Global and sectoral aspects.* Working Group II contribution to the fifth assessment report of the intergovernmental panel in climate change.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An introduction to statistical learning* (Vol. 8). Springer.

Jarzębowski, S., Bourlakis, M., & Bezat-Jarzębowska, A. (2020). Short food supply chains (sfsc) as local and sustainable systems. *Sustainability*, *12*(11), 4715.

Jensen, L., Nielsen, N., Nadeau, E., Markussen, B., & Nørgaard, P. (2015). Evaluation of five models predicting feed intake by dairy cows fed total mixed rations. *Livestock Science*, *176*, 91–103.

Kennedy, E., O'Donovan, M., Murphy, J.-P., Delaby, L., & O'Mara, F. (2005). Effects of grass pasture and concentrate-based feeding systems for spring-calving dairy cows in early spring on performance during lactation. *Grass and Forage Science*, *60*(3), 310–318.

Key, N., & Sneeringer, S. (2014). Potential effects of climate change on the productivity of us dairies. *American Journal of Agricultural Economics*, *96*(4), 1136–1156.

Klevmarken, N. A. (1989). Panel studies: what can we learn from them? *European Economic Review*, *33*, 523–529.

Knuck, J. (2020). Erfolgsfaktoren alternativer Vermarktungswege für Milch. *Austrian Journal of Agricultural Economics and Rural Studies*, *29*(6), 39–45.

Kolver, E., & Muller, L. (1998). Performance and nutrient intake of high producing holstein cows consuming pasture or a total mixed ration. *Journal of Dairy Science*, *81*(5), 1403–1411.

Köster, U. (2010). *Grundzüge der landwirtschaftlichen Marktlehre* (Vol. 4). Vahlen.

Kotler, P., Keller, K. L., & Bliemel, F. (2007). *Marketing-Management: Strategien für wertschaffendes Handeln* (Vol. 12). Pearson.

Koutsoukas, A., Monaghan, K. J., Li, X., & Huan, J. (2017). Deep-learning: investigating deep neural networks hyper-parameters and comparison of performance to shallow methods for modeling bioactivity data. *Journal of Cheminformatics*, *9*(1), 1–13.

Kraatz, S. (2012). Energy intensity in livestock operations – modeling of dairy farming systems in germany. *Agricultural Systems*, *110*, 90–106.

Kroeber-Riel, W., & Weinberg, P. (2019). *Konsumentenverhalten* (Vol. 11). Vahlen.

Kuhnert, H., & Wirthgen, B. (1997). Die Bedeutung der Direktvermarktung als Einkommensalternative für landwirtschaftliche Betriebe in der Bundesrepublik Deutschland: eine Situationsanalyse. *Schriftenreihe des Bundesministeriums für Ernährung, Landwirtschaft und Forsten*, *460*.

Langenberg, J., Schasse, I., & Theuvsen, L. (2017). Entwicklung des Bodenmarktes: Einflussfaktoren auf Kauf-und Pachtpreise. *Jahrbuch der Österreichischen Gesellschaft für Agrarökonomie*, *26*, 199–208.

Larcher, M., Schönhart, M., Schmid, E., & Vogel, S. (2015). Intensivierungspläne Österreichischer MilchproduzentInnen angesichts der auslaufenden Milchquotenregelung 2015 – ein empirisches Modell zur Erklärung von Verhaltensintentionen. *German Journal of Agricultural Economics*, *64*, 148–168.

La Trobe, H., & Acott, T. (2000). Localizing the global food system. *International Journal of Sustainable Development & World Ecology*, *7*(4), 309–320.

LeDell, E., Gill, N., Aiello, S., Fu, A., Candel, A., Click, C., ... Malohlava, M. (2021). h2o: R interface for the 'h2o' scalable machine learning platform [Computer

software manual]. Retrieved from `https://CRAN.R-project.org/package=h2o` (R package version 3.32.1.3)

Lesk, C., Rowhani, P., & Ramankutty, N. (2016). Influence of extreme weather disasters on global crop production. *Nature*, *529*(7584), 84–87.

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, *18*(8), 2674.

Liu, Y., & Just, A. (2020). Shapforxgboost: Shap plots for 'xgboost' [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=SHAPforxgboost`

Lobell, D. B., Hammer, G. L., McLean, G., Messina, C., Roberts, M. J., & Schlenker, W. (2013). The critical role of extreme heat for maize production in the United States. *Nature Climate Change*, *3*(5), 497–501.

Lones, M. A. (2021). How to avoid machine learning pitfalls: a guide for academic researchers. *arXiv preprint: 2108.02497*.

Lopez-Pumarejo, T. A., & Bassell, M. (2009). The renaissance of outdoor advertising: From harlem to hong kong. *American Journal of Business*, *24*(2), 33–40.

LSN. (2020). *Statistische Berichte Niedersachsen, CIII 1 – hj 1 / 2020.* Berichte des Landesamt für Statistik Niedersachsen.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., ... Lee, S.-I. (2020). From local explanations to global understanding with explainable ai for trees. *Nature Machine Intelligence*, *2*(1), 56–67.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*.

Mader, T. L., Davis, M., & Brown-Brandl, T. (2006). Environmental factors influencing heat stress in feedlot cattle. *Journal of Animal Science*, *84*(3), 712–719.

Mankiw, N. G., Taylor, M. P., Wagner, A., & Herrmann, M. (2008). *Grundzüge der Volkswirtschaftslehre* (Vol. 4). Schäffer-Poeschel Stuttgart.

Meffert, H., Burmann, C., Kirchgeorg, M., & Eisenbeiß, M. (2018). *Marketing: Grundlagen marktorientierter Unternehmensführung Konzepte–Instrumente–Praxisbeispiele* (Vol. 13). Springer.

Megersa, B., Markemann, A., Angassa, A., Ogutu, J. O., Piepho, H.-P., & Zaráte, A. V. (2014). Impacts of climate change and variability on cattle production in southern ethiopia: Perceptions and empirical evidence. *Agricultural Systems*, *130*, 23–34.

Mehlhose, C., Hunecke, C., Spiller, A., & Brümmer, B. (2019). Der Markt für Milch und Milcherzeugnisse im Jahr 2018. *German Journal of Agricultural Economics*, *68*, 62–84.

Mellin, M., Spiller, A., & Zühlsdorf, A. (2006). Kundenzufriedenheit in der Direktvermarktung. In *Das Unternehmen nachhaltige Ernährungskultur* (pp. 203–235). Metropolis-Verlag.

Mullainathan, S., & Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106.

Müller-Hagedorn, L., Toporowski, W., & Zielke, S. (2012). *Der Handel: Grundlagen-Management-Strategien* (Vol. 2). Kohlhammer Verlag.

Murphy, M. D., O'Mahony, M. J., Shalloo, L., French, P., & Upton, J. (2014). Comparison of modelling techniques for milk-production forecasting. *Journal of Dairy Science*, *97*(6), 3352–3363.

Neuenfeldt, S., Gocht, A., Heckelei, T., & Ciaian, P. (2019). Explaining farm structural change in the European agriculture: a novel analytical framework. *European Review of Agricultural Economics*, *46*(5), 713–768.

Nguyen, Q. T., Fouchereau, R., Frenod, E., Gerard, C., & Sincholle, V. (2020). Comparison of forecast models of production of dairy cows combining animal and diet parameters. *Computers and Electronics in Agriculture*, *170*.

Nielsen. (2020). *Plant-based foods in Europe: How big is the market? Smart protein plant-based food sector report by Smart Protein Project, European Union's horizon 2020 research and innovation programme (No 862957).* Report from the Smart Protein Project.

Nielsen, M. (2015). *Neural networks and deep learning* (Vol. 25). Determination Press San Francisco.

Ogutu, J. O., Schulz-Streeck, T., & Piepho, H.-P. (2012). Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. In *Bmc proceedings* (Vol. 6, pp. 1–6).

Parsons, D. J., Rey, D., Tanguy, M., & Holman, I. P. (2019). Regional variations in the link between drought indices and reported agricultural impacts of drought. *Agricultural Systems*, *173*, 119–129.

Pavlov, I. P. (1927). *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex*. Oxford University Press.

Pepels, W. (1994). *Werbung und Absatzförderung: Kommunikation, Werbeziel, Werbeobjekt, Werbebudget, Absatzquelle, Zielpersonengruppe, Kampagne, Marktsegmentierung, Werbemedien, Werbeeffizienz*. Springer.

Peremans, K., & Van Aelst, S. (2018). Robust inference for seemingly unrelated regression models. *Journal of Multivariate Analysis*, *167*, 212–224.

Perez-Mendez, J. A., Roibas, D., & Wall, A. (2019). The influence of weather conditions on dairy production. *Agricultural Economics*, *50*(2), 165–175.

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Taieb, S. B., ... Ziel, F. (2022). Forecasting: theory and practice. *International Journal of Forecasting*, *38*(3), 705–871.

Pottebaum, P. (1989). *Handbuch Direktvermarktung–Neue Wege für den Absatz landwirtschaftlicher Produkte* (Vol. 2). Landwirtschaftsverlag Münster.

Qi, L., Bravo-Ureta, B., & Cabrera, V. (2015). From cold to hot: Climatic effects and productivity in Wisconsin dairy farms. *Journal of Dairy Science*, *98*(12), 8664–8677.

R Core Team. (2021). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, *31*(2), 350–371.

Regionalstatistik. (2023). *Durchschnittsalter der Bevölkerung – Stichtag 31.12. – regionale Ebenen (12411-07-01-5-B)*. Retrieved from `https://www.regionalstatistik.de/genesis/online` (Last accessed on 13-04-2023)

Roche, J., Turner, L., Lee, J., Edmeades, D., Donaghy, D., Macdonald, K., ... Berry, D. (2009). Weather, herbage quality and milk production in pastoral systems. inter-relationships and associations between weather variables and herbage

growth rate, quality and mineral concentration. *Animal Production Science*, *49*(3), 211–221.

RStudio Team. (2020). Rstudio: Integrated development environment for R [Computer software manual]. Boston, MA. Retrieved from `http://www.rstudio.com/`

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215.

Rutz, C., Steil, C., von Kutzleben, N., Schramek, J., Gehrlein, U. D., & Schmidt, M. (2016). Bietet der demografische Wandel Chancen für die landwirtschaftliche Diversifizierung? *Berichte über Landwirtschaft – Zeitschrift für Agrarpolitik und Landwirtschaft*, *94*(3).

Sanker, C., Lambertz, C., & Gauly, M. (2013). Climatic effects in central Europe on the frequency of medical treatments of dairy cows. *Animal*, *7*(2), 316–321.

Schaper, C., Deimel, M., & Theuvsen, L. (2011). Determinanten der Wettbewerbsfähigkeit „erweiterter Familienbetriebe“ – Ergebnisse einer Betriebsleiterbefragung. *German Journal of Agricultural Economics*, *60*(1), 1–16.

Schapire, R. E. (1990). The strength of weak learnability. *Machine Learning*, *5*, 197–227.

Schapire, R. E. (1999). A brief introduction to boosting. In *Proceedings of the sixteenth international joint conference on artificial intelligence* (Vol. 16, pp. 1401–1406).

Schlenker, W., & Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to us crop yields under climate change. *Proceedings of the National Academy of Sciences*, *106*(37), 15594–15598.

Schnell, R., Hill, P. B., & Esser, E. (2018). *Methoden der empirischen Sozialforschung* (Vol. 11). De Gruyter Oldenbourg.

Schöpe, M. (2011). Diversifizierung in der Landwirtschaft. *ifo Schnelldienst*, *64*(14), 43–47.

Shapley, L. S. (1953). A value for n-person games. In *Contributions to the theory of games* (pp. 307–318). Princeton University Press Princeton.

Sharma, R., Kamble, S. S., Gunasekaran, A., Kumar, V., & Kumar, A. (2020). A systematic literature review on machine learning applications for sustainable agriculture supply chain performance. *Computers & Operations Research*, *119*.

Sinha, S., Singh, T., Singh, V., & Verma, A. (2010). Epoch determination for neural network by self-organized map (som). *Computational Geosciences*, *14*, 199–206.

Sjaunja, L. (1990). A nordic proposal for an energy-corrected milk (ecm) formula. *27th Session International Committee for Recording and Productivity of Milk Animals; 2-6 July 1990, Paris, France*.

Smit, H. J., Metzger, M. J., & Ewert, F. (2008). Spatial distribution of grassland productivity and land use in europe. *Agricultural Systems*, *98*(3), 208–219.

Smith, A. (1776). *An inquiry into the nature and causes of the wealth of nations*. W. Strahan and T. Cadell.

Smith, S., Kindermans, P.-J., Ying, C., & Le, Q. V. (2017). Don't decay the learning rate, increase the batch size. *arXiv preprint: 1711.00489*.

Soyeurt, H., Grelet, C., McParland, S., Calmels, M., Coffey, M., Tedde, A., ... Gengler, N. (2020). A comparison of 4 different machine learning algorithms to predict lactoferrin content in bovine milk from mid-infrared spectra. *Journal of Dairy Science*, *103*(12), 11585–11596.

Statista Consumer Market Insights. (2023). *Milchersatzprodukte.* Retrieved from `https://de.statista.com/outlook/cmo/lebensmittel/milchprodukte-eier/milchersatzprodukte/deutschland` (Last accessed on 31-03-2023)

Steinshamn, H. (2010). Effect of forage legumes on feed intake, milk production and milk quality–a review. *Animal Science Papers and Reports*, *28*(3), 195–206.

Steinwidder, A., Starz, W., Pfister, R., Pötsch, E., Schwab, E., Schwaiger, E., ... Kirner, L. (2008). *Untersuchungen zur Vollweidehaltung von Milchkühen unter alpinen Produktionsbedingungen.* Abschlussbericht des Forschungsprojektes Nr. 10271 des BMLFUW.

Storm, H., Baylis, K., & Heckelei, T. (2020). Machine learning in agricultural and applied economics. *European Review of Agricultural Economics*, *47*(3), 849–892.

Sultan, P., Wong, H. Y., & Azam, M. S. (2021). How perceived communication source and food value stimulate purchase intention of organic food: An examination

of the stimulus-organism-response (SOR) model. *Journal of Cleaner Production*, *312*.

Tekerli, M., Akinci, Z., Dogan, I., & Akcan, A. (2000). Factors affecting the shape of lactation curves of holstein cows from the Balikesir province of Turkey. *Journal of Dairy Science*, *83*(6), 1381–1386.

Thomas, E., Mandebvu, P., Ballard, C., Sniffen, C., Carter, M., & Beck, J. (2001). Comparison of corn silage hybrids for yield, nutrient composition, in vitro digestibility, and milk yield by dairy cows. *Journal of Dairy Science*, *84*(10), 2217–2226.

Thünen-Landatlas. (2023). *Raumstruktur – Ländlichkeit.* Retrieved from `https://karten.landatlas.de/app/landatlas/` (Last accessed on 13-04-2023)

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Timlin, M., Tobin, J. T., Brodkorb, A., Murphy, E. G., Dillon, P., Hennessy, D., ... O'Callaghan, T. F. (2021). The impact of seasonality in pasture-based production systems on milk composition and functionality. *Foods*, *10*(3).

Umweltbundesamt. (2023). *Beitrag der Landwirtschaft zu den Treibhausgas-Emissionen.* Retrieved from `https://www.umweltbundesamt.de/daten/land-forstwirtschaft/beitrag-der-landwirtschaft-zu-den-treibhausgas#treibhausgas-emissionen-aus-der-landwirtschaft` (Last accessed on 31-03-2023)

Van der Laan, M. J., Polley, E. C., & Hubbard, A. E. (2007). Super learner. *Statistical Applications in Genetics and Molecular Biology*, *6*(1).

Van Soest, P. J. (2018). *Nutritional ecology of the ruminant.* Cornell University Press.

Varian, H. R., & Varian, H. R. (1992). *Microeconomic analysis* (Vol. 3). W.W. Norton.

Verbeek, M. (2012). *A guide to modern econometrics* (Vol. 4). John Wiley & Sons.

Vitali, A., Felici, A., Esposito, S., Bernabucci, U., Bertocchi, L., Maresca, C., ... Lacetera, N. (2015). The effect of heat waves on dairy cow mortality. *Journal of Dairy Science*, *98*(7), 4572–4579.

Wangsness, P. J., & Muller, L. D. (1981). Maximum forage for dairy cows. *Journal of Dairy Science*, *64*(1), 1–13.

Weinberg, Z., & Muck, R. (1996). New trends and opportunities in the development and use of inoculants for silage. *FEMS Microbiology Reviews*, *19*(1), 53–68.

Weinrich, R., Kühl, S., Zühlsdorf, A., & Spiller, A. (2014). Consumer attitudes in Germany towards different dairy housing systems and their implications for the marketing of pasture raised milk. *International Food and Agribusiness Management Review*, *17*, 205–222.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, 817–838.

Wille, S. C., Spiller, A., & von Meyer-Höfer, M. (2018). *Lage, Lage, Lage? Welche Rolle spielt der Standort für die landwirtschaftliche Direktvermarktung?* Department of Agricultural Economics and Rural Development Discussion Paper.

Wirthgen, B., & Maurer, O. (2000). *Direktvermarktung: Verarbeitung, Absatz, Rentabilität* (Vol. 2). Ulmer.

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, *5*(2), 241–259.

Wood, P. (1967). Algebraic model of the lactation curve in cattle. *Nature*, *216*(5111), 164–165.

Wood, P. (1969). Factors affecting the shape of the lactation curve in cattle. *Animal Science*, *11*(3), 307–316.

Woodworth, R. S. (1929). *Psychology*. W.W. Horton.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (Vol. 2). MIT press.

Xia, Y., Liu, C., Li, Y., & Liu, N. (2017). A boosted decision tree approach using bayesian hyper-parameter optimization for credit scoring. *Expert Systems with Applications*, *78*, 225–241.

Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, *415*, 295–316.

Young, G., & Young, G. (2016). Stimulus–organism–response model: Soring to new heights. In *Unifying causality and psychology: Being, brain, and behavior* (pp. 699–717). Springer.

Zeileis, A. (2006). Object-oriented computation of sandwich estimators. *Journal of Statistical Software*, *16*(9), 1–16.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, *57*(298), 348–368.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Methodological)*, *67*(2), 301–320.

Zscheischler, J., Orth, R., & Seneviratne, S. I. (2017). Bivariate return periods of temperature and precipitation explain a large fraction of European crop yields. *Biogeosciences*, *14*(13), 3309–3320.

Supplementary Material

Dominik Baaken

Lupinenstr. 9, 70599 Stuttgart | dominik.baaken@uni-hohenheim.de | +49 162 97 18 120

Education

PhD Candidate Agricultural Economics since Apr 2019
University of Kiel & University of Hohenheim, Supervisor: Prof. Dr.Sebastian Hess
Thesis: Using Machine Learning for Supply and Demand Predictions in the German Milk Market

Master of Science Agricultural Economics Oct 2016 - Mar 2019
Georg-August University Goettingen
Thesis: Economic Impacts of a Potential Outbreak of African Swine Fever in Germany

Semester Abroad Management, Economics and Consumer Studies Sep 2017 - Mar 2018
Wageningen University, Netherlands

Bachelor of Science Agricultural Sciences (Focus: Economics) Oct 2013 - Sep 2016
Georg-August University Goettingen
Thesis: Sustainability in Feed Production - Identification of Criteria along the Dairy Supply Chain

Work Experience

Research Associate / PhD Candidate since Feb 2020
University of Hohenheim, Institute of Agricultural Policy and Markets

Research Associate / PhD Candidate Apr 2019 - Jan 2020
Christian-Albrechts-University Kiel, Chair of Dairy and Food Industry Economics

Student Research Assistant May 2015 - Jan 2017
University of Goettingen, Department of Agricultural Economics and Rural Development

Technical Organization EuroTier 2016 Sep 2016 - Nov 2016
German Agricultural Society (DLG), Frankfurt

Online Marketing Intern Sep 2015 - Oct 2015
Eurofins Scientific, Luxembourg

Agriculture Intern Aug 2014 - Oct 2014
Farm Wilhelm Puellen, Noervenich

Public Relations Intern Mar 2014
Association of Oilseed Processing Industry (OVID), Berlin

Volunteer Social Year Aug 2012 - Sep 2013
Malteser Hilfsdienst & Johanna-Etienne Hospital, Neuss

Talks and Presentations

2022 **GEWISOLA** (Gesellschaft für Wirtschafs- und Sozialwissenschaften des Landbaus e.V.), Stuttgart; Kieler Milchtage, Kiel; Weihenstephaner Herbsttagung, Freising
2021 **ICAE** International Conference of Agricultural Economists, New-Delhi; **GEWISOLA**, Berlin
2019 Transatlantic Sync Conf, San Francisco

Stuttgart, 28th June 2023

Declaration in lieu of an oath on independent work

according to Sec. 18(3) sentence 5 of the University of Hohenheim's Doctoral Regulations for the Faculties of Agricultural Sciences, Natural Sciences, and Business, Economics and Social Sciences

1. The dissertation submitted on the topic 'Using Machine Learning for Supply and Demand Predictions in the German milk market' is work done independently by my.

2. I only used the sources and aids listed and did not make use of any impermissible assistance from third parties. In particular, I marked all content taken word-for-word or paraphrased from other works.

3. I did not use the assistance of a commercial doctoral placement or advising agency.

4. I am aware of the importance of the declaration in lieu of oath and the criminal consequences of false or incomplete declarations in lieu of oath.

I confirm that the declaration above is correct. I declare in lieu of oath that I have declared only the truth to the best of my knowledge and have not omitted anything.

Dominik Baaken

Stuttgart, 28th June 2023