Studies on the Agricultural and Food Sector
in Transition Economies

Mathias Kloss

Factor productivity in EU agriculture:
A microeconometric perspective

Leibniz Institute of Agricultural Development
in Transition Economies

Factor productivity in EU agriculture:

A microeconometric perspective

Studies on the Agricultural and Food Sector
in Transition Economies

Edited by
Leibniz Institute of Agricultural Development
in Transition Economies
IAMO

Volume 89

Factor productivity in EU agriculture:

A microeconometric perspective

by

Mathias Kloss

IAMO

2017

“Models are never true; but there is truth in models.”

– Dani Rodrik

Bibliografische Information Der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.ddb.de abrufbar.

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists the publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the internet at: http://dnb.ddb.de.

This thesis was accepted as a doctoral dissertation in fulfillment of the re-quire-ments for the degree "doctor agriculturarum" by the Faculty of Natural Sciences III at Martin Luther University Halle-Wittenberg on 19.04.2016.

Date of oral examination:	14.11.2016
Supervisor and Reviewer:	Prof. Dr. Martin Petrick
Co-Reviewer:	Prof. Dr. Silke Hüttel

Diese Veröffentlichung kann kostenfrei im Internet unter
<*www.iamo.de/dok/sr_vol89pdf*> heruntergeladen werden.

This publication can be downloaded free from the website
<*www.iamo.de/dok/sr_vol89pdf*>.

Leibniz-Institut für Agrarentwicklung in Transformationsökonomien (IAMO)
Theodor-Lieser-Straße 2
06120 Halle (Saale)
Tel.: 49 (345) 2928-0
e-mail: iamo@iamo.de
http://www.iamo.de
ISSN 1436-221X
ISBN 978-3-95992-054-4

Acknowledgements

This thesis is the outcome of my doctoral studies at the Leibniz Institute of Agricultural Development in Transition Economies (IAMO). Part of the research was funded by EU's 7th Research Framework Programme within the project "Factor Markets" (245123-FP7-KBBE-2009-3). At this stage, I would like to express my gratitude to everybody who supported and encouraged me throughout the completion of my thesis. Without you, this endeavour would not have been possible.

First, I thank my supervisor Martin Petrick. I gratefully acknowledge his constant support and encouragement. Throughout, his door was always open to me. It was also Martin who encouraged me to start this Ph.D. project at the end of my Master studies.

Furthermore, I thank my colleagues at the Chair of Statistics at Martin-Luther-University Halle-Wittenberg, where I also worked as a doctoral student: Claudia Becker, Lisiane Cardoso Jacobs, Thomas Kirschstein, Juliane Müller, and Steffen Liebscher. I especially thank Claudia for inviting me to work at her chair, and Thomas and Steffen for fruitful research discussions throughout this time. I further thank Silke Hüttel (University of Rostock) for agreeing to review this thesis.

This work benefitted from discussions at international scientific conferences and seminars. In particular, I thank the participants of the following events: *2nd Factor Markets Workshop in Ljubljana, Slovenia (2012); 53th GEWISOLA Conference in Berlin, Germany (2013); Statistische Woche in Berlin, Germany (2013); Wageningen Cluster of Economics Seminar in Wageningen, The Netherlands (2013); 14th EAAE Congress in Ljubljana, Slovenia (2014); 54th GEWISOLA Conference in Göttingen, Germany (2014);* and *6th EAAE PhD Workshop in Rome, Italy (2015).* I also thank Jack Peerlings for the invitation to conduct an extended research stay at the Wageningen University "Agricultural and Rural Policy" group in Fall 2013. In addition, I am grateful that IAMO made available the funding for all these events.

My thanks also go to the entire "Agricultural Policy" department. Throughout, I enjoyed a familiar atmosphere that everyone helped to build and maintain. I especially appreciate the help of Gabriele Mewes with administrative issues. Apart from research, I enjoyed daily lunches with her, Thomas Herzfeld, Martin Petrick, Axel Wolz, and Antje Jantsch.

Finally, I thank my parents (Martina and Hartmut) and parents-in-law (Doreen and Tom). My Mum and Dad made all this possible. I deeply acknowledge their interest, trust, and support for my path of life. Ultimately, I am grateful for the ongoing support of my wife, Saskia, especially during the final phase of this thesis.

Tapadh leibh!

Halle (Saale), November 2016 Mathias Kloss

Abstract

Measuring factor productivity has been important in economics since its early days as a scientific discipline for a number of reasons. The first is the availability of systematically collected agricultural data after World War I, which allowed researchers to motivate and test newly developed methods. This data was collected to fulfill the societal need to learn more about the farming sector, which was stuck in a deep economic crisis at that time. In addition, economists stressed that agricultural technologies approximate the key assumptions of production theory particularly well. To measure agricultural productivity the analyst must deal with tangible (land, labour, and capital) as well as intangible (e.g., management abilities or unexpected weather shocks) production factors. Separating these two types of inputs and appropriately accounting for the latter is at the core of understanding agricultural production. Recent developments such as rising food prices and the decline in global productivity growth indicate that there is a societal need to understand and raise agricultural productivity again. Interestingly, these trends are accompanied by a new debate among econometricians about basic methodological issues in measuring firm level productivity.

This debate is based on the fundamental idea of a mathematical relationship between the various inputs and output – the production function. While early work focused on the measurement of productivity itself, soon after the question was raised whether statistical methods exist to identify the individual contribution of inputs to the joint product. Because farmers usually control the input levels they want to apply, the different production factors are subject to an *endogeneity problem*. In addition, there is also a *collinearity* problem because the standard identifying assumptions underlying production function estimation are usually not powerful enough to measure the productivities of different variable inputs at all. A third, often carelessly treated, problem arises because some parts of the data deviate significantly from the majority of observation. This *outlier problem* as well as the two identification issues bias production function estimates. Against this background, I take on a factor markets perspective to study the productivity of European Union agriculture. To this end, I exploit a panel data set of field crop farms originating from the EU's Farm Accountancy Data Network.

In the methodological portion, I first examine the plausibility of four established and innovative identification strategies within the field of agriculture. Recently suggested control function and dynamic panel data approaches provide enticing conceptual improvements over traditional within and duality models. Second, I analyse the general practice of coping with outliers. Usually only simple procedures based on a single model variable are applied. However, I argue by example that a multivariate detection method controlling for all model dimensions effectively removes outliers. Therefore, with the identification issues in mind, I propose a practical two-step approach to encounter these issues. First, I decontaminate the data with a multivariate non-parametric outlier detection procedure and second, I consistently estimate the parameters of the production function.

In the empirical portion of this study, I start with an assessment of the traditional as well as the recent advanced identification approaches. I estimate production elasticities and factor shadow prices for a set of six EU member states. Even so, empirical implementation of new developments in production function analysis do not always live up to expectations, especially in the case of the dynamic panel estimator. In most instances this estimator failed to identify reasonable elasticities for the (quasi-) fixed inputs. Fortunately, proxy approaches which are less demanding in terms of identification represent an interesting alternative for agricultural applications. In the EU sample available to me, high production elasticities for materials prevail. Generally, output reacts most elastically to materials inputs. Through further investigation of this factor, I find different rationing regimes in different EU member states. Shadow interest rates of materials is low in Denmark and the United Kingdom. However, they are significantly higher than typical market interest rates in France, Germany, Italy, and Spain. In all countries they also increased toward the end of the observed period. This finding is consistent with a view of tightening access to short-term capital, which was possibly induced or reinforced by the onset of the recent financial crisis. Marginal returns to land, labour, and fixed capital are typically low. To conclude, functioning factor markets play an important role in fostering productivity growth. Nevertheless, factor market operations are considerably heterogeneous across different EU member states.

In a further empirical study, I apply a particularly suitable multivariate outlier decontamination to a panel of East and West German field crop data. Results show that this procedure detects outliers outside the main bulk of observations as well as those located within the production set. I estimate and compare production functions for different subsamples of the data. These include such subsamples without any outlier removal as well as univariate and multivariate outlier control. In general, the multivariate outlier control delivers more reasonable results with a higher precision in the estimation of some parameters and seems to mitigate the effects of multicollinearity – a feature of the input variables that I also analyse and discuss in-depth throughout this study.

In the final empirical study, the focus is on the role of labour and the effect of its different qualities on productivity. In particular, I test for heterogeneous effects of family and hired labour for a set of eight EU member states. To this end, I estimate augmented production functions using "Farm Accountancy Data Network (FADN) data for the years 2001-2008. The results reject the notion that hired labour is generally less productive than family workers. In fact, farms with a higher share of hired workers are more productive than pure family farms in countries traditionally characterised by family labour, namely France, West Germany, and Poland. Here, an increase in reliance on hired labour or the shift of family labour to more productive tasks could raise productivity. This finding calls into question a main pillar of the received family farm theory, namely that the growth of hired labour force beyond family members is inhibited by rising supervision costs. For the United Kingdom, I find the classical case with family farms being more productive than those relying on hired labour. In this situation supervision by family members could increase productivity. In the other countries of the sample, there are no statistically different effects of either type of labour.

Following my analyses, a number of policy implications unfold. As it turned out, materials is the most important input in European field crop farming. Hence, improving the availability of working capital is the most promising way to increase agricultural productivity. This finding is also underlined by the shadow price analysis, which indicated that in a number of countries the estimated return on working capital is much above observed market interest rates. Therefore, policy reforms should aim to ease access to short-term credit. With regards to agricultural labour markets the results indicate for France, West Germany and Italy that hired workers perform the highly specialised tasks leading to an increase in agricultural productivity. Consequently, policy makers should focus on creating incentives for farms to hire such specialised labour. For instance, programs to qualify and hire specialised labour could improve their inflow into agricultural labour markets.

Keywords: Agricultural factor productivity; Labour productivity; Production function estimation; Outlier detection; Farm Accountancy Data Network; European Union

Table of content

List of tables

List of tables in appendices

List of figures

List of figures in appendices

List of abbreviations

BB	Blundell/Bond
CD	Cobb-Douglas
Coeff	Coefficient
ESU	Economic size unit
EU	European Union
FACEPA	FARM Accountancy Cost Estimation and Policy Analysis of European Agriculture
FADN	Farm Accountancy Data Network
FD	First difference
FDH	Free disposable hull
GMM	General method of moments
GSOEP	German Socio-Economic Panel
Ha	Hectare
IID	Identically and independently distributed
IQR	Interquartile range
IV	Instrumental variables
LSMS	Living Standard Measurement Survey
LP	Levinsohn/Petrin
MCD	Minimum covariance determinant
MST	Minimum spanning tree
OID	Over-identification test
OLS	Ordinary least squares
PMST	Pruning the minimum spanning tree
SD	Standard deviation
SE	Standard error
SGM	Standard gross margin
TF	Type of farming
WLP	Wooldridge/Levinsohn/Petrin

1 Introduction

Productive tasks are at the heart of every economy. Therefore, it is only natural that economists have been interested in the evaluation of factor productivity – the weighted contribution of production factors to a marketable product – from its early days and onward. Adam Smith (1776/2010), one of the founding fathers of modern economic sciences, stressed the importance of labour specialisation in performing productive tasks. With regard to the measurement of factor productivity, agricultural applications drove many methodological advancements in economics. These, among others, include the introduction of marginal productivity theory by Johann Heinrich von Thünen in Germany around 1820, or early empirical estimations of agricultural technology by Tolley et al. (1924) in the United States. Within the subfield of econometrics, one might name the formal derivation of a test for constant returns to scale by Tintner (1944) and the invention of fixed-effects regression by Hoch (1955) and Mundlak (1961). Generally, the first half of the 20th century is marked by new methodologies developed with the agricultural sector in mind. This development is underlined in a variety of ways. The first is the availability of suitable microeconomic data after World War I that allowed motivation and testing of new approaches. At that time, statistical agencies started to systematically collect farm data to learn more about productivity and productivity growth in the farming sector. As this sector became stuck in a deep economic crisis, there was a societal need for investigations in the same. Finally, many economists stressed the special suitability of agricultural technologies in approximating the key assumptions of production theory such as diminishing marginal product of inputs and the substitutability of the same (Chambers, 1988).

Rising prices on world food markets in recent years have shown impressively that resources for agricultural production, on a global scale, are scarce (FAO, 2009; Djuric, 2014: 21-23). For instance, as of late there was news about exploding sugar prices in the Ukraine (Top agrar online, 2016). This trend is most likely reinforced by a decline in the global rate of productivity growth (Alston and Pardey, 2014). In addition, how farm productivity could be raised has recaptured attention of global media outlets such as the Economist (Parker, 2011), and food riots have been reported from several developing countries (e.g. Burkina Faso, Egypt, and Yemen). These recent developments signal that the aforementioned societal need is persistent. Interestingly, at more or less the same time, econometricians revived a discussion about basic methodological problems in identifying firm productivity.

This debate is based on the assumption that there is a continuous mathematical relationship, the production function, between the various inputs (e.g., land, labour, seed and fertiliser, and output), an idea that has been well known and widely built upon since the days of Cobb and Douglas (1928). Because this function is a theoretical construct and cannot readily be observed, it needs to be estimated. While the early work focused on the measurement of productivity itself (i.e., collecting data and constructing appropriate indexes of production factors) (cf. Tolley et al., 1924; Cobb and Douglas, 1928), soon after the question was raised whether statistical tools are available that can identify how much the various factors actually contribute to the joint product.[1] Marschak and Andrews (1944) pointed out that production is subject to unobserved factors such as managerial abilities or unexpected weather shocks. Their presence impedes the estimation of production function parameters having causal interpretation. At the core of this ongoing debate is the issue of giving such intangible factors a tractable structure so that they can be separated from the tangible inputs of land, labour and capital. Obtaining valid figures of the structure of agricultural production is important for farmers and policy makers. It enables the researcher as well as these interest groups to understand and impose agricultural productivity growth.

Two issues fuel the recent debate. In the first issue the level of inputs applied is a control variable to the farmer which is potentially determined simultaneously with other unobserved events, or may depend on unobserved, omitted, variables. This *endogeneity problem* again moved to centre stage after Olley and Pakes (1996) (OP) suggested a non-parametric control function to proxy these unobservables, thus starting a new line of research on dealing with this classical identification problem. Furthermore, according to Bond and Söderbom (2005) as well as Ackerberg et al. (2007), typical production function identifying assumptions are unable to carve out exogenous variation amenable to estimation from the different variable inputs. Hence, to induce such variation some sort of adjustment cost is necessary. The most recent contributions to tackle this so-called *collinearity problem* are papers by Wooldridge (2009) and Gandhi et al. (2011). Both authors try to solve the two identification problems simultaneously. The former takes the Olley and Pakes (1996) identification strategy as a starting point and modifies it as well as extends the identifying assumptions. The latter relies on an idea that has been around for many decades, the factor share regression.

1 The retrospective by Biddle (2012) points this out nicely.

However, even though if the statistical identification is secured, other issues in estimation might emerge from the data itself. One such problem might be *multicollinearity* – high correlation between two or more variables. However, there is little that can be done about this other than re-specifying the model of interest (Wooldridge, 2006: 104). A second and by applied researchers more carelessly treated problem that might bias parameter estimates, is the presence of *outliers* in the data set, which is a ubiquitous feature of many real world data sets, and hence an issue in many empirical applications. For instance, such outliers might occur due to measurement errors, variations in the data-generating process, or misreporting. If they are not dealt with, estimators can be obscured arbitrarily by the presence of as little as one single outlier in the data set – an extreme case but perfectly possible, as many introductory textbooks in statistics suggest – thus leading to biased estimates. Moreover, in terms of estimating factor productivity, accounting for outliers ensures that the assumption of a homogenous production technology is maintained.

In general, other than by being very pragmatic and not performing any outlier decontamination, there are two concurrent views on how to define and, consequently, identify outliers. On the one hand, there are methods which assume that outlying observations follow a different distribution other than the one in which the researcher is interested. Methods in this category commonly aim at estimating some features of the target distribution while outlier detection is usually not their primary objective. However, such methods might produce biased results if the assumed distributional assumptions are not fulfilled. Many robust estimators for various models can be found in this category. For instance, see Rousseeuw and Leroy (1987); Barnett and Lewis (2000); Hampel et al. (2005); Maronna et al. (2006); and Huber and Ronchetti (2009) for an overview of such methods.

On the other hand, other approaches follow a more sample-oriented view on how to define an outlier. Universally, all of these methods interpret outliers as observations that differ substantially from the target observations (cf. Johnson, 1992; Barnett and Lewis, 2000). Obviously, there are numerous proposals on how that difference can actually be quantified. A large number of these methods is based on some type of distance, either between any two observations or with respect to some reference point, but there are also other approaches (e.g., based on depth or on the empirical density) (Chandola et al., 2009). With these methods, the primary goal is indeed the identification of outliers. An additional analysis can then be carried out on the identified non-outliers. Another major advantage of such approaches is their applicability without being limited to certain distributional assumptions of the data.

1.1 Outline and summary of the thesis

1.1.1 Issues in estimating factor productivity

To provide a comprehensive assessment the above-mentioned issues, the identification and outlier problems, I take on the following approach. First, the various methodological identification approaches and their properties within an agricultural context are discussed and evaluated (chapter 2.1). To this end, I review the central identifying assumptions maintained by six traditional and recent approaches to the estimation of production functions. A panel data perspective is assumed throughout because the data used in this work is a panel of firms observed for several years. Second, I start with an extensive survey of outlier treatment in empirical economic practice to assess the outlier issue. This survey includes studies exploiting microeconomic data sources from general, development, and agricultural economics. I proceed, by example, with outlining the effects of outliers on productivity analysis and provide a solution accounting for aggregate output and all input dimensions in agricultural production (chapter 2.2). Insights from chapter 2 are subsequently applied throughout the empirical chapters.

In the first empirical study, I apply the various methodological identification strategies to a panel data set on European agriculture. Throughout, I discuss whether theoretical considerations carry over to empirical production function estimation (chapter 4). Their plausibility in an agricultural context is the core of my discussion. The estimation techniques applied are the calculation of factor shares in farm revenue, ordinary least squares (OLS) as the "naïve" estimation standard providing baseline estimates, fixed-effects (within) regression, the dynamic panel data estimator by Blundell and Bond (2000) (BB), as well as the control function approaches by Levinsohn and Petrin (2003) (LP) and Wooldridge (2009) (WLP). I estimate all models for a Cobb-Douglas production function specification. In addition, a translog functional form is explored for the OLS, within, and WLP models. Hence, I arrive at a total of nine estimated models.

Recently, there has been considerable research activity on new approaches in production function estimation and there have been comparative evaluations on such approaches using simulated data (cf. Van Biesebroeck, 2007). However, most researchers refrain from comparative evaluations employing real world data. If estimator developers provide empirical applications at all, they do so from highly specific contexts. For instance, Blundell and Bond (2000) use data on R&D-performing US manufacturing firms covering the years 1982 to 1989 which has been employed by other researchers for methodological elaborations.

A similar situation may be observed for Levinsohn and Petrin (2003), who utilise data from Chilean firms. Use of this data set goes back to the early 1990s. Later, it was also utilised by Ackerberg et al. (2006) as well as Kasahara and Rodrigue (2008). The latter apply various panel data estimators to this data set, including dynamic panel and proxy approaches. By holding the data set constant, researchers control for variation induced by the same while evaluating new estimation strategies. Such an approach may be beneficial if the research interest is mainly methodological. Nevertheless, the true value only unfolds after application to data that is also of topical or political interest. In the present study, I attempt to fill this gap as it is among the first to apply a whole set of recently discussed estimators to a dataset that is highly relevant in a policy context.

The unique European database available for this study covers firm-level data from all EU member states that was collected by national agencies and consolidated centrally following the same harmonised procedure in all countries, thus making cross-country comparisons particularly meaningful. This is one of the first agricultural productivity studies to use this micro data to analyse factor productivity for several EU countries using a variety of techniques to attack different sources of biases.

Surprisingly, while agriculture is a classical field of production function estimation, few analyses have used the primal production function approach lately. This observation is attributed to the popularity of duality theory in agricultural economics starting in the 1970s Mundlak (2001). In contrast to the primal approach, in which the output elasticities of inputs are estimated, this approach recovers price elasticities of factor demand. Mundlak (2001) further notes the strong theoretical assumptions and methodological problems of duality, both of which restrict its usefulness. I elaborate on these issues below as well. Shumway (1995) argues that a key advantage of duality was the possibility to model more flexible functional forms or production technologies such as the translog. However, my results indicate that making the Cobb-Douglas specification more flexible by adding second-order terms of inputs does not provide any further insight. Whereas OLS and Wooldridge (2009) produced highly implausible results, there was little difference for the fixed-effects regression compared to the Cobb-Douglas case.

The empirical estimates paint a picture of low land and fixed capital output elasticities throughout the European subsamples (chapter 4). On the other hand, the materials elasticity is quite high, around 0.7. This outcome is particularly prominent for the estimators basing their identification strategy on adjustment costs, namely LP, WLP, and BB. The estimate for the labour output elasticity is somewhere in between. In addition, because fertiliser and land are highly collinear, I adjust the materials definition to attack and mitigate multicollinearity problems while maintaining a correctly specified production function specification.

The results obtained in this study are consistent with other recent work on EU agriculture. For instance, in an analysis of Finnish dairy farms, Heikkilä et al. (2012) find that materials and fixed capital inputs are the most important factors in driving productivity. Rizov et al. (2013) confirms and extends this view to 15 EU member states. In an aggregation over different types of farms, they recover materials elasticities ranging between 0.59 (for Greece) and 0.87 (for Sweden) while labour and fixed capital output elasticities are rather low. In contrast, estimates by Mundlak et al. (2012) utilising a cross-country sample of developing and developed countries suggest that output reacts most elastically to land and fixed capital.

The shadow price analysis reveals a heterogeneous picture across our EU sample of countries. In France, Spain, Italy, and Germany (East and West), I observe shadow interest rates of materials much above typical market interest rates, especially toward the end of the observed period – a view that is consistent with constrained access to funding. A possibly explanation for this finding might be the unfolding global financial crisis in 2007 and 2008 (Petrick and Kloss 2013c). The cross-country variation in these figures also reflects characteristics and inner mechanisms of agricultural banking sectors in the different member states. For Denmark and the United Kingdom, returns to materials are low, suggesting an over-utilisation of such inputs. The remuneration of labour is generally below 9 EUR/hour throughout the European sub-samples. The value of land fluctuates around 0 EUR/ha land, while fixed capital always displays a negative shadow interest rate. An exception to these general trends is to some extent Denmark.

In the conceptual section, I state that the established within and duality approaches pose (too) strong assumptions about the structure of agricultural production and that the control function and dynamic panel data models provide more plausible identification strategies. Both approaches relate identification to adjustment costs that occur as a reaction to past productivity shocks. This adjustment cost is used to distill the necessary exogenous variation from the different inputs and make it useable for estimation purposes. The adjustment process is assumed to be completed in one period in the Levinsohn/Petrin and Wooldridge/Levinsohn/Petrin models while the Blundell/Bond model implies a multiperiod adjustment, making the latter more compelling in an agricultural context, even though the former is easier to implement. However, Blundell/Bond often did not pass major specification tests and produced unreasonable parameter estimates for land and fixed capital inputs in most of the country subsamples. In summary, my analysis showed that the theoretical plausibility and empirical robustness of the different identification strategies delineate a trade-off.

In a second empirical study, the focus is on the bias induced by outliers while maintaining the primal production function perspective. To this end, I analyse the consequences of productivity estimates in light of their existence. I start with a working definition of the term "outlier" to have a meaningful delimitation to non-outliers, which provides necessary and valuable information for the subsequent analysis. A typical way to deal with outliers is to exclude them from the estimation. Generally, as I discovered by evaluating an exhaustive sample of empirical economics papers, if decontamination is conducted it is almost always done on a single variable, i.e., univariate. Often ad hoc and handmade methods focusing on one variable of a more complex multivariate model are employed prior to the follow-up analysis (chapter 2.2.2). Therefore, the general picture that was obtained from this survey suggests that outlier problems are somewhat "second rank" issues that are often neglected. However, simple univariate approaches do not consider the multivariate nature of the model at hand. This is a real drawback, as a simulated example, employed to illustrate the effects of outliers and different contamination schemes on productivity estimates, suggests (chapter 2.2.3). Therefore, I utilise a multivariate approach outlined in chapter 2.2.4.

Throughout, I propose a robust two-stage approach for estimating production functions, first by performing the outlier decontamination, and second by estimating the parameters of the production function. I apply this approach to a data set of East and West German field crop farms. While the literature on outlier detection methods is vast (Maronna et al., 2006) and other methods might be feasible as well (e.g., cluster analysis), I resort to a multivariate decontamination procedure following Kirschstein et al. (2013) for identifying the outliers. This method, which falls into the second category of how outliers are viewed and defined, has already proven its effectiveness in an application to unsuccessful warship designs (Liebscher and Kirschstein, 2012). Moreover, it is a non-parametric approach and therefore does not require special distributional assumptions that the data needs to fulfil and it offers computational advantages (especially for large data sets as the one analysed here), which makes it my preferred choice. After decontamination, production function parameters are estimated by the Wooldridge (2009) instrumental variable estimator to account for common identification issues. In particular, by holding the production function estimator constant, meaningful insights are gained by comparing output elasticity estimates arising after the application of the different decontamination schemes – uni- and multivariate – as well as no decontamination at all.

The proposed two-step approach allows for robust and consistent estimation of production functions. By applying a multivariate assessment of outliers, I am able to consider the whole spectrum (i.e., output and inputs) of agricultural production. Therefore, I am not only able to find conventional outliers outside the main bulk of observations but also such outliers located within the production technology (chapter 5). The majority of outliers lies within this region. This is a major advantage compared to the common univariate strategies. By just dropping observations beyond some threshold in a single dimension, the univariate procedure tends to misestimate the number of outliers in my specific application and consequently drops valuable information. Results suggest that small farms comprise a considerable amount of outliers. The advantage of the multivariate detection also carries over to the estimation, at least to some extent. Estimates based on the multivariately decontaminated sample result in improved results compared to those for the other samples (e.g., it allows for greater precision in the estimation of some parameters and seems to further mitigate the effects of multicollinearity). The latter observation is also confirmed by simulated example. In general, productivity estimates following the elimination of outliers convey that material inputs are the most important production factors in German agriculture. In summary, the main insight from this exercise is that a multivariate model of interest should comprise a multivariate outlier assessment.

Throughout this dissertation, I attempt to make methodological as well as empirical contributions to the body of economic literature. My methodological contribution is the first comparative evaluation of a number of recently proposed production function estimators within an agricultural context. My empirical contribution is a unique and updated set of estimated output elasticities as well as returns on factor use exploiting firm-level data for a sample of seven EU countries. Finally, following an assessment of the generally empirical practice as well as my own analysis and the evidence it produced, I propose and advocate an approach to productivity analysis that considers outliers in a multivariate manner prior to estimation.

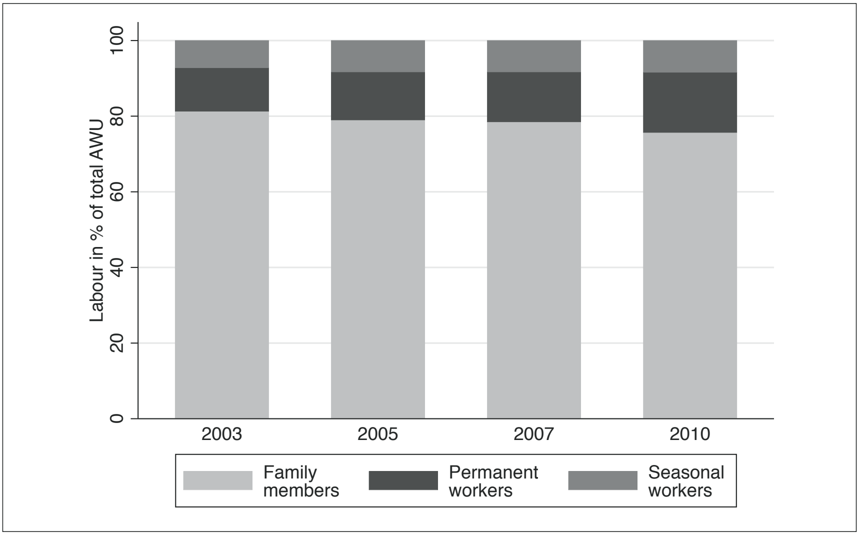
1.1.2 The impact of labour force composition on farm productivity

Following the general analysis of factor productivity in EU agriculture, I turn my attention to the role of labour and its impact on productivity. How different types of labour affect farm output is the subject of the final empirical study of this dissertation. The analysis builds again on my theoretical production function model.

According to a widely accepted view, large-scale farming operations with many workers under a centralised management authority are economically inferior to smaller family-run businesses, at least in temperate zones (Hayami, 2010). The two maintained hypotheses of the underlying "family farm theory" are that (1) technological scale economies are typically exhausted before farm size exceeds the labour capacity of a family, and that (2) growth of the labour force beyond family members is inhibited by rising supervision costs. These hypotheses used to be supported by a large body of empirical literature from developed and developing countries (Brewster, 1950; Schmitt, 1991; Hayami and Otsuka, 1993; Allen and Lueck, 1998; Eastwood et al., 2010). For many decades after World War II, the economic and social superiority of family farms over agriculture based on hired labour was a widely held notion among researchers, governments, and international organisations.

However, even in agricultural regions traditionally dominated by small to medium family farm operations, such as Western Europe or the US, farm sizes have been

Figure 1.1: Change in the composition of the agricultural labour force in EU-27.



Notes: AWU, Annual Working Units. Family consists of permanent family labour including holders and members of the sole holder's family. Permanent workers consist of regularly employed non-family members. Seasonal workers are comprised of irregularly employed non-family members.

Source: Author compilation based on European Commission (2012, 2013).

growing and, more importantly, the share of hired workers in total labour force has been slowly but steadily increasing (Blanc et al., 2008; Darpeix et al., 2014). According to the latest figures by the European Commission (2013), regularly employed non-family members on average contributed 14.7 per cent of the total agricultural workload in the EU-27 in 2010, whereas irregularly employed non-family members contributed another 7.7 per cent. This share has been on the rise for years, especially with regard to regularly employed hired labour. These workers replace family members on a year-round basis rather than complementing them during harvest time, a fact that calls into question the validity of hypothesis (2) outlined above (Figure 1.1).

The typical argument for different productivities of hired and family labour is based on the idea that both have diverging incentives. Hired labour are usually not residual claimants and their effort cannot commonly be observed because of the idiosyncracy of agricultural production (e.g., seasonality, weather effects). Therefore, hired labourers have incentives to "shirk", resulting in effort levels that are only a fraction of those achieved by family labour. As a result, neither kind of labour is easily substituted. This perceived problem can be mitigated by hired labour supervision. Hence, transaction costs in the form of supervision costs arise, making farm production based on hired labour more expensive. On the other hand, the following argument in favour of hired labour is often overlooked: growing farms with a larger stock of workers may allow more specialisation and the division of labour into distinct tasks (Allen and Lueck, 1998; Kimhi, 2009). For example, family members might concentrate on management and/or supervision tasks, while hired labourers specialise in non-managerial tasks. To the extent that modern farming technologies allow such specialisation benefits, the productivity of hired labour may well exceed that of a family member who is a "jack of all trades but the master of none".[2]

Given these conflicting views, this study aims to revisit the relative superiority of family over hired labour by confronting the accepted wisdom with new empirical evidence (chapter 6). In exploring the relative productivity of family versus hired labour, I follow Bardhan (1973), Deolalikar and Vijverberg (1983, 1987), and Frisvold (1994), who investigated this question in the context of the developing country of India. Whereas these authors found evidence in favour of both argu-

[2] Productivity differences may also be due to family members and workers possessing different levels of education and technical expertise. While the positive effects of farmers' human capital on production decisions have been analysed, relatively little is known about the effects of workers' education (Huffman and Orazem, 2007).

ments presented in the preceding paragraph, supervision versus specialisation effects, I am primarily interested in their methodological approach. I follow these authors in using a parametric production function specification that accounts for heterogeneous labour impacts. This approach focuses on a single parameter of relative labour productivity and thus allows straightforward interpretation. Yet, my estimation technique goes beyond the received estimators used by the previous authors in tackling potential endogeneity problems. Deolalikar and Vijverberg as well as Frisvold resorted to traditional household/farm fixed effects approaches. Following my methodological and empirical insights on production function estimation, I focus on the estimation procedure introduced by Wooldridge (2009). My database is a panel originating from the Farm Accountancy Data Network (FADN) of eight EU member states: Denmark, France, Germany, Italy, Poland, Slovakia, Spain, and the United Kingdom. Germany is split into East and West. Data is available for the years 2001-2008. The sample of countries includes full-time arable farms and reflects the diverse farm structures prevalent in different member states. It comprises countries with traditional family-type farming (e.g., France and Italy) as well as a high share of hired labour (e.g., East Germany and Slovakia). The variability in farm structures across countries provides the necessary variance to study the influence of hired labour. I limit the analysis to arable farms to justify the assumption of a homogenous production technology. I compare the main results with the received ordinary least squares (OLS) approach. To my knowledge, there are no comparable studies for EU agriculture in the area of labour force heterogeneity to date.

My results reject the notion that farms with a higher share of hired labour are generally less productive than those with more (or only) family workers, everything else being equal. The most striking outcome is that farms with more hired labour are more productive than farms with less hired workers in countries traditionally characterised by family farms, namely France, West Germany, and Poland. In the rest of the countries, there are no statistically different effects of the composition of labour. As a side result, I find little evidence of non-constant technical returns to scale. Thus, farm growth in Europe may indeed be increasingly driven by scale-neutral technologies which allow the realisation of gains from labour specialisation.

1.2 Organisation of the thesis

In this thesis, I assume a factor market perspective to analyse productivity in EU agriculture. I proceed as follows. In chapter 2, I evaluate issues in identifying factor productivity. In chapter 2.1, I discuss the key identification problems (econometric in nature) that have motivated much of the methodological debate in production function estimation as well as the four main assumptions invoked in the literature to address them. I proceed by outlining the statistical-in-nature outlier problem (chapter 2.2). This is accompanied by a survey of the general treatment of outliers in empirical economics. At this stage, I illustrate by example the consequences on production function parameter estimates in the presence of outlying observations in the data and review my decontamination approach. In chapter 3, I discuss the FADN database. Chapters 4 and 5 are comprised of an empirical assessment of factor productivity in EU agriculture, in which I provide a new view on factor output elasticities as well as returns on input use by utilising traditional as well as novel identification and outlier decontamination strategies in approaching the issues outlined in chapter 2. In chapter 6, I take the novel control function identification approach to study the impact of a farms' labour force composition on its productivity to further extend the view on the role of labour in EU agriculture. Chapter 7 concludes. Throughout, this dissertation builds upon results from revised and extended versions of Petrick and Kloss (2013a) and Kloss and Petrick (2014) as well as Kloss et al. (2015).

2 Selected problems in estimating factor productivity

Throughout this chapter, I discuss major problems faced by analysts in estimating production functions within an agricultural context. I start with a presentation of the identification problems and evaluate traditional and recent solutions to these problems within this field of application. Next, I present the impact of outlying observations in the data on the estimation of production functions and solutions. I argue for a multivariate assessment of outliers because univariate methods are not necessarily able to capture all relevant outliers.

2.1 Identification problems in production function estimation and approaches to their solution[3]

In this subchapter, I focus on the methodological identification problems of production functions. I essentially provide a typology of production factors, given the existence of appropriate data, dividing them into observable and unobservable inputs. The unobservables are a source of two identification problems to be discussed. Making them tangible and separating them from the observables are the core aims for the researcher. Subsequently, I evaluate traditional and recent approaches to tackle these problems.

2.1.1 A typology of production factors

Agricultural production serves as a useful illustration for the different nature of production factors. For the ensuing discussion, two characteristics of these factors are of particular importance:

a) their variability or the speed with which they can be adjusted, and
b) whether they are observed by the econometrician.

Table 2.1 differentiates three categories of variability. Among the highly variable factors are intermediate inputs such as seed, fertiliser or concentrate fodder. These factors are typically included in farm-level datasets and thus observed by the econometrician (type I factors). In economic lingo, they are also called control variables because the decision maker (the farmer) can manipulate their

---

[3] This section is based on Petrick and Kloss (2013a). In particular, this previous research has been extended to a more complete treatment of the identification strategy presented in 2.1.6.

level to achieve her objectives. Other highly variable control variables, such as work effort (type IV factors), may be hard to observe from the outside or difficult to measure.

Other important factors are much less variable and are subject to adjustment costs (type II and V factors, depending on whether they are observed). For example, land is often only available in limited quantities and subject to long-term rental agreements. Agriculture in Europe is typically organised in family farms on which labour is often highly immobile (TOCCO et al., 2012) and may be influenced significantly by life cycle considerations of the farm family (GLAUBEN et al., 2009). Agricultural credit markets suffer from informational asymmetries and may be characterised by rationing and high transaction costs (e.g., BENJAMIN and PHIMISTER 2002; PETRICK and LATRUFFE, 2006). Management has long been recognised as an important factor of production that is nevertheless difficult to measure (MUNDLAK, 1961).

A final group includes factors that are completely fixed in the long run, such as the geographic location of the farm or the quality of its soils (type III and VI factors). All the less variable factors (type II, III, V and VI) are called state variables because their value cannot be modified within a short-term planning horizon.

Table 2.1: A typology of production factors in agriculture.

	Highly variable	Subject to adjustment costs	Fixed
Observed by econometrician & farmer	**Type I** Seed, fertiliser, chemicals, concentrate, livestock numbers	**Type II** Land, labour, machinery, buildings	**Type III** Geographical location
Typically unobserved by econometrician but known to the farmer	**Type IV** Farmer's effort, reaction to environmental shocks	**Type V** Management knowledge and abilities, human capital of labour force, availability of a farm successor	**Type VI** Soil quality, climatic conditions
Unobserved by econometrician & unanticipated by the farmer	**Type VII** Weather events, rainfall, diseases, legal requirements	--	--

Source: PETRICK and KLOSS (2013a).

As indicated in Table 2.1, there is an important distinction between the highly variable and unobserved factors type IV and VII. Some of these also come as a surprise to the farmer. They represent exogenous states (shocks) of the environment (type VII factors). However, the farmer's reaction to these shocks will be endogenous (type IV factors) and are characterised by a high variability. This variability results from the fact that every (policy or environmental) shock is different by nature and requires a matching reaction.

2.1.2 Two problems of identification

To illustrate the involved problems, I start with a simple model of a farmer wishing to produce an aggregate output. Denote y_{it} the natural logarithm of farm i's output Y at time t, A_{it} land use of this farm, L_{it} labour, K_{it} fixed capital, and M_{it} materials or working capital. These four production factors are observed by the econometrician. The variable ω_{it} is an aggregate, farm-specific, time-varying factor that is anticipated by the farmer at the time of decision making about current production, but unobserved by the econometrician. Without further specification, it compounds the effects of factors categorised as type IV to VI in Table 2.1. The ε_{it} is a productivity shock not anticipated by the farmer (and not observed, thus type VII), or simply measurement error. Assuming a log-linear structure of the model and the availability of panel data containing the observed output and inputs, the econometrician's problem is to recover farm productivity determined by the following equation:

$$y_{it} = f(A_{it}, L_{it}, K_{it}, M_{it}) + \omega_{it} + \varepsilon_{it}, \tag{2-1}$$

where $f(\cdot)$ is the production function. It may best be understood as a farms' technology to produce a marketable good. This concept extends naturally to whole industries by assuming homogeneity of technologies across farms. For instance, it is feasible to assume that field crop farms rely on more or less the same production technology.

Because will likely be correlated with the other input choices, estimation of (2-1) is subject to an *endogeneity problem* (Marschak and Andrews, 1944). The production elasticities of the observed factors are not identified because the compound error term $\omega_{it} + \varepsilon_{it}$ is, in statistical parlance, not identically and independently distributed (IID). Choosing an appropriate functional form for $f(\cdot)$ and regressing output on observed input levels using OLS will then produce biased estimates because this estimator technically neglects the presence of the ω_{it}. In particular, input coefficients will be upward biased if there is serial correlation in ω_{it}. This effect will be stronger the easier it is to adjust input use (Levinsohn and PETRIN 2003: 332). A typical OLS result may be that the coefficients of labour and materials are upward biased, while those of land and capital are downward biased. Much of the methodological literature on production function estimation is concerned with precisely this issue. For further reading, see the survey and review by GRILICHES and MAIRESSE (1998).

According to the implicit theoretical setup so far, all observed factors are assumed to be control variables and are treated as being fully flexible (as if they all belonged to type I). The typical assumption in the literature (e.g., CHAMBERS, 1988) then is that output and all factors are traded on perfectly competitive markets so that on each of the markets all farmers face the same price for the traded good. If farmers maximise profits defined as revenues from the sale of output minus costs of all inputs and $f(\cdot)$ is a monotonous and concave function, the canonical decision rule for allocating inputs is identical for all inputs and says that the marginal revenue product of each factor should equal its factor price. For example, this decision rule is as follows for materials:

$$p^Y \frac{\partial f}{\partial M} = p^M, \qquad (2\text{-}2)$$

with p^Y denoting the price of output and p^M that of materials, respectively. Estimation of (2-1) requires the assumption that the technology represented by $f(\cdot)$ is identical for all farmers included in the estimating sample. If all farmers also face the same price on each of the output and input markets, there is nothing in the model that induces heterogeneous factor use across farms except for the unobserved ω_{it}. This is the *collinearity problem* pointed out recently by BOND and SÖDERBOM (2005) and ACKERBERG et al. (2007).[4] Factor use across firms varies only with the unobserved , so that again the different production elasticities are not identified.

I now review the main approaches found in the literature to deal with either of these identification problems. The discussion is guided by Table 2.2, which summarises the four approaches I distinguish. After introducing each approach, I ask how plausible the specific identifying assumption is in the context of agriculture. I then evaluate the extent to which the two key identification problems presented before are addressed and how the resulting estimator can be applied in practice.

4 A very detailed exposition is ACKERBERG ET AL. (2006).

Table 2.2: Identifying assumptions in production function estimation.

	(A) ω_{it} is additively separable & time invariant	(B) Profit maximization & perfect competition on product & factor markets	(C) Heterogeneous frictions in factor adjustments	(D) ω_{it} evolves monotonously with an observed characteristic of the firm
If correct, does the assumption solve the endogeneity problem?	Yes.	Yes if prices can be used as instruments.	Yes, if adjustment costs are sufficiently heterogeneous across inputs.	Yes.
Does it solve the collinearity problem?	Not without further assumptions.	Yes if there is only one free input (GANDHI et al. 2011).	Yes, if adjustment costs are sufficiently heterogeneous across inputs.	Not without further assumptions (ACKERBERG et al. 2006; WOOLDRIDGE 2009).
Practical implementation	Within regression to sweep out fixed effect.	Share regression, approaches based on duality.	Typically combined with assumption (A) in a dynamic panel data regression model using first differences.	Semiparametric control function approaches using investment or intermediate inputs as proxies.
Remaining problems	Remaining variance may be too small to allow precise parameter estimation.	Prices with sufficient variation may not be observed. Heterogeneous firm-specific prices may not be exogenous.	Weak instruments, small variance of differenced variables.	Zero observations for proxies (e.g., investment). Slowly changing unobserved effects are not captured.
Plausibility in agriculture	Limited plausibility as farm- & time-specific effects are likely (e.g., reactions to weather shocks).	Limited plausibility as market imperfections on labour, land, & capital markets are widespread in agriculture.	Plausible for land, labour, fixed capital; less for seed, fertiliser, plant protection, concentrate, energy.	Plausible for annually fluctuating shocks; less for slowly changing unobservables such as soil or management quality.
Examples in the literature	Widely used. See MUNDLAK (1961); overview in GRILICHES & MAIRESSE (1998).	Widely used. See overview in MUNDLAK (2001) and BONNIEUX (1989) on French agriculture.	BLUNDELL & BOND (2000); HEMPELL (2005). No agricultural applications so far.	OLLEY & PAKES (1996); LEVINSOHN & PETRIN (2003); KAZUKAUSKAS et al. (2010) on Irish dairy farms; PETRIN and LEVINSOHN (2012).

Source: PETRICK and KLOSS (2013a), Author for additions in column D.

2.1.3 Additively separable, time-invariant firm characteristics

The key idea of this approach is that ω_{it} can be further decomposed into:

$$\omega_{it} = \gamma_t + \eta_i + v_{it}, \tag{2-3}$$

where γ_t is a time-specific shock that is identical for all farms in t (likely a type VII event), η_i is a farm-specific fixed effect that does not vary over time (a type VI factor), and v_{it} is the remaining farm- and time-specific productivity shock (type VII). Think of γ_t as representing common weather or policy shocks andcapturing soil quality or time-invariant preferences of the manager. In a farming context, v_{it} may represent local weather conditions that vary between farms and years. If they are not anticipated by the manager, v_{it} is subsumed into ε_{it}. If the production function is linearly separable in the logs of observed and unobserved factors, a commonly used functional form is Cobb-Douglas, so that the function can be written as $y_{it} = \alpha^A a_{it} + \alpha^L l_{it} + \alpha^K k_{it} + \alpha^M m_{it} + \gamma_t + \eta_i + \varepsilon_{it}$, with lowercase letters denoting logs, α^X the coefficients to be estimated, and X a shorthand for the observed production factors $X \in \{A, L, K, M\}$. Using panel data, a *within transformation* expresses all values as deviations from farm-specific means and thus eliminates η_i and all levels from the equation:

$$y_{it} - \bar{y}_i = \textstyle\sum_X \alpha^X (x_{it} - \bar{x}_i) + \gamma_t + (\varepsilon_{it} - \bar{\varepsilon}_i), \tag{2-4}$$

where $\bar{y}_i$, $\bar{x}_i$, and $\bar{\varepsilon}_i$ denote farm-specific log means over time. The fixed effect is hence swept out of the equation. This was introduced by HOCH (1955) and MUNDLAK (1961) in a farming context to eliminate "management bias" from the production function equation. This model has found widespread application at different levels of aggregation. The effect of γ_t is typically taken into account by including $T - 1$ time dummies in the model; T is the total number of observed time periods.[5] An alternative to the within transformation is to estimate the model in first differences, as discussed by WOOLDRIDGE (2010: 321-326). In this way variables are transformed by subtracting the preceding or lagged value in econometric parlance value from the current one.

MUNDLAK et al. (2012: 146) present a recent application to agricultural productivity at the country level where the farm-specific fixed and year effects alone explained 98.5% of output variation. Even so, the question remains whether it is legitimate to assume that v_{it} is an innovation that is orthogonal (i.e., uncorrelated) to observed factor use so that all unobserved factors are indeed either

5 This treatment is also applied in the subsequently outlined estimation strategies.

time invariant or the same for all farms. Table 2.1 suggests that farm- and time-specific unobserved effects *which the farmer still takes into account when making input decisions* (type IV and V) are very likely to be relevant. Examples include annual fluctuations in rainfall or pest occurrence as well as patterns of livestock health. Furthermore, applications in practice have found that the within transformation removes (too) much variance from some of the variables, particularly those which display little variation over time. In agriculture, input levels of the type II production factors land, labour, and fixed capital often vary only little in time. As a consequence, the signal-to-noise ratio with regard to these factors is reduced and the estimated coefficients are biased downward (Griliches and Mairesse 1998: 180-185). Finally, without further assumptions, the collinearity problem is not addressed at all by this approach.

2.1.4 Profit maximisation and perfect competition

This approach imposes further microeconomic theory upon the data, including its main assumptions of profit maximisation and perfect competition on product and input markets. A key result of this theory is the first-order condition (2-2), which multiplied through with $\frac{M}{p^Y Y}$ yields (for the case of materials):

$$\frac{\partial f}{\partial M}\frac{M}{Y} = \frac{p^M M}{p^Y Y}. \qquad (2\text{-}5)$$

If one further assumes constant returns to scale, (2-5) says that the production elasticity of each input (left-hand side) is equal to its value share in revenue (right-hand side). All value shares add up to one. Given these assumptions, revenue shares of inputs are valid estimators of production elasticities. For the simple Cobb-Douglas technology, the problem of estimating production elasticities has thus been "solved" by the imposition of strong theoretical assumptions. However, production function estimates of elasticities in agriculture were often found to differ from observed revenue shares (Mundlak, 2001). These differences may even be an object of investigation, for example in studies of credit rationing (Petrick, 2005). Such studies thus require productivity estimation independent of the revenue share.

For more flexible functional forms, (2-5) has led to the widely applied share regression model. For example, if the production function is assumed to be translog, thus also including quadratic and cross terms of the variable inputs in logs, the first order condition yields the following *share regression* (again for the case of materials):

$$s_{it}^{M} = \alpha^{M} + \alpha^{MM} m_{it} + \alpha^{MA} a_{it} + \alpha^{ML} l_{it} + \alpha^{MK} k_{it} + \omega_{it}^{M} + \varepsilon_{it}^{M}, \qquad (2\text{-}6)$$

with $s_{it}^{M} = \dfrac{p_{it}^{M} M_{it}}{p_{it}^{Y} Y_{it}}$ the revenue share of materials of firm i at time t, α^{X} the direct and cross-elasticities of the involved inputs, ω_{it}^{M} the part of the unobserved productivity characteristic that affects s_{it}^{M}, and ε_{it}^{M} an IID error term. Such an equation can be derived for all production factors, thus constituting a system of equations amenable to estimation by imposing the parameter restrictions derived from theory (Berndt and Christensen, 1973; see Bonnieux, 1989 for an application to French agriculture). The estimation itself can then be carried out by a system estimator such as three-stage least squares (see Greene, 2011: 369-374 for an introduction to system estimators).

Note that (2-6) is still subject to the endogeneity and collinearity of factors. The way out of these problems typical to this approach is finding appropriate instruments for the input levels. The role of the instruments would be to distill the part out of m, a, l, and k that is not correlated with ω_{it}^{M} (i.e., the exogenous part). In the given theoretical framework, the most natural candidates are factor prices, which were used to estimate systems of share equations such as (2-6) by two- and three-stage least squares (Antle and Capalbo, 1988). Given the possibility to also recover technology parameters from profit and cost functions by means of duality theory (Chambers, 1988), there is now a large body of empirical literature with agricultural applications of this approach (see the critical review in Mundlak, 2001).

Despite the applications in the literature, theoretical conditions that link primal production to dual profit or cost functions are often not met empirically, such as convexity in the former and concavity in the latter. Moreover, the use of prices to solve the two identification problems must be questioned on both theoretical and empirical grounds. To qualify as instruments, prices must not be endogenous to the decision problem of the farmer. This condition is usually ensured by the assumption of perfectly competitive markets on which atomistic agents have no price-setting power. In agriculture, it may hold for a number of output markets, but is very unlikely to prevail on most factor markets. For example, in many European countries farmland markets are known to be characterised by spatial oligopolies and strong government regulation (Huettel and Margarian, 2009; Ciaian et al., 2012). As noted before, agricultural labour is usually very immobile due to life-cycle considerations and specific human capital. Agricultural credit may be due to a rationing regime that depends on the credit history of the farmer or his ability to provide collateral. Hence, factor prices may not be exogenous and may depend on the farmer's past and current decisions. Under such conditions, the theoretical model underlying this approach is clearly too simplistic to allow straightforward identification of the production function.[6]

On the other hand, if factor markets were at least approximately working as postulated by the theoretical ideal, there should be little price variation across farms so that the value of prices for solving the endogeneity and collinearity problems is doubtful. In the first place, this is a theoretical argument – in perfect markets, there is no price variation across firms and so the different flexible factors are not identified by the data generating process. In fact, empirical applications have shown that price variation is indeed often small and may be due to quality differentials (Griliches and Mairesse, 1998: 189). With regard to agricultural labour or land, it may be hard to find appropriate price series at all.

2.1.5 Heterogeneous frictions in factor adjustment

If prices are problematic instruments, another option is to look for a different source of exogenous variation that has explanatory power for productivity analysis. Past decisions on factor use is one such source now routinely employed; it is based on the literature on dynamic panel data modeling (Arellano and Bond, 1991; Blundell and Bond, 1998). This literature suggests that current variation in input use is caused by lagged adjustment to past productivity shocks. It thus introduces the history of input use as a source of identification which is plausible if modifications of input levels are subject to adjustment costs (Bond and Söderbom, 2005). This approach effectively turns observed input levels into state variables (type II) and subjects them to an intertemporal optimisation problem. One way to account for costly adjustment is to allow serial correlation of the unobserved productivity characteristic of the firm, so that it could be written as:

$$v_{it} = \rho v_{it-1} + e_{it}, \text{ with } |\rho| < 1, \text{ with }, \qquad (2\text{-}7)$$

where ρ denotes the autoregressive parameter and e_{it} is an independent mean zero innovation. Substituting (2-7) as well as (2-3) into a Cobb-Douglas specification of (2-1), Blundell and Bond (2000) suggest a dynamic production function specification that can be estimated with a *dynamic panel data estimator*:

---

6 An important step to relax the rigid assumptions of this approach was the introduction of dynamic duality in studies of agricultural production (e.g., Thijssen, 1994; Sckokai and Moro, 2009). Conceptually, these studies build a bridge to the approaches described in subsequent sections. The empirical interest is often no longer on recovering factor productivities, however.

$$y_{it} = \sum_X (\alpha^X x_{it} - \alpha^X \rho x_{it-1}) + \rho y_{it-1} + (\gamma_t - \rho \gamma_{t-1}) +$$

$$(1-\rho)\eta_i + \varepsilon_{it}^*. \tag{2-8}$$

Alternatively, this model can be written as:

$$y_{it} = \sum_X \pi^{1X} x_{it} + \sum_X \pi^{2X} x_{it-1} + \pi^3 y_{it-1} + \gamma_t^* + \eta_i^* + \varepsilon_{it}^*, \tag{2-9}$$

subject to the common factor restrictions that $\pi^{2X} = -\pi^{1X}\pi^3$ for all X.

Blundell and Bond (2000) use lagged levels and differences of inputs as instruments in a general methods of moments (GMM) framework to estimate (2-8). If the η_i are removed by first differencing (FD), this estimator allows the consistent recovery of all input elasticities in (2-1) as well as ρ. Blundell and Bond (2000) suggest the method of minimum distance to test whether the parameters estimated by the unrestricted model (2-8) conform to the restrictions imposed by (2-9). A minimum distance estimate for (2-9) is chosen such that the distance between the coefficient estimates of the unrestricted and restricted model is minimal (Wooldridge 2010: 545-547).

Note that the within transformation (chapter 2.1.3) assumes *strict* exogeneity of inputs which means that ω_{it}^M must not be transmitted to any future period (contrary to what is assumed in (2-7)). First differencing to eliminate fixed effects only assumes that input levels are *sequentially* exogeneous (i.e., transmission of ω_{it}^M to the next but one and subsequent periods are allowed) (Chamberlain, 1982; Wooldridge, 2010: 321-326). FD is thus the typical approach to eliminate time invariant heterogeneity in GMM applications, as it allows input levels lagged more than two periods to be used as instruments for contemporaneous differences (Arellano and Bond, 1991). Of course, these instruments will only have power if there actually *is* such a transmission (e.g. motivated by adjustment costs). To increase the power of the GMM approach, Blundell and Bond (1998) have shown that in addition to past levels, lagged differences of inputs also can be used as instruments if they are orthogonal to the fixed effects (η_i) – an assumption which will hold if their variance is assumed to be, in the broadest sense, stationary Roodman (2009: 114-115). This leads to the systems GMM estimator for production functions presented in Blundell and Bond (2000) and applied by Hempell (2005). Hempell uses data on German service firms from 1994 to 1999. In the empirical application of Blundell and Bond (2000), the preferred systems estimator produces a lower employment coefficient and a higher capital coefficient than OLS or within estimators, thus correcting the expected bias.

If factor levels can suitably be instrumented by this approach, it addresses both the endogeneity and the collinearity problems. Contrary to the duality approach presented in section 2.1.4, it is much more plausible that the instruments proposed here are actually valid in an agricultural context. There are important production factors in agriculture which are subject to adjustment costs (or "transaction costs"; type II variables in Table 2.1) and such costs should be an element in any plausible theory of agricultural factor markets. As the nature of these costs is likely to differ among factors (see section 2.1.1), it is also plausible that different factors of production display different dynamic paths of adjustment. This is a favourable condition for identification (BOND and SÖDERBOM, 2005). It is only with regard to some intermediate inputs such as seed, fertiliser, plant protection, concentrate, or energy that factor use appears to be more flexible so that the assumption of adjustment costs may be harder to justify (type I factors). In sum, this estimator is a promising candidate for agricultural applications.

2.1.6 Monotonous coevolution of unobserved productivity shocks with observed firm characteristics

The final method to be discussed here avoids the main disadvantage of any fixed effects approach to unobserved heterogeneity, which is the typically low variance of the transformed variables. However, it also does not rely on the strong a-priories about market structure of duality theory to identify the productivity parameters of interest. Rather, it attempts to proxy ω_{it} (as a compound type IV to VI production factor) by a *non-parametric control function* which itself contains only observed firm characteristics. OLLEY and PAKES (1996) were the first to suggest log investment (i_{it}) as an observed characteristic driven by ω_{it}:

$$i_{it} = i_t(\omega_{it}, k_{it}), \tag{2-10}$$

where k_{it} is the pre-determined level of capital use at time t. The latter is assumed to evolve according to $k_{it+1} = (1-\delta)k_{it} + i_{it}$, with δ the depreciation rate.

The function $i_t(\cdot)$ can vary over time and is not parametrically restricted except that it needs to be monotonous in ω_{it}. This latter trait allows inversion of this function, so that:

$$\omega_{it} = h_t(i_{it}, k_{it}),$$

where h_t is now potentially observable and acts as a proxy for ω_{it}. Furthermore, it is assumed that unobserved productivity follows a first-order Markov process:

$$\omega_{it} = E[\omega_{it}|\omega_{it-1}] + \xi_{it}, \tag{2-11}$$

where ξ_{it} is an innovation (a type VII factor) uncorrelated with k_{it}, but possibly correlated with the other factors in the production function. Because k_{it} is a type II factor, the moment condition $E[\, k_{it}\xi_{it}] = 0$ can be used to identify α^K.

Given this setup, estimation proceeds in two stages. The basic idea is to jointly control for the influence of k and ω in the first stage and to recover the true coefficient of k as well as ω in the second. Referring again to the Cobb-Douglas example, all observed factors except capital are assumed to be fully variable type I factors. Their elasticities are determined in the *first stage* by substituting $h(\cdot)$ into the production function and estimating:

$$y_{it} = \alpha^A a_{it} + \alpha^L l_{it} + \alpha^M m_{it} + \phi_t(i_{it}, k_{it}) + \varepsilon_{it}, \tag{2-12}$$

where $\phi_t = \alpha^K k_{it} + h_t(i_{it}, k_{it})$. In practice, ϕ_t is approximated by a low-order polynomial of i and k which controls for ω_{it}. Equation (2-12) shows that ϕ_t is assumed to be additively separable from the remaining variable inputs. Flexible functional forms involving interactions of all variable and fixed inputs (such as the translog) thus cannot be implemented with this procedure.

In the *second stage*, α^K is determined in a series of steps (e.g., Petrin et al., 2004). First, using the parameters of ϕ_t and a candidate value for α^K, a prediction $\widehat{\omega}_{it}$ is computed for all periods. Next, $\widehat{\omega}_{it}$ is regressed on its lagged values to obtain a consistent predictor of that part of ω that is free of the innovation ξ. Finally, using the parameters of the variable factors from the first stage together with the prediction of the "clean" ω_{it} and the moment condition $E[\, k_{it}\xi_{it}] = 0$, a consistent estimate of α^K can be obtained by minimum distance.[7] In their original application to the US telecommunications equipment industry, Olley and Pakes (1996) show how this procedure yields lower labour coefficients than OLS and higher capital coefficients than within. In the only application to agriculture known to us, Kazukauskas et al. (2010) found for Irish dairy farms that the materials coefficient estimated with an OP procedure was lower than the OLS result.

7 This is the algorithm used in literature subsequent to Olley and Pakes (1996). In the original paper, it was combined with an exit and entry mechanism for firms which I ignore to simplify the exposition. Furthermore, the data does not allow to model such a mechanism because exit and entry are random in the FADN data base.

One problem that arises from using investment as a proxy is zero observations for certain years and firms. LEVINSOHN and PETRIN (2003) therefore suggested materials instead of investment as a proxy of ω_{it} in the previous algorithm. Again, the assumption is that materials evolve monotonously with the unobserved productivity characteristic, so that the effect of the latter can be inverted out. Materials is assumedly a type I factor and thus part of the production function. However, in the LP approach, its elasticity cannot be estimated in the first stage, because it is now part of $h(\cdot)$. Therefore, the additional moment condition $E[\,m_{it-1}\xi_{it}] = 0$ is postulated to obtain α^M in the second stage.

If the control function fully captures the influence of ω_{it}, it solves the endogeneity problem and provides a useful alternative to the fixed effects approaches described above. However, in agriculture, the assumptions on monotonicity and dynamic evolution of the productivity shock must be considered with caution. A key question is *what exactly ω_{it} is representing and whether investment or material use are good proxies for it*. If ω_{it} stands for annually fluctuating, unobserved factors (type IV) such as management effort or reaction to environmental conditions, there may be cases in which the "right behaviour" of the farmer (i.e., positive ω_{it}) does not lead to more investment. The same is true for materials. The productivity enhancing reaction to environmental shocks in crop production may sometimes be less input use (fertiliser, chemicals) rather than more. In all these cases, neither investment nor materials will be good proxies of ω_{it}. Furthermore, the "memoryless" first-order Markov process appears unconvincing if ω_{it} actually represents unobserved type V factors which are subject to adjustment costs. They evolve slowly and will typically have implications for the intertemporal optimisation problem, so that k_{it} is also affected by them and (2-10) is misspecified. Investment may not be a good proxy for ω_{it} if there are other important determinants of it beyond k_{it}. In a farming context, this is likely to be the case, because investment decisions are usually influenced by long-term business strategies and/or the availability of a farm successor.

Another problem with the procedure suggested by OP and LP is that it does not solve the collinearity problem. As discussed at length by ACKERBERG et al. (2006), unless one is willing to make very unintuitive assumptions on measurement error or timing, there is no data generation process that separately identifies the coefficients of the type I factors in either of the two approaches. ACKERBERG et al. therefore suggest giving up estimation of these coefficients in the first stage altogether, and invoke additional timing assumptions that justify moment conditions for estimating these coefficients in the second stage. In the framework of a translog specification, GANDHI et al. (2011) propose to estimate the coefficient for one free input from a share regression akin to equation (2-6) and then to proceed in a similar way as described in this section to recover the other elasticities. WOOLDRIDGE (2009) suggests a simple procedure that borrows the identification strategy from OP and LP and modifies as well as extends the moment conditions to resolve the collinearity problem. Hence, this approach is referred to as the WOOLDRIDGE/LEVINSOHN/PETRIN (WLP) estimator (PETRIN and LEVINSOHN, 2012). Returning to the Cobb-Douglas example, it is again assumed that some function $h(\cdot)$ to proxy ω exists such that $\omega_{it} = h(m_{it}, k_{it})$.[8] Unlike LP who state $E[\varepsilon_{it}|a_{it}, l_{it}, k_{it}, m_{it}] = 0$, Wooldridge allows for:

$$E\big[\varepsilon_{it}|a_{it}, l_{it}, k_{it}, m_{it}, a_{it-1}, l_{it-1}, k_{it-1}, m_{i,t-1}, \dots, a_{i1}, l_{i1}, k_{i1}, m_{i1}] = 0. \tag{2-13}$$

Therefore, ε_{it} is assumed to be orthogonal not only to current but also all past values of a, l, k and m. In practical implementation as proposed by WOOLDRIDGE (2009), (2-13) is weakened in that only current realisations and one lag of the inputs are assumed to be uncorrelated with the ε_{it}.

Furthermore, again the dynamics of the ω_{it} are fully described by (2-11). A further condition about them is proposed by Wooldridge as follows:

$$E\big[\omega_{it}|k_{it}, a_{it-1}, l_{it-1}, k_{it-1}, m_{it-1}, \dots, a_{i1,}l_{i1}, k_{i1}, m_{i1}\big]$$

$$= E[\omega_{it}|\omega_{it-1}] = g(\omega_{it-1}) \equiv g[h(m_{it-1}, k_{it-1})], \tag{2-14}$$

where g is an unknown productivity function. Equation (2-14) together with (2-11) provides some deeper insight into the innovation ξ_{it}. It states that this innovation is uncorrelated with current and past realisations of k and past values of a, l and m. Note, these assumptions are necessary to obtain a consistent estimate of α^K and α^M in the second stage of the LP procedure.

Now, the problem can be formulated in terms of two equations. The first is given by:

$$y_{it} = \alpha^A a_{it} + \alpha^L l_{it} + \alpha^K k_{it} + \alpha^M m_{it} + h(m_{it}, k_{it}) + \varepsilon_{it}. \tag{2-15}$$

The second can be obtained by plugging the last identity of (2-14), $\omega_{it} = g[h(m_{it-1}, k_{it-1})] + \xi_{it}$, into the production function:

[8] Wooldridge in contrary to Levinsohn/Petrin assumes that h is time invariant.

$$y_{it} = \alpha^A a_{it} + \alpha^L l_{it} + \alpha^K k_{it} + \alpha^M m_{it} + g[h(m_{it-1}, k_{it-1})] + e_{it}, \tag{2-16}$$

where $e_{it} = \xi_{it} + \varepsilon_{it}$. The moment conditions that hold for (15) are given by (13) and the ones that hold for (16) are:

$$E[e_{it}|k_{it}, a_{it-1}, l_{it-1}, k_{it-1}, m_{it-1}, \dots, a_{i1}, l_{i1}, k_{i1}, m_{i1}] = 0. \tag{2-17}$$

Hence, in (2-15) and (2-16), current and past values of k; past values of a, l, and m as well as functions of these can be used as instruments. Additionally, in (2-15) contemporaneous proxy variables and current realisations of a and l are valid instruments. Given this setup, the two equations (2-15) and (2-16) together with the moment conditions in (2-13) and (2-17) can be estimated within a GMM framework. Alternatively, one can identify the production function parameters by estimating (2-16) using IV estimation with instruments for a, l, and m (WOOLDRIDGE, 2009: 113). PETRIN and LEVINSOHN (2012) employ this second approach. The control function h is approximated by low-order polynomials of first-order lags of m and k. The function g is assumed to be a random walk with drift (WOOLDRIDGE, 2009: 114).[9]

In (2-16), I proxy for h with lags rather than contemporaneous values of m and k, as is done in the traditional control function setup. This treatment allows for easy implementation of the translog functional form (PETRIN and LEVINSOHN, 2012: 718).

The assumption of costly factor adjustment is a cornerstone of both the dynamic panel data approach described in section 2.1.5 and the present one. In both cases, this assumption provides moment conditions necessary for consistent estimation of the parameters. The main difference is that the former approach allows time-invariant fixed effects, whereas the latter does not. The former imposes a linear structure on the dynamic process, while it can be arbitrary in the latter. Even so, factor adjustment is assumed to occur in a single period in OP and followers, whereas the process covers many periods in the dynamic panel data models. In the light of agricultural applications, this may be one key advantage of the dynamic panel data approach.[10]

9 If $\omega_{it} = \lambda + \omega_{it-1} + \xi_{it}$, (2-16) becomes $y_{it} = \lambda + \alpha^A a_{it} + \alpha^L l_{it} + \alpha^K k_{it} + \alpha^M m_{it} + h(m_{it-1}, k_{it-1}) + e_{it}$.

10 Other subtle differences between the two approaches are discussed in ACKERBERG et al. (2006).

2.1.7 Interim evaluation of estimation approaches

The previous discussion has displayed the variety of assumptions invoked for addressing the endogeneity and collinearity problems inherent to production function estimation. In my opinion, the assumptions underlying within regression and the duality approach are fairly strong and implausible for the case of agriculture. Perhaps not surprisingly, they also often have not performed well in estimation practice. This insight shifts my attention to the promising new approaches using heterogeneous frictions in factor adjustment. I regard the presence of adjustment costs as particularly relevant for the production factors that are of key interest in agricultural applications. They also provide an interesting link to more sophisticated theories of business structures in agriculture, which usually embody some form of adjustment frictions in agricultural factor use (such as Allen and Lueck, 2002 or Pollak, 1985). So far, there are almost no applications to agricultural data of these new estimators. In the empirical chapters I aim to fill this void.

2.2 Outlier problems and their solution in production function estimation

In this subchapter, I examine the consequences on estimation in presence of outliers in the data and evaluate different solutions to this problem. Even though I consider the estimation of production functions, the insights should also be relevant to other model-driven empirical applications. After discussing the definition of an outlier, I survey the empirical practice for important data sources relevant in economics. This is accompanied by a simple empirical example illustrating the effects of outliers and different decontamination schemes on production function estimates. Finally, given my insights, I propose and outline the multivariate outlier decontamination procedure following Kirschstein et al. (2013).

2.2.1 The definition and identification of outliers

The presence of outliers in data is a severe problem because it biases estimates of parameters of interest. In many standard statistics textbooks there is an example of the arithmetic mean being biased because of the occurence of outliers (see, for instance, Fahrmeir et al., 2004: 55). This problem carries over to more complex models (e.g. the specification and estimation of the conditional mean of a function of interest). Hence, it also affects my main interest, the identification and estimation of production functions.

Outliers occur for for a variety of reasons including measurement errors, variations in the data generating process, or misreporting. In general, other than by being very pragmatic and not performing any outlier decontamination, there are two concurrent views on how to define and, consequently, identify outliers.[11] On the one hand, there are methods which assume a rigorous statistical model, in a sense that outliers are thought of as data coming from a different distribution (i.e., from something other than what the researcher is actually interested in). The whole data can then be understood as a mixture of two or more distributions, where the target distribution should comprise the majority of the probability mass. In contrast, if the outliers would constitute the majority, it would not be feasible to speak of outliers at all. In such cases the outliers should rather be considered as the distribution of interest. Methods falling into that category commonly aim at estimating some features of the target distribution while outlier detection is usually not their primary objective. Hence, it is more of an inherent by-product. If outliers are detected, they are identified with that precise model in mind, meaning that the methods in that category cannot be reasonably applied when the underlying (distributional) assumptions are not fulfilled (at least no useful results can be expected in that case). In this category, many robust estimators for various models can be found (see, e.g., Rousseeuw and Leroy (1987); Barnett and Lewis (2000); Hampel et al. (2005); Maronna et al. (2006); Huber and Ronchetti (2009) for an overview of such methods).

On the other hand, other methods follow a more sample-oriented view on how to define an outlier. Universally, all of these methods interpret outliers as observations that differ substantially from the target observations (cf. Johnson, 1992; Barnett and Lewis, 2000). Obviously, there are numerous proposals on how that difference can actually be quantified. Many of these methods are based on some type of distance, either between any two observations or with respect to some reference point, but there are also other approaches (e.g., based on depth or on the empirical density). Members of this class of approaches include nearest neighbour or clustering based methods. Furthermore, these and other representatives such as the classification-based neural and Bayesian network approaches mostly originate from the computer science and data mining literature (Chandola et al., 2009). The primary goal is indeed the identification of outliers with these methods. An additional analysis might still be carried out on the identified non-outliers, though. The biggest advantage of methods falling into that category is that they can generally be applied without being limited to situations where certain distributional assumptions are fulfilled.

11 Not performing any outlier control is a common practice (see chapter 2.2.2).

2.2.2 Going practice to outlier treatment in empirical economics

To obtain a general idea of how outliers have been treated within the field of (micro-)economics, I surveyed studies from two additional micro data sources that have approximately the same importance as the "Farm Accountancy Data Network" (FADN), my data source, in agricultural economics. For the FADN data, I reviewed studies from two recent research projects funded by the European Commission under FP7 which heavily employed this data set: "Farm Accountancy Cost Estimation and Policy Analysis of European Agriculture" (FACEPA) and "Factor Markets". The former ran from 2008 to 2011, and the latter comprised the years 2010 to 2013. I reviewed working papers from the World Bank's "Living Standard Measurement Survey" (LSMS) studies for the field of development economics, and I analysed studies employing data from the "German Socio-Economic Panel" (GSOEP) for the field of general economics. The GSOEP's "SOEPapers" series is a collection of such work. I summarise the going practice of outlier treatment for a sample of work from these three data sources in Table 2.3. I reviewed empirical studies (i.e., studies including a data analysis component) for this sample.

Generally, the three data sources leave the assessment as well as handling of outliers to the researcher (Grosh and Munoz, 1996; Haisken-DeNew and Frick, 2005; European Commission, 2010a). For the LSMS, Grosh and Munoz (1996, p. 125) are explicitly open about this view, explaining that "further treatment of these problems should be left to analysts, since there is no universally acceptable solution to these problems". This argument opens up many possibilities for researchers to deal with outliers. In addition, it explains why there are so many different ways to deal with outliers.

In empirical economic literature emerging from these sources, a large group of authors assess outliers by visual observation - which is usually only feasible for smaller data sets and two dimensions - and contextual reasoning. Based on these modes of operation they drop implausible cases (e.g., Deaton, 1981, 1988; Crosetto and Filippin, 2012; Obschonka et al., 2013; Oltmanns et al., 2014; Auer and Danzer, 2014). Other approaches involve the transformation of variables (e.g., logarithmisation), the application of influence measures or censoring of extreme values (cf. Schneck, 2011; Olper et al., 2012; Liverpool-Tasie et al., 2015). Some authors remove outliers without stating their method of detection (e.g., Bauernschuster et al., 2011; Ciaian et al., 2011; Lang, 2012; Kemptner, 2013; Arnold et al., 2014).

More structured approaches apply a two-step procedure by first identifying and removing extreme observations for a target variable (univariate outliers)

Table 2.3: Outlier treatment in empirical economics.

	FADN	LSMS	GSOEP
Sample composition	FACEPA (2009-2011); Factor Markets (2011-2013)	World Bank Living Standard Measurement Study (1980-2002) & Policy Research Working Paper (2010-2015)	SOEPapers (2011-2015)
Number of studies surveyed	36	129	396
Number of studies that deal with outliers	11 (≈30.6%)	23 (≈17.8%)	34 (≈8.6%)
Dominant decontamination approach	Univariate	Univariate	Univariate
Frequently used methods and examples	*Trimming* (Guastella and Moro, 2013; Petrick and Kloss, 2013a), *No method stated* (Bakucs et al., 2010; Ciaian, 2011)	*Visual observation* (Deaton, 1981, 1988); *Trimming* (Oseni et al., 2014; Backiny-Yetna and McGee, 2015), *Censoring of outliers* (Liverpool-Tasie et al., 2015)	*Visual observation* (Crosetto and Filippin, 2012; Obschonka et al., 2013), *Trimming* (Pfeiffer and Schulz, 2011; Murphy and Oesch, 2015), *No method stated* (Lang, 2012; Arnold et al., 2014)

Notes: FACEPA: Farm Accountancy Cost Estimation and Policy Analysis, FADN: Farm Accountancy Data Network, GSOEP: German Socio-Economic Panel, LSMS: Living Standards Measurement Survey.

Source: Author.

prior to the desired follow-up analysis. Therefore, some sort of threshold value is defined to separate outliers from the non-outliers. What distinguishes these methods is their statistical definition of such thresholds. The approach most commonly used among all empirical data sources is 'trimming', either by removing some percentage of the data (usually 1% or 5%) at the top and/or bottom of the distribution of a single univariate measure central to the analysis or by applying a quantile-based rule (e.g., upper/lower quartile ± $s \cdot IQR$ with s being some scaling factor and IQR the interquartile range; e.g., Pfeiffer and Schulz, 2011; Zibrowius, 2012; Schmitt, 2013; Sorgner and Fritsch, 2013; Guastella and Moro, 2013; Oseni et al., 2014; Murphy and Oesch, 2015; Avdic and

BÜNNINGS, 2015; BACKINY-YETNA and MCGEE, 2015). However, univariate outlier identification approaches neglect the multivariate nature of many models that researchers specify.

Most strikingly, the group of papers that does not mention any outlier control is by far the largest (e.g., PITT, 1995; ALDERMAN, 1998; BAUER et al., 2011; CARLETTO ET AL., 2011; HEADEY et al., 2012; DUSTMANN and GÖRLACH, 2015; SCHURER, 2015). For instance, an outlier control was mentioned in only 8.6% of the sample comprising 396 studies that I reviewed from the SOEPapers data base (period 2011 - 2015).

Another way is to incorporate the outlier problem directly into the estimator by developing or applying a robust estimation procedure. HÜBLER (2012) applies a simple quantile regression estimator. This way, outliers are accounted for by using an appropriate estimation method. However, with the estimation of causal relationships in mind, this method is not feasible because it does not account for endogeneity and collinearity issues. Therefore, while this approach mitigates the outlier problem, it would not identify the causal relationship of interest. To develop a robust estimator capable of dealing with identification problems would require extensive work which is beyond the scope of this work.

Another two-step approach, and so far the only one to my knowledge which accounts for outliers in a multivariate manner using FADN data, has been used within the FACEPA project (BAHTA et al., 2011). Their detection procedure employs the minimum covariance determinant (MCD) algorithm following ROUSSEEUW (1985). This approach is widely used in many fields of applied statistics as a simple search on Google Scholar. However, this approach imposes strict distributional assumptions on the data. Furthermore, their subsequently applied estimation procedure does not account for farm-specific heterogeneity.

In summary, no outlier control is mentioned in the majority of studies. In those studies that do mention a decontamination procedure, univariate methods prevail.

2.2.3 A simulated example

To demonstrate the effects of outliers on non-robust estimations, I discuss a simplified example. Therefore, I simulate a data set with 100 farms over seven periods. The data-generating process of the majority of observations is, hence, as follows:

$$y_{it} = 0.4 \cdot l_{it} + 0.6 \cdot k_{it} + \omega_i + \varepsilon_{it}, \tag{2-18}$$

where y, l and k are the natural logarithm of output, labour, and capital; ω_i represents unobserved heterogeneity with $\omega_i \sim N(0, 25)$; and ε_{it} is the remaining disturbance following $N(0,1)$.[12] The labour and capital inputs are random variables with $N(0, 4)$.

As outliers, two data sets with 20 small farms over the 7 time periods with $\omega_i \sim N(-5, 4)$ and $l_{it}, k_{it} \sim N(-5, 9)$ are generated. Finally, I arrive at two different outlier-contaminated data samples by adding the outliers to the raw sample. In the first outlier data set, I set $cor(l_{it}, k_{it}) = 0$ (sample I) and in the second $(l_{it}, k_{it}) \approx -1$ (sample II), so that labour and capital are (almost perfect) substitutes. This assumption has been chosen for demonstrative purposes. Please note, I further differentiate in this way to illustrate the effects of multicollinearity on outlier-infested data and how outlier decontamination is able to mitigate such effects. As I will demonstrate later, multicollinearity is an important feature in the FADN data employed in my analyses. In both samples the production function for the outlier sets is:

$$y_{it} = 0.99 \cdot l_{it} + 0.01 \cdot k_{it} + \omega_i + \varepsilon_{it}, \tag{2-19}$$

i.e., in both sets $7 \cdot 20 = 140$ observations are generated by (2-19) which are regarded as outlying from the process assumed in (2-18).

Table 2.4: Results of simulated production function estimation.

Decontamination scheme	Raw sample			Sample I			Sample II		
	N	Labour	Capital	N	Labour	Capital	N	Labour	Capital
Without	700	0.41***	0.61***	840	0.28***	0.73***	840	0.22***	0.80***
Univariate	526	0.39***	0.62***	677	0.24***	0.74***	711	0.20***	0.78***
Multivariate	498	0.40***	0.61***	539	0.42***	0.62***	534	0.41***	0.62***

Notes: *** (**, *) significant at the 1% (5%, 10%) level. The univariate decontamination is based on the average capital productivity per farm. I exclude values outside [Q1-1.5·IQR;Q3+1.5·IQR]. The multivariate decontamination is based on the pMST procedure discussed below. Estimation samples include farms with a minimum panel representation of four years.

Source: Author.

---

12 Throughout, I assume $N(\mu, \sigma^2)$, i.e. $N(mean, variance)$.

I present fixed effects regression results to control for unobserved time constant farm-specific effects (ω_i) for all three data sets, applying a) no decontamination, b) univariate outlier decontamination, and c) multivariate decontamination using the pMST method outlined below in the next subchapter. The estimated output elasticities for labour and capital input as well as final sample size are summarised in Table Table 2.4.

Given an appropriate estimator that controls for unobserved heterogeneity, I am able to recover the 'true' production function parameters, as given in (2-18), rather precisely for both contaminated samples with the multivariate decontamination method. Univariate decontamination yields no improvement in estimation accuracy. In fact, results are close to the sample without decontamination. Hence, this simple procedure fails to detect the meaningful outliers in a sense that it cannot detect outliers in dimensions other than the one under consideration. Therefore, without even applying multivariate decontamination, a relatively small number of outliers (about 16.7 per cent in this example) severely biases the estimates.[13]

Multicollinearity (sample II) increases the problem which results in even more biased estimates. However, multivariate decontamination is able to identify the outliers correctly in this case, too. In particular, this example demonstrates that the presence of multicollinearity does not interfere with the multivariate detection procedures' ability to identify outliers correctly, even in this extreme case of almost perfect correlation between the inputs. Moreover, to put it differently, the pMST procedure might also mitigate multicollinearity. This, as illustrated in this example, works especially well if the outliers are the (major) source of multicollinearity in the data.

In summary, only after controlling for both unobserved heterogeneity and the effects of outliers I am capable of obtaining reliable output elasticity estimates. Hence, these two issues have to be treated separately. Given the results of this simple example, an effective outlier decontamination can only be conducted if all model dimensions are considered (i.e., multivariate outlier detection is conducted).

---

[13] This is close to the number of outliers that is observed for West Germany (see chapter 5.2.1).

2.2.4 Decontamination by pruning the minimum spanning tree

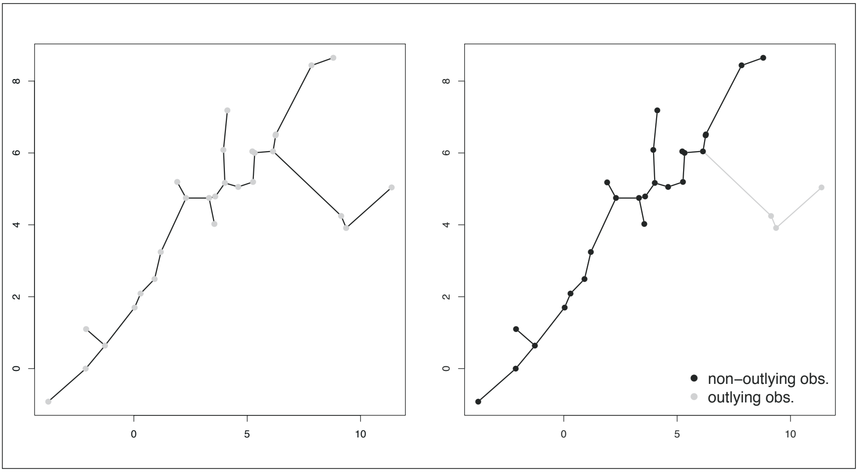
Robust statistical methods are designed to deliver unbiased estimates of measures of interest in the presence of contaminated data sets. A general two-step-approach is, therefore, to identify an outlier-free subsample first. This step can be called decontamination. In turn, observations not belonging to the uncontaminated subsample are suspected to be outliers. In contrast to outlier detection approaches, this procedure tries to find an outlier-free subsample instead of outliers. This way, I accept the risk to falsely discard non-outlying observations in favor of a (most likely) outlier-free subsample. A robust estimator relying on a non-parametric decontamination procedure is the pMST estimator which is a robust estimator of multivariate location and scatter (Kirschstein et al., 2013).[14] The decontamination results from **p**runing of the so called **M**inimum **S**panning **T**ree (MST) of a data set. The idea is that each observation of a data set represents a point in Euclidean space whereby its coordinates are the observation's values in each dimension. Note that in a panel data context, an observation corresponds to an entry in the data base (e.g., a farm's record of labour, land, materials, and capital use in a certain year).[15] For qualifying observations to be outlying, the pMST procedure implies that outliers are isolated with respect to similar observations. Similarity is here defined as the Euclidiean distance between two observations. Similar observations can be interpreted as neighbours. The spanning tree concept is used to identify the observations' neighbourhoods. See Figure 2.1 for a graphic illustration in two dimensions. In this illustration one can see that this concept is intuitive to the outlier definition as observations differing substantially from target observations. The aim is to select a minimal set of connections between observations such that all observations are connected with each other. Among all spanning trees, the minimum spanning tree has smallest weight; this means the sum of the lengths of all connections is minimal. For decontamination, the pMST procedure then iteratively deletes the longest connections in the MST until a certain threshold is reached (see below). The largest (still connected) fraction of the original MST is finally retained as the non-outlying part of the original data set. A formal description of the algorithm follows.

---

14 The code is reproduced in appendix D.

15 In the formal presentation, I resign from using a time index to prevent a too cluttered notation.

Figure 2.1: Illustration of pMST procedure.



Notes: Left panel: minimum spanning tree. Right panel: pruned minimum spanning tree.
Source: Author.

Formally, given a data set $\mathbf{X}$ of points in dimension p, i.e. $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, and all pairwise links (edges) $\mathbf{E}$ (i.e.$\mathbf{E} \ni e_{ij} = \{\mathbf{x}_i, \mathbf{x}_j\}$ with $i, j = 1, \dots, n$ and $i \neq j$), the MST is defined as the graph $\mathbf{G} = (\mathbf{X}, \mathbf{E}^*)$ connecting all points of $\mathbf{X}$ such that its total length is minimised, i.e. arg $\min_{\mathbf{E}^* \subset \mathbf{E}} \left(\sum_{e_{ij} \in \mathbf{E}^*} w(e_{ij})\right)$. Typically, the weight (w) of an edge $e_{ij} = \{\mathrm{x_i}, \mathrm{x_j}\}$ is the Euclidean distances between $\mathbf{x}_i$ and $\mathbf{x}_j$ such that $w(e_{ij}) = \sqrt{(\mathbf{x}_i - \mathbf{x}_j) \cdot (\mathbf{x}_i - \mathbf{x}_j)'}$. From this fact follows that Euclidean distances are meaningful for the data set (i.e. all variables are estimated on the same, or at least a similar, scale). The term connected refers to the property that there must be a path – a sequence of edges – in $\mathbf{G}$ between any two points of $\mathbf{X}$. It can be proven that the number of edges in $\mathbf{G}$ is always $|\mathbf{E}^*| = n - 1$ (JUNGNICKEL, 2008). Moreover, it can be shown that the MST is unique if all edges in $\mathbf{E}$ have unique weights. The MST can be efficiently computed even for large data sets. See e.g. JUNGNICKEL (2008) for a review of efficient algorithms.

Given the MST, pruning is realised by successively deleting edges in $\mathbf{G}$ according to their length. Therefore, in the first iteration the longest edge is removed, in the second iteration the second longest edge, and so on. This way $\mathbf{G}$ is split into several subgraphs which are not connected among each other. During the pruning process, the subgraphs' cardinality (i.e., the corresponding number of observations) declines. The pruning process is stopped if, by deleting the next edge, the cardinality of the largest subgraph would fall below $[(n+p+1)/2]$. Stopping at this bound assures that subsequently applied estimators achieve a maximum breakdown point (i.e., they are resistant against a maximum level of contamination. The largest subgraph at this point contains more than or exactly $[(n+p+1)/2]$ observations and is denoted by $\mathbf{G}^* = (\mathbf{X'}, \mathbf{E'})$ with $\mathbf{X'} \subset \mathbf{X}$ and $\mathbf{E'} \subset \mathbf{E}^*$. This approach was first proposed by Bennett and Willemain (2001). A discussion on the robustness properties of the associated estimators can be found in Kirschstein et al. (2013). Problematically, in most real-world data sets much less than $n - [(n+p+1)/2]$ outliers occur. To avoid low efficiency of robust estimators, reweighting procedures were proposed to enlarge the uncontaminated subsample.

For MST-based outlier decontamination, reweighting implies that a critical edge length w_α^{crit} has to be determined up to which (with a certain probability α) the MST consists of uncontaminated observations only. To estimate w_α^{crit}, a nonparametric approach relying on a finite sample version of Chebychev's inequality is described in Liebscher and Kirschstein (2014). The main idea is to estimate w_α^{crit} based on the mean edge length μ^w and the edge lengths' standard deviation σ^w. The parameters μ^w and σ^w are estimated based on the edge set $\mathbf{E'}$ of the initial robust subset of $\mathbf{G'}$, for w_α^{crit} follows $w_\alpha^{crit} = \hat{\mu}^w + \sqrt{\frac{(m^2-1)}{m^2 \cdot (1-\alpha) - m}} \cdot \hat{\sigma}^w$

where denotes the cardinality of $\mathbf{E'}$. Once w_α^{crit} is determined, $\mathbf{G'}$ is rebuilt by attaching all edges of $\mathbf{G}$ with edge weights smaller than or equal to w_α^{crit}. This way, a still robust but larger subgraph (say $\mathbf{G''}$) is determined whose associated observations are considered as the outlier-free subsample used in further analyses.

3 Database

The FADN provides a farm level data set that holds accountancy data for 25 of the 28 EU member states. Each year about 80,000 farms are sampled. They represent a population of about 5 million farms in the member states. In each member state a liaison agency is responsible for the data collection and transmission, which consists of about 1,000 variables including structural, economic, and financial data. Many aggregate variables can be partitioned into its various components as the data additionally includes these. A stratified sample is obtained to account for the heterogeneity of farms and to maintain representativeness for agriculture in the European Union. The stratification criteria are region, economic size, and type of farming.

The farm universe consists of all farms with more than one hectare or those with less than one hectare that provide the market with a specified amount of output. All non-commercial farms are excluded from this universe to arrive at the field of observation from which the data will be sampled. A farm must exceed a certain economic size to be classified as a commercial farm. It is measured in economic size units (ESU). One ESU represents a certain amount in euros and is periodically adjusted for inflation. The concept of standard gross margin (SGM) is used to determine the economic size of farms. To this end, liaison agencies assign SGM coefficients "on a regional basis for more than 90 separate crop and livestock items" (European Commission, 2010a). The SGM of a crop farm is then the sum of the product of a particular SGM coefficient times the respective number of hectares (European Commission, 2014). In addition, farms are classified by type of farming (TF).

Even though the FADN is a stratified sample, there is no need to use a sample weighting scheme for estimation. In my analyses, I employ a structural modeling approach, that is, I estimate the parameters of a theoretically motivated model. Hence, as long as the conditional mean is correctly specified and stratification is not based on the dependent variable of the model (total output) my estimates have causal interpretation regardless whether weighting is applied or not (Cameron and Trivedi, 2005: 820-823). While the latter is clearly not the case, the former will be achieved by the appropriate estimation strategies outlined in chapter 2.1, as with these I am able to carve out and utilise exogenous variation for the purpose of estimation. Virtually all empirical work employing FADN data implicitly adopts this perspective; see, for instance, SWINNEN and KNOPS (2013).[16]

In the present work, I only use field crop farms (TF1) (i.e., the farm's operation is the core activity for the farmer with at least 40 operating hours per week) (EUROPEAN COMMISSION, 2014; FADN data). I do this to justify the assumption of a homogenous state of technology across farms. Now one might argue that focusing only on field crop farms might introduce selection bias in my estimates. However, the statements I make are only with regards to the population of field crop farms and not the population including all possible types of farms. Therefore, my results should not suffer from sample selection bias. Furthermore, sample selection bias only occurs if sampling is based on a variable endogenous to the model under consideration (cf. CAMERON and TRIVEDI, 2005: 42-43). In this study, I sampled on farm types, which is exogenous to the production function model.

The sample of countries is selected to reflect the diverse farm sizes and structures in EU agriculture (Table 3.1). The range is from small-scale family farms in Italy, Spain, and West Germany to medium-sized commercial farms in Denmark, France, and the UK to large-scale and mostly corporate farms in East Germany (EUROPEAN COMMISSION, 2012). It is the heterogeneity of these structures across countries that makes comparisons with regard to productivity particularly insightful. As East and West Germany are structurally so distinct (MATHIJS and SWINNEN, 2001), I treat them separately throughout the analysis. East Germany contains the five federal states of Mecklenburg-West Pomerania, Brandenburg, Saxony-Anhalt, Thuringia, and Saxony. West Germany contains all of the other states except Berlin and Bremen, which are not represented in the FADN data. Therefore, throughout the empirical chapters, I can choose from the following sample of countries:

- Denmark (DK),
- France (FR),
- Germany East (DEE),
- Germany West (DEW),
- Italy (IT),
- Poland (POL),
- Slovakia (SVK),
- Spain (ES), and the
- United Kingdom (UK).

---

16 SWINNEN and KNOPS (2013) is an edited volume which consists of empirical studies utilising FADN data, among others.

The raw data provided by FADN was arranged in a way that panel data estimators can be applied. For every country (region in case of Germany) in the study, I created a panel data set covering the years from 2001 up to 2008. Essentially, the data was available in yearly cross sections (i.e., one data file per year and country). To arrive at the final data file for every country, I appended the individual cross sections and sorted the data by farm identification number (variable A3) and year. A small number of duplicates in the data were dropped.[17] The panels for Poland and Slovakia cover only five years as FADN data collection for these countries started only in 2004.

Table 3.1: Farm structures in the selected countries.

Country	Farm structures	Degree of farm commercialisation
Denmark	Medium-scale farms	High
France	Medium-scale farms	Medium
Germany	Small- to medium-scale farms (West)	Medium (West)
	Large corporate farms (East)	High (East)
Italy	Small-scale family farms	Medium
Poland	Small-scale family farms	Low
Slovakia	Large corporate farms	High
Spain	Small- to medium-scale farms	Medium
UK	Medium-scale farms	High

Source: Author compilation based on EUROPEAN COMMISSION (2010b), FADN data.

All monetary values were deflated to real values in 2005 prices using respective price indices. Price indices were extracted from the Eurostat online database and merged with the country panels. Output was deflated by the agricultural output price index. Fixed capital was deflated by the agricultural input price index for goods and services contributing to agricultural investment, and materials by the agricultural input price index for goods and services currently consumed in agriculture. Revenue shares were all calculated in nominal terms. Detailed variable definitions and descriptive statistics for the data employed are given in the upcoming empirical chapters.

17 Duplicates likely arise because farms run legally separate operations in several regions.

4 Assessment of factor productivities in EU agriculture

In this section, I present the empirical production function estimates for the estimators outlined and discussed in chapter 2.1. I produce separate results for Denmark, France, Germany (East and West), Italy, Spain, and the United Kingdom. In the following chapter 4.1, I briefly present the model specification before I move on to a discussion of the actual data set employed for estimation purposes. In chapter 4.2, I proceed with an extensive exposition of the results.

4.1 Model specification and data

To estimate a production function given by (2-1), I need to specify a functional form. Initially, I assume a Cobb-Douglas functional form that is successively extended to a translog specification by adding interaction terms of the different inputs with each other as well as itself to the model equation.

The variables and their measurement are readily available in the codebooks provided by FADN (European Commission, 2008, 2011). Output is measured as the total farm output in euros. Labour is measured by the time worked in hours by total labour input on the farm, including both hired and family labour. The total utilised agricultural area is my land input in ha. It includes owned and rented land, and land in sharecropping.

A persistent issue in estimating production functions has been the specification of the capital variable. Typically, some simple measures of input quantities (such as fertilisers or pesticides) and machinery use (such as fuel expenses or tractor hours) are used in cross-sectional studies. In this study, the materials or working capital input is proxied by total intermediate consumption in euros. It consists of total specific costs, including costs for seeds and seedlings, crop protection and other crop specific costs, as well as overheads arising from production in the accounting year. In particular, they consist of costs for fuel and electricity. I do not include the costs for fertiliser in the materials input. Land and fertiliser are highly correlated. This observation implies that these two inputs are applied in more or less fixed ratios on the majority of farms, which, in return, might induce a multicollinearity problem in the estimations. I discuss and elaborate on this perceived problem below. To this end, I show that multicollinearity is an issue and demonstrate the effects of such problems on estimation by referring to a previous specification of Petrick and Kloss (2013a) that includes fertiliser.

Nevertheless, even though fertiliser inputs are not included, the effect of this input is captured by the land input.

Furthermore, in order to be a suitable proxy, materials should be increasing in unobserved productivity at least for a subset of the data (OLLEY and PAKES, 1996: 1265). This is generally the case for my data sample. I show this for every country of the sample in the appendix of this chapter (Figure A1 – Figure A7). All sample countries display extended periods of increasing materials use (e.g., Denmark from 2001 to 2005 or France from 2003 to 2007). For Italy (from 2001 to 2003) and the United Kingdom (from 2001 to 2004) these periods are short. However, for Italy, the sample size to calculate these figures is rather low as there are many missing values in the individual components of the materials input used for this illustration.[18] The monotonicity condition is particularly well met in Spain and Germany (East and West). Consistent with most of the recent literature on production function estimation with firm level data (such as OLLEY and PAKES, 1996; BLUNDELL and BOND, 2000; LEVINSOHN and PETRIN, 2003), I approximate fixed capital inputs by using the opening valuation of assets. In this case, I took the asset value of machinery and buildings from the FADN data.

To calculate revenue shares, I needed factor prices for labour, land, and capital. These were taken from the actually paid wage to hired farm workers, the actually paid rent per hectare of rented land, and the actually paid interest per debt capital. Because there were many missing values, I calculated median factor prices per FADN region (variable A1) and imputed these to all farms in that region. There is zero variation (in terms of standard deviation in the factor prices because Denmark is an FADN region on its own (Table A1). Table 4.1 summarises the variable definitions and gives the actual FADN codes.

Outliers were identified on the basis of the fixed capital productivity per farm (real SE131/(real (L.SE450 + L.SE455))). Observations were dropped for the production function estimation if their value was beyond the upper or lower quartile ± 1.5 times the interquartile range (IQR).[19] Furthermore, as the novel control function and dynamic panel data identification assume dynamic factor adjustment from one period to the other, I only included farms which had some minimum panel representation in the data. Farms had to be present in the data for at

18 In order to compass this issue in the estimation stage, I take the "full" materials input (SE275) and subtract the cost for fertiliser inputs (SE295).

19 A detailed explanation and motivation of this rule is given in chapters 2 and 5, respectively.

Table 4.1: Definition of variables.

FADN code	Variable description
Outputs	
SE131	Total output (EUR)
Inputs	
SE011	Labour input (hours)
SE025	Total utilised agricultural area (ha) = land
F72 + SE300 + SE305 + SE336	Costs for seed and seedlings + crop protection + other crop-specific costs + overheads (EUR) = materials
L.SE450 + L.SE455	Opening valuation of machinery and buildings (EUR) = fixed capital
Factor prices	
SE370/SE021	Wage per hour (EUR)
SE375/SE030	Land rent per ha (EUR)
SE380/SE485*100	Interest on capital (%)

Note: L. denotes the one-year lag.
Source: Author, FADN data.

least four consecutive years. In total, 23,942 observations were included in the EU-wide sample. Descriptive statistics including the data patterns of the panels are given in the appendix (Table A1). The numbers on output and land generally confirm the picture conveyed about farm size and agricultural structures in the considered member states. In particular, the figures for East Germany are extremely large compared to the other countries. Hence, here the legacy of state collective farms still prevails.

4.2 Results[20]

4.2.1 Overview

For this study, I estimated nine models per country: Output shares, OLS Cobb-Douglas, OLS translog, within Cobb-Douglas, within translog, LP Cobb-Douglas, WLP Cobb-Douglas, WLP translog, and BB Cobb-Douglas. The within translog was obtained by interacting the groupwise demeaned logs of factors and using an appropriate degree of freedom correction in order to get cluster robust standard errors. Other than by simply calling a build-in fixed effects panel estimation command with the interacted variables in logs, this procedure ensures that levels are effectively eliminated from the regression.[21]

Table 4.2 displays a summary evaluation of the estimators with regard to the estimated production elasticities and returns to scale. The performance of the translog specifications and the dynamic panel data model is given particular attention. Generally, the objective was to detect systematic differences across estimators and countries, and to assess their practical implementation. Detailed results tables are presented in appendix A, which includes an overview table for each country containing the results for the eight models, plus an additional table for each country including more in-depth diagnostic results for the BB model.

All estimations were performed with Stata 12. For the LP estimator, I used the user-written routine `levpet` (Petrin et al., 2004). To implement the WLP estimator the `ivreg2` routine by Baum et al. (2007b) was utilised as demonstrated in Petrin and Levinsohn (2012). This procedure includes lags of inputs up to the second order. Therefore, the panel length is reduced by two years. The BB estimator was implemented with `xtabond2` by Roodman (2009) using the `h(2)` option, and combined with Söderbom's (2009) `md _ ar1` minimum distance estimator. For the Cobb-Douglas within regression, I used the build-in `xtreg` command. To accommodate the necessary changes for the Translog within regression, I applied a hand coded solution as outlined previously. The implementation of theses estimators is outlined in appendix D. To maintain a maximum

---

20 This section is based in part on previous research of Petrick and Kloss (2013a, d). It has been extended in various directions; among others it includes insights concerning the multicollinearity of inputs, the mitigation of this problem, extended empirical applications of the control function estimation approach, and a more elaborate view on shadow prices in EU crop farming.

21 Note, the build-in Stata routine would first interact and then demean the resulting interaction.

Table 4.2: Summary evaluation of estimator performance.

	DK	FR	DEE	DEW	IT	ES	UK
Factor elasticities	All OLS below shares; Materials below shares throughout CD, insignificant in LP and WLP (=0); Capital=0 in Within	Land & labour =0 in BB; Materials above shares in OLS, LP, WLP, BB; Capital<0.1 in shares, Within, BB	Labour=0 throughout; Land=0 in OLS, LP, WLP; Materials>0.8 in OLS, LP, WLP (>1); Capital≈0 in Within, LP, BB	Materials above shares in OLS, LP WLP, BB, lower in Within; Capital≈0 in Within&WLP, higher in LP&BB	Land=0 in OLS, LP, WLP, <0 in BB; Materials above shares throughout CD; Capital<0.1 in OLS, >0.1 BB, =0 in all other CD	Materials above shares throughout CD; Capital<0.1 in WLP, =0 in all other CD	Materials above shares in OLS, LP, BB; Capital≤0.1 throughout
Returns to scale	Shares add up to 2.07; OLS, LP, lower but still >1; Within, WLP, BB close to 1	OLS >1; Close to 1.0 for the other estimators	Close to 1.0 in OLS, Within, LP, WLP; ≈0.8 in BB	1.1 in OLS, <1 in Within, BB; Close to 1.0 in LP, WLP	Shares add up to 1.61; OLS ≈1.1; Within, LP, WLP <0.9; BB=0.5	Shares add up to 1.42; Within, LP close to 1; BB <1, OLS ≈1.15; WLP larger	OLS, Within, LP≈1.2; WLP≈1.1; BB≈1.5
Performance of Translog	OLS unreasonable; Within close to CD; WLP unreasonable; Interactions not sig. in Within & WLP	OLS unreasonable; Within close to CD; WLP unreasonable; Interactions not sig. in WLP	OLS unreasonable; Within in part close to CD; WLP unreasonable; Interactions not sig in WLP	OLS unreasonable; Within close to CD; WLP unreasonable; Interactions not sig. in Within & WLP	OLS unreasonable; Within in part close to CD; WLP unreasonable; Interactions not sig. in WLP	OLS unreasonable; Within close to CD; WLP unreasonable; Interactions not sig. throughout	OLS unreasonable; Within close to CD; WLP unreasonable; Interactions not sig. in OLS & WLP
Blundell/ Bond estimator	Specification tests ok; levels better instrumented than diff.; relatively poor instrumentation	OID not passed; Land, mat, capital, output highly persistent; levels better instrumented than diff.	Specification tests ok; Labour, land, mat highly persistent; Output explosive; levels better instrumented than diff.	OID not passed; Labour, land, mat highly persistent; Capital, output explosive; levels better instrumented than diff.	OID not passed; Labour, capital highly persistent; Land & output explosive; poor instrumentation	Specification tests not passed; Labour, land, materials, capital highly persistent, levels better instrumented than diff.	Specification tests ok; Labour, land, capital highly persistent; Materials & output explosive; poor instrumentation

Notes: BB: Blundell/Bond, CD: Cobb-Douglas, LP: Levinsohn/Petrin, OID: over-identification test, OLS: ordinary least squares, WLP: Wooldridge/Levinsohn/Petrin.

Source: Author based on Petrick and Kloss (2013a).

of comparability and homogeneity of the estimation samples as well as utilising the highest amount of data as possible for estimation, I proceeded as follows. Estimations for all estimators, except for the WLP estimator, are based on the BB estimation sample. Since this estimator implies a dynamic specification with first order lags of inputs and the dependent variable, the effective panel length is reduced by one year. I did not impose this restriction for the WLP procedure, which includes lags of inputs up to the second order. Hence, the difference between the WLP estimation sample and the sample employed for all other estimators is one round of observations as depicted by the tables in the appendix.

As a general tendency, factor elasticities were found to be low for land and capital, high for materials, and somewhere in between for labour (Table 4.2 and Table 4.3). Hence, materials are the key drivers of crop productivity. Estimates for the first two of these factors are in the range of 0.2 and lower, sometimes not significantly different from zero. The production elasticity of materials is typically between 0.7 and 1.0. Estimates for the labour elasticities usually fluctuate at around 0.2.

The estimates support the conventional wisdom that OLS tends to be upward biased for particularly variable factors. In the present data, this primarily applies to materials, the OLS estimate of which is (except for Denmark) higher than its revenue share. It may be taken as evidence for the existence of serially correlated, unobservable factors (Olley and Pakes 1996: 1274). The opposite downward bias is found for capital in the within estimator, which is typically below the revenue share. This tendency is also in line with previous studies and can be attributed to the low variance of capital over time (Griliches and Hausman, 1986).

The LP estimator commonly produces a lower elasticity for materials than OLS, the only exception being the United Kingdom. The WLP estimator excepts France, East Germany, and Spain. LP and WLP estimates are typically very similar, which makes me confident of the proxy variable identification strategy. These models may thus be taken as plausible alternatives to the received estimators. However, the WLP estimator has a few more advantages, giving this estimator an edge over the LP model. First, the former is occasionally more successful in identifying the capital coefficient (East Germany and Spain) or estimates said coefficient with a lower standard error, that is, higher precision (West Germany, Italy and the United Kingdom). Second, by applying the WLP estimator it is possible to rely on analytic standard errors rather than on bootstrapped errors.[22] Such standard errors are obtained by repeated sampling from the data (Greene, 2011: 651-655). Finally, on theoretical grounds, the WLP model further corrects for collinearity, a trait that is not available in the LP model (see section 2.1.6).

Estimated elasticities of scale fluctuate around 1.0, with a higher value observed only for Spain. In this case, the hypothesis of constant returns to scale is rejected at the 5 per cent significance level based on WLP results. Given the previous findings on production elasticities, OLS returns to scale estimates tend to be higher than within estimates. Overall, the scale elasticity in European crop farming appears to be close to one. Therefore, there is evidence that this type of farming may be characterised by constant returns to scale.

I report the production elasticities estimated by the WLP procedure for all subsamples in Table 4.3 and compare them with two rather distinct agricultural benchmark studies. Heady and Dillon (1961) is an early collection of OLS Cobb-Douglas production function estimates. It is based on farm-level data from 32 countries around the world, with a focus on North America, Australia, and India, and represents one of the most comprehensive collections of production elasticity estimates ever published. Table 4.3 simply reports the overall mean elasticities of all 32 studies. It should be noted that these studies display considerable variation among themselves (see the extensive discussion in Heady and Dillon (1961: 585-643)). Mundlak et al. (2012) is a recent cross-country regression of a Cobb-Douglas production function based on the within estimator. The authors use data from 30 developing and developed countries from 1972 to 2000. Against these benchmarks from the literature, Table 4.3 illustrates a number of interesting tendencies:

- A comparatively low production elasticity of labour prevails throughout the EU samples and was also found by Heady and Dillon as well as Mundlak et al. Denmark is an exception to that.

- The production elasticity of land is much lower in the EU than in the benchmark studies.

[22] Note, this reliance is possible because the WLP estimator is implemented by utilising the well-established instrumental variable estimation framework. See, for example, Cameron and Trivedi (2005: 95-103).

- The production elasticity of materials is much higher in the EU than in the benchmark studies.

- The MUNDLAK et al. study reveals remarkably low elasticities for labour and materials. Despite the use of the within approach, the capital elasticity is surprisingly high. The low materials coefficient can be explained by the fact that the dependent variable in their model is value added.

Table 4.3: Agricultural production elasticities in comparison.

	DK	FR	DEE	DEW	IT	ES	UK	Heady Dillon (1961)	Mund-lak et al. (2012)
Labour	0.62	0.17	0.04#	0.22	0.32	0.42	0.19	0.21	0.01#
Land	0.23	0.04	-0.03#	-0.01#	-0.01#	0.08	0.17	0.38	0.44
Materials	0.00#	0.80	1.08	0.77	0.51	0.73	0.62#	0.39	0.10
Capital	0.10#	0.12	0.08	0.09	0.02#	0.08	0.10#	--	0.46
Ret. to Scale	0.95	1.13	1.17	1.08	0.84	1.30	1.09	0.98	1.00*

Notes: Results for field crop farms in EU countries based on WOOLDRIDGE/LEVINSOHN/PETRIN (WLP) estimator. HEADY AND DILLON (1961) represents mean elasticities from a sample of 32 cross-sectional Cobb Douglas estimates originating from various countries (their table 17.15). MUNDLAK et al. (2012) based on a cross country regression of 30 countries for 1972-2000, using value added as dependent variable and the within estimator (their table 2, first column). * imposed on model. # not significantly different from zero at conventional confidence levels.

Source: Author.

4.2.2 The roles of materials and land in EU field crop farms

Throughout this chapter, I use a materials definition that does not include fertiliser inputs. While it would certainly be desirable to have a variable of total intermediate consumption that includes this input, such a specification occasionally produces inconsistent results from a theoretical point of view. To elaborate on this, I refer to the previous work of PETRICK and KLOSS (2013a), who utilise a materials input including fertiliser. As I employ the same data as these authors, direct comparisons are possible and particularly meaningful. Furthermore, I am able to perform additional diagnostics. The results of these analyses are summarised in Table 4.4.

Several EU countries in the PETRICK and KLOSS (2013a: 17) study display negative and statistically significant estimates for the land output elasticity in conjunction with relatively high materials coefficients, particularly in France, Germany (East and West) and Italy;the results presented in Table 4.3 and the appendix do not suffer from that problem. In addition, partial correlations between these two inputs are the highest among all partial correlations for all countries of my EU sample (Table 4.4). Variance inflation factors (VIF) suggest that there is a negligible degree of multicollinearity among the land and materials input in Spain; a slight degree of multicollinearity in France, West Germany, Italy, and the United Kingdom; as well as considerable multicollinearity among these inputs in Denmark and East Germany.[23] This multicollinearity problem has differing effects on the estimates, but France, Italy, and Germany display statistically significantly negative land coefficients. From a statistical point of view, it seems that much variation of the land input is captured by the materials input. The parameter estimates for the materials input should be more or less stable after dropping the land input from the production function in a situation characterised by multicollinearity; remember both inputs are applied in an almost fixed ratio.

To further investigate this issue, I re-estimated the production functions without the land input. Indeed, materials estimates are similar in all cases. Absolute differences in materials estimates between the specification with and without the land input fluctuate around 0.08 among countries; the only exception is East Germany with a difference of 0.29 (Table 4.4). The negative land coefficient thus appears to be an artefact of multicollinearity among materials and land.

[23] There are many rules of thumb in the analysis of multicollinearity. Here, I resort to the rule that the largest VIF is larger than 10 and that the average VIF is bigger than 1 (Chatterjee and Hadi, 2006: 236-238).

Table 4.4: Evaluation of Multicollinearity in production function estimates.

	DK	FR	DEE	DEW	IT	ES	UK
Is the point estimate in Petrick and Kloss (2013a) for the land input negative and statistically significant?	No	Yes	Yes	Yes	Yes	No	No
Correlation coefficient between land and full materials input	0.91	0.71	0.92	0.68	0.76	0.56	0.86
Is this the highest partial correlation?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Production factor with highest VIF	Materials (10.61)	Materials (3.81)	Materials (10.97)	Materials (3.19)	Materials (3.89)	Materials (1.89)	Materials (5.53)
Average VIF	7.95	2.42	6.70	2.15	2.62	1.50	3.81
Materials coefficient before and after dropping land (#)	0.48* 0.52*	0.93*** 0.83***	1.40*** 1.11***	0.85*** 0.78***	0.54*** 0.48***	0.76*** 0.79***	0.74* 0.92**

Notes: *** (**, *) significant at the 1% (5%, 10%) level. Materials definition including fertiliser inputs as in Petrick and Kloss (2013a). VIF: variance inflation factor. (#) First line: materials output elasticity for the Cobb-Douglas production function as a reference. Second line: re-estimated model without the land input. Parameters based on WLP estimator.

Source: Author.

A straightforward interpretation of this finding is that EU field crop farms typically employ closely intertwined packages of land and materials in their production process. Particularly in France, Germany and Italy, the intensity of these inputs is relatively homogeneous across farms, which makes the statistical isolation of separate land and materials effects on output difficult. The analysis revealed that fertiliser is the main driver of the multicollinearity problem. Therefore, it is a feasible solution to exclude this input from the materials specification to mitigate multicollinearity of land and materials. At the same time, the high correlation between these two inputs implies that one can control for the effect of fertiliser by including the land input in the production function. Therefore, my results should not be affected by any omitted variable bias.[24]

4.2.3 Functional form: Cobb Douglas vs. translog

The results on the translog specification display remarkably uniform features across countries. The within translog elasticities were at sample means typically close to the within Cobb Douglas, and the interaction terms of the translog were often not jointly different from zero. The OLS translog, on the other hand, produced unreasonable results throughout, for instance, reflected in the coexistence of negative production elasticities for some factors and elasticities bigger than one for others (at sample means). Similarly unreasonable results are observed for the WLP translog. In this model, not a single country displayed interaction terms that were jointly significantly different from zero. Additionally, I applied the Kleibergen and Paap (2006) under-identification test to the WLP translog model. This produces a lagrange multiplier test statistic which tests whether the model equation is identified (i.e., the excluded instruments are correlated with the endogenous regressors).[25] It can be interpreted as a test whether the lags of land, labour, and materials are suitable instruments for their contemporaneous counterparts. Failing to reject the null hypothesis that the equation is unidentified implies an increased bias in the estimated coefficients. They are then biased in the same direction as the OLS estimator (Baum et al., 2007a). While the null hypothesis was always rejected at the 5 per cent significance level in

24 Omitted variable bias is the result of omitting a relevant regressor. Such an omission will result in the error term picking up the effect of the omitted variable, which, in return, will violate the IID assumption of the same.

25 “Excluded” means that these instruments are not part of the model equation. Note, tests for over-identification restrictions could not be performed because the WLP model is just identified, meaning that the number of instruments is equal to the number of endogenous regressors (cf. Cameron and Trivedi, 2005: 100).

the Cobb-Douglas model, that was not the case with Denmark, East Germany, and the United Kingdom in the translog model (Table 4.5).

Table 4.5: Results for the Kleinbergen/Paap under-identification test.

Country	Cobb-Douglas	Translog
Denmark	<0.001	0.410
France	<0.001	0.012
East Germany	<0.001	0.105
West Germany	<0.001	<0.001
Italy	<0.001	<0.001
Spain	<0.001	<0.001
United Kingdom	<0.001	0.624

Notes: P-values for the Kleibergen and Paap (2006) under-identification test.
Source: Author.

To summarise, the translog specification does not perform well. The findings are in line with other recent studies utilising FADN data with this functional form (cf. Zhengfei et al., 2006; Latruffe and Nauges, 2013). The prime reason for these difficulties might again be multicollinearity, which supposedly inflames again in the translog model, as many more parameters that include different variants of inputs must be estimated. While I cannot ultimately decide whether the true data generation process followed a translog technology, I can say that farm-level data typically does not allow estimating its parameters. This makes the translog a less credible functional form for applied work.

4.2.4 Dynamic panel data estimation

I examined the performance of the BB estimator in some detail. I present results for the unrestricted and the restricted model along with Arellano-Bond tests for serial correlation of error terms. If the model is correctly specified, the test should reject autocorrelation of order one but not of order two (Arellano and Bond, 1991). I also apply Hansen's over-identification (OID) test for instrument validity (Hansen, 1982). While serial correlation of the error terms for the models was only a problem in Spain and the common factor restriction was never rejected, the Hansen OID test of instrument validity was not passed in four instances – France, West Germany, Italy, and Spain.[26]

To allow further diagnosis, simple autoregressive models of order one (AR(1)) were estimated separately for all factors and output, following Blundell and Bond (2000). Labour and land were found to be highly persistent, which makes dynamic panel data estimation a natural option. Moreover, I regressed the differences of the latest available year on the lagged levels of all available previous years and the latest available levels on all available lagged differences of previous years. The reported p-values and coefficients of determination allow an insight into the explanatory power of the instrument sets. Generally, the instrument performance was better for levels (instrumented by differences) than for differences (instrumented by levels). System GMM approaches which do not only use differences but also levels for instrumentation (Blundell and Bond, 1998) are thus warranted. Even so, the elasticities of the persistent factors labour, land, and capital could often not be identified. For Spain and Italy, even the estimated land coefficient is statistically significantly different from zero – probably a direct cause of the major specification tests failing.[27] Parameters were very sensitive to the composition of the sample and the precise specification of the estimator. Occasionally, dynamic factor evolution apparently followed an explosive process, as the AR(1) coefficient was estimated to be bigger than one. On the other hand, the estimates for materials appear very reasonable throughout, as they were typically somewhere between the OLS and within results. It is here that the BB estimator can likely claim some superiority.

There are some noteworthy findings for Denmark compared to the other countries and estimators. Here the materials elasticity was lower than the materials' revenue share. Shares add up to the extremely high value of 2.07 (which is actually inconsistent with the interpretation as shares). This outcome may be an artifact of systematically higher imputed factor prices than in other countries. The unbalanced panel pattern of Denmark made it difficult to perform the diagnostic regressions on the explanatory power of the lagged instruments in the BB approach. Admittedly, the capital coefficient in such regressions is relatively low (0.52), a result that is only undercut by the figure for East Germany. Even

[26] That is assuming a 5 per cent significance level.

[27] For both countries, the Hansen OID test fails, while for Spain the Arellano-Bond test for second order serial correlation also gives rise to concern. The latter result might also explain the negative estimate for ρ and hints at some serious problems of the dynamic panel data estimation.

so, compared to the control function estimation approaches the BB estimator is able to identify a materials output elasticity. The reason LP and WLP estimators leaving the materials elasticity unidentified might be explained by the non-parametric control function utilised in these estimators. In this function higher order and interaction terms of materials enter so that the same captures much of the explaining variance. Hence, there is not enough variation left for the sole materials input. Furthermore, there is also an extreme materials coefficient decrease in size from LP to WLP which might be explained by the lower sample size of the latter and consequently reinforces the problem. Given that specification tests do not fail and the materials coefficient is close to the within regression, the BB coefficient for this parameter is a more plausible candidate.

4.2.5 Analysis of shadow prices of production factors

4.2.5.1 Theoretical considerations

A simple theoretical model of farm production that is subject to a generalised input constraint can usefully illustrate the factor market perspective on agricultural productivity. It serves as a motivation for the upcoming empirical analysis. Assume a farmer maximises profit by producing one output with one input. Profit is then defined as revenue minus the costs of the input. Following Petrick (2003: 174-175), one can write:

$$max_x\, \pi = p^y f(x) - p, \text{ subject to} \tag{4-1}$$

$$\bar{x} - x \geq 0, \tag{4-2}$$

where π is profit, p^y is the output price, f the production function, x input use, p the input price observed in the market, and $\bar{x}$ the generalised input constraint. This input constraint captures the general observation that most agricultural production factors cannot be adjusted instantaneously but rather are subject to more or less pronounced adjustment costs. For example, land is often available in limited quantities only and subject to long-term rental agreements. Agricultural credit markets suffer from informational asymmetries and may be characterised by rationing and high transaction costs (see, e.g., Benjamin and Phimister, 2002; Petrick and Latruffe, 2006; Curtiss, 2012).

I assume that f is monotonically increasing and concave in x. Solving the optimisation problem given by (4-1) and (4-2) through the Lagrangean L yields $L = p^y f(x) - px + \lambda(\bar{x} - x)$, where λ is the Lagrange multiplier. Assuming that (4-2) is binding, I obtain the first-order condition $\frac{\partial L}{\partial x} = p^y \frac{\partial f}{\partial x} - (p + \lambda) = 0$. Rearranging leads to:

$$p^y \frac{\partial f}{\partial x} = p^* > p, \text{ with } p^* \equiv p + \lambda. \tag{4-3}$$

I define p^* as the shadow price of the on-farm production factor. It represents the willingness to pay for this input.[28] With a more severe input constraint, the decision price for input use is increasing and use of that factor is reduced. For instance, farmers may face a credit constraint if they are unable to provide sufficient collateral which would help to mitigate the above mentioned informational asymmetries.

The above model serves as a useful motivation for the empirical measurement of factor productivity and factor market imperfections in agriculture. The practical implementation involves the use of an estimate of the shadow price to study drivers and impacts of factor use. It requires a consistent estimate of the production function as well as reliable data on input use and factor prices. The empirical relation $p^* > p$ is a measure of on-farm input productivity and the severity of supply rationing.

By similar reasoning, a release constraint also could be modelled. $p^* < p$ would then be evidence of a resource over-utilisation. This may, for example, be due to non-pecuniary benefits of input use (tractors as prestige objects) or the wish to provide safeguards against production risk (use of insurance contracts, precautionary investment in powerful machinery to mitigate production peaks; Witzke, 1993: 157). Aurbacher et al. (2011) have recently shown that farmers trapped in small agricultural structures may be unable to coordinate on machinery sharing and thus may hold inefficiently high stocks of machinery. Furthermore, agriculture in Europe is typically organised in family farms on which labour is often highly immobile (Tocco et al., 2012). Furthermore, on-farm labour may also be influenced significantly by life cycle considerations of the farm family (Glauben et al., 2009).

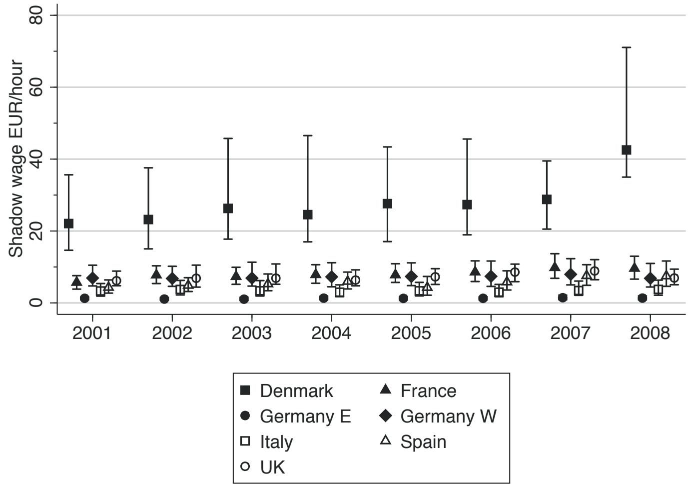
4.2.5.2 Empirical results

Based on the theoretical results from the preceding section, I arrive at estimates of the farm-individual shadow prices ($\hat{p}^*$) for the different inputs by multiplying the production elasticities obtained from the Wooldridge/Levinsohn/Petrin

---

28 Note, λ might be interpreted as the "shadow component" in p^*.

Figure 4.1: Distribution of shadow wages per country and year.



Notes: Results based on Wooldridge/Levinsohn/Petrin estimator. Dots (squares, triangles, circles, diamonds) denote median. The bars represent the lower and upper quartiles, respectively.

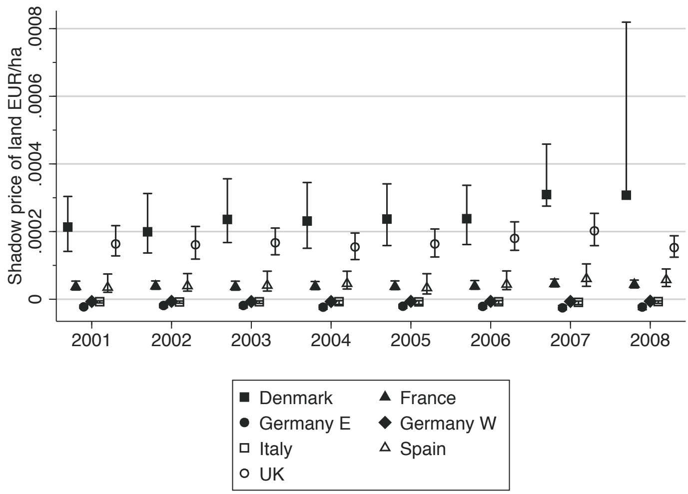
Source: Author based on PETRICK and KLOSS (2013d).

estimator with the farm-specific average factor productivities.[29] Following CARTER and WIEBE (1990), this gives the respective shadow prices holding all other production factors constant for profit maximizing farmers. Net returns equal to the marginal value product minus one were calculated for the materials and capital variable, so that they can be compared with common market interest

[29] The choice of estimator for calculating shadow prices is important as biases in estimates of output elasticities carry over to the calculation of shadow prices. See PETRICK and KLOSS (2013d: 330-331) for an example of how a biased estimate leads to wrong conclusions regarding constraint access to capital.

rates for credit.[30] The distribution of the shadow prices for the four input factors is illustrated in Figures 4.1-4.5 by using plots displaying the median and first and third quartiles of the distribution. The plots depict the evolution of these marginal returns to input use for every country in the sample by year to allow for cross country as well as dynamic comparisons.

Figure 4.2: Distribution of shadow land rents per country and year.

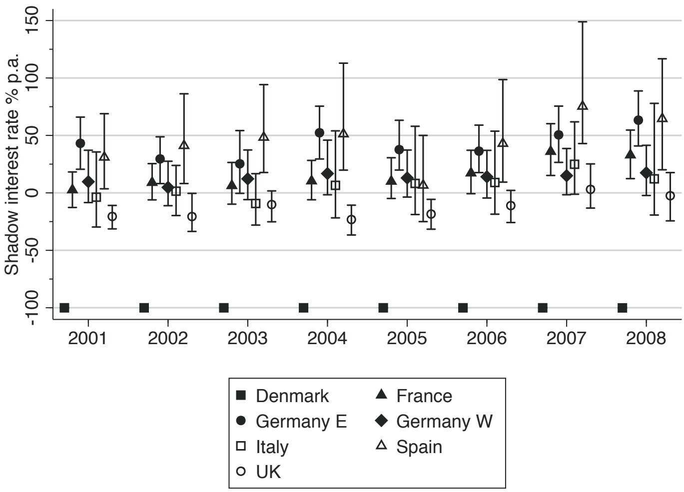


Notes: Results based on Wooldridge/Levinsohn/Petrin estimator. Dots (squares, triangles, circles, diamonds) denote median. The bars represent the lower and upper quartiles, respectively.

Source: Author based on PETRICK and KLOSS (2013d).

30 This treatment is motivated by a number of studies, including SIAL and CARTER (1996: 783) and PETRICK (2004: 81-83). In the theoretical model presented in 4.2.5.1, one can account for it by assuming that the output is produced by a single credit financed input. This assumption is plausible in light of the sequential nature of agricultural production due to which some materials inputs must be financed upfront. Again, profit is defined as revenue minus costs. Hence, the costs are now the repayment of credit and the interest. Then in (4-1), I write $p = 1 + r$, where r is the market interest rate. The solution (4-3) is: $p^y \frac{\partial f}{\partial x} = 1 + r^* > 1 + r$, with $r^* \equiv r + \lambda$. r^*, the shadow interest rate, is then $r^* = p^y \frac{\partial f}{\partial x} - 1$, the marginal value product minus 1.

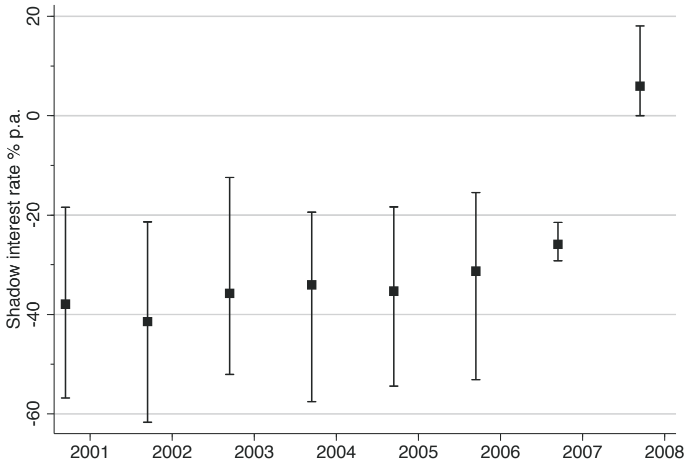
Figure 4.3: Distribution of shadow interest rates of materials per country and year.



Notes: Results based on Wooldridge/Levinsohn/Petrin estimator. Dots (squares, triangles, circles, diamonds) denote median. The bars represent the lower and upper quartiles, respectively.

Source: Author based on PETRICK and KLOSS (2013d).

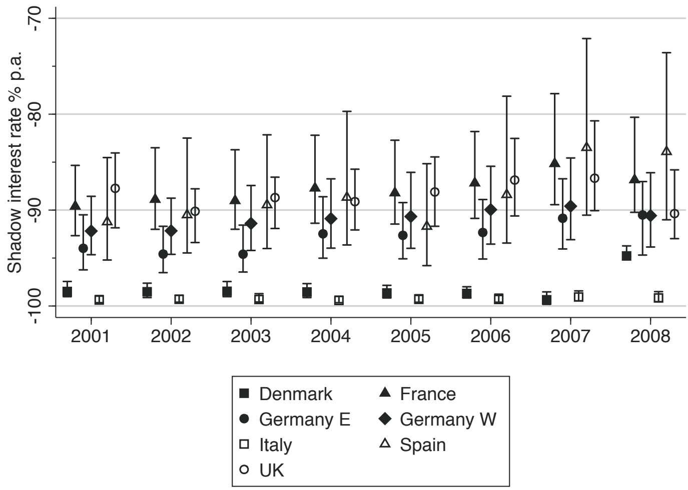
Figure 4.4: Distribution of shadow interest rates of materials for Denmark.



Notes: Results based on Blundell/Bond estimator. Dots denote median. The bars represent the lower and upper quartiles, respectively.

Source: Author.

Figure 4.5: Distribution of shadow interest rates of capital per country and year.



Notes: Results based on Wooldridge/Levinsohn/Petrin estimator. Dots (squares, triangles, circles, diamonds) denote median. The bars represent the lower and upper quartiles, respectively.

Source: Author based on PETRICK and KLOSS (2013d).

The findings from the plots are not too surprising given the results on output elasticities presented in the previous subchapters. The shadow prices of the factors labour, land, and fixed capital tend to be quite low. The median shadow wage in agriculture is below 9 EUR/hour in France, West Germany, Spain, and the UK throughout the years; in Italy it is below 5 EUR/hour for most of the period. East Germany displays particularly low values with numbers below 2 EUR/hour. Denmark stands out with a value fluctuating at around 30 EUR/hour. Shadow land rents are only minimally different from zero throughout. Shadow prices of fixed capital are negative in all subsamples, with medians per country and year in the range of -85 to -100 per cent. Furthermore, there is considerable variation for some of the country subsamples.

The distributional plots on the marginal return to materials deserve a closer look (Figure 4.3 and Figure 4.4). As materials use is variable on a short-term basis, it reacts quickly to fluctuations in the economic environment. In the observed study period, the financial crisis was epitomised by the emerging US subprime crisis in 2007 and the collapse of the investment bank Lehman Brothers in 2008. The shock waves of the crisis hurt the various EU member countries quite differently, and there is little analysis available so far regarding how they affected access to working capital in agriculture. Indeed, both the cross country as well as the dynamic variation reveal interesting patterns in this regard.[31] Across countries, Denmark and the United Kingdom are the only countries where the median farm exhibited negative marginal returns on working capital throughout most of the periods. This is consistent with an excess capital use and the absence of funding constraints, and possibly reflects the strong position of the Danish agricultural banking sector, which is based on a mortgage-banking model, during the crisis. A similarly strong banking sector is present in the UK. For both countries, I observe increasing shadow rates right before or during the crisis. However, Danish farms are typically much more highly leveraged than their European counterparts (Petrick and Kloss, 2013c). Danish farms were thus hit harder by the emerging financial crisis, consistent with a sharp increase in shadow rates in the year 2008 in Figure 4.4. At the other end of the spectrum, farms in Spain and East Germany

[31] Because the WLP estimator was not able to identify a materials output elasticity in the case of Denmark, the shadow price of working capital of this country is somewhat predetermined. Remember, a coefficient close to zero will also produce a marginal value product close to zero; subtracting one will then result in a shadow interest rate close to minus one (-100 per cent). Indeed, this is exactly what is observed in my calculations (Figure 4.3). Given that the BB estimator produces a plausible materials coefficient (see section 4.2.4), I interpret and compare the shadow interest rate based on this estimator for Denmark (Figure 4.4).

show high shadow rates on working capital, with an upward tendency over the observed period. Also many Italian farms are in the range above 50 per cent of interest. Spain and Italy are countries with very low leverage in the agricultural sector, but also with banks suffering from the crisis. Thus, farms may have been forced to reduce their use of working capital, particularly after the onset of the crisis. East German agriculture is dominated by corporate farms which are often based on rented land. Capital access is less easy to obtain for them than for West German family farms, and may have become more difficult during the crisis. This argument is further underlined by the very large difference in shadow rates between East and West Germany. In addition, with regard to West Germany, shadow rates remain more or less stable during crises, which possibly reflects the strong position of its agricultural banking sector. France is somewhere in the middle of the field.

Finally, with typical market interest rates for capital in mind, such as those reported in PETRICK and KLOSS (2013c), shadow prices of materials are notably higher than the market rate in many countries, especially in the two crisis years 2007 and 2008.[32] This finding supports the view that quantity rationing on the market for short-term capital was prevalent in these years. However, this view does not hold with regard to fixed capital because these numbers are negative throughout.

[32] Ideally one would compare the shadow prices with the marginal interest rate. However, these are not observed. Therefore, they have to be compared to the average interest rates, which the figures observed essentially are.

5 Outlier Robust productivity analysis: an application to German FADN data

In this chapter, I produce and discuss results for East and West Germany. Both regions are treated separately in the following because they are, as discussed earlier, so different in their agrarian structures. East German agriculture can be characterised by large-scale corporate farms, whereas West Germany is dominated by small- to medium-scale family farms. In addition, after the removal of outliers such a treatment allows carving out similarities and differences more precisely since both German regions are under the same jurisdiction while having historically different forms of agricultural organisation. In the following subchapter I discuss the model specification and data; chapter 5.2 presents and discusses the results. I also provide a synthesis outlining the general implications of my results in the latter.

5.1 Model specification and data

I proceed in two steps to perform an unbiased estimation. First, I decontaminate the sample from outliers by pruning the minimum spanning tree as outlined in chapter 2.2.3 (Kirschstein et al., 2013). After eliminating the outliers from the data, I proceed by estimating the production function using the Wooldridge (2009) production function estimator discussed in 2.1.6. These estimation results are then compared to a variety of alternative decontamination schemes. I recover the parameters of the following log-linearised production function specification:

$$y_{it} = \alpha^A a_{it} + \alpha^L l_{it} + \alpha^K k_{it} + \alpha^M m_{it} + \omega_{it} + \varepsilon_{it}, \tag{5.1}$$

where y is the natural logarithm of output Y, A is land use, L is labour, K fixed capital, M (working capital), and i and t are farm and time indices. The α^x are parameters to be estimated, and $X \in \{A,\ L,\ K,\ M\}$ refers to the production factors. The ω_{it} are farm- and time-specific factors known by the farmer but unobserved by the analyst, while ε_{it} are the remaining independent and identically distributed (IID) errors.

Again, I resort to this Cobb-Douglas specification because making it more flexible by adding second order terms – quadratics and interactions of the different inputs – did not yield any additional insights. Furthermore, as demonstrated in chapters 4.2.3 and 6.3, such a translog specification produced unreasonable results. Furthermore, virtually any data structure can be modeled with increasing degrees of powers in polynomials of inputs. This could possibly also mitigate the outlier problem but it also implies adding more and more regressors to the estimating equation. As a result, problems such as multicollinearity are potentially amplified.

I measure output as the total farm output in euros, labour as the total of on-farm hired and family labour working time, and land as the utilised agricultural area in ha, including owned and rented land as well as land in sharecropping. Following the discussion on the capital variable in chapter 4.1, the material or working capital input is proxied by total intermediate consumption in euros. It consists of the total specific costs and overheads arising from production in the accounting year. Among others, it includes costs for fuel, lubricants, water, electricity, and seed. Fixed capital inputs are approximated by depreciation of capital assets estimated at replacement value in euros. This input accounts for different depreciation rates of the various capital assets. It includes depreciation for plantations of permanent crops, buildings and equipment, land improvements, machinery, and forest plantations (European Commission, 2011). Table 5.1 summarises the variable definitions and gives the actual FADN codes.

Table 5.1: Selection of variables.

FADN code	Variable description
Outputs	
SE131	Total output (EUR)
Inputs	
SE011	Labour input (hours)
SE025	Total utilised agricultural area (ha)
F72 + SE300 + SE305 + SE336	Costs for seed and seedlings + crop protection + other crop-specific costs + overheads (EUR) = materials
SE360	Depreciation (EUR) = fixed capital

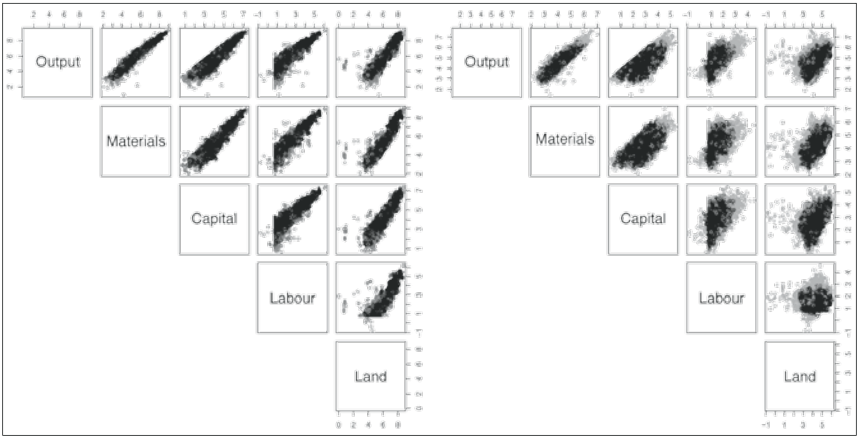
Source: Author.

5.2 RESULTS

5.2.1 Outlier identification

The data sets which were finally used for the outlier identification contain a total of 3,610 observations (in the case of the East German field crop data) and 8,490 observations (in the case of the West German field crop data). Data has been logarithmised to compensate for different scales and heavy tails. Figure 5.1 depicts the corresponding scatterplot matrices. Neither the plot for East nor West Germany gives rise to serious concern. In most cases variables show a high pairwise correlation (especially for East Germany). The major portion of the data forms a large cluster with some observations scattered around. An interesting effect can be observed for the labour input where the observations seem to be compressed for values somewhere between 0 and 1.[33] This effect results from a substantial amount of "one-man-companies" in the data sets which all unanimously stated 2,210 hours (i.e., 52 weeks · 5 days · 8.5 hours) as their working input.

Figure 5.1: Scatterplot matrix of field crop data.



Notes: Left panel: East Germany. Right panel: West Germany. Outliers coloured in grey

Source: Author.

[33] Note, this variable is expressed in thousand hours. Therefore, taking the natural logarithm results in values between 0 and 1.

When applying the pMST procedure with critical edge length probability $\alpha = 0{,}95$ (as described in chapter 2.2.4) on the data (i.e. output and inputs that enter the production function), 322 (East), and 1,121 (West) potential outliers are identified. Figure 5.1 shows the scatterplot matrices with the identified outliers coloured in grey.

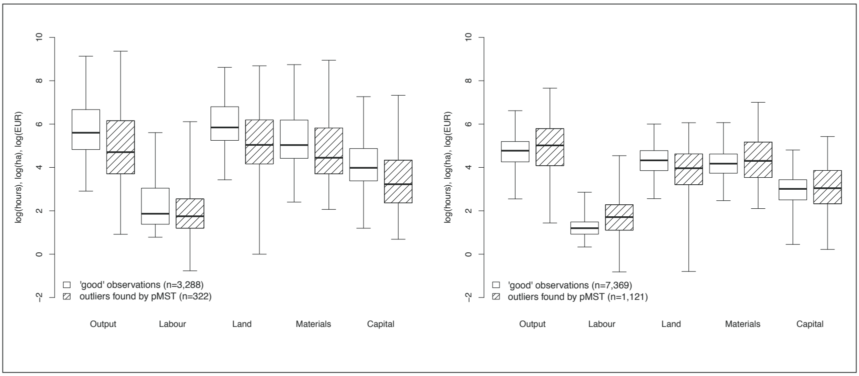
The following results are revealed:

- All scattered observations (lying around the main bulk of the data) are identified as outliers (as expected).

- It seems that observations lying *within* the main bulk are also identified as outliers, which – at a first glance – might be understood as an undesired result. However, this view leaves the multivariate nature of the data out of consideration (i.e., an observation which is outlying in one scatterplot might be very well an outlier in another bivariate plot or higher dimensional displays). By tendency, outliers seem to be made up of small farms (with low levels of inputs and output) because they mainly occur in the lower left corner of the scatterplots (this is especially visible for East Germany in Figure 5.1).

The latter point also can be confirmed by having a closer look at the outlier characteristics by means of parallel boxplots; see Figure 5.2.

Figure 5.2: Characteristics of outliers and non-outliers.



Notes: Left panel: East Germany. Right panel: West Germany.

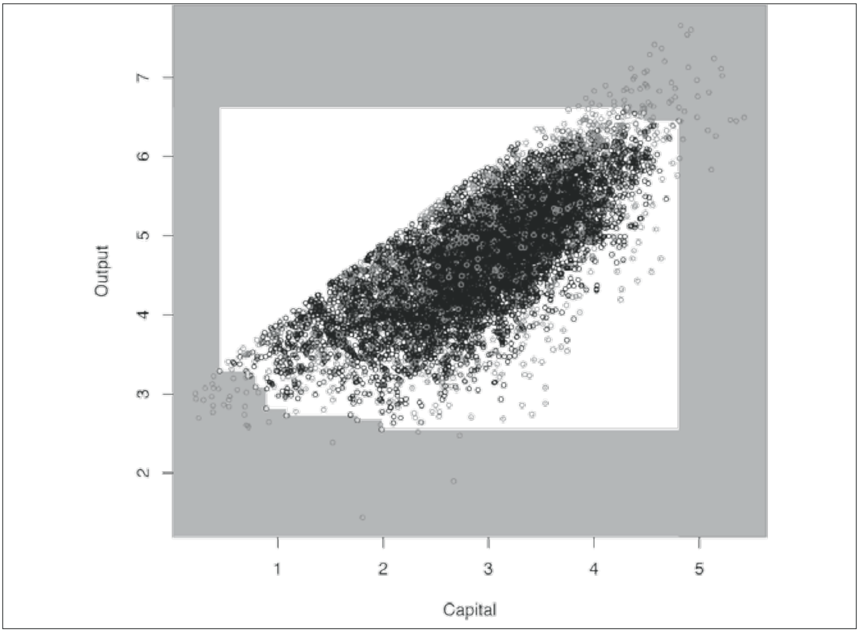
Source: Author.

Figure 5.2 (left panel) shows that the outliers are primarily made up of small farms (presumably the one-man-operations mentioned earlier) because the boxes and the medians are clearly situated below those of the non-outliers.[34] Additionally, the outliers also cover those observations with very large values in each variable as the (upper) whiskers reach out beyond those of the non-outliers. The situation is not as clear for the West German data (Figure 5.2, right panel). While the outliers are still made up of the largest and smallest observations in each variable (as expected), those identified seem to be more evenly distributed.

The approach described in Liebscher and Kirschstein (2012) can be applied to gain a better understanding of how the outliers divide into small and large farms: Based on the identified *non*-outliers and their values in the five variables, a frontier/boundary (in concept similar to the Free Disposal Hull (FDH), see, e.g., Cooper et al., 2007) can be constructed. This involves the formulation of a linear program which is subsequently solved. Those observations among the *non*-outliers which dominate the remaining non-outliers (i.e., in the sense that they possess a higher value in at least one variable while being at least equal in the remaining variables) constitute the upper boundary (see Figure 5.3 for an example when considering only the variables capital and output of the West German data set). Likewise, a lower boundary is constructed by switching the sign of the variables. The identified *outliers* can now be assigned to one of the two groups (large/small) by examining their position in relation to these boundaries. Outliers lying beyond the upper boundary are considered as large companies, whereas observations lying below the lower boundary are considered as "small" companies. There might also be outliers which lie within the region encapsulated by both boundaries. Table 5.2 shows the result of this analysis by giving the number of outliers falling in each region.

[34] Note, it is feasible to talk in terms of farms (companies) as, for instance, a farm reporting low values in one period likely reports such values in other periods. Consequently, the whole farm is characterised as an outlier. Nevertheless, I leave open the possibility that only part of the observations per farm are considered as outliers, thus preserving valuable information.

Figure 5.3: Illustration of constructed boundary for two dimensions.



Notes: Based on West German data. Boundaries (filled dots with black circles) are given by the identified non-outliers (black dots). Outliers (grey dots) lying in the dark- grey region are considered as large farms and those lying in the light-grey region as small farms.

Source: Author.

Table 5.2: Multivariate outliers divided into small and large farms.

	small farms	large farms	neither nor	∑
East Germany	79	6	237	322
West Germany	229	236	656	1121

Notes: Number of observations belonging to respective farm category according to the FDH procedure.

Source: Author.

Note that all five variables have been jointly considered for these results. Therefore, the regions are actually hypercuboidal and not planar as Figure 5.3 might suggest. The results generally support the conclusions already drawn from the parallel boxplots. Interestingly, for both East and West Germany, there is a substantial number of outliers within the encapsulated region. These observations, while being identified as outliers, must be located close to the bulk of non-outlying observations. Hence, if one is interested in a more conservative approach in outlier identification, one may consider including these observations in the set of non-outliers again for any follow-up production function analysis.

5.2.2 Production function estimation

I estimate the parameters of the production function by the WOOLDRIDGE (2009) (WLP) estimator. It is applied to various samples: a) the full sample without any outlier identification (no-out), b) the cleaned subsample resulting after univariate outlier identification (uni-out), c) the cleaned subsample resulting after multivariate outlier identification (full-out), and d) the cleaned subsample resulting after removing only small and large multivariate outliers (small-large). For case b), observations were dropped if the fixed capital productivity per farm was beyond the upper/lower quartile ± 1.5 times the interquartile range (IQR). I resort to such a trimming rule because it is prominent (i.e., widely used) in the literature outlined in chapter 2.2.2. According to this rule, 298 outliers have been detected for East Germany and 1,429 for West Germany, which differs from the numbers observed in the multivariate case (see Table 5.2). Even though the differences in figures are not that large, especially in the case of East Germany, there is little overlap in the detected outliers between the uni- and multivariate detection in both German regions.[35] This result is not too surprising because both ways to detect outliers are inherently different. The former only considers one separate dimension, whereas the latter evaluates the whole multidimensional "tree" of data. The final estimation samples include farms that have a minimum panel representation of four consecutive years to justify the assumption that factor adjustment drives unobserved heterogeneity. This treatment ensures that the panels do not become jagged after applying the outlier removal procedure because the detection is done on observations. In addition, it copes well with the assumption that costly factor adjustment drives unobserved heterogeneity (PETRICK and KLOSS, 2013a). Finally, it also has implications for the farms in the

[35] This overlap amounts to 8 per cent in the eastern part and 3 per cent in the western part of Germany, measured in percentage of the total number of multivariate outliers.

sample. For instance, if a farm over-reports in one or several non-consecutive periods, it will not be included in the estimation sample.

5.2.2.1 Results for East Germany

Comparing the results in Table 5.3, it turns out that for the majority of samples the labour coefficient is insignificant and close to zero. Generally, materials is the most important input as it displays output elasticities fluctuating at around 1.0. Furthermore, the hypothesis of constant returns to scale cannot be rejected for any sample.

As expected, the estimates based on the small-large subsample and the uni-out subsample are quite similar to the no-out estimates because only relatively few outliers are discarded from the full sample, especially in the former subsample. By removing the full set of multivariate outliers, only farms relying, on average, on a more (fixed) capital intensive production are left in the sample as indicated by the highest capital coefficient among all subsamples. Furthermore, in this case results display a higher precision as the standard errors of this parameter are smallest among all subsamples. Likewise, the remaining farms after multivariate decontamination are less working capital intensive. In addition, the scale elasticity is closest to one.

Interestingly, the results for the full-out subsample are somewhat different. Remember that throughout a modified specification of the materials input is used to mitigate multicollinearity issues (chapter 4.2.2). However, the general tendency with a relatively high materials coefficient in conjunction with a low land coefficient prevails. Compared to the other subsamples, it was possible to identify the land output elasticity. Moreover, the materials coefficient slightly decreased. Hence, the multivariate outlier decontamination procedure mitigates the multicollinearity problem even further, a result also observed for West Germany (see below). This observation further implies that much of the multicollinearity is originating from the multivariate outliers. Unfortunately, a negative and significant labour coefficient is observed.

Summing up the findings, there seems to be a trade-off. On the one hand, estimates after the full multivariate outlier decontamination provide reasonable and sometimes more precise results for the three inputs of land, materials, and capital. On the other hand, I observe a significantly negative labour output elasticity. Nevertheless, estimation results for the other subsamples, including the more conservatively multivariately cleaned subsample d), propose a consensus suggesting that labour is an abundant input factor.

5.2.2.2 Results for West Germany

The production function estimates for West Germany are summarised in Table 5.4. In general, the structure of farms in East and West Germany is quite different (as Figure 5.2 suggests). While there are many very large farms in East Germany (including some farms identified as outlying), in West Germany primarily small and medium scale family farms are prevalent. Hence, one can presume that production technologies in both regions are different to some extent. However, there are also some similarities. Material inputs are also the single most important production factors in West German arable farming; this is true to a lesser extent than in East Germany. Another similarity to the results for East Germany is the insignificant and negative land elasticity for estimation samples not relying on multivariate decontamination. Again, one can observe that after removing all multivariate outliers the parameter estimate increases, which confirms the indication that the multivariate outlier detection alleviates effects of multicollinearity. Note this is the only instance for which a positive (but still insignificant) land coefficient is estimated.

Furthermore, while the assumption of constant returns to scale cannot be rejected for all subsamples, this result is much clearer for estimation samples based on multivariate detection methods. The no-out subsample even rejects this assumption at the 10 per cent significance level. Orders of magnitude of estimated labour and capital coefficients are smaller in the multivariate compared to the no-out and univariate cases, hinting at upward biased coefficients in the latter two. This drives the results more into the direction of returns to scale approaching unity. Additionally, a consistent difference between East and West Germany is that the labour elasticity is significantly positive throughout, indicating that this input is to some extent a scarce factor in West Germany.

Generally, for the case of West Germany, the latest methodological advancements in production function estimation together with multivariate outlier decontamination are able to fully embrace their benefits. As outlined above, this is signaled by plausible estimates which exhibit positive output elasticities throughout all inputs and, in the case of capital inputs, the highest parameter precision.

5.2.3 Evaluation of Results

In summary, the two-step approach presented allows for robust and consistent estimation of production functions. By applying a multivariate assessment of outliers, I am able to treat all considered dimensions of agricultural production, resulting in a global assessment of outliers. Perhaps unsurprisingly, the univariate decontamination procedure is not capable in detecting the meaningful outliers.

Table 5.3: Results of production function estimation for East Germany.

	No-out		Uni-out		Full-out		Small-large	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Labour	0.000	0.050	0.002	0.049	-0.116***	0.040	-0.004	0.056
Land	-0.044	0.042	-0.044	0.034	0.125***	0.042	-0.019	0.062
Materials	1.015***	0.136	1.034***	0.158	0.884***	0.147	1.001***	0.145
Capital	0.125***	0.047	0.123**	0.050	0.146***	0.039	0.110**	0.049
N	1340		1305		1262		1327	
Elasticity of scale	1.096		1.115		1.039		1.089	
p-value const. ret. to scale	0.395		0.381		0.757		0.463	
p-value coeff. jointly zero	<0.001		<0.001		<0.001		<0.001	

Notes: Year dummies included in all models. *** (**, *) significant at the 1% (5%, 10%) level, based on standard errors robust to clustering in groups.
Source: Author.

Table 5.4: Results of production function estimation for West Germany.

	No-out		Uni-out		Full-out		Small-large	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Labour	0.186***	0.019	0.184***	0.019	0.167***	0.023	0.178***	0.021
Land	-0.010	0.016	-0.011	0.016	0.019	0.020	-0.000	0.018
Materials	0.802***	0.086	0.773***	0.083	0.803***	0.091	0.786***	0.089
Capital	0.159***	0.023	0.169***	0.023	0.129***	0.022	0.156***	0.023
N	3382		3355		3053		3245	
Elasticity of scale	1.137		1.115		1.118		1.120	
p-value const. ret. to scale	0.061		0.105		0.135		0.112	
p-value coeff. jointly zero	<0.001		<0.001		<0.001		<0.001	

Notes: Year dummies included in all models. *** (**, *) significant at the 1% (5%, 10%) level, based on standard errors robust to clustering in groups.
Source: Author.

This procedure can only detect conventional outliers beyond some threshold in a single dimension, while the multivariate algorithm at hand may, in addition, also detect such outliers located within the production technology. This is a major advantage compared to the common univariate approach employed. Results indicate that the pMST procedure indeed detected all scattered observations located around the main bulk of observations as well as those within the production technology. In light of these findings, the univariate procedure tends to mis-estimate the number of outliers and consequently drops (keeps) valuable (valueless) information, which is also indicated by the low overlap in observations between the two procedures.

The advantage of the multivariate detection also carries over to the estimation, at least to some extent. It results in improved estimates compared to the results for the other samples (e.g., it allows for a higher precision in the estimation of some parameters). In addition, multivariate detection seems to mitigate the effects of multicollinearity further, however, with better results for West Germany. Discarding only extreme multivariate outliers leads to more conservative results closer to the full samples estimates.

To further illustrate the multicollinearity mitigating effects of the multivariate detection procedure utilised, I additionally evaluate an analysis based on a materials specification including fertiliser inputs (Table B1-Table B3). The difference in identified outliers does not only occur because of the alternate specification itself as the distance between observations is changed but also because it was not possible to construct the materials specification without fertiliser inputs due to missing values for a few farms. Hence, in the full materials specification one can build on a slightly extended data set. Now, if the multivariate detection procedure indeed mitigates multicollinearity, the general tendencies outlined above should also prevail in such a specification. As outlined in chapter 4.2.2, the consideration of fertiliser in the materials input leads to a severe multicollinearity problem characterised by negative and statistically significant land coefficients (Table B2 and Table B3). By moving from uni- to multivariate decontamination, land output elasticities increase, reaching the value closest to zero for both German regions. However, this time they remain significantly negative. Parameters after multivariate detection are estimated with a higher precision, as well, this time around not only for capital but also for materials. Generally, the tendencies of moving from univariate to multivariate outlier decontamination for a materials specification without fertiliser may also be observed on a more dampened level for a materials specification including this input. Thus, one may conclude that the multivariate decontamination further mitigates the multicollinearity problem after refining the precise definition of the materials input.

6 The productivity of family and hired labour in EU arable farming

In this chapter, I utilise the primal production function framework to answer the question whether on-farm labour force composition (i.e., the ratio between family and hired labour) has a gaugeable impact on productivity of a particular field crop farm. Because the production function model employed so far needs to be modified to measure the effects of labour force composition, I start with a discussion of the model specification regarding issues of functional form as well as my strategy to identify the parameters of that function and the parameter of interest. In chapter 6.2, I discuss the data sample, in 6.3 the results.

6.1 Model specification

In the following, I explore the core hypothesis that the composition of the labour force (family versus hired) affects the productivity of agricultural labour input in European field crop farming. Testing this hypothesis empirically involves a number of specification decisions. First, I want my empirical model to be consistent with microeconomic production theory, which requires the specification of a production technology. Moreover, my empirical strategy must make sure it identifies the parameters I am interested in for testing the core hypothesis while being empirically tractable given the farm-level panel data available to me. As I discuss in the following, my preferred choice is a semi-parametric production function model that combines a parametrically specified technology with a robust, moment-based estimator controlling for unobserved heterogeneity. In addition, following the insights from chapter 4.2.2., several modifications to the original analyses must be made with respect to the specification of variables to accommodate the effects of multicollinearity.

6.1.1 Production technology

A key dilemma in modelling production concerns the choice between functional flexibility and empirical tractability. On the one hand, researchers want to impose as little a-priori structure on the data as possible. On the other, less structure typically implies less precise estimates, less meaningful statistical tests, and potential inconsistencies with theoretical assumptions such as concavity or monotonicity. At one extreme, technology could be estimated in an entirely non-parametric fashion. However, a disadvantage of such methods is that estimation with real-world data sets is rarely possible if the number of covariates is higher than two or three (the “curse of dimensionality”, Ichimura and Todd, 2007).

Another shortcoming is that fully non-parametric methods that can handle complex identifying assumptions are not well developed. I therefore resort to a parametric technology specification that allows a straightforward implementation of my core hypothesis. I start with the conventional Cobb-Douglas technology as a workhorse model, which is then extended in various directions to accommodate my assumptions concerning labour force heterogeneity and identification.

Suppose the production technology can be described by the following expression:

$$y_{it} = \alpha^A a_{it} + \alpha^E e_{it} + \alpha^K k_{it} + \alpha^M m_{it} + \omega_{it} + \varepsilon_{it}, \tag{6-1}$$

where yit is the natural logarithm of output Y, A is land use, E is the effective labour effort, K fixed capital, M materials (working capital), lower case letters denote the natural logarithms of these variables, the αs are parameters to be estimated, and i and t are farm and time indices. ω_{it} are farm- and time-specific factors known by the farmer but unobserved by the analyst (unobserved productivity). ε_{it} are the remaining independent and identically distributed errors.

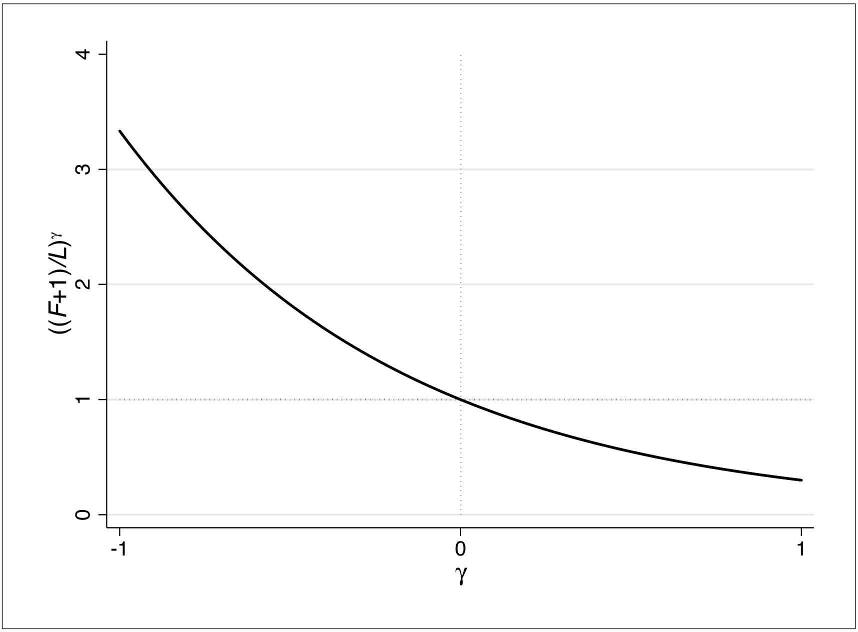
A key idea in my strategy to test the influence of labour force heterogeneity is to substitute E by an effective labour function determined by the share of family labour in total labour input of the farm. I suggest a specification introduced by FRISVOLD (1994):

$$E = L\left(\frac{F+1}{L}\right)^{\gamma}, \tag{6-2}$$

where E is the effective labour input in efficiency units, L is total labour time (i.e., the sum of hired and family labour time), F is family labour time, and γ is a parameter measuring effective labour effort, which is to be estimated. DEOLALIKAR and VIJVERBERG (1983, 1987) experimented with CES and general linear and generalised quadratic effective labour functions, while BARDHAN (1973) also employed an exponential specification.

As equation (6-2) shows, the exponential expression $((F+1)/L)^{\gamma}$ acts as a scaling factor for total labour time input. Following this model, the productivity of each hour of labour supplied to the farm depends on the share of family labour in total labour input and the parameter γ. The latter measures how a farm's labour force composition affects productivity. If $\gamma > 0$, a higher share of family labour increases farm productivity. If $\gamma < 0$, productivity is decreased by a higher share of family labour. A given ratio of family to hired labour can decrease or increase farm productivity, depending on whether γ is positive or negative (Figure 6.1). If $\gamma = 0$, there are no effects of labour force composition. An advantage is that

Figure 6.1: Effective labour as a function of .



Notes: The ratio $((F+1)/L) \in [0,1]$ has been set to 0.3.

Source: Author.

this specification of E allows for farms entirely run by family or hired labour because a 1 is added in the numerator. Furthermore, the exponential form of (6-2) allows for direct estimation of γ in the framework of a Cobb-Douglas function. Applying basic logarithm rules to (6-2) and inserting it into (6-1) gives:

$$y_{it} = \alpha^A a_{it} + \alpha^E l_{it} + \theta r_{it} + \alpha^K k_{it} + \alpha^M m_{it} + \omega_{it} + \varepsilon_{it}, \tag{6-3}$$

where r and l are the natural logarithm of $R = ((F + 1)/L)$ and L, respectively, and $\theta = \alpha^E \gamma$. Given this formulation, γ is equal to θ/α^E. I thus arrive at an empirically tractable technology specification that allows a direct test of the effect of labour force composition.

A further refined model could try to directly estimate even more specific aspects of labour composition, such as time spent on supervision or relative education or technological skills of the different groups of workers (see FRISVOLD (1994) for some steps in the former direction). In my application, these could not be implemented due to data limitations. Even so, my estimates of γ might indeed reflect different qualifications of family and hired labour.

The Cobb-Douglas technology has maintained its status as the workhorse of applied production function analysis up until the present (see Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Petrin and Levinsohn, 2012 for some recent examples). However, it imposes a lot of structure on the production technology, including strong separability, constant output elasticities, a constant scale elasticity, and substitution elasticities between all input pairs which are always constant (= 1) as well (Chambers, 1988). This rigidity can be overcome by adding quadratic and interaction terms of inputs, leading to the more flexible translog formulation. I test this extension below (section 6.3). To support the assumption of a homogeneous production technology, I restrict the empirical analysis to full-time farmers specialising in crop production (see section 6.2).

6.1.2 Identification

Factor use across firms is usually under the control of the farmer and decided simultaneously with unobserved events or may depend on such events. Therefore, the inputs in (6-3) are subject to an *endogeneity problem*. For instance, the farmer's and workers' reactions to environmental shocks are clearly endogenous because they may depend on omitted variables such as technological skills or the experience with past comparable shocks. In return, adjustment to these shocks also affects the other input choices. The unobserved productivity (ω_{it}) might further represent factors such as natural resource endowments of the farm (e.g., soil quality). As a result, the ω_{it} will likely be correlated with the other inputs. The standard OLS estimator will produce biased estimates of output elasticities as it neglects the presence of ω_{it}. This endogeneity problem typically leads to upward biased elasticities for variable inputs (e.g., labour and materials; Levinsohn and Petrin, 2003). As Ackerberg et al. (2007) pointed out, the standard OLS approach also lacks the necessary information that allows separate identification of the production elasticities, leading to a *collinearity problem*. Factor use across farms varies only with the unobserved ω_{it}, so that the different production elasticities are not identified.[36]

[36] See Petrick and Kloss (2013a) as well as chapter 2.1 for a general discussion of these endogeneity and collinearity issues in the context of agricultural production function estimation.

To tackle these problems, one must control for ω_{it} and provide identifying information for the inputs. Returning to the identification strategy presented in chapter 2.1.6, I do this by inserting a non-parametric control function ω_{it} into (6-3), ending up with a partially linear, semi-parametric model first proposed by OLLEY and PAKES (1996: 1275). Moreover, I use the identification approach suggested by WOOLDRIDGE (2009), who uses orthogonality assumptions about past and present levels of input use in the framework of an instrumental variables estimator. This latter approach is consistent with the idea of adjustment costs in input provision that vary across inputs. With regard to my core hypothesis, it assumes that today's labour composition of a farm is affected by past endowments with factors. For example, contemporary labour composition may be driven by past decisions on land purchases.

The rationale behind this approach may be best understood by comparing it with the traditional way to control ω_{it}, the within or fixed effects approach (MUNDLAK, 1961). Suppose ω_{it} can be decomposed further in:

$$\omega_{it} = \lambda_t + \eta_i + v_{it},$$

where λ_t is a time-specific shock identical for all farms in t, η_i is a farm-specific fixed effect that is constant over time, and v_{it} is the remaining farm- and time-specific productivity shock unanticipated by the farmer and unobserved by the analyst. The usual approach then is to purge the fixed effects (η_i) by the so called within transformation. To do so, farm-specific means are subtracted from all the variables. The λ_t are usually controlled for by incorporating time dummies into the model. However, the question remains whether the assumption of time-constant fixed effects is plausible. If η_i represents factors such as management or soil quality, they can be considered as time-varying over a sufficiently long period. Therefore, this assumption is likely to hold only for panels that cover rather short periods of time. Furthermore, the within transformation is known for removing too much variance from variables that exhibit little variation over time, such as land, labour, and fixed capital, resulting in downward biased estimates for these factors (GRILICHES and MAIRESSE, 1998: 180-185). Especially with the effective labour function in mind, this can potentially lead to wrong conclusions.

In contrast, the approach utilised here controls for ω_{it} by a function of observed firm characteristics (OLLEY and PAKES, 1996). LEVINSOHN and PETRIN (2003) proposed the level of materials input to be used as a proxy. Therefore, I assume ω_{it} evolves according to:

$$\omega_{it} = h(m_{it}, k_{it}),$$

where h is a non-parametric function. Furthermore, it is assumed that unobserved productivity follows a first-order Markov process:

$$\omega_{it} = E[\omega_{it}|\omega_{it-1}] + \xi_{it}, \quad (6\text{-}4)$$

where ξ_{it} is an innovation uncorrelated with k_{it}, but possibly correlated with the other factors in the production function. Following WOOLDRIDGE (2009), I additionally assume that:

$$E[\omega_{it}|k_{it}, a_{it-1}l_{it-1}, r_{it-1}, k_{it-1}, m_{it-1}, \dots, a_{i1}, l_{i1}, r_{i1}, k_{i1}, m_{i1}]$$

$$= E[\omega_{it}|\omega_{it-1}] = g(\omega_{it-1}) \equiv g[h(m_{it-1}, k_{it-1})], \quad (6\text{-}5)$$

where g is an unknown productivity function. Equation (6-4) together with (6-5) provide some deeper insight in the innovation . It asserts that this innovation is uncorrelated with current and past realisations of and past realisations of a, l, r, and m. These assumptions are necessary to obtain a consistent estimate of α^K and α^M.

For the ε_{it}, Wooldridge proposes:

$$E\big[\varepsilon_{it}|a_{it}, l_{it}, r_{it}, k_{it}, m_{it}, a_{it-1}, l_{it-1}, r_{it-1}, k_{it-1}, m_{i,t-1}, \dots$$

$$a_{i1}, l_{i1}, r_{i1}, k_{i1}, m_{i1}] = 0 \quad (6\text{-}6)$$

Therefore, the residuals are assumed to be orthogonal not only to current but also all past values of a, l, r, k and m.

Now, starting from (6-3), the problem can be formulated in terms of two equations. The first is given by:

$$y_{it} = \alpha^A a_{it} + \alpha^E l_{it} + \theta r_{it} + \alpha^K k_{it} + \alpha^M m_{it} + h(m_{it}, k_{it}) + \varepsilon_{it}, \quad (6\text{-}7)$$

where (6-6) provides the moment conditions holding for this equation. The second can be obtained by plugging ω_{it} = $g[h(m_{it-1}, k_{it-1})]$ + ξ_{it} into the production function:

$$y_{it} = \alpha^A a_{it} + \alpha^E l_{it} + \theta r_{it} + \alpha^K k_{it} + \alpha^M m_{it} +$$

$$g[h(m_{it-1}, k_{it-1})] + \epsilon_{it}, \tag{6-8}$$

where $\epsilon_{it} = \xi_{it} + \varepsilon_{it}$. The moment conditions holding for this equation are:

$$E[\epsilon_{it} | k_{it}, a_{it-1}, l_{it-1}, r_{it-1}, k_{it-1}, m_{it-1}, \ldots, a_{i1}, l_{i1}, r_{i1}, k_{i1}, m_{i1}] = 0. \tag{6-9}$$

Hence, in (6-7) and (6-8), current and past values of k, past values of a, l, r and m as well as functions of these can be used as instruments. Additionally, in (6-7), contemporaneous proxy variables and current realisations of a, l and r are valid instruments.

The two equations (6-7) and (6-8) together with the moment conditions in (6-6) and (6-9) can be estimated within a generalised methods of moments (GMM) framework by imposing the accompanying moment conditions. In empirical practice, these orthogonality conditions are usually weakened in that only lags up to order one are included. Alternatively, the production function parameters can be identified by estimating (6-8) using instrumental variable estimation with instruments for a, l, r and m (Wooldridge, 2009: 113). The function h is approximated by low-order polynomials of first-order lags of m and k which act as their own instruments. According to this theoretical set-up so far, m needs to be instrumented by its second order lag while a, l and r are instrumented by their first-order lags. Recently, PETRIN and LEVINSOHN (2012) employed this second approach, which I also follow here. The function g is assumed to follow a random walk with drift (WOOLDRIDGE, 2009: 114).

In my agricultural application, the intuition of this approach may be as follows (cf. LEVINSOHN and PETRIN, 2003: 322). Consider ω_{it} to represent a farm-specific stock of management knowledge. Any positive shift of ω_{it} assumedly increases the marginal productivity of m_{it} and possibly all other production factors. As m can be readily adjusted, a profit-maximising farmer increases the level of m_{it} in response to the shift, thus motivating the use of m as a proxy for ω_{it}. The same process may also work in the other direction, so that farms with negative shocks reduce material inputs. If ω is persistent, the farm-specific over- or under-application of material inputs is likely to be correlated over time, so that past levels can be used as proxies for current productivity shifts. Consistent with primarily positive shifts is the empirical observation that, on average, both farm output and materials input increase over the years. This is precisely what the data confirms.

Given this theoretical framework, the WOOLDRIDGE (2009) estimation procedure does not only control for endogeneity problems but also solves the collinearity issue raised by Ackerberg et al. (2007). This is in contrast to former versions of these so-called control function approaches (cf. Bond and Söderbom, 2005; Ackerberg et al., 2007). Petrick and Kloss (2013a) demonstrate that such approaches behave robustly in empirical practice, making them interesting alternatives to the traditional within approach. In the following, I present results for a set of estimators that involves OLS as a baseline as well as the preferred semi-parametric estimator due to Wooldridge (2009).

6.2 Data

I utilise data from the FADN described in chapter 3. Output is measured as the total farm output in euros. The total utilised agricultural area is my land input in hectares. It includes owned and rented land, and land in sharecropping. Material or working capital input is proxied by total intermediate consumption in euros. It consists of total specific costs and overheads arising from production in the accounting year. The former consists of costs for seeds and seedlings, crop protection and other crop-specific costs. Overheads are comprised of "supply costs linked to production activity" and are usually the single largest position in the materials input (European Commission, 2011). They include, among others, costs for energy such as fuel and electricity. I do not include the costs for fertiliser in the materials input. Land and fertiliser are highly correlated, suggesting that land and fertiliser inputs are utilised by farmers in an (almost) fixed ratio.[37] I capture this package by including land input in hectares. Fixed capital is approximated by using the opening valuation of assets which is consistent with most of the recent literature on production function estimation with firm level data such as Olley and Pakes (1996), Blundell and Bond (2000), and Levinsohn and Petrin (2003). In this case, I took the asset value of machinery and buildings from the FADN data. This measure accounts for different depreciation rates of machinery and buildings which are estimated at replacement value of these inputs (European Commission, 2011). Separate information on hired and family labour working time is needed as well as the total labour hours to estimate the effective labour function (2) within a production function framework (i.e. estimating (3)). All this information is readily available in the FADN data. Having this data available, one can construct the additional covariate r. To this end, I calculate $R = ((F + 1)/L)$ and take its natural logarithm. Table 6.1 gives definitions of the variables needed as well as their FADN codes. As discussed in chapter 3, the sample of countries

[37] This shows and confirms the view presented in chapter 4.2.2. In this section I present further analyses regarding the, statistically speaking, closely related multicollinearity problem.

Table 6.1: Description of variables.

FADN code	Variable description
Left-hand side	
SE131	Total output (EUR)
Right-hand side	
Inputs	
SE025	Total utilised agricultural area (ha) = land
F72 + SE300 + SE305 + SE336	Costs for seed and seedlings + crop protection + other crop specific costs + overheads (EUR) = materials
L.SE450 + L.SE455	Opening valuation of machinery and buildings (EUR) = capital
Effective labour effort	
SE011	Total labour input (hours)
SE016	Unpaid labour input, generally family (hours)
SE021	Paid labour input (hours)

Source: Author, European Commission (2011).

is selected to reflect the diverse farm sizes and structures in EU arable farming, ranging from small-scale family farms in Italy, Poland, Spain, and West Germany, to medium-sized commercial farms in Denmark, France, and the UK, to large-scale and mostly corporate farms in East Germany and Slovakia (European Commission, 2012). It is the variability of these structures across countries that makes comparisons of labour force heterogeneity particularly insightful.

The panels for Poland and Slovakia cover only five years because FADN data collection for these countries began in 2004. Moreover, the effective panel length is reduced to four years because I use the opening valuation of fixed assets which is taken from the previous year of observations as my capital proxy. Therefore, my European database consists of 34,896 observations. To be included in the estimating sample, farms had to be present for at least four consecutive years (three years for Poland and Slovakia). Similar to Petrick and Kloss (2013a), outlier analysis was performed on the basis of the fixed capital productivity per farm. Observations were excluded from the estimation if their value exceeded the interval given by $[Q_1 - 1.5 \cdot IQR;\ Q_3 + 1.5 \cdot IQR\ ]$, where Q_1/Q_3 is the lower/upper quartile and IQR the interquartile range.

Table 6.2 summarises the number of farms for every country in my sample, the labour force composition (average percentage of family labour), and the other variable means. My data sample covers a total of 6,546 farms. The numbers on output reflect the different forms of agricultural organisation outlined as discussed in chapter 3 (e.g., small scale, family farming in Spain or Poland up to large scale, corporate faming in East Germany and Slovakia). A full set of descriptive statistics is given in Table C1. Furthermore, according to the table, the dominant type of labour in EU arable farming is family labour. Only Slovakia displays numbers well below 50 per cent. In addition, there are farms entirely run on hired or family labour (e.g., in Germany, Italy, and Poland). To get a more dynamic view of these figures, I graph their evolution over the sample period (Figure C1). In the majority of countries, the percentage of family labour declined between 2001 and 2008. The drop is most remarkable in East Germany and Slovakia. Rather than converging to a family farming model as expected by some observers in the early 1990s, these former socialist countries are now characterised by a firmly established corporate farm sector based on hired labour. Exceptions to the general tendency are Poland and Spain, where the numbers remained more or less constant.

Table 6.2: Sample size and variable means.

Country	Farms	Family labour in % of total labour	Total labour (ths hours)	Output (ths EUR)	Land (ha)	Materials (ths EUR)	Capital (ths EUR)
Denmark	208	84.52	2.8	180.4	122.7	98.0	840.0
France	1030	84.11	3.2	155.8	143.5	85.1	160.1
Germany (East)	271	55.86	15.6	545.6	538.9	345.8	519.4
Germany (West)	566	84.70	4.2	150.9	92.3	84.7	153.9
Italy	1322	88.55	3.6	60.6	44.7	23.8	125.2
Poland	1518	87.07	4.7	39.8	48.6	17.4	78.9
Slovakia	55	28.85	39.1	514.0	768.9	342.1	940.2
Spain	1388	90.06	2.7	40.4	72.7	15.0	31.1
United Kingdom	188	64.76	6.3	278.1	248.7	157.2	239.3

Source: Author based on FADN data.

6.3 Results

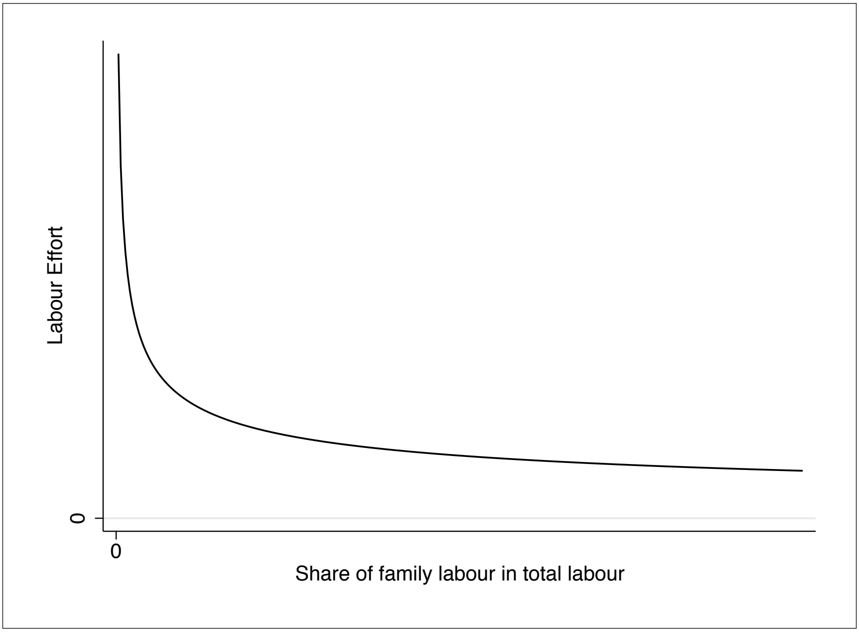
To infer the effective labour effort parameter γ, I estimate equation (6-3) employing two estimators per country. These are 1) OLS as a baseline and 2) Wooldridge (2009), hereafter Wooldridge/Levinsohn/Petrin (WLP). All estimations were performed with Stata 12. To implement the WLP estimator I employed the `ivreg2` routine by Baum et al. (2007) as shown in Petrin and Levinsohn (2012). The code is reproduced in appendix D.

As I mentioned in the identification subsection, the WLP estimation procedure incorporates lags of inputs up to the second order, which reduces the panel length for every country by two years. For Poland and Slovakia the panel size is reduced to two years. In the case of Slovakia, an already small sample is reduced even further. Therefore, the results for this country should be treated with caution. But I kept it as the only representative of an East European country dominated by hired labour. I also use the resulting estimation sample for the OLS estimates to ensure comparability. To recover the standard error of γ, I use the 'delta method' (Greene, 2011: 1123-1124); this method needs to be applied because γ is a non-linear function of θ and α To this end, a linear function approximating the non-linear counterpart is obtained by a so called Taylor series expansion of the non-linear function. Afterward, one can easily calculate the variance of this linear approximation. Returns to scale was measured as the sum of the direct production elasticities of labour, land, materials, and capital. Given sufficiently developed factor markets for these four inputs in the countries studied, it seems reasonable to assume that all factors can be adjusted at some cost and with some delay (see chapter 6.1.2).

In Table C2 and Table C3, I report detailed results of the production function estimation per estimator and country. I prefer the WLP estimator because on theoretical grounds, it corrects the biases induced by the endogeneity and collinearity problems present in production function estimation. Empirically, the results look very plausible. In contrast to the OLS estimates, which reject the assumption of constant technical returns to scale for every country but West Germany and Slovakia at the 5 per cent significance level, the WLP estimates reject this assumption only for Poland and Spain.

Table 6.3 gives the sample size and the point estimate as well as the standard error of γ per country and estimator. Regarding its significance in the different member states and regions, the following picture unfolds. In Denmark, East Germany, Italy, Slovakia, and Spain, the coefficient of γ is not significantly different from zero, meaning the null hypothesis of perfect substitution between hired and family labour cannot be rejected. Generally, labour does not seem to be a scarce factor in Slovakia and East Germany because their labour coefficients are not statistically different from zero (Table C3). Both exhibit large scale farming structures. The small and medium scale agricultural structures of West Germany, Poland, and France exhibit negative γs that are significantly different from zero. This result implies that effective labour effort is a monotonically decreasing function of the share of family labour (Figure 6.2). Hence, farms relying on hired labour are more productive than family farms, and farms relying almost completely on hired labour are particularly productive. In other words, the productivity loss created by a higher share of family labour declines as the ratio of family to total labour expressed in (6-2) increases (cf. BARDHAN, 1973: 1381). This is likely to be instances of hired labour specialising in high productivity tasks and/or family labour focusing on low productivity tasks.

Figure 6.2: Effective labour as a function of family labour share in total labour if γ < 0.



Notes: Graphical representation of hired labour being more productive than family labour ($\gamma < 0$) and constant total labour input.

Source: Authors' elaboration.

Regarding the size of γ, West Germany exhibits the largest productivity differential between family and hired labour. An example calculation illustrates the effects. In West Germany, an increase in the share of hired labour time from 20 per cent to 30 per cent in total labour time amounts to an average increase in labour productivity by 0.56 EUR/hour or about 2.4 per cent – up to 23.50 EUR/hour.[38]

The classical case of family members being more productive than hired labour is only observed for the United Kingdom, and the extent is moderate. Here, there is an argument for labour supervision. Finally, the distribution of labour force heterogeneity across the sample countries suggests that mainly small to medium scale agrarian structures display differing effects on productivity for the two types labour.

The direct output effects of the land input is small in most countries, with the exception of Denmark and the United Kingdom, and often it is not significantly different from zero. East Germany, Italy, and Slovakia even display negative parameter estimates, though also not statistically significantly different from zero (Table C3). Therefore, they are assumed to be zero. This finding is consistent with the plausible view that, while holding all other factors constant, expanding land does not raise output. The rationale behind this observation is as follows. Most farms utilise inputs, particularly materials, in such abundance that an additional hectare of land does not have a positive output effect. This result holds even more in the WLP model (i.e., when unobserved productivity differences are controlled for).

Levinsohn and Petrin (2003) argue that the bias direction in OLS estimates of the capital input depends on the degree of correlation between this input and the unobserved productivity (ω_{it}). In general, it tends to be upward biased in my application, thus contributing to an upward bias in returns to scale. The OLS estimates of materials and the ratio variable are also often larger than their WLP counterparts, which is consistent with prior theoretical predictions and empirical observations.

Compared to the WLP estimates, the OLS estimator detects labour force heterogeneity in one more case, Denmark, whereas it does not detect such heterogeneity in Poland. The reason for the former result is probably an upward biased OLS

38 Calculation based on the sample means of the different inputs and the WLP production function estimates. Similarly, the average labour productivity increase for France amounts to 0.9 per cent; for Poland it amounts to 1.2 per cent.

estimate of the ratio r in absolute terms leading to an upward biased estimate of γ while the argument for the latter is an upward biased OLS estimate of the labour coefficient. Such a result is commonly observed for the OLS estimator in the presence of endogeneity. Therefore, these estimates are most likely biased.

There is one noteworthy finding for the case of Denmark,. In this instance, the WLP estimator was not able to identify a parameter estimate for the materials input. The reason for this result is twofold. First, there is probably not enough identifying information (variance) left in this input after the exclusion of fertiliser inputs.[39] Second, the non-parametric control function that also houses the materials inputs most likely captures huge parts of the explaining variation that is left. This second argument explains why the WLP estimation procedure is affected, at least in this particular case.

Table 6.3: Effective labour effort parameter (γ) in comparison.

Country	OLS			'Wooldridge/Levinsohn/Petrin'		
	N	γ	SE	N	γ	SE
Denmark	605	0.290***	0.108	605	0.117	0.090
France	4289	-0.554**	0.217	4289	-0.481**	0.214
Germany (East)	1047	0.380	0.269	1047	0.212	0.166
Germany (West)	2408	-1.641**	0.698	2408	-1.517**	0.615
Italy	3545	-0.274	0.221	3545	-0.165	0.198
Poland	2534	-0.123	0.130	2534	-0.848**	0.374
Slovakia	89	-0.874	2.937	89	-0.247	0.288
Spain	6393	-0.014	0.025	6393	0.001	0.033
United Kingdom	612	0.190**	0.090	612	0.215**	0.095

Notes: Year dummies included in all models. *** (**, *) significant at the 1% (5%, 10%) level, based on standard errors robust to clustering in groups.

Source: Author.

[39] Remember, the cost for fertiliser inputs was not included to mitigate multicollinearity problems between fertiliser (i.e., materials) and land inputs.

To understand how far the assumption of a Cobb-Douglas technology restricts my findings, I also experimented with a translog production function.[40] This specification simultaneously produced unreasonable results that exhibited (at sample means) negative production elasticities for some factors and elasticities greater than one for others. In addition, the null hypothesis of joint insignificance of the interaction terms was never rejected for any country. Hence, interaction terms do not add any meaningful insights to the Cobb-Douglas specification. I do not consider the translog functional form to be a suitable approximation to the data. My findings are in line with other recent studies based on FADN data (cf. Zhengfei et al., 2006; Latruffe and Nauges, 2013).

[40] I estimated the translog using total labour input (Table C4). A translog specification incorporating the effective labour input could only be implemented insofar that interaction terms incorporating E were excluded.

7 Conclusions

Throughout this work, I employed the primal approach to assess agricultural factor productivities. Methodological as well as data issues inhibit the unbiased estimation of production functions. Therefore, I attacked this classical field of application from different sides. At first, dealing with identification issues, I discussed and evaluated traditional as well as recent estimation approaches within the context of agriculture. This assessment was conducted from a theoretical as well as empirical point of view. Second, estimates might also be biased due to the presence of outliers in the data. By departing from the general practice in empirical economics of only using simple univariate measures to identify outliers, I outlined the consequences to estimation of applying uni- and multivariate decontamination schemes. In the empirical section, I applied one particularly suitable multivariate outlier detection procedure to my FADN data base. Subsequently, estimation results for the different decontamination schemes were compared. Finally, after evaluating the available toolbox, I modified and scrutinised my production function model specification to measure the impact of a farms' labour force composition on productivity in European field crop farming.

7.1 Main findings and policy recommendations

The first major goal of this dissertation was to give a conceptual as well as applied comparison of a set of received and innovative production function estimators. The empirical comparison is facilitated by a firm-level dataset representing the agricultural sectors of six EU countries. In this study, I depart from the recently reignited scientific discourse on identification problems – endogeneity and collinearity – and their solution in estimating production functions. Throughout, I argue that the choice of an appropriate estimator depends on (1) the adjustment flexibility over time and (2) the potential observability of production factors by the econometrician. To this end, I have developed and referred to a typology of the relevant tangible and intangible inputs.

In my theoretical considerations, I state that the underlying assumptions accompanying the within and duality approaches are too strong and implausible for the case of agriculture. Within approaches assume that unobserved effects are either constant over time or individuals. Hence, it is not possible to control for potentially important time-varying unobserveables. Duality is based on short-term profit maximisation of economic entities as well as output and factor markets that may be described by perfect competition. Such conditions are unlikely to be found in agriculture. Perhaps unsurprisingly, approaches based on duality have not performed well in empirical applications.

These observations brought my attention to identification strategies employing heterogeneous frictions in factor adjustment. With the comprehensive literature on adjustment frictions of land, labour, and capital markets in mind, adjustment costs may be particularly relevant for key production factors in agricultural production. Olley and Pakes (1996), Blundell and Bond (2000), Levinsohn and Petrin (2003), as well as Wooldridge (2009) base identification precisely on this idea that such adjustment frictions drive factor allocation – an a-priori plausible approach. The main distinction between the control function and BB approach is that the latter allows time-invariant fixed effects, whereas the former does not. Moreover, in the former it is assumed that factor adjustment is completed within a single period, while the process might potentially cover many periods in the dynamic panel data models. In agricultural applications, this is a conceptual advantage as adjustment processes of land, labour, and capital are generally of an intertemporal nature, which cannot be adequately covered by a one-period lag. Another issue is that the collinearity problem is not satisfactorily addressed by OP and LP. Both approaches treat labour and land as fully flexible inputs, thus leaving no source of exogenous variation for identification purposes across observations (Ackerberg et al., 2006). However, Wooldridge (2009) proposes a solution to this issue by modifying and extending the central identifying assumptions of OP and LP.

Following the general insight, that estimates might also be biased due to outlying observations in the data, the second aim of the study was to provide a comprehensive assessment of the effects of outliers on production function estimates. To this end, I first conducted an exhaustive survey of studies from empirical economics. To get a general picture, I employed empirical work utilising three important data sources – the Farm Accountancy Data Network, the World Bank Living Standard Measurement Surveys, and the German Socio-Economic Panel. According to my survey, the general mode of operation is to apply a univariate outlier decontamination, which focuses on a single model variable. However, such approaches neglect the multivariate nature of the model at hand. In addition, they are often ad hoc and handmade. Moreover, in the majority of studies no outlier control is carried out. This leads us to conclude that outlier problems are somewhat second rank issues and are also often neglected. Second, I demonstrated the consequences on production function estimation that the presence of outliers can have by introducing a simulated example. This example illustrates the bias on estimated output elasticities if outliers are not taken care of. More importantly, it shows that a univariate decontamination mechanism might not mitigate the outlier induced bias. Therefore, I argue for a multivariate decontamination procedure.

In my application, I utilise a non-parametric multivariate decontamination procedure based on pruning the minimum spanning tree of a data set to determine an outlier-free subsample (Kirschstein et al., 2013). In contrast to univariate extreme value detection, this method assesses outliers considering all observed dimensions of agricultural production. This decontamination procedure is applied prior to production function estimation – effectively resulting in a two-step procedure.

In the empirical chapter focusing on the evaluation of production function identification strategies, results are given for revenue shares, OLS Cobb-Douglas and translog, within Cobb-Douglas and translog, WLP Cobb-Douglas and translog, as well as OP, LP, and BB Cobb-Douglas models. A separate model for panels of field crop farms in Denmark, France, East and West Germany, Italy, Spain, and the United Kingdom was estimated. OLS and within display the biases expected from the literature compared to the revenue shares. Generally, the former overestimated the variable factor materials, whereas the latter underestimated the relatively fixed factor capital.

Concerning the control function approaches, LP may be taken as an alternative to the received OLS and within estimators because it allowed easy implementation and produced plausible results. However, it is only the second-best choice because of problems in identifying the supposedly flexible factors labour and land. With the exception of BB and WLP, the other estimators share this issue. Overall, results produced by LP and WLP were very similar, which strengthens my confidence in the proxy approach. However, not only does the latter provide a theoretical advantage in identifying land and labour coefficients, it is also sometimes more successful in identifying the capital output elasticity and provides analytic standard errors. Consequently, these arguments give the WLP an edge over the LP estimator.

The performance of the BB estimator was not always satisfactory. By combining a fixed effects – by means of first-differencing – and instrumental variable approach, this estimator must travel far to try to control for all the factors impeding an unbiased productivity estimation. The BB estimator poses assumptions on adjustment costs that are theoretically very plausible. In the empirical domain this could be supported for labour, land, and capital. Because adjustment costs are very high and factor evolution is so persistent in agriculture, this approach somewhat overshoots the target. Even though the system GMM approach of Blundell and Bond (1998) is utilised, often there is not enough explaining variation available for identification. Nevertheless, this estimator produced plausible results with regard to materials inputs.

no further insights were gained by moving from the workhorse Cobb-Douglas to a translog specification. The results were either implausible, for the OLS and WLP estimators, or they differed little from Cobb-Douglas as in case of the fixed-effects regression. Supposedly, the implausible results are a consequence of a deepened multicollinearity issue. Hence, the more parsimonious Cobb-Douglas specification is not only a pragmatic but also empirically well-supported alternative.

Additionally, I analyse the special role of materials and land – two highly correlated inputs – in EU field crop farming. By using previous research and performing additional diagnostic analyses, I reveal that the definition of the materials variable is the crucial point in mitigating multicollinearity problems between materials and land. Nevertheless, my final specification is able to capture all relevant inputs.

My estimates unfold the following picture. Throughout, I find very low production elasticities for labour, land, and fixed capital, whereas the elasticity of materials is above 0.7. This result indicates that an improved access to working capital is most promising in increasing agricultural productivity.

Mundlak et al. (2012) report output elasticities of land and fixed capital in a cross-country sample of developing and developed countries which display much larger orders of magnitude than my results. In contrast to other parts of the world, EU field crop technologies might be characterised by a strong sensitivity to the level of applied variable inputs such as fuel, fertiliser, and chemicals. Thus, policy makers attempting to increase EU agricultural productivity with a given level of the other inputs and technology must apply the lever to this factor. I calculated shadow prices of production factors to assess whether farmers exploit the returns to the utilised inputs. The analysis revealed considerable heterogeneity across EU countries. Because the shadow prices are based on the estimated output elasticities obtained in this study, the remuneration to labour, land, and fixed capital is also, as expected, quite low. An exception to some extent is Denmark. With regard to short-term lending, I find marginal returns to materials much above typical market interest rates in France, Spain, Italy, and Germany, especially toward the end of the observed period. Such a finding is consistent with the perception of constrained access to short-term credit, possibly induced by the emerging financial crisis (Petrick and Kloss, 2013c). This makes a strong case for recommending improved access to short-term funding to release the perceived credit constraint.

Overall, the methodological insights suggest that the recently proposed approaches basing their identification on the presence of adjustment costs provide important conceptual improvements compared to within and duality models. Building on adjustment frictions in land, labour, and capital inputs within the agricultural sector is particularly plausible because such adjustment costs have long been recognised. In any event, the conceptual improvements do not always carry over empirical implementation. This especially holds for the Blundell and Bond (2000) dynamic panel data estimator, which often did not identify plausible results for the (quasi-) fixed inputs. However, less demanding, in terms of assumptions and computability, proxy approaches present a viable alternative for agricultural productivity analysis.

In the empirical chapter on outlier robust production function estimates, I proceeded in two-steps. First, I applied the multivariate decontamination procedure. In the second step, production functions were estimated using my preferred panel data estimator by Wooldridge (2009). This procedure is applied to East and West German agricultural field crop data from the FADN data base.

The analysis reveals that many outlying observations are made up of relatively small farms. In addition, outliers also were detected within the main bulk of observations, which cannot be detected with a univariate approach. Hence, future empirical analyses utilising FADN data should apply a multivariate outlier decontamination procedure. The estimated production functions show that the WLP estimator delivers convincing results with respect to input elasticities and returns to scale when it is applied to outlier-free subsamples derived by the multivariate pMST approach. The estimates for East and West Germany show similarly positive elasticities for capital (working and fixed) as well as scale elasticities close to one. By moving from the univariate to the multivariate decontaminated sample a picture unfolds in which, on average, more fixed capital and less working capital farms remain in the East German sample; it is the reverse for the West German sample. A key advantage of the multivariate decontamination is that it helps to further mitigate multicollinearity problems. Generally, estimations based on the multivariately decontaminated sample provide parameter estimates of previously unidentified coefficients (land in East Germany) or a higher precision in estimation. In the case of East Germany, this does come at some cost as I observed a negative and significant labour coefficient. However, results for the other subsamples, including the more conservative multivariate solution, suggest that labour is not a scarce factor in East Germany. In essence, the proposed two-step approach reveals that production technology in East and West Germany is not as different as it seems at the first glance. In both regions, farms rely heavily on material and capital inputs.

The most remarkable remaining difference between East and West Germany is the labour elasticity, which is significantly positive for West Germany. This can be interpreted as an indication of a shortage of labour force in West Germany, whereas East Germany does not suffer from such a restriction. Similarly, for both regions fixed capital constitutes an equally restrictive input factor whereby the materials input is the far most important input. In general, the results reflect the tendencies from chapter 4 with regard to orders of magnitude of estimated coefficients. Therefore, policy reforms should aim to ease the access of agricultural companies to capital and labour force, particularly in West Germany.

In the empirical chapter on the productivity of family and hired labour, I assessed the heterogeneity of these two types of labour in European field crop farming. To this end, I took a sample of eight EU countries and estimated augmented production functions that allow testing for labour force heterogeneity using farm-level FADN data. The results unveil a diverse picture.

Contrary to the received wisdom, I find that farms with a higher share of hired labour are more productive than family farms in the small- and medium-scale agrarian structures of France, West Germany, and Poland. According to my estimates, hired labour performs the high productivity tasks in these countries. Therefore, an increased reliance on hired labour or the shift of family labour to more productive tasks raises productivity. Hence, labour market reforms should aim to provide incentives to hire specialised labour. For instance, programs to train and hire skilled labour could improve their inflow into agricultural labour markets. In the majority of countries I found no evidence for labour force heterogeneity. For the United Kingdom, I observe that total labour productivity is higher when there are more family members in the labour force. In this case, supervision by family members apparently does increase productivity.

The results have implications for future theoretical and empirical work. Most importantly, my results call into question the general validity of one of the received family farm theory's main pillars (i.e., the dominant effect of supervision costs on hired labour productivity). Countries regarded as traditional strongholds of the family farm have apparently crossed a technological threshold where specialisation of hired labour overcompensates the negative effects of workers' moral hazard. Factors such as the increasing importance of non-traditional and non-agricultural sources of farm household income likely reinforce this trend. On the other hand, the assumption of constant technical returns to scale is confirmed.

In classical production function estimation, labour input is measured as the sum of both hired and family workers. Given the evidence on labour force heterogeneity in some countries, their heterogeneity should not be ignored. Such a treatment will improve model fit and avoid misspecification.

Finally, this work is also a plea for refined methods that control for the problems in production function estimation. Endogeneity and collinearity problems potentially lead to misleading results. The OLS estimator neglecting the presence of endogeneity does not always seem to detect labour force heterogeneity correctly. A possible alternative which has been extensively used in prior empirical work is the fixed effects regression. However, it is notorious for removing too much variance from variables that exhibit little variation over time. Hence, not enough variation is left in the data for estimation purposes (Petrick and Kloss, 2013a). This shifted my attention to the control function framework introduced by Olley/Pakes and then further refined by Levinsohn/Petrin and Wooldridge. The latter is an especially promising alternative to traditional OLS and within approaches. The results obtained from the Wooldridge/Levinsohn/Petrin approach seem to strengthen their validity on empirical grounds, in addition to being plausible in the theoretical domain.

7.2 Recommendations for future research

The analyses conducted in this study have provided new insights regarding factor productivity in EU agriculture. Through the support of state-of-the-art methods, previously not applied within an agricultural context, I managed to draw a picture as precise as possible. However, new issues emerged throughout the research process. I outline those issues below and give recommendations for future research.

First, I regard the analysis of alternative functional forms in conjunction with FADN data as an interesting starting point for future research. For instance, Zhengfei et al. (2006) proposed augmented translog specifications that incorporate agronomic principles. Unfortunately, up until now, there has been a trade-off in applied empirical work between more flexible functional forms for production functions and methodological sophistication in terms of estimation methods. Therefore, combining this methodological sophistication with more flexible functional forms seems to be desirable.

Second, control function approaches to estimate factor productivity remain a vital field of research. Kim et al. (2016) further extend this approach to the case when inputs are measured with error, thus accounting for a further potential identification issue. This might be a useful extension for empirical applications, depending on data quality.

Third, the calculated shadow prices give rise to the question about their main determinants. For instance, with regard to (working) capital it would be interesting to ask what are the differences in driving forces of their shadow interest rate between rationed and non-rationed farms. Such factors could be structural variables such as the availability of collateral, the experience of the farm manager or the workforce composition of an operation, or environmental variables (e.g., regional conditions such as the average farm size or rising/declining prices in crop products). While the exogenous market interest rate should be the only determining factor for non-rationed farms in absence of market frictions, the given examples might be driving shadow rates on rationed farms. This treatment would require a suitable criterion to split farms between rationed and non-rationed farms. In addition, the specification of a sample selection model is methodologically necessary (Petrick, 2004: 141-142). These methods are well established within a cross-sectional data context. However, this is still a challenge in a panel data world (see, e.g., Wooldridge (1995) and Semykina and Wooldridge (2010) for an illustration of and solution to the problem).

Fourth, the data used for analysis revealed a considerable amount of multicollinearity between land and materials inputs, thus impeding the separate identification of the two input coefficients. Prior research using the same data, as a direct result of this multicollinearity problem, reported negative land input coefficients for former input. To deal with this issue, I re-defined the materials input by excluding the most collinear factor – fertiliser – from it. However, a more general analysis of multicollinearity issues and their driving forces in the FADN data is desirable, and thus left to future research.

Fifth, statistical outlier analysis in general has its limitations. As it is only a statistical tool, it will not absolve the researcher from thinking about the scope of research. With regard to production function estimation, statistical outlier analysis is a mean to homogenise production technologies. Therefore, the statements made in this work are with respect to the prevalent production technology in both German regions. However, many of the outliers detected are comprised of small farms. If such farms are of interest to the research question at hand, the outlier analysis, uni- or multivariate, needs to be reconsidered. Hence, the contextualisation of (alleged) outliers seems to be a fruitful starting point for further work.

Sixth, farming in France, West Germany, and Poland demonstrates that a higher share of hired labour leads to increased on-farm productivity. Hence, future research should try to find out why hired labour in arable farming is so productive in these countries, because they are traditionally dominated by small family farms. One possible explanation is that farm technology (e.g., modern tractors and other field machinery using precision farming methods) have become so complex that benefits from labour specialisation can be reaped.

Finally, my data did not allow me to explicitly address the effects of skills and technical expertise of family versus hired workers. However, I am confident that my identification strategy is able to account for such effects. Nevertheless, I regard a more thorough treatment in which these human capital effects are quantified as an important area for future research, too. Given the current structure of the FADN data this can only be accomplished by design and conduction of a separate survey.

References

Ackerberg, D., Lanier Benkard, C., Berry, S., & Pakes, A. (2007). Econometric Tools for Analyzing Market Outcomes, in J.J. Heckman & E.E. Leamer (eds.), *Handbook of Econometrics. Volume 6A* (pp. 4171–4276) Amsterdam: Elsevier.

Ackerberg, D., K. Caves & Frazer, G. (2006). *Structural Identification of Production Functions.* Working Paper, Los Angeles, CA.

Alderman, H. (1998). *Social assistance in Albania : decentralization and targeted transfers*. Living Standard Measurement Survey Working Paper No. 134,Washington, DC.

Allen, D. W., & Lueck, D. (1998). The Nature of the Farm. *Journal of Law and Economics, 41*(2), 343–386. doi:10.1086/467393

Allen, D.W., & Lueck, D. (2002). *The Nature of the Farm. Contracts, Risk, and Organization in Agriculture*. Cambridge: MIT Press.

Alston, J. M., & Pardey, P. G. (2014). Agriculture in the Global Economy. *Journal of Economic Perspectives*, 28(1), 121–146. doi:10.1257/jep.28.1.121

Andersen, M. A., Alston, J. M., & Pardey, P. G. (2011). Capital Services in U.S. Agriculture: Concepts, Comparisons, and the Treatment of Interest Rates. *American Journal of Agricultural Economics*, *93*(3), 718–738. doi:10.1093/ajae/aar010

Antle, J.M., & Capalbo, S.M. (1988). An Introduction to Recent Developments in Production Theory and Productivity Measurement. In S.M. Capalbo & J.M. Antle (eds.), *Agricultural Productivity. Measurement and Explanation (*pp. 17–95). Washington D.C.: Resources for the future.

Arellano, M., & Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*, *58*, 277–297.

Arnold, D., Brändle, T., & Goerke, L. (2014). *Sickness Absence and Works Councils: Evidence from German Individual and Linked Employer-Employee Data*. SOEPapers on Multidisciplinary Panel Data Research No. 691, Berlin.

Auer, W., & Danzer, N. (2014). *Fixed-Term Employment and Fertility: Evidence from German Micro Data.* SOEPapers on Multidisciplinary Panel Data Research No. 719, Berlin.

Aurbacher, J., Lippert, C., & Dabbert, S. (2011). Imperfect markets for used machinery, asynchronous replacement times, and heterogeneity in cost as path-dependent barriers to cooperation between farmers. *Biosystems Engineering*, *108*(2), 144–153. Doi:10.1016/j.biosystemseng.2010.11.007

Avdic, D., & Bünnings, C. (2015). *Does the Burglar Also Disturb the Neighbor?: Crime Spillovers on Individual Well-being.* SOEPapers on Multidisciplinary Panel Data Research No. 737, Berlin.

Backiny-Yetna, P., & McGee, K. (2015). *Gender differentials and agricultural productivity in Niger.* World Bank Policy Research Working Paper No. WPS7199, Washington, DC.

Bahta, S. T., Kleinhanss, W., & Berner, A. (2011). Results for Germany. In W. Kleinhanss (Ed.), *Implementation, validation and results of the cost of production model using the EU FADN, FACEPA Deliverable No. 3.2*. Braunschweig.

Bakucs, Z., Ferto, I., Fogarasi, J., Forgacs, C., Toth, J., Matveev, E., Sepp, M., Desjeux, Y., Latruffe, L., Cesaro, L., Marongiu, S., Zanoli, A., Dolman, M., & Soboh, R. (2010). *EU farms' technical efficiency, allocative efficiency, and productivity change in 1990-2006*. FACEPA Deliverable 5.2.

Bardhan, P. K. (1973). Size, Productivity, and Returns to Scale: An Analysis of Farm-Level Data in Indian Agriculture. *Journal of Political Economy*, *81*(6), 1370–1386. doi:10.2307/1830745

Barnett, V., & Lewis, T. (2000). *Outliers in statistical data* (3rd edition). Wiley.

Bauer, T., Fertig, M., & Vorel, M. (2011). *Neighborhood Effects and Individual Unemployment.* SOEPapers on Multidisciplinary Panel Data Research No. 409, Berlin.

Bauernschuster, S., Falck, O., & Wössmann, L. (2011). *Surfing Alone? The Internet and Social Capital: Evidence from an Unforeseen Technological Mistake.* SOEPapers on Multidisciplinary Panel Data Research No. 392, Berlin.

Baum, C., Schaffer, M.E., & Stillman S. (2007a). Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *Stata Journal*, *7*(4), 465–506.

Baum, C. F., Schaffer, M. E., & Stillman, S. (2007b) IVREG2: Stata module for extended instrumental variables/2SLS and GMM estimation. Retrieved from http://ideas.repec.org/c/boc/bocode/s425401.html

Benjamin, C., & Phimister, E. (2002). Does Capital Market Structure Affect Farm Investment? A Comparison Using French and British Farm-Level Panel Data. *American Journal of Agricultural Economics*, *84*(4), 1115–1129.

Bennett, M., & Willemain, T. (2001). Resistant estimation of multivariate location using minimum spanning trees. *Journal of Statistical Computation and Simulation*, *69*, 19–40.

Berndt, E.R., & L.R. Christensen (1973). The Translog Function and the Substitution of Equipment, Structures, and Labor in U.S. Manufacturing 1929-68. *Journal of Econometrics*, *1*, 81–114.

Biddle, J. (2012). Retrospectives: The Introduction of the Cobb–Douglas Regression. *Journal of Economic Perspectives*, *26*(2), 223–236. doi:10.1257/jep.26.2.223

Blanc, M., Cahuzac, E., Elyakime, B., & Tahar, G. (2008). Demand for on-farm permanent hired labour on family holdings. *European Review of Agricultural Economics*, *35*(4),493–518. doi:10.1093/erae/jbn032

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, *87*, 115–143.

Blundell, R., & Bond, S. (2000), "GMM Estimation with Persistent Panel Data: An Application to Production Functions", *Econometric Reviews*, *19*, 321–340.

Bond, S., & Söderbom, M. (2005). *Adjustment Costs and the Identification of Cobb Douglas Production Functions*. Nuffield College Economics Working Papers, No. 2005-W04, University of Oxford.

Bonnieux, F. (1989). Estimating regional-level input demand for French agriculture using a translog production function. *European Review of Agricultural Economics*, *16*, 229–241.

Brewster, J. M. (1950). The Machine Process in Agriculture and Industry. *Journal of Farm Economics*, *32*, 69–81.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press. http://doi.org/10.1016/S0304-4076(00)00050-6

Carletto, C., Savastano, S., & Zezza, A. (2011). *Fact or artefact : the impact of measurement errors on the farm size - productivity relationship.* World Bank Policy Research Working Paper No. WPS5908, Washington, DC.

Carter, M. R., & Wiebe, K. D. (1990). Access to Capital and Its Impact on Agrarian Structure and Productivity in Kenya. *American Journal of Agricultural Economics*, *72*(5), 1146–1150. http://doi.org/10.2307/1242523

Chamberlain, G. (1982). Multivariate Regression Models for Panel Data. *Journal of Econometrics*, *18*, 5–46.

Chambers, R.G. (1988), *Applied production analysis: A dual approach*, Cambridge: Cambridge University Press.

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, *41*(3), 1–58.

Chatterjee, S., & Hadi, A. S. (2006), *Regression Analysis by example,* 4th ed., Wiley, New York.

Ciaian, P., Falkowski, J., Kancs, D., & Pokrivcak, J. (2011). *Productivity and Credit Constraints: Firm-Level Evidence from Propensity Score Matching* Factor Markets Working Paper, No. 32, CEPS, Brussels.

Ciaian, P., Kancs d'A., Swinnen, J.F.M., Van Herck, K. & Vranken, L. (2012). *Institutional Factors Affecting Agricultural Land Markets,* Factor Markets Working Paper, No. 16, CEPS, Brussels.

Cobb, C.W., & Douglas, P.H. (1928). A Theory of Production. *American Economic Review*, *18*, 139–165.

Cooper, W., Seiford, L. M., Tone, K., & Zhu, J. (2007). Some models and measures for evaluating performances with dea: past accomplishments and future prospects. *Journal of Productivity Analysis*, *38*(3), 151–163.

Crosetto, P., & Filippin, A. (2012). *The "Bomb" Risk Elicitation Task.* SOEPapers on Multidisciplinary Panel Data Research No. 517, Berlin.

Curtiss, J. (2012). *Determinants of Financial Capital Use: Review of theories and implications for rural businesses*. Factor Markets Working Papers, No. 19, CEPS, Brussels.

Darpeix, A., Bignebat, C., & Perrier-Cornet, P. (2014). Demand for Seasonal Wage Labour in Agriculture: What Does Family Farming Hide? *Journal of Agricultural Economics*, *65*(1), 257–272. doi:10.1111/1477-9552.12019

Deaton, A. (1981). *Three essays on a Sri Lanka household survey.* Living Standards Measurement Study No. LSM11, Washington, DC.

Deaton, A. (1988). *Quality, quantity, and spatial variation of price : estimating price elasticities from cross sectional data.* Living Standards Measurement Study No. LSM30, Washington, DC.

Deolalikar, A. B., & Vijverberg, W. P. M. (1983). The heterogeneity of family and hired labor in agricultural production : a test using district level data from India. *Journal of economic development*, *8*(2), 45–69.

Deolalikar, A. B., & Vijverberg, W. P. M. (1987). A Test of Heterogeneity of Family and Hired Labour in Asian Agriculture. *Oxford Bulletin of Economics and Statistics*, *49*(3), 291–305. doi:10.1111/j.1468-0084.1987.mp49003003.x

Djuric, I. (2014). *Impact of policy measures on wheat-to-bread supply chain during the global commodity price peaks: The case of Serbia*. Studies on the Agricultural and Food sector in Transition Economies, Vol. 76. Halle (Saale): IAMO.

Dustmann, C., & Görlach, J. (2015). *The Economics of Temporary Migrations.* SOEPapers on Multidisciplinary Panel Data Research No. 729, Berlin.

Eastwood, R., Lipton, M., & Newell, A. (2010). Farm Size. In P. L. Pingali & R. E. Evenson (Eds.), *Handbook of Agricultural Economics. Volume 4* (pp. 3323–3397). Amsterdam: Elsevier.

European Commission (2008). *Farm Return Data Definitions*, Brussels.

European Commission. (2010a). *Farm Accounting Data Network: An A to Z of methodology*. Brussels. Retrieved from http://ec.europa.eu/agriculture/rica//pdf/site_en.pdf.

European Commission (2010b). *Rural Development in the EU: Statistical and Economic Information Report 2009*, Brussels.

European Commission. (2011). *Definitions of Variables used in FADN standard results*. Brussels.

European Commission. (2012). *Agriculture in the European Union. Satistical and economic information. Report 2012.* Luxembourg.

European Commission. (2013). *Agriculture in the European Union. Satistical and economic information. Report 2013.* Luxembourg.

European Commission. (2014). FADN - Field of Survey. Retrieved October 27, 2015, from http://ec.europa.eu/agriculture/rica/methodology1_en.cfm#tuof

Fahrmeir, L., Künstler, R., Pigeot, I., & Tutz, G. (2004). *Statistik : Der Weg zur Datenanalyse* (5., verb.). Berlin [u.a.]: Springer.

Food and Agriculture Organization (2009). *High food prices and the food crisis. Experiences and lessons learned,* The State of Agricultural Commodity Markets, FAO, Rome.

Frisvold, G. B. (1994). Does supervision matter? Some hypothesis tests using Indian farm-level data. *Journal of Development Economics*, *43*(2), 217–238. doi:10.1016/0304-3878(94)90005-1

Gandhi, A., Navarro, S., & Rivers, D. (2011). *On the identification of Production Functions: How Heterogeneous is Productivity?*. Madison, WI.

Glauben, T., Petrick, M., Tietje, H., & Weiss, C. (2009). Probability and timing of succession or closure in family firms: a switching regression analysis of farm households in Germany. *Applied Economics*, *41*, 45–54.

Greene, W. H. (2011). *Econometric analysis* (7. ed.). Boston, Mass. [u.a.]: Pearson.

Griliches, Z., & Hausman, J. (1986). Errors in Variables in Panel Data. *Journal of Econometrics*, *31*, 93-118.

Griliches, Z., & Mairesse, J. (1998). Production Functions: The Search for Identification. In S. Strom (Ed.), *Econometrics and economic theory in the 20th century* (pp. 169–203). Cambridge, MA: Cambridge University Press.

Grosh, M., & Munoz, J. (1996). *A Manual for Planning and Implementing the Living Standards Measurement Study Surve.* Living Standards Measurement Study No. LSM126, Washington.

Guastella, G., & Moro, D. (2013). Investment behaviour of EU arable crop farms in selected EU countries and the impact of policy reforms. Factor Markets Working Paper, No. 42, CEPS, Brussels.

Haisken-DeNew, J., & Frick, J. (2005). *Desktop Companion to the German Socio-Economic Panel*. Berlin.

Hampel, F., Ronchetti, E., Rousseeuw, P., & Stahel, W. (2005). *Robust Statistics: The Approach based on Influence Functions*. Wiley.

Hansen, L.P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, *50*(4), 1029–1054.

Hayami, Y. (2010). Plantation Agriculture. In P. L. Pingali & R. E. Evenson (Eds.), *Handbook of Agricultural Economics. Volume 4* (pp. 3305–3322). Amsterdam: North-Holland.

Hayami, Y., & Otsuka, K. (1993). *The economics of contract choice: An agrarian perspective*. Oxford: Clarendon Press.

Headey, B., Muffels, R., & Wagner, G. (2012). *Parents Transmit Happiness along with Associated Values and Behaviors to Their Children: A Lifelong Happiness Divided.* SOEPapers on Multidisciplinary Panel Data Research No. 492, Berlin.

Heady, E.O., & J.L. Dillon, J.L. (1961). *Agricultural Production Functions*. Ames, IA: Iowa State University Press.

Heikkilä, A.-M., Myrrä, S., & Pietola, K. (2012). *Effects of Economic Factors on Adoption of Robotics and Consequences of Automation for Productivity Growth*. Factor Markets Working Paper, No. 32, CEPS, Brussels.

Hempell, T. (2005). What's spurious, what's real? Measuring the productivity impacts of ICT at the firm-level. *Empirical Economics*, *30*, 427–464.

Hoch, I. (1955). Estimation of production function parameters and testing for efficiency. *Econometrica*, *23*, 325–326.

Huber, P., & Ronchetti, E. (2009). *Robust statistics*. John Wiley & Sons Inc.

Hübler, O. (2012). *Are Tall People Less Risk Averse than Others?* SOEPapers on Multidisciplinary Panel Data Research No. 457, Berlin.

Huettel, S., & Margarian, A. (2009). Structural change in the West German agricultural sector. *Agricultural Economics*. *40*, 759–772.

Huffman, W. E., & Orazem, P. F. (2007). Agricultural Development: Farmers, Farm Production and Farm Markets. In R. Evenson & P. Pingali (Eds.), *Handbook of Agricultural Economics* (Vol. 3, pp. 2281–2341). Amsterdam: Elsevier.

Ichimura, H., & Todd, P. (2007). Implementing Nonparametric and Semiparametric Estimators. In J. Heckman & E. Leamer (Eds.), *Handbook of Econometrics, Volume 6B* (pp. 5369–5468). Amsterdam: Elsevier.

Johnson, R. (1992). *Applied multivariate Statistical Analysis*. Prentice Hall.

Jungnickel (2008). *Graphs, networks and algorithms* (3rd edition). Springer.

Kasahara, H., & Rodrigue, J. (2008). Does the use of imported intermediates increase productivity? Plant-level evidence. *Journal of Development Economics*, *87*, 106–118.

Kazukauskas, A., Newman, C., & Thorne, F. (2010). Analysing the Effect of Decoupling on Agricultural Production: Evidence from Irish Dairy Farms using the Olley and Pakes Approach. *German Journal of Agricultural Economics*, *59*, 144–157.

Kemptner. (2013). *Health-Related Life Cycle Risks and Public Insurance.* SOEPapers on Multidisciplinary Panel Data Research No. 583, Berlin.

Kim, K. il, Petrin, A., & Song, S. (2016). Estimating production functions with control functions when capital is measured with error. *Journal of Econometrics*, *190*(2), 267–279. http://doi.org/10.1016/j.jeconom.2015.06.016

Kimhi, A. (2009). Demand for on-farm permanent hired labour on family holdings: a comment. *European Review of Agricultural Economics*, *36*(3), 447–452. doi:10.1093/erae/jbp024

Kirschstein, T., Liebscher, S., & Becker, C. (2013). Robust estimation of location and scatter by pruning the minimum spanning tree. *Journal of Multivariate Analysis*, *120*(0), 172–184.

Kleibergen, F., & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, *133*(1), 97–126. doi:10.1016/j.jeconom.2005.02.011

Kloss, M., Kirschstein, T., Liebscher, S., Petrick, M. (2015). Robust Productivity Analysis: An application to German FADN data. Italian Association of Agricultural and Applied Economics, Department of Economics of Roma Tre University, 6th EAAE PhD Workshop, Rom / Italien, 08.06.2015 - 10.06.2015.

Kloss, M., & Petrick, M. (2014). The Productivity of Family and Hired Labour in EU Arable Farming. In *German Association of Agricultural Economists (GEWISOLA): 54th Annual Conference, Goettingen, Germany, September 17-19, 2014*. Retrieved from http://ageconsearch.umn.edu//handle/187353

Lang, J. (2012). *The Aims of Lifelong Learning: Age-Related Effects of Training on Wages and Job Security.* SOEPapers on Multidisciplinary Panel Data Research No. 478, Berlin.

Latruffe, L., & Nauges, C. (2013). "Technical efficiency and conversion to organic farming: the case of France", *European Review of Agricultural Economics*, *41*(2), 227–253.

Levinsohn, J., & Petrin, A. (2003). Estimating Production Functions Using Inputs to Control for Unobservables. *Review of Economic Studies*, *70*(2), 317–341.

Liebscher, S., & Kirschstein, T. (2012). Identification of unbalanced warship designs using multivariate outlier detection procedures. *Milittary Operations Research*, *17*(1), 31–42.

Liebscher, S., & Kirschstein, T. (2014). Efficiency of the pmst and rdela location and scatter estimators. *Advances in Statistical Analysis*, 1-20.

Liebscher, S., & Kirschstein, T. (2015). restlos: Robust Estimation of Location and Scatter. R package Version 0.2-2. https://cran.r-project.org/web/packages/restlos/index.html

Liverpool-Tasie, L. S. O., Sanou, A., Ogunleye, W., & Omonona, B. T. (2015). *Is increasing inorganic fertilizer use in Sub-Saharan Africa a profitable proposition ? Evidence from Nigeria.* World Bank Policy Research Working Paper No. WPS7201, Washington, DC.

Maronna, R., Martin, D., & Yohai, V. (2006). *Robust Statistics: Theory and Methods*. John Wiley and Sons.

Marschak, J., & Andrews, W.H. (1944). Random Simultaneous Equations and the Theory of Production. *Econometrica*, 12, 143–205.

Mathijs, E., & Swinnen, J.F.M. (2001). Production Organization and Efficiency during Transition: An Empirical Analysis of East German Agriculture. *Review of Economics and Statistics*, *83*, 100–107.

Mundlak, Y. (1961). Empirical Production Function Free of Management Bias. *Journal of Farm Economics*, *43*(1). doi:10.2307/1235460

Mundlak, Y. (2001). Production and supply. In B.L. Gardner & G.C. Rausser (eds.), *Handbook of Agricultural Economics. Volume 1A* (pp. 3–85). *Agricultural Production.* Amsterdam: Elsevier.

Mundlak, Y., Butzer, R., & Larson, D.F. (2012). Heterogeneous technology and panel data: The case of the agricultural production function. *Journal of Development Economics*, *99*, 139–149.

Murphy, E., & Oesch, D. (2015). *The Feminization of Occupations and Change in Wages: A Panel Analysis of Britain, Germany and Switzerland,* SOEPapers on Multidisciplinary Panel Data Research No. 731, Berlin.

Obschonka, M., Schmitt-Rodermund, E., Silbereisen, R., Gosling, S., & Potter, J. (2013). *The Regional Distribution and Correlates of an Entrepreneurship-Prone Personality Profile in the United States, Germany, and the United Kingdom: A Socioecological Perspective.* SOEPapers on Multidisciplinary Panel Data Research No. 550, Berlin.

Olley, G. S., & Pakes, A. (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, *64*(6), 1263–1297. doi:10.2307/2171831

Olper, A., Raimondi, V., Vigani, M., & Cavicchioli, D. (2012). Does the Common Agricultural Policy Reduce Farm Labour Migration? Panel Data Analysis Across EU Regions. Factor Markets Working Paper, No. 28, CEPS, Brussels.

Oltmanns, E., Braakmann, A., & Schmidt, J. (2014). *Monitoring Subjective Well-Being: Some New Empirical Evidence for Germany.* SOEPpapers on Multidisciplinary Panel Data Research No. 696, Berlin.

Oseni, G., McGee, K., & Dabalen, A. (2014). *Can agricultural households farm their way out of poverty ?* World Bank Policy Research Working Paper No. WPS7093, Washington, DC.

Parker, J. (ed.) (2011). *The 9 billion-people question. A special report on feeding the world*. The Economist supplement, February 26.

Petrick, M. (2003). Sind Polens Landwirte kreditrationiert? Eine mikroökonomische Analyse von Marktversagen in Transformationsprozessen. *Proceedings "Schriften der Gesellschaft für Wirtschafts- und Sozialwissenschaften des Landbaues e.V."*, *38*, 173–183.

Petrick, M. (2004). *Credit rationing of Polish farm households – A theoretical and empirical analysis*. Halle (Saale): IAMO.

Petrick, M. (2005). Empirical measurement of credit rationing in agriculture: a methodological survey. *Agricultural Economics*, *33*, 191–203.

Petrick, M., & Kloss, M. (2013a). *Identifying Factor Productivity from Micro-data. The case of EU agriculture*. Factor Markets Working Papers, No. 34, CEPS, Brussels.

Petrick, M., & Kloss, M. (2013b). *Synthesis Report on the Impact of Capital Use*. Factor Markets Working Papers, No. 57, CEPS, Brussels.

Petrick, M., & Kloss, M. (2013c). Exposure of EU Farmers to the Financial Crisis. *Choices*, *28*(2), 1–6.

Petrick, M., & Kloss, M. (2013d). Factor Market Imperfections and Productivity in EU Agriculture. In J. Swinnen & L. Knops (Eds.), *Land, Labour and Capital Markets in European Agriculture: Diversity under a Common Policy* (pp. 316–333). Brussels: Centre for European Policy Studies.

Petrick, M. & Latruffe, L. (2006). Contractual relations in agricultural credit markets: a hedonic pricing approach with application to Poland. *Journal of Agricultural Economics*, *57*, 49–63.

Petrin, A., & Levinsohn, J. (2012). Measuring aggregate productivity growth using plant-level data. *RAND Journal of Economics*, *43*(4), 705–725. doi:10.1111/1756-2171.12005

Petrin, A., Poi, B.P., & Levinsohn, J. (2004). Production function estimation in Stata using inputs to control for unobservables. *Stata Journal*, 4, 113–123.

Pfeiffer, F., & Schulz, N. (2011). *Gregariousness, Interactive Jobs and Wages.* SOEPapers on Multidisciplinary Panel Data Research No. 363, Berlin.

Pitt, M. (1995). *Women's schooling, the selectivity of fertility, and child mortality in sub-Saharan Africa*. Living Standards Measurement Study No. LSM119, Washington, DC.

Pollak, R.A. (1985). A Transaction Cost Approach to Families and Households. *Journal of Economic Literature*, *23*, 581–608.

Rizov, M., Pokrivcak, J., & Ciaian, P. (2013). *CAP Subsidies and the Productivity of EU Farms*. Factor Markets Working Papers, No. 34, CEPS, Brussels.

Roodman, D. (2009). How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal*, *9*(1), pp. 86–136.

Rousseeuw, P. (1985). Multivariate estimation with high breakdown point. In W. Grossmann, G. Pflug, I. Vincze, & W. Wertz (Eds.), (pp. 283–297). Dordrecht: Reidel Publishing Company.

Rousseeuw, P., & Leroy, A. (1987). *Robust regression and outlier detection*. New York: John Wiley and Sons.

Schmitt, G. (1991). Why is the agriculture of advanced Western economies still organized by family farms? Will this continue to be so in the future? *European Review of Agricultural Economics*, *18*(3-4), 443–458. doi:10.1093/erae/18.3-4.443

Schmitt, M. (2013). *Subjective Well-Being and Air Quality in Germany.* SOEPapers on Multidisciplinary Panel Data Research No. 541, Berlin.

Schneck, S. (2011). *The Effect of Relative Standing on Considerations about Self- Employment.* SOEPapers on Multidisciplinary Panel Data Research No. 427, Berlin.

Schurer, S. (2015). *Lifecycle Patterns in the Socioeconomic Gradient of Risk Preferences.* SOEPapers on Multidisciplinary Panel Data Research No. 742, Berlin.

Sckokai, P. & Moro, D. (2009). Modelling the impact of the CAP Single Farm Payment on farm investment and output. *European Review of Agricultural Economics*, *36*, 395–423.

Semykina, A., & Wooldridge, J. M. (2010). Estimating panel data models in the presence of endogeneity and selection. *Journal of Econometrics*, *157*(2), 375–380. http://doi.org/10.1016/j.jeconom.2010.03.039

Shumway, C.R. (1995). Recent Duality Contributions in Production Economics. *Journal of Agricultural and Resource Economics*. *20*, 178–194.

Sial, M. H., & Carter, M. R. (1996). Financial market efficiency in an agrarian economy: Microeconometric analysis of the Pakistani Punjab. *The Journal of Development Studies*, *32*(5), 771–798. http://doi.org/10.1080/00220389608422439

Smith, A. (2010). *The Wealth of Nations: The Economics Classic - A selected edition for the contemporary reader.* (T. Butler-Bowdon, Ed.). London: Capstone Publishing Ltd. (Original work published 1776)

Söderbom, M. (2009). *md_ar1*, User-written Stata routine, University of Gothenburg, Gothenburg.

Sorgner, A., & Fritsch, M. (2013). *Occupational Choice and Self-Employment: Are They Related?* SOEPapers on Multidisciplinary Panel Data Research No. 533, Berlin.

StataCorp. 2011. *Stata Statistical Software: Release 12*. College Station, TX: StataCorp LP.

Swinnnen, J., & Knops, L. (Eds.). (2013). *Land, Labour and Capital Markets in European Agriculture: Diversity under a Common Policy*. Brussels: Centre for European Policy Studies.

Thijssen, G. (1994). Supply response and dynamic factor demand of Dutch dairy farms. *European Review of Agricultural Economics*, *21*, 241–258.

Tintner, G. (1944). A Note on the Derivation of Production Functions from Farm Records. *Econometrica*, *12*(1), 26–34. doi:10.2307/1905565

Tocco, B., Davidova, S., & Bailey, A. (2012). *Key Issues in Agricultural Labour Markets: A Review of Major Studies and Project Reports on Agriculture and Rural Labour Markets.* Factor Markets Working Paper, No. 20, CEPS, Brussels.

Tolley, H., Black, J, & Ezekiel, M. (1924). *Inputs as Related to Output.* Bulletin, No. 1277, U.S.Department of Agriculture, Washington, D.C.

Top agrar online (2016). Verbraucherpreise für Lebensmittel explodieren. Retrieved February 22, 2016, from http://www.topagrar.com/news/Markt-Marktnews-Verbraucherpreise-fuer-Lebensmittel-explodieren-2817797.html

Van Biesebroeck, J. (2007). Robustness of Productivity Estimates. *Journal of Industrial Economics*, *55*(3), 529–569. doi:10.1111/j.1467-6451.2007.00322.x

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics*, *126*, 25–51.

Witzke, H. P. (1993). *Mikrotheorie in der Agrarsektoranalyse. Neoklassischer Standard, Konflikte mit der Realität und Versöhnungsversuche mit Komplikationen*. Volkswirtschaftliche Schriften, No. 431, Berlin: Duncker & Humblot.

Wooldridge, J. M. (1995). Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics*, *68*(1), 115–132. http://doi.org/10.1016/0304-4076(94)01645-G

Wooldridge, J. M. (2006). *Introductory econometrics* (3. ed.). Mason, Ohio [u.a.]: Thomson South-Western.

Wooldridge, J. M. (2009). On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters*, *104*(3), 112–114. doi:10.1016/j.econlet.2009.04.026

Wooldridge, J.M. (2010). *Econometric analysis of cross section and panel data* (2nd. ed.). Cambridge, Mass: MIT Press.

Zhengfei, G., Lansink, A., Van Ittersum, M., & Wossink, A. (2006). Integrating Agronomic Principles into Production Function Specification: A Dichotomy of Growth Inputs and Facilitating Inputs. *American Journal of Agricultural Economics*, *88*(1), 203–214.

Zibrowius, M. (2012). *Convergence or Divergence?: Immigrant Wage Assimilation Patterns in Germany.* SOEPapers on Multidisciplinary Panel Data Research No. 479, Berlin.

APPENDICES

Appendix A: The identification of factor productivity: An application to EU Micro-data

Table A1: Descriptive statistics.

	Denmark				France				Germany (East)			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Output (ths EUR)	184.2	278.7	3.1	2733.4	158.6	115.9	5.2	1574.7	585.8	1054.1	11.5	9242.1
Labour (ths hours)	2.8	3.9	0.1	49.0	3.1	2.3	1.2	38.2	16.1	30.7	2.2	268.1
Land (ha)	124.0	174.3	10.1	1760.0	145.0	83.2	3.6	647.4	562.1	667.0	2.3	5155.9
Materials (ths EUR)	97.7	153.4	5.5	1683.4	85.2	56.1	4.7	618.6	350.9	653.8	13.9	6264.1
Capital (ths EUR)	872.2	1434.8	42.3	21381.0	158.9	127.4	2.8	1379.8	529.5	748.5	14.7	6591.7
Wage (EUR / hour)	17.4	0.0	17.4	17.4	10.6	0.6	9.3	14.1	9.4	0.9	8.1	10.4
Land rent (EUR / ha)	370.7	0.0	370.7	370.7	137.4	36.7	99.9	949.7	146.6	34.1	89.2	179.3
Interest on capital (%)	5.8	0.0	5.8	5.8	3.6	0.4	2.9	4.5	3.8	0.3	3.4	4.5
No. of observations	813				5321				1334			
No. of farms	209				1031				292			

Denmark Pattern	Denmark Frequency	France Pattern	France Frequency	Germany (East) Pattern	Germany (East) Frequency
...1111.	2	1111	79	1111	16
...11111	2	...1111.	12	...1111.	6
..1111..	31	...11111	44	...11111	4
.1111...	13	..1111..	23	..1111..	13
.11111..	27	..11111.	19	..11111.	35
.1111111	1	..111111	74	..111111	18
1111....	40	.1111...	16	.1111...	11
11111...	28	.11111..	17	.11111..	3
111111..	64	.111111.	8	.111111.	5
1111111.	1	.1111111	64	.1111111	14
		1111....	107	1111....	6
		11111...	90	11111...	43
		111111..	74	111111..	9
		1111111.	53	1111111.	55
		11111111	351	11111111	54

Notes: Descriptive statistics based on BLUNDELL/BOND estimation sample.
Source: Author based on FADN data

Descriptive statistics continued.

	Germany (West)				Italy				Spain			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Output (ths EUR)	153.9	140.6	14.8	2114.7	62.4	126.5	1.1	2165.2	41.7	37.2	0.0	768.1
Labour (ths hours)	4.2	3.4	1.1	93.9	3.6	4.6	0.0	98.7	2.7	1.5	0.1	23.9
Land (ha)	93.9	61.1	0.5	429.5	44.4	75.3	0.6	723.3	72.5	66.8	2.5	897.6
Materials (ths EUR)	85.1	68.3	11.9	737.5	23.9	52.3	0.5	938.4	15.3	14.1	0.8	237.3
Capital (ths EUR)	152.8	126.2	11.1	1008.0	122.6	230.6	2.7	4360.4	30.7	34.0	0.8	551.7
Wage (EUR / hour)	7.6	0.9	6.3	10.5	7.4	1.5	5.0	11.5	5.6	0.6	4.3	8.9
Land rent (EUR / ha)	254.8	57.5	63.4	314.3	187.0	95.7	61.1	500.0	126.7	68.7	1.8	1090.9
Interest on capital (%)	4.2	0.4	3.4	4.9	6.3	2.2	3.0	13.3	5.5	2.0	1.3	10.3
No. of observations*	2977				4890				7807			
No. of farms*	573				1362				1400			

Germany (West) Pattern	Frequency	Italy Pattern	Frequency	Spain Pattern	Frequency
....1111	55	1111	94	1111	31
...1111.	5	...1111.	410	...1111.	15
...11111	28	...11111	667	...11111	56
..1111..	6	..1111..	11	..1111..	8
..11111.	4	..11111.	10	..11111.	6
..111111	43	..111111	16	..111111	42
.1111...	16	.1111...	5	.1111...	9
.11111..	17	.11111..	5	.11111..	9
.111111.	8	.111111.	4	.111111.	10
.1111111	77	.1111111	7	.1111111	28
1111....	32	1111....	37	1111....	80
11111...	38	11111...	27	11111...	139
111111..	29	111111..	30	111111..	166
1111111.	27	1111111.	19	1111111.	216
11111111	188	11111111	20	11111111	585

Notes: Descriptive statistics based on BLUNDELL/BOND estimation sample.* Except for wage, land rent and interest on capital in Spain.

Source: Author based on FADN data.

Descriptive statistics continued.

	United Kingdom			
	Mean	SD	Min	Max
Output (ths EUR)	282.9	343.2	8.7	3548.6
Labour (ths hours)	6.2	5.2	0.3	51.8
Land (ha)	250.5	182.4	17.8	1178.5
Materials (ths EUR)	155.8	151.7	10.4	1475.8
Capital (ths EUR)	237.7	215.2	10.0	1522.9
Wage (EUR / hour)	10.9	0.3	9.1	11.2
Land rent (EUR ha)	197.4	13.6	172.9	206.9
Interest on capital (%)	4.7	0.3	4.2	5.2
No. of observations	800			
No. of farms	189			

Pattern	Frequency
....1111	34
...1111.	8
...11111	29
..1111..	10
..11111.	7
..111111	17
.1111...	14
.11111..	14
.111111.	9
.1111111	23
1111....	3
11111...	4
111111..	1
1111111.	1
11111111	15

Notes: Descriptive statistics based on BLUNDELL/BOND estimation sample.
Source: Author based on FADN data

Table A2: Results production function estimations, Denmark.

	Output shares		**OLS** Cobb Douglas		**OLS** Translog		**Within** Cobb Douglas		**Within** Translog		**Levinsohn/Petrin** Cobb Douglas		**Wooldridge/ Levinsohn/Petrin** Cobb Douglas		**Wooldridge/ Levinsohn/Petrin** Translog	
	Mean	SD	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Labour	0.476***	0.313	0.482***	0.053	0.888**	0.376	0.207***	0.047	0.196***	0.029	0.472***	0.053	0.617***	0.091	0.548	1.440
Land	0.412***	0.230	0.273***	0.060	0.606	0.491	0.287***	0.066	0.282***	0.041	0.267***	0.054	0.232***	0.080	0.163	0.988
Materials	0.716***	0.402	0.495***	0.047	0.332	0.509	0.481***	0.047	0.477***	0.027	0.365	0.226	-0.001	0.269	0.132	2.208
Capital	0.469***	0.367	0.134***	0.041	0.722*	0.413	-0.023	0.042	-0.001	0.026	0.106	0.068	0.096	0.075	0.610	1.282
N	813		813		813		813		813		813		605		605	
Elasticity of scale			1.384***	0.026			0.952***	0.072			1.209***	0.234	0.945***	0.229		
p-value const. ret. to scale			<0.001				0.504				<0.001		0.809			
R^2			0.948		0.957		0.564		0.562							
p-value coeff. jointly zero			<0.001		<0.001		<0.001		<0.001		<0.001		<0.001		<0.001	
p-value interact. terms jtly. zero					<0.001				0.315						0.824	

Notes: Year dummies included in all models. *** (**, *) significant at the 1% (5%, 10%) level, based on standard errors robust to clustering in groups. Standard errors in Levinsohn/Petrin based on bootstrapping with 20 replications.

Source: Author.

Table A3: Results production function estimations, France.

	Output shares		**OLS** Cobb Douglas		**OLS** Translog		**Within** Cobb Douglas		**Within** Translog		**Levinsohn/Petrin** Cobb Douglas		**Wooldridge/ Levinsohn/Petrin** Cobb Douglas		**Wooldridge/ Levinsohn/Petrin** Translog	
	Mean	SD	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Labour	0.243***	0.171	0.169***	0.015	-0.232	0.336	0.109***	0.031	0.106***	0.015	0.181***	0.020	0.168***	0.018	-0.325	0.494
Land	0.134***	0.063	0.060***	0.017	0.284	0.209	0.341**	0.066	0.345***	0.030	0.048***	0.015	0.041**	0.017	0.250	0.267
Materials	0.559***	0.181	0.749***	0.018	1.323***	0.290	0.547***	0.038	0.500***	0.017	0.716***	0.057	0.804***	0.084	0.794	0.894
Capital	0.037***	0.025	0.158***	0.012	0.066	0.167	0.038***	0.012	0.033***	0.006	0.114***	0.014	0.119***	0.015	-0.632	0.612
N	5321		5321		5321		5321		5321		5321		4289		4289	
Elasticity of scale			1.136***	0.014			1.035***	0.058			1.059***	0.058	1.132***	0.077		
p-value const. ret. to scale			<0.001				0.549				0.309		0.086			
R^2			0.865		0.873		0.500		0.485							
p-value coeff. jointly zero			<0.001		<0.001		<0.001		<0.001		<0.001		<0.001		<0.001	
p-value interact. terms jtly. zero					0.132				0.004				0.086		0.397	

Notes: Year dummies included in all models. *** (**, *) significant at the 1% (5%, 10%) level, based on standard errors robust to clustering in groups. Standard errors in Levinsohn/Petrin based on bootstrapping with 20 replications.

Source: Author.

Table A4: Results production function estimations, Germany (East).

	Output shares		**OLS** Cobb Douglas		**OLS** Translog		**Within** Cobb Douglas		**Within** Translog		**Levinsohn/Petrin** Cobb Douglas		**Wooldridge/ Levinsohn/Petrin** Cobb Douglas		**Wooldridge/ Levinsohn/Petrin** Translog	
	Mean	SD	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Labour	0.306***	0.242	-0.011	0.045	-0.025	0.278	0.053*	0.030	0.043***	0.016	0.065	0.059	0.036	0.051	-0.400	0.422
Land	0.196***	0.102	0.056	0.064	1.655***	0.264	0.371***	0.064	0.468***	0.032	0.028	0.064	-0.028	0.055	1.491***	0.407
Materials	0.631***	0.201	0.905***	0.040	0.283	0.372	0.539***	0.043	0.531***	0.024	0.840***	0.164	1.077***	0.170	0.136	1.344
Capital	0.050***	0.036	0.122***	0.028	-0.223	0.277	0.016	0.024	-0.034***	0.012	0.132	0.121	0.083*	0.047	0.013	0.461
N	1334		1334		1334		1334		1334		1334		1047		1047	
Elasticity of scale			1.072***	0.016			0.979***	0.060			1.065***	0.234	1.168***	0.140		
p-value const. ret. to scale			<0.001				0.729				0.781		0.229			
R^2			0.950		0.959		0.552		0.516							
p-value coeff. jointly zero			<0.001		<0.001		<0.001		<0.001		<0.001		<0.001		<0.001	
p-value interact. terms jtly. zero					<0.001				0.003						0.979	

Notes: Year dummies included in all models. *** (**, *) significant at the 1% (5%, 10%) level, based on standard errors robust to clustering in groups. Standard errors in LEVINSOHN/PETRIN based on bootstrapping with 20 replications.

Source: Author.

Table A5: Results production function estimations, Germany (West).

	Output shares		OLS Cobb Douglas		OLS Translog		Within Cobb Douglas		Within Translog		Levinsohn/Petrin Cobb Douglas		Wooldridge/ Levinsohn/Petrin Cobb Douglas		Wooldridge/ Levinsohn/Petrin Translog	
	Mean	SD	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Labour	0.256***	0.162	0.208***	0.023	-0.667*	0.379	0.082***	0.026	0.080***	0.013	0.222***	0.022	0.222***	0.026	-1.066**	0.491
Land	0.177***	0.102	0.025	0.017	1.360***	0.315	0.273***	0.050	0.293***	0.023	0.022*	0.012	-0.005	0.019	1.046***	0.329
Materials	0.583***	0.173	0.799***	0.023	0.904**	0.363	0.476***	0.026	0.469***	0.013	0.652***	0.048	0.770***	0.086	1.848**	0.810
Capital	0.045***	0.031	0.120***	0.018	0.203	0.323	0.044***	0.015	0.038***	0.007	0.155***	0.029	0.088***	0.024	-0.007	0.438
N	2977		2977		2977		2977		2977		2977		2408		2408	
Elasticity of scale			1.152***	0.023			0.875***	0.060			1.051***	0.048	1.075***	0.080		
p-value const. ret. to scale			<0.001				0.038				0.283		0.344			
R^2			0.854		0.865		0.332		0.331							
p-value coeff. jointly zero			<0.001		<0.001		<0.001		<0.001		<0.001		<0.001		<0.001	
p-value interact. terms jtly. zero					0.008				0.350						0.171	

Notes: Year dummies included in all models. *** (**, *) significant at the 1% (5%, 10%) level, based on standard errors robust to clustering in groups. Standard errors in Levinsohn/Petrin based on bootstrapping with 20 replications.

Source: Author.

Table A6: Results production function estimations, Italy.

	Output shares		OLS Cobb Douglas		OLS Translog		Within Cobb Douglas		Within Translog		Levinsohn/Petrin Cobb Douglas		Wooldridge/ Levinsohn/Petrin Cobb Douglas		Wooldridge/ Levinsohn/Petrin Translog	
	Mean	SD	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Labour	0.866***	0.917	0.308***	0.018	-0.530***	0.199	0.094***	0.019	0.100***	0.013	0.289***	0.014	0.318***	0.025	-0.300	0.424
Land	0.153***	0.182	0.001	0.017	-0.021	0.145	0.352***	0.056	0.335***	0.030	-0.008	0.016	-0.006	0.022	-0.031	0.294
Materials	0.404***	0.232	0.675***	0.021	1.345***	0.191	0.443***	0.028	0.445***	0.017	0.514***	0.040	0.508***	0.108	1.033	0.637
Capital	0.188***	0.249	0.098***	0.013	-0.111	0.147	-0.012	0.026	-0.037**	0.016	0.018	0.055	0.016	0.031	-0.407	0.376
N	4890		4890		4890		4890		4890		4890		3545		3545	
Elasticity of scale			1.082***	0.013			0.877***	0.059			0.813***	0.075	0.836***	0.094		
p-value const. ret. to scale			<0.001				0.037				0.012		0.081			
R^2			0.843		0.855		0.327		0.328							
p-value coeff. jointly zero			<0.001		<0.001		<0.001		<0.001		<0.001		<0.001		<0.001	
p-value interact. terms jtly. zero					<0.001				0.012						0.138	

Notes: Year dummies included in all models. *** (**, *) significant at the 1% (5%, 10%) level, based on standard errors robust to clustering in groups. Standard errors in LEVINSOHN/PETRIN based on bootstrapping with 20 replications.

Source: Author.

Table A7: Results production function estimations, Spain.

	Output shares		OLS Cobb Douglas		OLS Translog		Within Cobb Douglas		Within Translog		Levinsohn/Petrin Cobb Douglas		Wooldridge/ Levinsohn/Petrin Cobb Douglas		Wooldridge/ Levinsohn/Petrin Translog	
	Mean	SD	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Labour	0.589***	2.595	0.378***	0.023	-0.959***	0.363	0.183***	0.024	0.188***	0.011	0.372***	0.019	0.417***	0.028	-1.288**	0.577
Land	0.320***	1.351	0.102***	0.015	-1.022***	0.250	0.320***	0.044	0.331***	0.020	0.098***	0.018	0.080***	0.017	-1.353***	0.300
Materials	0.452***	0.522	0.663***	0.016	2.013***	0.273	0.510***	0.024	0.525***	0.011	0.562***	0.070	0.727***	0.060	2.520***	0.529
Capital	0.060***	0.148	0.017	0.013	0.716***	0.230	0.010	0.021	-0.020**	0.010	0.013	0.041	0.075***	0.027	1.001***	0.319
N	7807		7807		7807		7807		7807		7807		6393		6393	
Elasticity of scale			1.161***	0.024			1.022***	0.053			1.045***	0.059	1.299***	0.063		
p-value const. ret. to scale			<0.001				0.676				0.449		<0.001			
R^2			0.699		0.709		0.453		0.441							
p-value coeff. jointly zero			<0.001		<0.001		<0.001		<0.001		<0.001		<0.001		<0.001	
p-value interact. terms jtly. zero					0.907				0.083						0.657	

Notes: Year dummies included in all models. *** (**, *) significant at the 1% (5%, 10%) level, based on standard errors robust to clustering in groups. Standard errors in LEVINSOHN/PETRIN based on bootstrapping with 20 replications.

Source: Author.

Table A8: Results production function estimations, United Kingdom.

	Output shares		**OLS** Cobb Douglas		**OLS** Translog		**Within** Cobb Douglas		**Within** Translog		**Levinsohn/Petrin** Cobb Douglas		**Wooldridge/ Levinsohn/Petrin** Cobb Douglas		**Wooldridge/ Levinsohn/Petrin** Translog	
	Mean	SD	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Labour	0.293***	0.179	0.203***	0.039	0.158	0.320	0.242***	0.055	0.209***	0.029	0.187***	0.048	0.191***	0.050	1.230	1.138
Land	0.204***	0.097	0.165***	0.046	1.575***	0.403	0.387***	0.105	0.392***	0.053	0.158***	0.031	0.174***	0.061	2.383*	1.388
Materials	0.623***	0.227	0.729***	0.046	-0.096	0.442	0.594***	0.072	0.524***	0.041	0.731***	0.123	0.622	0.399	-0.230	5.644
Capital	0.043***	0.030	0.077**	0.032	-0.361	0.253	0.017	0.030	-0.017	0.016	0.106	0.066	0.099	0.065	-1.261	2.035
N	800		800		800		800		800		800		612		612	
Elasticity of scale			1.174***	0.027			1.240***	0.093			1.182***	0.156	1.086***	0.380		
p-value const. ret. to scale			<0.001				0.010				0.244		0.820			
R^2			0.903		0.909		0.578		0.567							
p-value coeff. jointly zero			<0.001		<0.001		<0.001		<0.001		<0.001		<0.001		<0.001	
p-value interact. terms jtly. zero					0.732				0.003						0.921	

Notes: Year dummies included in all models. *** (**, *) significant at the 1% (5%, 10%) level, based on standard errors robust to clustering in groups. Standard errors in LEVINSOHN/PETRIN based on bootstrapping with 20 replications.

Source: Author.

Table A9: Results Blundell/Bond Cobb Douglas estimator Denmark.

	Production function estimates	
Unrestricted model	Coeff	SE
Labour	0.241	0.248
- lagged	-0.064	0.186
Land	0.334*	0.193
- lagged	-0.018	0.172
Materials	0.460***	0.171
- lagged	0.015	0.188
Capital	-0.014	0.150
- lagged	0.196	0.190
Output lagged	0.093	0.128
p-val. coeff. jointly zero	<0.001	
Arellano-Bond test (1)	<0.001	
Arellano-Bond test (2)	0.494	
Hansen OID test	0.070	
Restricted model		
Labour	0.260	0.225
Land	0.309*	0.162
Materials	0.464***	0.166
Capital	0.031	0.129
ρ	0.160***	0.056
Elasticity of scale	1.020***	0.211
Common factors	0.986	

	Diagnosis of model specification									
	Labour		Land		Materials		Capital		Output	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
AR(1) model	0.161	0.31	0.737	0.58	0.634*	0.35	0.515	0.60	0.168	0.52
Instruments differences										
p-value coeff. jointly zero	<0.001		0.987		0.309		0.172		0.601	
R^2	0.305		<0.001		0.026		0.038		0.011	
Instruments levels										
p-value coeff. jointly zero	0.005		0.738		0.356		0.039		0.060	
R^2	0.111		0.007		0.023		0.069		0.061	

Notes: N=813. *** (**, *) significant at the 1% (5%, 10%) level. Year dummies included in production function and AR(1) models. In production function and AR(1) models, lags of order two back to the maximum possible are used as GMM-type instruments for the lagged dependent variable in the differenced equation using the two-step procedure. Lagged differences used as GMM-type instruments for the lagged dependent variable in the level equation. First differences of year dummies used as standard instruments in the production function. Standard errors adjusted using WINDMEIJER (2005) procedure. Minimum distance estimation due to SÖDERBOM (2009). “Instruments differences” based on regression of first difference for year=2006 on lagged levels three and four years back. “Instruments levels” based on regression of lagged level for year=2006 on lagged first differences two and three years back, using OLS.

Source: Author.

Table A10: Results Blundell/Bond Cobb Douglas estimator France.

	Production function estimates	
Unrestricted model	Coeff	SE
Labour	0.028	0.120
- lagged	0.003	0.095
Land	0.001	0.162
- lagged	-0.073	0.162
Materials	0.960***	0.141
- lagged	-0.073	0.147
Capital	-0.024	0.077
- lagged	0.116**	0.058
Output lagged	0.123	0.082
p-val. coeff. jointly zero	<0.001	
Arellano-Bond test (1)	<0.001	
Arellano-Bond test (2)	0.841	
Hansen OID test	<0.001	
Restricted model		
Labour	0.068	0.068
Land	0.008	0.092
Materials	0.926***	0.090
Capital	-0.034	0.073
ρ	0.225***	0.050
Elasticity of scale	0.966	0.157
Common factors	0.585	

	Diagnosis of model specification									
	Labour		Land		Materials		Capital		Output	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
AR(1) model	-0.062	0.34	0.887***	0.25	0.922***	0.18	0.859***	0.11	0.965***	0.25
Instruments differences										
p-value coeff. jointly zero	0.641		0.399		0.448		0.035		0.108	
R^2	0.010		0.015		0.014		0.034		0.026	
Instruments levels										
p-value coeff. jointly zero	<0.001		0.272		0.004		<0.001		<0.001	
R^2	0.119		0.018		0.048		0.176		0.096	

Notes: N=5321. *** (**, *) significant at the 1% (5%, 10%) level. Year dummies included in production function and AR(1) models. In production function and AR(1) models, lags of order two back to the maximum possible are used as GMM-type instruments for the lagged dependent variable in the differenced equation using the two-step procedure. Lagged differences used as GMM-type instruments for the lagged dependent variable in the level equation. First differences of year dummies used as standard instruments in the production function. Standard errors adjusted using Windmeijer (2005) procedure. Minimum distance estimation due to Söderbom (2009). "Instruments differences" based on regression of first difference for year=2008 on lagged levels three to seven years back. "Instruments levels" based on regression of lagged level for year=2008 on lagged first differences two to six years back, using OLS.

Source: Author.

Table A11: Results Blundell/Bond Cobb Douglas estimator Germany (East).

	Production function estimates	
Unrestricted model	Coeff	SE
Labour	0.082	0.118
- lagged	-0.009	0.091
Land	0.211	0.184
- lagged	-0.270*	0.145
Materials	0.501***	0.107
- lagged	0.237**	0.114
Capital	0.022	0.093
- lagged	0.036	0.077
Output lagged	0.217***	0.080
p-val. coeff. jointly zero	<0.001	
Arellano-Bond test (1)	<0.001	
Arellano-Bond test (2)	0.513	
Hansen OID test	0.100	
Restricted model		
Labour	0.111	0.093
Land	0.232*	0.124
Materials	0.507***	0.104
Capital	0.006	0.080
ρ	0.353***	0.038
Elasticity of scale	0.816***	0.199
Common factors	0.411	

	Diagnosis of model specification									
	Labour		Land		Materials		Capital		Output	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
AR(1) model	0.978***	0.02	0.934***	0.08	0.941***	0.29	0.290	0.56	1.282***	0.36
Instruments differences										
p-value coeff. jointly zero	0.111		0.350		0.044		0.076		0.207	
R^2	0.165		0.107		0.241		0.183		0.135	
Instruments levels										
p-value coeff. jointly zero	<0.001		0.412		0.683		0.645		0.001	
R^2	0.567		0.097		0.074		0.066		0.358	

Notes: N=1334. *** (**, *) significant at the 1% (5%, 10%) level. Year dummies included in production function and AR(1) models. In production function and AR(1) models, lags of order two back to the maximum possible are used as GMM-type instruments for the lagged dependent variable in the differenced equation using the two-step procedure. Lagged differences used as GMM-type instruments for the lagged dependent variable in the level equation. First differences of year dummies used as standard instruments in the production function. Standard errors adjusted using Windmeijer (2005) procedure. Minimum distance estimation due to Söderbom (2009). "Instruments differences" based on regression of first difference for year=2008 on lagged levels three to seven years back. "Instruments levels" based on regression of lagged level for year=2008 on lagged first differences two to six years back, using OLS.

Source: Author.

Table A12: Results Blundell/Bond Cobb Douglas estimator Germany (West).

	Production function estimates	
Unrestricted model	Coeff	SE
Labour	-0.006	0.118
- lagged	0.083	0.103
Land	-0.042	0.202
- lagged	0.015	0.183
Materials	0.681***	0.087
- lagged	0.209*	0.115
Capital	0.094	0.103
- lagged	0.026	0.095
Output lagged	0.047	0.091
p-val. coeff. jointly zero	<0.001	
Arellano-Bond test (1)	<0.001	
Arellano-Bond test (2)	0.104	
Hansen OID test	<0.001	
Restricted model		
Labour	-0.007	0.103
Land	-0.045	0.072
Materials	0.671***	0.079
Capital	0.103**	0.049
ρ	0.092***	0.013
Elasticity of scale	0.727***	0.234
Common factors	0.991	

Diagnosis of model specification	Labour		Land		Materials		Capital		Output	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
AR(1) model	0.860***	0.12	0.929***	0.04	0.865***	0.13	1.042***	0.06	1.213***	0.11
Instruments differences										
p-value coeff. jointly zero	0.171		0.388		0.315		0.556		0.760	
R^2	0.041		0.028		0.032		0.021		0.014	
Instruments levels										
p-value coeff. jointly zero	<0.001		<0.001		<0.001		<0.001		<0.001	
R^2	0.281		0.179		0.154		0.232		0.171	

Notes: N=2977. *** (**, *) significant at the 1% (5%, 10%) level. Year dummies included in production function and AR(1) models. In production function and AR(1) models, lags of order two back to the maximum possible are used as GMM-type instruments for the lagged dependent variable in the differenced equation using the two-step procedure. Lagged differences used as GMM-type instruments for the lagged dependent variable in the level equation. First differences of year dummies used as standard instruments in the production function. Standard errors adjusted using Windmeijer (2005) procedure. Minimum distance estimation due to Söderbom (2009). "Instruments differences" based on regression of first difference for year=2008 on lagged levels three to seven years back. "Instruments levels" based on regression of lagged level for year=2008 on lagged first differences two to six years back, using OLS.

Source: Author.

Table A13: Results Blundell/Bond Cobb Douglas estimator Italy.

	Production function estimates	
Unrestricted model	Coeff	SE
Labour	-0.072	0.060
- lagged	0.263***	0.078
Land	-0.218	0.189
- lagged	0.033	0.176
Materials	0.665***	0.089
- lagged	0.010	0.082
Capital	0.148	0.122
- lagged	-0.037	0.105
Output lagged	0.156**	0.062
p-val. coeff. jointly zero	<0.001	
Arellano-Bond test (1)	<0.001	
Arellano-Bond test (2)	0.401	
Hansen OID test	<0.001	
Restricted model		
Labour	-0.066	0.056
Land	-0.199**	0.080
Materials	0.621***	0.077
Capital	0.186***	0.071
ρ	0.257***	0.030
Elasticity of scale	0.522***	0.229
Common factors	0.428	

	Diagnosis of model specification									
	Labour		Land		Materials		Capital		Output	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
AR(1) model	0.726***	0.20	1.007***	0.05	0.770	0.32	0.985***	0.02	1.242**	0.50
Instruments differences										
p-value coeff. jointly zero	0.086		0.839		0.180		0.617		0.873	
R^2	0.466		0.126		0.433		0.205		0.111	
Instruments levels										
p-value coeff. jointly zero	0.175		0.275		0.561		0.007		0.712	
R^2	0.393		0.337		0.254		0.651		0.173	

Notes: N=4890. *** (**, *) significant at the 1% (5%, 10%) level. Year dummies included in production function and AR(1) models. In production function and AR(1) models, lags of order two back to the maximum possible are used as GMM-type instruments for the lagged dependent variable in the differenced equation using the two-step procedure. Lagged differences used as GMM-type instruments for the lagged dependent variable in the level equation. First differences of year dummies used as standard instruments in the production function. Standard errors adjusted using Windmeijer (2005) procedure. Minimum distance estimation due to Söderbom (2009). "Instruments differences" based on regression of first difference for year=2008 on lagged levels three to seven years back. "Instruments levels" based on regression of lagged level for year=2008 on lagged first differences two to six years back, using OLS.

Source: Author.

Table A14: Results Blundell/Bond Cobb Douglas estimator Spain.

Production function estimates		
Unrestricted model	Coeff	SE
Labour	0.156	0.145
- lagged	0.619***	0.142
Land	-0.496**	0.206
- lagged	0.460**	0.199
Materials	1.134***	0.107
- lagged	-0.355***	0.102
Capital	0.017	0.179
- lagged	0.014	0.181
Output lagged	-0.048	0.045
p-val. coeff. jointly zero	<0.001	
Arellano-Bond test (1)	<0.001	
Arellano-Bond test (2)	<0.001	
Hansen OID test	<0.001	
Restricted model		
Labour	0.189	0.123
Land	-0.463***	0.105
Materials	1.128***	0.097
Capital	0.018	0.063
ρ	-0.094***	0.009
Elasticity of scale	0.811***	0.259
Common factors	0.878	

Diagnosis of model specification	Labour		Land		Materials		Capital		Output	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
AR(1) model	0.951***	0.23	0.881***	0.11	0.732***	0.08	0.997***	0.02	0.353**	0.15
Instruments differences										
p-value coeff. jointly zero	0.099		0.001		0.004		0.007		<0.001	
R^2	0.016		0.037		0.030		0.027		0.060	
Instruments levels										
p-value coeff. jointly zero	<0.001		<0.001		<0.001		<0.001		<0.001	
R^2	0.194		0.051		0.105		0.167		0.040	

Notes: N=7807. *** (**, *) significant at the 1% (5%, 10%) level. Year dummies included in production function and AR(1) models. In production function and AR(1) models, lags of order two back to the maximum possible are used as GMM-type instruments for the lagged dependent variable in the differenced equation using the two-step procedure. Lagged differences used as GMM-type instruments for the lagged dependent variable in the level equation. First differences of year dummies used as standard instruments in the production function. Standard errors adjusted using Windmeijer (2005) procedure. Minimum distance estimation due to Söderbom (2009). "Instruments differences" based on regression of first difference for year=2008 on lagged levels three to seven years back.

Source: Author.

Table A15: Results Blundell/Bond Cobb Douglas estimator United Kingdom.

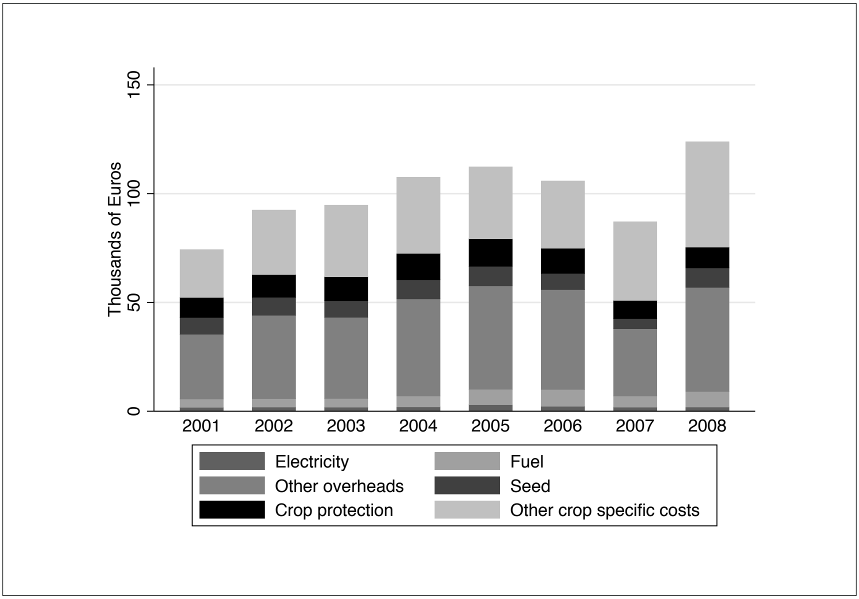
	Production function estimates	
Unrestricted model	Coeff	SE
Labour	0.269**	0.123
- lagged	-0.123	0.127
Land	0.498*	0.282
- lagged	-0.243	0.269
Materials	0.664***	0.136
- lagged	-0.176	0.181
Capital	0.016	0.070
- lagged	-0.002	0.064
Output lagged	0.225**	0.116
p-val. coeff. jointly zero	<0.001	
Arellano-Bond test (1)	<0.001	
Arellano-Bond test (2)	0.943	
Hansen OID test	0.708	
Restricted model		
Labour	0.241**	0.095
Land	0.389***	0.120
Materials	0.640***	0.093
Capital	0.020	0.059
ρ	0.297***	0.080
Elasticity of scale	1.448***	0.217
Common factors	0.929	

Diagnosis of model specification	Labour		Land		Materials		Capital		Output	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
AR(1) model	0.812***	0.26	0.929***	0.06	1.166***	0.22	0.902***	0.13	1.194***	0.23
Instruments differences										
p-value coeff. jointly zero	0.379		0.646		0.274		0.653		0.355	
R^2	0.402		0.276		0.494		0.273		0.414	
Instruments levels										
p-value coeff. jointly zero	0.186		0.206		0.447		0.194		0.152	
R^2	0.516		0.502		0.399		0.511		0.543	

Notes: N=807. *** (**, *) significant at the 1% (5%, 10%) level. Year dummies included in production function and AR(1) models. In production function and AR(1) models, lags of order two back to the maximum possible are used as GMM-type instruments for the lagged dependent variable in the differenced equation using the two-step procedure. Lagged differences used as GMM-type instruments for the lagged dependent variable in the level equation. First differences of year dummies used as standard instruments in the production function. Standard errors adjusted using Windmeijer (2005) procedure. Minimum distance estimation due to Söderbom (2009). "Instruments differences" based on regression of first difference for year=2008 on lagged levels three to seven years back. "Instruments levels" based on regression of lagged level for year=2008 on lagged first differences two to six years back, using OLS.

Source: Author.

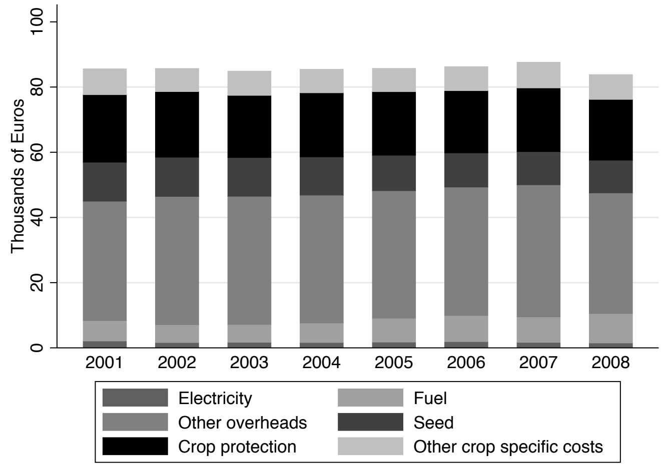
Figure A1: Evolution of materials input over observed sample period (Denmark).



Notes: Breakdown of average materials input use into its different factors. Smallest possible split available in the FADN data.

Source: Author based on FADN data.

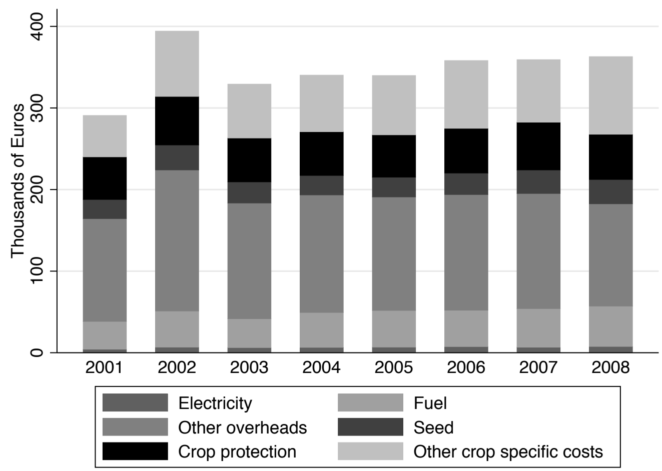
Figure A2: Evolution of materials input over observed sample period (France).



Notes: Breakdown of average materials input use into its different factors. Smallest possible split available in the FADN data.

Source: Author based on FADN data.

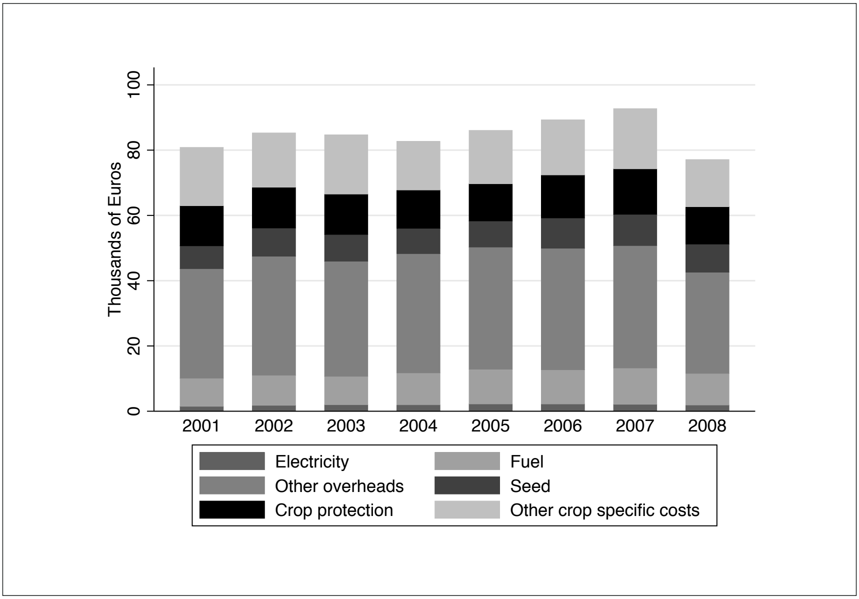
Figure A3: Evolution of materials input over observed sample period (East Germany).



Notes: Breakdown of average materials input use into its different factors. Smallest possible split available in the FADN data.

Source: Author based on FADN data.

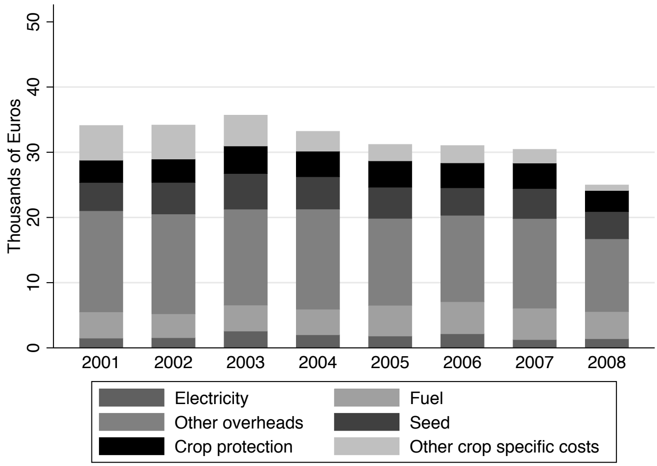
Figure A4: Evolution of materials input over observed sample period (West Germany).



Notes: Breakdown of average materials input use into its different factors. Smallest possible split available in the FADN data.

Source: Author based on FADN data.

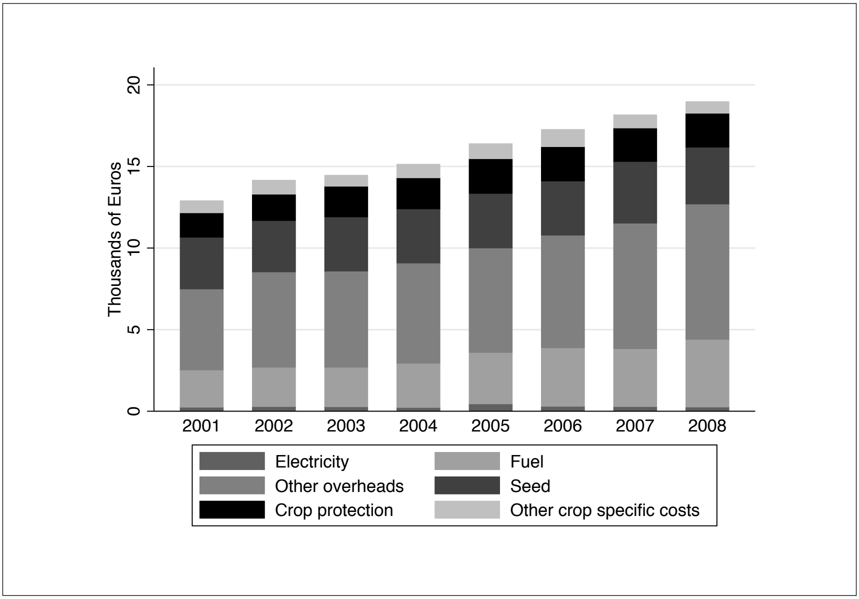
Figure A5: Evolution of materials input over observed sample period (Italy).



Notes: Breakdown of average materials input use into its different factors. Smallest possible split available in the FADN data.

Source: Author based on FADN data.

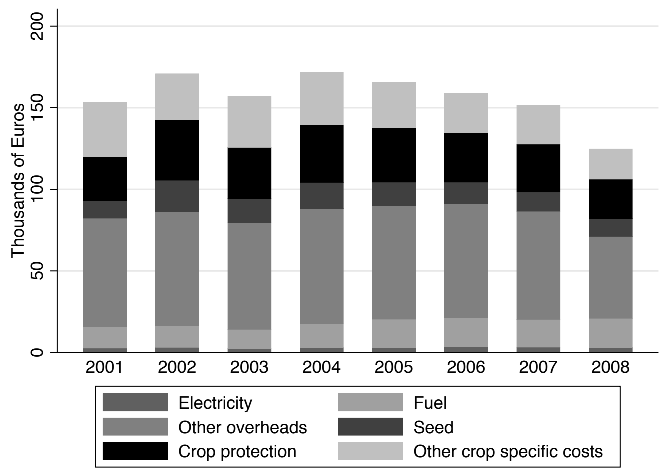
Figure A6: Evolution of materials input over observed sample period (Spain).



Notes: Breakdown of average materials input use into its different factors. Smallest possible split available in the FADN data.

Source: Author based on FADN data.

Figure A7: Evolution of materials input over observed sample period (United Kingdom).



Notes: Breakdown of average materials input use into its different factors. Smallest possible split available in the FADN data.

Source: Author based on FADN data.

Appendix B: Outlier Robust productivity analysis: An application to German FADN data

The following illustrations depict a repeated analysis for a materials specification that incorporates inputs.

Table B1: Multivariate Outliers divided into small and large farms.

	Number of Observations	Univariate Outliers	Multivariate Outliers			
			small farms	large farms	neither nor	$\sum$
East Germany	3791	375	85	6	236	327
West Germany	8691	1554	292	250	701	1243

Notes: Outliers following an analysis incorporating a materials specification that includes fertiliser inputs. Number of observations belonging to respective farm category according to the 'FDH' procedure.

Source: Author.

Table B2: Results of production function estimation for East Germany.

	'No-out'		'Uni-out'		'Full-out'	
	Coeff	SE	Coeff	SE	Coeff	SE
Labour	-0.009	0.046	0.003	0.045	-0.066*	0.037
Land	-0.189***	0.048	-0.168***	0.045	-0.090*	0.046
Materials	1.115***	0.147	1.103***	0.146	1.055***	0.130
Capital	0.127**	0.050	0.125***	0.048	0.136***	0.034
N	1478		1461		1378	
Elasticity of scale	1.045		1.063		1.035	
p-value const. ret. to scale	0.726		0.622		0.757	
p-value coeff. jointly zero	<0.001		<0.001		<0.001	

Notes: Year dummies included in all models. *** (**, *) significant at the 1% (5%, 10%) level, based on standard errors robust to clustering in groups. Multivariate outliers following an analysis incorporating a materials specification that includes fertiliser inputs.

Source: Author.

Table B3: Results of production function estimation for West Germany.

	'No-out'		'Uni-out'		'Full-out'	
	Coeff	SE	Coeff	SE	Coeff	SE
Labour	0.194***	0.020	0.192***	0.019	0.160***	0.023
Land	-0.098***	0.018	-0.097***	0.018	-0.072***	0.021
Materials	0.962***	0.112	0.957***	0.112	0.909***	0.097
Capital	0.138***	0.023	0.142**	0.023	0.134***	0.022
N	3461		3452		3078	
Elasticity of scale	1.195		1.194		1.132	
p-value const. ret. to scale	0.043		0.044		0.108	
p-value coeff. jointly zero	<0.001		<0.001		<0.001	

Notes: Year dummies included in all models. *** (**, *) significant at the 1% (5%, 10%) level, based on standard errors robust to clustering in groups. Multivariate outliers following an analysis incorporating a materials specification that includes fertiliser inputs.

Source: Author.

Appendix C: The productivity of family and hired labour in EU arable farming

Table C1: Descriptive statistics.

Country	No. of farms	Output (ths EUR)				Total labour (ths hours)				Land (ha)				Materials (ths EUR)				Capital (ths EUR)			
		Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Denmark	208	180.4	277.1	3.1	2733.4	2.8	4.0	0.1	49.0	122.7	172.2	3.3	1760.0	98.0	158.6	5.5	1824.4	840.0	1402.7	42.3	21381.0
France	1030	155.8	114.5	5.2	1574.7	3.2	2.4	1.2	38.2	143.5	82.6	3.6	647.4	85.1	56.2	4.7	618.6	160.1	127.7	2.8	1379.8
Germany (East)	271	545.6	1018.3	5.5	9242.1	15.6	30.0	2.2	271.8	538.9	654.4	2.3	5155.9	345.8	649.8	13.9	6264.1	519.4	739.8	14.7	6591.7
Germany (West)	566	150.9	137.1	12.8	2114.7	4.2	3.3	1.1	93.9	92.3	60.1	0.5	429.5	84.7	67.6	11.9	737.5	153.9	125.5	11.1	1013.3
Italy	1322	60.6	125.2	0.8	2165.2	3.6	4.5	0.0	98.7	44.7	75.5	0.6	723.3	23.8	53.8	0.5	1204.2	125.2	232.6	2.7	4692.2
Poland	1518	39.8	61.9	0.9	1464.0	4.7	4.6	0.5	107.2	48.6	72.9	1.6	1439.6	17.4	30.5	0.9	725.1	78.9	91.2	4.1	1684.0
Slovakia	55	514.0	547.7	11.5	2785.6	39.1	41.5	1.3	188.8	768.9	727.5	30.2	3299.6	342.1	321.3	7.5	1405.5	940.2	1396.2	7.6	9273.0
Spain	1388	40.4	36.3	0.0	768.1	2.7	1.5	0.1	23.9	72.7	67.6	2.5	897.6	15.0	13.9	0.7	237.3	31.1	34.3	0.8	565.2
United Kingdom	188	278.1	330.3	8.7	3548.6	6.3	5.1	0.3	51.8	248.7	182.4	17.8	1178.5	157.2	150.0	10.4	1475.8	239.3	216.5	8.4	1555.2

Notes: SD: standard deviation. Min: Minimum value. Max: Maximum value.

Source: Authors' calculations.

Table C2: Results of production function estimations for the Ordinary Least Squares estimator per country.

	Denmark		France		Germany (East)		Germany (West)		Italy	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Labour	0.535***	0.067	0.125***	0.019	0.066	0.053	0.106***	0.031	0.279***	0.022
Ratio (r)	0.155**	0.062	-0.069***	0.019	0.025***	0.006	-0.173***	0.030	-0.077	0.058
Land	0.229***	0.062	0.073***	0.018	0.075	0.063	0.045**	0.019	0.009	0.015
Materials	0.523***	0.052	0.744***	0.019	0.892***	0.041	0.777***	0.024	0.664***	0.020
Capital	0.155***	0.040	0.159***	0.013	0.126***	0.030	0.119***	0.019	0.093***	0.013
N	605		4289		1047		2408		3545	
Elasticity of scale	1.442	0.031	1.101***	0.018	1.159***	0.029	1.046***	0.031	1.045***	0.016
p-value const. ret. to scale	<0.001		<0.001		<0.001		0.137		0.035	
R^2	0.949		0.864		0.952		0.859		0.847	
p-value coeff. jointly zero	<0.001		<0.001		<0.001		<0.001		<0.001	

Table C2 continued.

	Poland		Slovakia		Spain		United Kingdom	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Labour	0.185***	0.021	-0.034	0.102	0.377***	0.023	0.219***	0.041
Ratio (r)	-0.023	0.023	0.030*	0.017	-0.005	0.009	0.041**	0.021
Land	0.142***	0.014	-0.137	0.104	0.103***	0.016	0.196***	0.046
Materials	0.609***	0.020	0.994***	0.127	0.662***	0.016	0.726***	0.048
Capital	0.250***	0.015	0.186***	0.062	0.020	0.013	0.073**	0.032
N	2534		89		6393		612	
Elasticity of scale	1.185***	0.019	1.009***	0.068	1.162***	0.024	1.214***	0.033
p-value const. ret. to scale	<0.001		0.897		<0.001		<0.001	
R^2	0.894		0.939		0.701		0.907	
p-value coeff. jointly zero	<0.001		<0.001		<0.001		<0.001	

Notes: Year dummies included in all models. *** (**, *) significantly different from zero at the 1% (5%, 10%) level, based on standard errors robust to clustering in groups.

Source: Authors' calculations.

Table C3: Results of production function estimations for the Wooldridge/Levinsohn/Petrin estimator per country.

	Denmark		France		Germany (East)		Germany (West)		Italy	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Labour	0.653***	0.098	0.136***	0.022	0.071	0.052	0.120***	0.032	0.308***	0.027
Ratio (r)	0.076	0.062	-0.065***	0.021	0.015**	0.007	-0.182***	0.032	-0.051	0.059
Land	0.220***	0.080	0.046***	0.018	-0.025	0.055	0.013	0.019	-0.009	0.021
Materials	-0.012	0.268	0.799***	0.083	1.109***	0.174	0.729***	0.086	0.501***	0.107
Capital	0.100	0.075	0.118***	0.015	0.083*	0.048	0.085***	0.024	0.018	0.031
N	605		4289		1047		2408		3545	
Elasticity of scale	0.962***	0.230	1.099***	0.076	1.238***	0.139	0.947***	0.082	0.818***	0.094
p-value const. ret. to scale	0.868		0.193		0.086		0.517		0.054	
p-value coeff. jointly zero	<0.001		<0.001		<0.001		<0.001		<0.001	

Table C3 continued.

	Poland		Slovakia		Spain		United Kingdom	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Labour	0.109***	0.025	-0.155	0.137	0.417***	0.028	0.203***	0.050
Ratio (r)	-0.092***	0.028	0.038*	0.017	0.001	0.014	0.044***	0.020
Land	0.065***	0.019	-0.189	0.163	0.080***	0.017	0.182***	0.060
Materials	0.987***	0.100	1.551***	0.411	0.727***	0.060	0.722***	0.379
Capital	0.066	0.045	-0.032	0.164	0.075***	0.027	0.107***	0.063
N	2534		89		6393		612	
Elasticity of scale	1.226***	0.087	1.174***	0.274	1.299***	0.063	1.213***	0.364
p-value const. ret. to scale	0.009		0.526		<0.001		0.557	
p-value coeff. jointly zero	<0.001		<0.001		<0.001		<0.001	

Notes: Year dummies included in all models. *** (**, *) significantly different from zero at the 1% (5%, 10%) level, based on standard errors robust to clustering in groups.

Source: Authors' calculations.

Table C4: Results of Translog production function estimations for the Wooldridge/Levinsohn/Petrin estimator per country.

	Denmark		France		Germany (East)		Germany (West)		Italy	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Labour	0.548	1.440	-0.325	0.494	-0.400	0.422	-1.066**	0.491	-0.300	0.424
Land	0.163	0.988	0.250	0.267	1.491***	0.407	1.046***	0.329	-0.031	0.294
Materials	0.132	2.208	0.794	0.894	0.136	1.344	1.848**	0.810	1.033	0.637
Capital	0.610	1.282	-0.632	0.612	0.013	0.461	-0.007	0.438	-0.407	0.376
N	605		4289		1047		2408		3545	
p-value const. ret. to scale	<0.001		<0.001		<0.001		<0.001		<0.001	
p-value coeff. jointly zero	0.824		0.397		0.979		0.171		0.138	

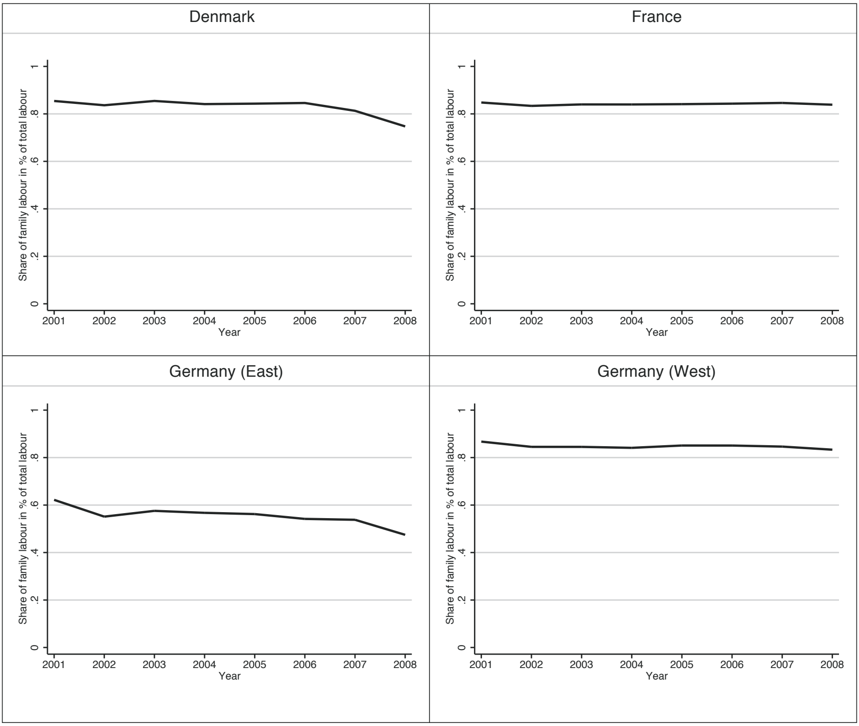
Table C4 continued.

	Poland		Slovakia		Spain		United Kingdom	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Labour	-2.977***	0.611	-2.710	9.028	-1.288**	0.577	1.230	1.138
Land	0.421	0.445	-5.529	6.779	-1.353***	0.300	2.383*	1.388
Materials	2.329*	1.380	2.597	8.357	2.520***	0.529	-0.230	5.644
Capital	0.010	0.445	1.228	2.672	1.001***	0.319	-1.261	2.035
N	2534		89		6393		612	
p-value const. ret. to scale	<0.001		<0.001		<0.001		<0.001	
p-value coeff. jointly zero	0.278		0.421		0.657		0.921	

Notes: Estimates at sample means. Year dummies included in all models. *** (**, *) significantly different from zero at the 1% (5%, 10%) level, based on standard errors robust to clustering in groups.

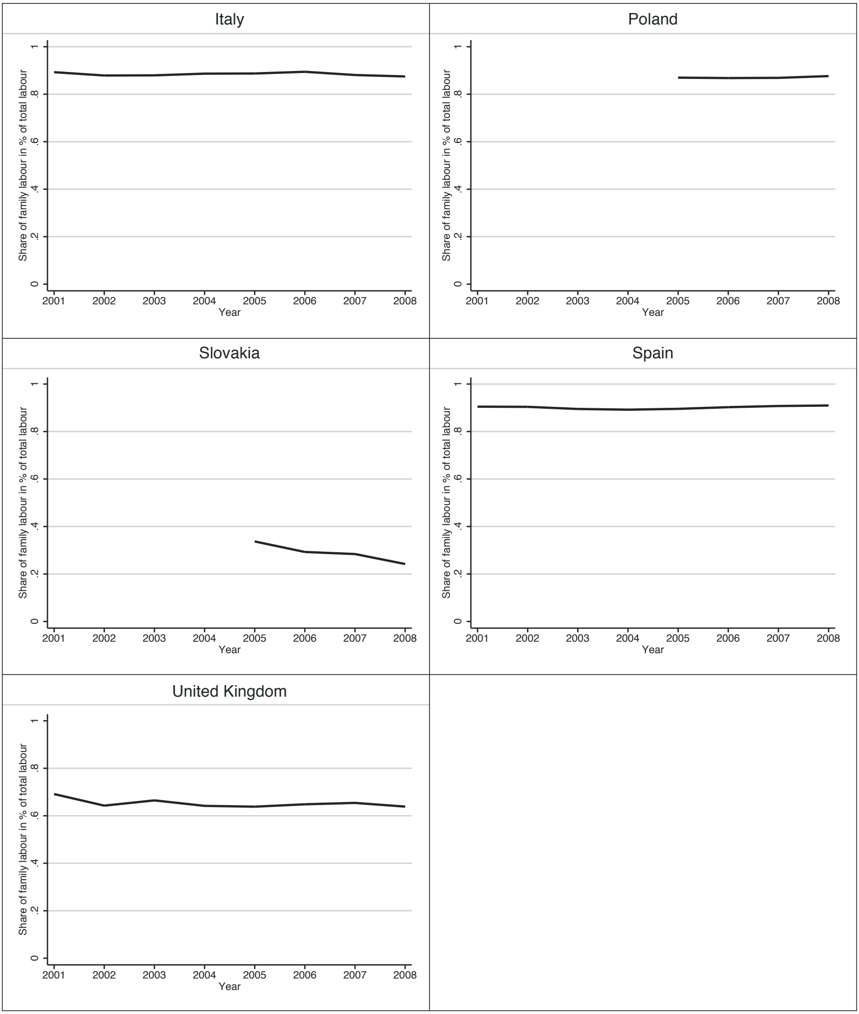
Source: Authors’ calculations

Figure C1: Evolution of the share of family labour over the sample period per country.



Source: Authors compilation based on FADN data.

Figure C1: continued.



Source: Authors compilation based on FADN data.

Appendix D: Core estimation commands and description

Below follows a description of the core Stata estimation commands used to produce the results for this dissertation.

Cobb-Douglas Production function

```
///description of variables
//logged variables
*output - log of output
*labour - log of labour
*land - log of land
*materials - log of materials
*capital - log of fixed capital
// additional variables
*td* - time dummy for period *
*a3 - unique farm identifier

//nominal output shares of inputs
* sh_labour -output share of labour
* sh_land -output share of land
* sh_materials - output share of materials
* sh_capital -output share of fixed capital

///estimation
//output shares
ttest sh_labour
ttest sh_land
ttest sh_materials
ttest sh_capital

//OLS with cluster robust standard errors
reg output labour land materials capital td*, vce(cluster a3)

*test for constant returns to scale
test labour+land+materials+capital=1

*estimate returns to scale
lincom labour+land+materials+capital

//fixed effects regression with cluster robust standard errors
xtreg output labour land materials capital td*, fe vce(cluster a3)

*test for constant returns to scale
test labour+land+materials+capital=1

*estimate returns to scale
lincom labour+land+materials+capital
```

```
//Levinsohn Petrin (2003) using materials as proxy
levpet output, free(labour land td*) proxy(materials) capital(capital)
reps(20) revenue

*test for constant returns to scale
test labour+land+materials+capital=1

*estimate returns to scale
lincom labour+land+materials+capital

//Wooldridge (2009) – LP using materials as proxy
sort a3 year
*generate lags of variables
gen labour_l1=L.labour
gen land_l1=L.land
gen capital_l1=L.capital
gen materials_l1=L.materials
gen materials_l2=L.materials_l1

*low order polynomials of capital and materials
gen capitalmaterials_l1=capital_l1*materials_l1
gen capital2_l1=capital_l1^2
gen materials2_l1=materials_l1^2

gen capital2materials_l1=capital_l1^2*materials_l1
gen capitalmaterials2_l1=capital_l1*materials_l1^2
gen capital3_l1=capital_l1^3
gen materials3_l1=materials_l1^3

*vectors of variables
*exog are the variables treated as exogenous, these represent the included
*instruments; endog are the endogenous variables; instr are the included
*instruments

global exog capital capital_l1 materials_l1 capitalmaterials_l1 ///
capital2_l1 materials2_l1 capital2materials_l1 capitalmaterials2_l1 ///
capital3_l1 materials3_l1 td*
global endog labour land materials
global instr labour_l1 land_l1 materials_l2

*estimation
ivreg2 output $exog ($endog=$instr), gmm2s cluster(a3)

*test for constant returns to scale
test labour+land+materials+capital=1

*estimate returns to scale
lincom labour+land+materials+capital
```

```
//system gmm estimation
sort a3 year

*options
*twostep – request two-step gmm
*robust – Windmeijer correction for standard errors
*gmm() – "gmm-style" instruments
*ivstyle() – standard instruments
*equation – suboption that defines equation using ivstyle instruments
*h(2) – option that defines structure of covariance matrix of iid errors

*estimation
xtabond2 output labour L.labour land L.land materials L.materials capital
L.capital L.output td*, ///
twostep robust gmm(L.output L2.(labour land materials capital)) ivstyle(td*,
equation(diff)) h(2)

*estimate common factors by minimum distance
md _ ar1, nx(4) beta(e(b)) cov(e(V))

*test for constant returns to scale
test labour+land+materials+capital=1

*estimate returns to scale
lincom labour+land+materials+capital

*Assess degree of persistence in factor series
xtabond2 labour L.labour td2-td7,twostep robust gmm(L3.labour, /// collapse)
ivstyle(td*, equation(diff)) h(2) nodiffsargan
xtabond2 land L.land td2-td7,twostep robust gmm(L3.land, collapse) ///
ivstyle(td*, equation(diff)) h(2) nodiffsargan
xtabond2 materials L.materials td2-td7,twostep robust gmm(L3.materials, ///
collapse) ivstyle(td2-td7, equation(diff)) h(2) nodiffsargan
xtabond2 capital L.capital td2-td7,twostep robust gmm(L3.capital, ///
collapse) ivstyle(td*, equation(diff)) h(2) nodiffsargan
xtabond2 output L.output td2-td7,twostep robust gmm(L3.output, /// collapse)
ivstyle(td*, equation(diff)) h(2) nodiffsargan


*Assess explanatory power of instrument sets
*First differences
reg D.labour L3.labour L4.labour L5.labour L6.labour L7.labour if ///
year==2008
reg D.land L3.land L4.land L5.land L6.land L7.land if year==2008
reg D.mat L3.materials L4.materials L5.materials L6.materials ///
L7.materials if year==2008
reg D.capital L3.capital L4.capital L5.capital L6.capital L7.capital if ///
year==2008
reg D.output L3.output L4.output L5.output L6.output L7.output if ///
year==2008
```

```
*Levels
reg L.labour L2D.labour L3D.labour L4D.labour L5D.labour L6D.labour if ///
year==2008
reg L.land L2D.land L3D.land L4D.land L5D.land L6D.land if year==2008
reg L.mat L2D.materials L3D.materials L4D.materials L5D.materials /// L6D.
materials if year==2008
reg L.capital L2D.capital L3D.capital L4D.capital L5D.capital /// L6D.capital
if year==2008
reg L.output L2D.output L3D.output L4D.output L5D.output L6D.output if ///
year==2008
```

Translog Production function

```
///OLS translog
//additional variables
gen labour2=labour^2
gen land2=land^2
gen materials2=matterials^2
gen capital2=capital^2
gen labourxland=labour*land
gen labourxmat=labour*mat
gen labourxcapital=labour*capital
gen landxmaterials=land*materials
gen landxcapital=land*capital
gen landxcows=land*cows
gen materialsxcapital=materials*capital
gen capitalxcows=capital*cows

*Demean (or center) all logged variables by using total sample means –
*this allows interpretation of non-interacted parameters as production
*elasticities at sample means
center output labour land materials capital labor2 land2 materials2
capital2 labourxland labourxmat labourxcapital landxmaterials landxcapital
materialsxcapital, prefix(d)

//estimation
reg d_output d_labour d_land d_materials d_capital d_labor2 d_
land2 /// d_materials2 d_capital2 d_laborxland d_laborxmaterials d_
laborxcapital /// d_landxmaterials d_landxcapital d_materialsxcapital
td*, vce(cluster a3)

*test for joint significance of interaction terms
test d_labour2+d_land2+d_materials2+d_capital2+d_laborxland+ /// d_
laborxmaterials+d_laborxcapital+d_landxmaterials+d_landxcapital+ ///
d_materialsxcapital=0

//fixed effects translog
*variable preparation
*express all variables in differences from farm mean over time
```

```
*(= „groupwise demeaning") - this eliminates fixed effects „by hand"
bysort a3 (year): center output labour land materials capital

*create interactions of variables in logs that are farmwise demeaned
gen c_labour2=c_labour^2
gen c_land2=c_land^2
gen c_materials2=c_materials^2
gen c_capital2=c_capital^2
gen c_labourxland=c_labour*c_land
gen c_labourxmat=c_labour*c_mat
gen c_labourxcapital=c_labour*c_capital
gen c_landxmaterials=c_land*c_materials
gen c_landxcapital=c_land*c_capital
gen c_materialsxcapital=c_materials*c_capital

*demean (or center) all logged and farmwise demeaned variables by using
*total sample means - this allows interpretation of non-interacted
*parameters as prod elasticities at sample means
foreach var of varlist c_output c_labour c_land c_materials c_capital
/// c_labour2 c_land2 c_materials2 c_capital2 c_cows2 c_labourxland
/// c_labourxmaterials c_labourxcapital c_landxmaterials c_landxcapital
/// c_materialsxcapital {
gen m`var'=`var'
}

center mc_output mc_labour mc_land mc_materials mc_capital mc_
labour2 /// mc_land2 mc_materials2 mc_capital2 mc_cows2 mc_labourxland
/// mc_laborxmaterials mc_labourxcapital mc_landxmaterials mc_
landxcapital /// mc_materialsxcapital, meansave

*estimate translog demeaned
*compute number of groups N_g (=farms) in regression sample for degrees of
*freedom correction
qui xtreg c_mc_output c_mc_labour c_mc_land c_mc_materials c_
mc_capital /// c_mc_labour2 c_mc_land2 c_mc_mat2 c_mc_capital2
c_mc_labourxland /// c_mc_labourxmaterials c_mc_labourxcapital c_
mc_landxmaterials /// c_mc_landxcapital c_mc_materialsxcapital td*, fe
vce(cluster a3)
scalar define groups=e(N_g)

*correct standard errors
local df=groups
mat b=e(b)
scalar vadj = e(df_r)/(e(N)-1-(e(df_m) + `df'))
matrix V = vadj*e(V)

*get estimation table with FE by hand results
ereturn post b V
display _newline
display „Adjusted standard errors"
ereturn display
```

```
* test for joint significance of interaction terms
test c _ mc _ labour2+c _ mc _ land2+c _ mc _ materials2+c _ mc _ capital2+ ///
c _ mc _ laborxland+c _ mc _ laborxmaterials+c _ mc _ laborxcapital+ ///
c _ mc _ landxmaterials+c _ mc _ landxcapital+c _ mc _ materialsxcapital=0

//Wooldridge (2009) translog

*low order polynomials of capital and materials

*additional variables needed for translog

gen capital2=capital^2
gen capital2 _ l2=L2.capital2
gen labour2=labour^2
gen land2=land^2
gen materials2=materials^2
gen laborland=labour*land
gen labourmaterials=labour*materials
gen landmaerialst=land*materials
gen capitallabour=capital*labor
gen capitalland=capital*land
gen capitalmat=capital*mat
gen labour2 _ l1=L.labour2
gen land2 _ l1=L.land2
gen mat2 _ l2=L.mat2 _ l1
gen laborland _ l1=labor _ l1*land _ l1
gen labourmaterials _ l1=labour _ l1*materials _ l1
gen landmaterials _ l1=land _ l1*mat _ l1
gen capitallabour _ l1=capital _ l1*labour _ l1
gen capitalland _ l1=capital _ l1*land _ l1
gen capital _ l2=L.capital _ l1
gen capitalmat _ l2=capital _ l2*materials _ l2


center output capital capital _ l1 materials _ l1 capitalmaterials _ l1 ///
capital2 _ l1 materials2 _ l1 capital2materials _ l1 capitalmaterials2 _ l1 ///
capital3 _ l1 materials3 _ l1 capital2 capital2 _ l2 labour land materials ///
labour2 land2 materials2 labourland labourmaterials landmaterials ///
capitallabor capitalland capitalmaterials labour _ l1 land _ l1 ///
materials _ l2 labour2 _ l1 land2 _ l1 materials2 _ l2 labourland _ l1 ///
labourmaterials _ l1 landmaterials _ l1 capitallabour _ l1 capitalland _ l1 ///
capitalmaterials _ l2, prefix(ce)
*vectors of variables


global exog ce _ capital ce _ capital _ l1 ce _ materials _ l1 ///
ce _ capitalmaterials _ l1 ce _ capital2 _ l1 ce _ materials2 _ l1 ///
ce _ capital2materials _ l1 ce _ capitalmaterials2 _ l1 ce _ capital3 _ l1 ///
ce _ materials3 _ l1 td* ce _ capital2 ce _ capital2 _ l2
```

```
global endog ce _ labour ce _ land ce _ materials ce _ labour2 ce _ land2 ///
ce _ materials2 ce _ labourland ce _ labourmat ce _ landmaterials ///
ce _ capitallabor ce _ capitalland ce _ capitalmaterials

global instr ce _ labour _ l1 ce _ land _ l1 ce _ materials _ l2 ce _ labour2 _ l1 ///
ce _ land2 _ l1 ce _ materials2 _ l2 ce _ labourland _ l1 c _ labourmaterials _ l1///
c _ landmaterials _ l1 ce _ capitallabour _ l1 ce _ capitalland _ l1 ///
ce _ capitalmaterials _ l2


ivreg2 ce _ output $exog ($endog=$instr), gmm2s cluster(a3)


*test interaction terms jointly zero
test c _ labour2+c _ land2+c _ materials2+c _ capital2+c _ laborland+ //
c _ labourmaterials+c _ landmaterials+c _ capitallabour+ ///
c _ capitaland+c _ capitalmaterials=0
```

Outlier identification

The univariate outlier detection was performed following the trimming rule upper/lower quartile +/- 1.5 times the interquartile range on the average capital productivity per farm.

```
* Drop outliers; acpf – average capital productivity per farm
qui sum acpf, det
scalar define iqr _ f=r(p75)-r(p25)
scalar define ub _ f=r(p75)+1.5*iqr _ f
scalar define lb _ f=r(p25)-1.5*iqr _ f
drop if acpf>ub _ f | acpf<lb _ f
```

The multivariate outlier analysis was performed using R package "restlos" (Liebscher and Kirschstein, 2015). This is as simple as calling the appropriate function and specifying the respective data set. Below, I reproduce the code of the function pmst.tsch that was utilised to produce the results in this dissertation.

```
pMST.tsch <- function (data, N=NULL, lmax = nrow(data) * 100, alpha=0.95)
{
   require(igraph, quietly=T)

     if (is.data.frame(data))
         data = as.matrix(data)
     if (!is.matrix(data))
         stop("at least two-dimensional data matrix required")
     if (mode(data) != "numeric")
         stop("numeric data required")
     if (dim(data)[1] <= dim(data)[2])
         stop("n > d required")
```

```
  ddmst <- function(dat) {
      ddat <- as.matrix(dist(dat, upper = T, diag = T))
      x.tmp <- graph.adjacency(ddat, weighted = TRUE, mode = „undirected")
      mstdat <- minimum.spanning.tree(x.tmp)
      mstdat <- matrix(as.numeric(get.edgelist(mstdat, names = F)),
          ncol = 2)
      o <- dim(dat)[1] - 1
      em <- numeric(o)
      k <- 0
      ddat <- as.matrix(ddat)
      for (i in 1:o) {
          k <- k + ddat[mstdat[i, 1], mstdat[i, 2]]
          em[i] <- ddat[mstdat[i, 1], mstdat[i, 2]]
      }
      emax <- max(em)
      return(k)
  }


m <- nrow(data)
d <- ncol(data)

  x1 <- as.matrix(dist(data, upper = T, diag = T))
  x.tmp <- graph.adjacency(x1, weighted = TRUE, mode = „undirected")
  x2 <- minimum.spanning.tree(x.tmp)
  x2 <- matrix(as.numeric(get.edgelist(x2, names = F)), ncol = 2)
  x1 <- as.matrix(x1)
  U1 <- diag(x1[x2[, 1], x2[, 2]])
  T1 <- order(U1)
  l <- 0
  LiB <- list(c())
  GeB <- c()
  LeB <- c()
  x6 <- matrix(c(0, 0, 0), ncol = 3)

stoppi <- FALSE
cuti <- max(U1)

  repeat {
      l <- l + 1
      T2 <- sapply(LiB, function(x) {
          any(x == x2[T1[l], 1] | x == x2[T1[l], 2])
      })
      x70 <- 0
      if (any(T2 == TRUE)) {
          if (sum(T2 == TRUE) > 1) {
              T4 <- which(T2 == TRUE)
              maxi <- which.max(sapply(LiB[T4], length))
```

```
        x3 <- colMeans(data[unlist(LiB[T4[maxi]]), ])
        LiB[[T4[1]]] <- unique(c(unlist(LiB[T4]), x2[T1[l],
          1], x2[T1[l], 2]))
        x4 <- mean(sapply(LiB[T4[-maxi]], function(x) {
          sqrt(sum((x3 - colMeans(data[x, ]))^2))
        }))
        LiB[T4[-1]] <- 0
        GeB[T4[1]] <- length(LiB[[T4[1]]])
        GeB[T4[-1]] <- NA
        LeB[T4[1]] <- sum(LeB[T4]) + U1[T1[l]]
        LeB[T4[-1]] <- 0
        x5 <- x4
        x7 <- LeB[which.max(sapply(LiB, length))]
    }
    else {
        T3 <- which(T2 == TRUE)
        x3 <- colMeans(data[LiB[[T3]], ])
        LiB[[T3]] <- unique(c(LiB[[T3]], x2[T1[l], 1],
          x2[T1[l], 2]))
        x4 <- colMeans(data[x2[T1[l], ], ])
        GeB[T3] <- length(LiB[[T3]])
        x5 <- sqrt(sum((x3 - x4)^2))
        LeB[T3] <- sum(LeB[T3]) + U1[T1[l]]
        x7 <- LeB[which.max(sapply(LiB, length))]
    }
    x6 <- rbind(x6, c(x7, U1[T1[l]], max(sapply(LiB,
        length))))
}
else {
    LiB[[l]] <- c(x2[T1[l], 1], x2[T1[l], 2])
    GeB[l] <- length(LiB[[l]])
    LeB[l] <- U1[T1[l]]
}

# calculate cut off value by Chebyshev bound
if(any(na.omit(GeB)>=floor((sum(dim(data))+1)/2))==TRUE & stoppi==F) {
  ini.set <- LiB[[which.max(sapply(LiB,length))]]

  tmp.x1 <- as.matrix(dist(data[ini.set,], upper = T, diag = T))
  tmp.x.tmp <- graph.adjacency(tmp.x1, weighted = TRUE, mode =
  „undirected")
  tmp.x2 <- minimum.spanning.tree(tmp.x.tmp)
  tmp.x2 <- matrix(as.numeric(get.edgelist(tmp.x2, names = F)), ncol =
  2)
  tmp.x1 <- as.matrix(tmp.x1)
  em <- diag(tmp.x1[tmp.x2[, 1], tmp.x2[, 2]])

  mm <- length(em)
  cuti <- mean(em) + sd(em)*sqrt((mm^2-1)/(mm^2*(1-alpha)-mm))
```

```
        stoppi <- TRUE
      }

      # stop procedure
       if(is.null(N)){
        if(U1[T1[l+1]] > cuti | any(na.omit(GeB)==m)) break
      }
      else{
        if(any(na.omit(GeB)>=N)==TRUE) break
      }
    }
    drin <- LiB[[which.max(sapply(LiB, length))]]
    pMST <- list(loc = colMeans(data[drin, ]), cov = cov(data[drin,
        ]), sam.fin=sort(drin), sam.ini=sort(ini.set), data = data, x6 = x6,
    cut.off = cuti)
    class(pMST) <- „pMST"
    return(pMST)
}
```

After conducting the outlier analysis one can proceed with the production function analysis in Stata (see above).

Assessing heterogeneous labour impacts

To incorporate heterogeneous labour impacts the Cobb-Douglas production function is augmented by the logarithmised share of family labour in total labour (ratio, see chapter 6). Below, I show the modified analysis.

```
///OLS
*estimation
reg output labor ratio land mat capital td*, vce(cluster a3)

*RTS
test labour+land+materials+capital=1
lincom labor+land+mat+capital

*recover gamma
di _b[ratio]/_b[labor]
nlcom _b[ratio]/_b[labor]

///Wooldridge (2009)
sort a3 year

*generate lags of variables
gen labour_l1=L.labour
```

```
gen land_l1=L.land
gen capital_l1=L.capital
gen materials_l1=L.materials
gen materials_l2=L.materials_l1
gen ratio_l1=L.ratio

*low order polynomials of capital and materials
gen capitalmaterials_l1=capital_l1*materials_l1
gen capital2_l1=capital_l1^2
gen materials2_l1=materials_l1^2

gen capital2materials_l1=capital_l1^2*materials_l1
gen capitalmaterials2_l1=capital_l1*materials_l1^2
gen capital3_l1=capital_l1^3
gen materials3_l1=materials_l1^3

*vectors of variables

global exog capital capital_l1 materials_l1 capitalmaterials_l1 ///
capital2_l1 materials2_l1 capital2materials_l1 capitalmaterials2_l1 ///
capital3_l1 materials3_l1 td*
global endog labour ratio land mat
global instr labour_l1 ratio_l1 land_l1 materials_l2

*estimation
ivreg2 output $exog ($endog=$instr), gmm2s cluster(a3)

*test for constant returns to scale
test labour+land+materials+capital=1

*estimate returns to scale
lincom labour+land+materials+capital

*recover gamma
di _b[ratio]/_b[labor]
nlcom _b[ratio]/_b[labor]
```

Studies on the Agricultural and Food Sector in Transition Economies
edited by Leibniz Institute of Agricultural Development in Transition Economies (IAMO)
ISSN 1436-221X

Vol. 1 **The importance of institutions for the transition in Central and Eastern Europe with emphasis on agricultural and food industry**
ed. by Klaus Frohberg and Witold-Roger Poganietz
1998, 137 pages, ISBN 3-8175-0258-3

Vol. 2 **The significance of politics and institutions for the design and formation of agricultural Policies**
ed. by Klaus Frohberg and Peter Weingarten
1999, 254 pages, ISBN 3-8175-0289-3

Vol. 3 **Food processing and distribution in transition countries. Problems and perspectives**
ed. by Monika Hartmann and Jürgen Wandel
1999, 349 pages, ISBN 3-8175-0293-1

Vol. 4 **Die private Nachfrage nach Nahrungsmitteln im Transformations-prozeß Tschechiens und Polens**
Stephan Brosig (PhD)
2000, 171 Seiten, ISBN 3-8175-0319-9

Vol. 5 **Integrating Estonia into the EU: Quantitative analysis of the agricultural and food sector**
Achim Fock (PhD)
2000, 286 pages, ISBN 3-8175-0320-2

Vol. 6 **Competitiveness of agricultural enterprises and farm activities in transition countries**
ed. by Peter Tillack and Frauke Pirscher
2000, 216 pages, ISBN 3-8175-0322-9

Vol. 7 **Конкурентоспособность сельскохозяйственных предприятий и фермерской деятельности в странах переходного периода**
под редакцией Петера Тиллака и Фрауке Пиршер
2000, 253 страницы, ISBN 3-8175-0324-5

Vol. 8 **Perspectives on agriculture in transition: Analytical issues, modelling approaches, and case study results**
ed. by Witold-Roger Poganietz, Alberto Zezza, Klaus Frohberg and Kostas G. Stamoulis
2000, 433 pages, ISBN 3-8175-0323-7

Vol. 9 **Land ownership, land markets and their influence on the efficiency of agricultural production in Central and Eastern Europe**
ed. by Peter Tillack and Eberhard Schulze
2000, 485 pages, ISBN 3-8175-0325-3

Vol. 10 **Landwirtschaft und Industrie in Russland – Der Transformationsprozeß in der Ernährungsindustrie**
Jürgen Wandel (PhD)
2000, 361 Seiten, ISBN 3-8175-0334-2

Vol. 11 **Food consumption in Russia. An econometric analysis based on household data**
Karin Elsner (PhD)
2001, 256 pages, ISBN 3-8175-0335-0

Vol. 12 **Alexander Wasiljewitsch Tschajanow – Die Tragödie eines großen Agrarökonomen**
hrsg. u. übers. von Eberhard Schulze
2001, 192 Seiten, ISBN 3-8175-0342-3

Vol. 13 **Analysis of food consumption in Central and Eastern Europe: Relevance and empirical methods**
ed. by Stephan Brosig and Monika Hartmann
2001, 253 pages, ISBN 3-8175-0349-0

Vol. 14 **Wettbewerbsprozesse und Firmenwachstum in der Transformation am Beispiel der polnischen Fleischindustrie**
Agata Pieniadz (PhD)
2002, 291 Seiten, ISBN 3-8175-0360-1

Vol. 15 **Agricultural enterprises in transition: Parallels and divergences in Eastern Germany, Poland and Hungary**
ed. by Ludger Hinners-Tobrägel and Jürgen Heinrich
2002, 455 pages, ISBN 3-8175-0366-0

Vol. 16 **Agricultural technology and economic development of Central and Eastern Europe. Results of the workshop in Halle, 2nd-3rd July 2001**
ed. by Peter Tillack and Ulrich Fiege
2002, 160 pages, ISBN 3-86037-199-1

Vol. 17 **Региональные аспекты аграрных преобразований: Политика, реструктуризация, рыночная адаптация**
под редакцией Петера Тиллака и Виталия Зиновчука
2003, 236 страницы, ISBN 3-928466-55-0

Vol. 18 **Alexander Vasilievich Chayanov – The tragedy of an outstanding agricultural economist**
ed. by Eberhard Schulze
2003, 188 pages, ISBN 3-86037-201-7

Vol. 19 **Development of agricultural market and trade policies in the CEE Candidate Countries**
by the Network of Independent Agricultural Experts in the CEE Candidate Countries
2003, 72 pages, ISBN 3-86037-212-2

Vol. 20 **Large farm management**
ed. by Alfons Balmann and Alexej Lissitsa
2003, 396 pages, ISBN 3-86037-213-0

Vol. 21 **Success and failures of transition – The Russian agriculture between fall and resurrection**
ed. by Eberhard Schulze, Elke Knappe, Eugenia Serova, Peter Wehrheim
2003, 521 pages, ISBN 3-9809270-1-6

Vol. 22 **Subsistence agriculture in Central and Eastern Europe: How to break the vicious circle?**
ed. by Steffen Abele and Klaus Frohberg
2003, 233 pages, ISBN 3-9809270-2-4

Vol. 23 **Pfadabhängigkeiten und Effizienz der Betriebsstrukturen in der ukrainischen Landwirtschaft – Eine theoretische und empirische Analyse**
Andriy Nedoborovskyy (PhD)
2004, 197 Seiten, ISBN 3-86037-216-5

Vol. 24 **Nichtmonetäre Transaktionen in der ukrainischen Landwirtschaft: Determinanten, Spezifika und Folgen**
Olena Dolud (PhD)
2004, 190 Seiten, ISBN 3-9809270-3-2

Vol. 25 **The role of agriculture in Central and Eastern European rural development: Engine of change or social buffer?**
ed. by Martin Petrick and Peter Weingarten
2004, 426 pages, ISBN 3-9809270-4-0

Vol. 26 **Credit rationing of Polish farm households – A theoretical and empirical analysis**
Martin Petrick (PhD)
2004, 254 pages, ISBN 3-9809270-6-7

Vol. 27 **Drei Jahrhunderte Agrarwissenschaft in Russland: Von 1700 bis zur Gegenwart**
Alexander Alexandrowitsch Nikonow und Eberhard Schulze
2004, 232 Seiten, ISBN 3-9809270-8-3

Vol. 28 **Russlands Weg vom Plan zum Markt: Sektorale Trends und regionale Spezifika**
Peter Voigt (PhD)
2004, 270 Seiten, ISBN 3-9809270-9-1

Vol. 29 **Auswirkungen des Transformationsprozesses auf die sozio-ökonomischen Funktionen ukrainischer Landwirtschafts-unternehmen**
Helga Biesold (PhD)
2004 182 Seiten, ISBN 3-938584-00-9

Vol. 30 **Agricultural policies and farm structures – Agent-based modelling and application to EU-policy reform**
Kathrin Happe (PhD)
2004, 291 pages, ISBN 3-938584-01-7

Vol. 31 **How effective is the invisible hand? Agricultural and food markets in Central and Eastern Europe**
ed. by Stephan Brosig and Heinrich Hockmann
2005, 361 pages, ISBN 3-938584-03-3

Vol. 32 **Erfolgsfaktoren von landwirtschaftlichen Unternehmen mit Marktfruchtanbau in Sachsen-Anhalt**
Kirsti Dautzenberg (PhD)
2005, 161 Seiten, ISBN 3-938584-06-8

Vol. 33 **Agriculture in the face of changing markets, institutions and policies: Challenges and strategies**
ed. by Jarmila Curtiss, Alfons Balmann, Kirsti Dautzenberg, Kathrin Happe
2006, 544 pages, ISBN 3-938584-10-6

Vol. 34 **Making rural households' livelihoods more resilient – The importance of social capital and the underlying social networks**
ed. by Gertrud Buchenrieder and Thomas Dufhues
2006, 106 pages, ISBN 3-938584-13-0

Vol. 35 **Außerlandwirtschaftliche Diversifikation im Transformations-prozess. Diversifikationsentscheidungen und -strategien ländlicher Haushalte in Slowenien und Mazedonien**
Judith Möllers (PhD)
2006, 323 Seiten, ISBN 3-938584-14-9

Vol. 36 **Accessing rural finance – The rural financial market in Northern Vietnam**
Thomas Dufhues (PhD)
2007, 166 Seiten, ISBN 3-938584-16-5

Vol. 37 **Страхование посевов в Казахстане: Анализ возможностей эффективного управления рисками**
Раушан Бокушева, Олаф Хайдельбах, Талгат Кусайынов
2007, 82 Seiten, ISBN 3-938584-17-3

Vol. 38 **Rethinking agricultural reform in Ukraine**
Zvi Lerman, David Sedik, Nikolai Pugachov, Aleksandr Goncharuk
2007, 167 Seiten, ISBN 3-938584-18-1

Vol. 39 **Sustainable rural development: What is the role of the agri-food sector?**
ed. by Martin Petrick, Gertrud Buchenrieder
2007, 293 pages, ISBN 3-938584-22-X

Vol. 40 **Efficiency of selected risk management instruments – An empirical analysis of risk reduction in Kazakhstani crop production**
Olaf Heidelbach (PhD)
2007, 223 Seiten, ISBN 3-938584-19-X

Vol. 41 **Marktstruktur und Preisbildung auf dem ukrainischen Markt für Rohmilch**
Oleksandr Perekhozhuk (PhD)
2007, 274 Seiten, ISBN 978-3-938584-24-8

Vol. 42 **Labor market behavior of Chinese rural households during transition**
Xiaobing Wang (PhD)
2007, 140 Seiten, ISBN 978-3-938584-25-5

Vol. 43 **Continuity and change: Land and water use reforms in rural Uzbekistan. Socio-economic and legal analyses for the region Khorezm**
ed. by Peter Wehrheim, Anja Schoeller-Schletter, Christopher Martius
2008, 211 Seiten, ISBN 978-3-938584-27-9

Vol. 44 **Agricultural economics and transition: What was expected, what we observed, the lessons learned (Vol I and II)**
ed. by Csaba Csáki, Csaba Forgács
2008, 634 Seiten, ISBN 978-3-938584-31-6

Vol. 45 **Theoretical and methodological topics in the institutional economics of European agriculture. With applications to farm organisation and rural credit arrangement**
Martin Petrick
2008, 223 Seiten, ISBN 978-3-938584-31-6

Vol. 46 **Agri-food business: Global challenges – Innovative solutions**
ed. by Thomas Glauben, Jon H. Hanf, Michael Kopsidis, Agata Pieniadz, Klaus Reinsberg
2008, 152 pages, ISBN 978-3-938584-33-0

Vol. 47 **Eine Analyse der Transformationsberatung für die "kollektive Landwirtschaft" während der ersten Transformationsphase (1989-1991) am Beispiel Ostdeutschlands: Lehren für Korea**
Jeong Nam Choi (PhD)
2009, 225 Seiten, ISBN 978-3-938584-36-1

Vol. 48 **Croatia's EU accession. Socio-economic assessment of farm households and policy recommendations**
Judith Möllers, Patrick Zier, Klaus Frohberg, Gertrud Buchenrieder and Štefan Bojnec
2009, 196 Seiten, ISBN 978-3-938584-35-4

Vol. 49 **Structural change in Europe's rural regions. Farm livelihoods between subsistence orientation, modernisation and non-farm diversification**
ed. by Gertrud Buchenrieder Judith Möllers
2009, 166 Seiten, ISBN 978-3-938584-39-2

Vol. 50 **Motive beim Weinkonsum – Unterschiede zwischen deutschen und ukrainischen Konsumenten**
Astrid Lucie Rewerts (PhD)
2009, 267 Seiten, ISBN 978-3-938584-40-8

Vol. 51 **Rural development as provision of local public goods: Theory and evidence from Poland**
Andreas Gramzow (PhD)
2009, 203 Seiten, ISBN 978-3-938584-41-5

Vol. 52 **Multi-level Processes of Integration and Disintegration. Proceedings of the Third Green Week Scientific Conference**
ed. by Franziska Schaft, Alfons Balmann
2009, 216 Seiten, ISBN 978-3-938584-42-2

Vol. 53 **Zur Bestimmung der Wettbewerbsfähigkeit des weißrussischen Milchsektors: Aussagefähigkeit von Wettbewerbsindikatoren und Entwicklung eines kohärenten Messungskonzepts**
Mikhail Ramanovich (PhD)
2010, 202 Seiten, ISBN 978-3-938584-44-6

Vol. 54 **Die Internationalisierung landwirtschaftlicher Unternehmen. Das Beispiel deutscher, dänischer und niederländischer Direkt-investitionen in den ukrainischen Agrarsektor**
Henriette Stange (PhD)
2010, 296 Seiten, ISBN 978-3-938584-45-3

Vol. 55 **Verhandlungsverhalten und Anspruchsanpassung im inter-nationalen Verhandlungsprozess: Die WTO-Agrarverhandlungen zum Abbau exportwettbewerbsfördernder Maßnahmen**
Ildiko Lajtos (PhD)
2010, 195 Seiten, ISBN 978-3-938584-48-4

Vol. 56 **Challenges of education and innovation. Proceedings of the Fourth Green Week Scientific Conference**
ed. by Kelly Labar, Martin Petrick, Gertrud Buchenrieder
2010, 155 Seiten, ISBN 978-3-938584-49-1

Vol. 57 **Agriculture in the Western Balkan Countries**
ed. by Tina Volk
2010, 249 Seiten, ISBN 978-3-938584-51-4

Vol. 58 **Perspectives on Institutional Change – Water Management in Europe**
ed. by Insa Theesfeld, Frauke Pirscher
2011, 127 Seiten, ISBN 978-3-938584-52-1

Vol. 59 **Der ukrainische Außenhandel mit Produkten der Agrar- und Ernährungswirtschaft: Eine quantitative Analyse aus Sicht traditioneller und neuer Außenhandelstheorien**
Inna Levkovych (PhD)
2011, 232 Seiten, ISBN 978-3-938584-53-8

Vol. 60 **Regional structural change in European agriculture: Effects of decoupling and EU accession**
Christoph Sahrbacher (PhD)
2011, 244 Seiten, ISBN 978-3-938584-58-3

Vol. 61 **Structural Change in Agriculture and Rural Livelihoods: Policy Implications for the New Member States of the European Union**
ed. by Judith Möllers, Gertrud Buchenrieder, Csaba Csáki
2011, 247 Seiten, ISBN 978-3-938584-59-0

Vol. 62 **Improving the functioning of the rural financial markets of Armenia**
Milada Kasarjyan (PhD)
2011, 121 Seiten, ISBN 978-3-938584-60-6

Vol. 63 **Integrierte Strukturen im Agrar- und Ernährungssektor Russlands: Entstehungsgründe, Funktionsweise, Entwicklungsperspektiven und volkswirtschaftliche Auswirkungen**
Jürgen Wandel
2011, 758 Seiten, ISBN 978-3-938584-61-3

Vol. 64 **Goal Achievement in Supply Chain Networks – A Study of the Ukrainian Agri-Food Business**
Taras Gagalyuk (PhD)
2012, 204 Seiten, ISBN 978-3-938584-63-7

Vol. 65 **Impacts of CAP reforms on farm structures and performance disparities – An agent-based approach**
Amanda Sahrbacher (PhD)
2012, 284 Seiten, ISBN 978-3-938584-64-4

Vol. 66 **Land fragmentation and off-farm labor supply in China**
Lili Jia (PhD)
2012, 143 Seiten, ISBN 978-3-938584-65-1

Vol. 67 **Ausprägung interregionaler Disparitäten und Ansätze zur Entwicklung ländlicher Räume in Mittel- und Osteuropa**
Sabine Baum (PhD)
2012, 214 Seiten, ISBN 978-3-938584-68-2

Vol. 68 **Patterns Behind Rural Success Stories in the European Union: Major Lessons of Former Enlargements**
ed. by Axel Wolz, Carmen Hubbard, Judith Möllers, Matthew Gorton, Gertrud Buchenrieder
2012, 190 Seiten, ISBN 978-3-938584-69-9

Vol. 69 **Motives for remitting from Germany to Kosovo**
Wiebke Meyer (PhD)
2012, 142 Seiten, ISBN 978-3-938584-70-5

Vol. 70 **Effizienz russischer Geflügelfleischproduzenten: Entwicklung und Determinanten**
Elena Epelstejn (PhD)
2013, 272 Seiten, ISBN 978-3-938584-72-9

Vol. 71 **Econometric impact assessment of the Common Agricultural Policy in East German agriculture**
Patrick Zier (PhD)
2013, 172 Seiten, ISBN 978-3-938584-73-6

Vol. 72 **Determinants of non-farm entrepreneurial intentions in a transitional context: Evidence from rural Bulgaria**
Diana Traikova (PhD)
2013, 136 Seiten, ISBN 978-3-938584-75-0

Vol. 73 **Human capital differences or labor market discrimination? The occupational outcomes of ethnic minorities in rural Guizhou (China)**
Bente Castro Campos (PhD)
2013, 266 Seiten, ISBN 978-3-938584-76-7

Vol. 74 **Identifying and understanding the patterns and processes of forest cover change in Albania and Kosovo**
Kuenda Laze (PhD)
2014, 152 Seiten, ISBN 978-3-938584-78-1

Vol. 75 **Flexibilität von Unternehmen. Eine theoretische und empirische Analyse**
Swetlana Renner (PhD)
2014, 194 Seiten, ISBN 978-3-938584-79-8

Vol. 76 **Impact of policy measures on wheat-to-bread supply chain during the global commodity price peaks: The case of Serbia**
Ivan Djuric (PhD)
2014, 160 Seiten, ISBN 978-3-938584-80-4

Vol. 77 **Marktwirtschaftliche Koordination: Möglichkeiten und Grenzen. Symposium anlässlich des 75. Geburtstages von Prof. Dr. Dr. h.c. mult. Ulrich Koester**
ed. by Jens-Peter Loy
2014, 94 Seiten, ISBN 978-3-938584-82-8

Vol. 78 **Participatory governance in rural development: Evidence from Ukraine**
Vasyl Kvartiuk (PhD)
2015, 200 Seiten, ISBN 978-3-938584-84-2

Vol. 79 **Agricultural transition in Post-Soviet Europe and Central Asia after 25 years. International workshop in honor of Professor Zvi Lerman**
ed. by Ayal Kimhi, Zvi Lerman
2015, 314 Seiten, ISBN 978-3-938584-95-8

Vol. 80 **Three essays on the Russian wheat export**
Zsombor Pall (PhD)
2015, 150 Seiten, ISBN 978-3-938584-86-6

Vol. 81 **Milchproduktion zwischen Pfadabhängigkeit und Pfadbrechung: Partizipative Analysen mit Hilfe des agentenbasierten Modells AgriPoliS**
Arlette Ostermeyer (PhD)
2015, 335 Seiten, ISBN 978-3-938584-88-0

Vol. 82 **Competitiveness and performance of EU agri-food chains**
ed. by Heinz Hockmann, Inna Levkovych, Aaron Grau
2016, ISBN 978-3-95992-006-3

Vol. 83 **Market uncertainty, project specificity and policy effects on bioenergy investments. A real options approach**
Lioudmila Chatalova (PhD)
2016, 202 Seiten, ISBN 978-3-95992-017-9

Vol. 84 **Too much but not enough: Issues of water management in Albania in light of climate change**
Klodjan Rama (PhD)
2016, 222 Seiten, ISBN 978-3-95992-034-6

Vol. 85 **Business groups in agriculture. Impact of ownership structures on performance: The case of Russia's agroholdings**
Andriy Matyukha (PhD)
2017, 128 Seiten, ISBN 978-3-95992-039-1

Vol. 86 **Russian demand for dietary quality: Nutrition transition, diet quality measurement, and health investment theory**
Christine Burggraf (PhD)
2017, 225 Seiten, ISBN 978-3-95992-040-7

Vol. 87 **Trade policy impacts on net wheat importers in the CIS: Three essays**
Nurzat Baisakova (PhD)
2017, 154 Seiten, ISBN 978-3-95992-041-4

Vol. 88 **Labour rationing of different farm types in Kazakhstan: A shadow price analysis**
Katharina Vantomme (PhD)
2017, 152 Seiten, ISBN 978-3-95992-043-8

Vol. 89 **Factor productivity in EU agriculture: A microeconometric perspective**
Mathias Kloss (PhD)
2017, 196 Seiten, ISBN 978-3-95992-054-4

In der Schriftenreihe *Studies on the Agricultural and Food Sector in Transition Economies* werden durch das IAMO Monographien und Tagungsberichte herausgegeben, die sich mit agrarökonomischen Fragestellungen zu Mittel- und Osteuropa beschäftigen. Wissenschaftlern, die in diesem Bereich forschen, steht die Schriftenreihe als Diskussionsforum offen.

In its series *Studies on the Agricultural and Food Sector in Transition Economies* IAMO publishes monographs and proceedings focusing on agricultural economic issues specific to Central and Eastern Europe. This series offers a forum to researchers studying this area.

ISSN 1436-221X
ISBN 978-3-95992-054-4