

Unobserved heterogeneity in productivity analysis of panel  
data: applications to meat chain firms and global growth in  
agriculture

Dissertation

zur Erlangung des Doktorgrades  
der Fakultät für Agrarwissenschaften  
der Georg-August-Universität Göttingen

vorgelegt von:

Jonathan Jakob Holtkamp  
geboren in Limburg an der Lahn

Göttingen, Februar 2015

D 7

1. Referent: Prof. Dr. Bernhard Brümmer

2. Korreferentin: JProf. Dr. Meike Wollni

Tag der mündlichen Prüfung: 12. Februar 2015



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Overview of research papers</b>	<b>4</b>
<b>3</b>	<b>Firm heterogeneity and divergent patterns of productivity change in European slaughtering and meat processing companies</b>	<b>7</b>
3.1	Introduction . . . . .	7
3.2	Econometric foundations . . . . .	8
3.3	Economic considerations . . . . .	12
3.4	Data and specification . . . . .	13
3.5	Results and discussion . . . . .	16
3.6	Conclusions . . . . .	24
<b>4</b>	<b>Productivity growth and potential intersectoral spillovers in the European meat supply chain</b>	<b>29</b>
4.1	Introduction . . . . .	29
4.2	Model of agricultural production . . . . .	30
4.3	Data and specification . . . . .	33
4.4	Model of meat chain productivity . . . . .	36
4.5	Results . . . . .	41
4.6	Discussion . . . . .	49
<b>5</b>	<b>Global TFP change in agriculture: consistent frontier estimation with country effects and time-varying inefficiency</b>	<b>59</b>
5.1	Introduction . . . . .	59
5.2	Estimation of the production frontier . . . . .	60
5.3	Data and specification . . . . .	63
5.4	Agricultural R&D and productivity . . . . .	68
5.5	Results . . . . .	70
5.6	Discussion . . . . .	76

<b>6</b>	<b>Discussion</b>	<b>86</b>
6.1	Findings on efficiency and productivity growth . . . . .	86
6.2	Consideration of unobserved heterogeneity . . . . .	90
<b>A</b>	<b>Stochastic Frontier Analysis using SFAMB for Ox</b>	<b>i</b>
A.1	Introduction . . . . .	i
A.2	Stochastic frontier production function estimation . . . . .	i
A.3	Data organisation and model formulation . . . . .	ix
A.4	Class member functions . . . . .	xi
A.5	Examples . . . . .	xvi
A.6	Future developments . . . . .	xxvii

## Acknowledgements

I'd like to start by thanking Prof. Dr. Bernhard Brümmer who made this project possible and offered so much guidance over the last years. This period has been associated with great experiences and has helped me develop a deeper insight into economics. It would have been far more difficult without his encouragement. His ideas and support have been invaluable in giving me motivation.

I'd like to thank Prof. Dr. Heinrich Hockmann and Prof. Dr. Inmaculada Martinez-Zarzoso for their helpful suggestions in the doctoral seminar. A separate acknowledgement must go to Mr. Hockmann who encouraged me to work on the topic of unobserved heterogeneity.

I'm grateful to all of the chair's colleagues for the jointly shared time. Notably, the exchange of ideas with the "efficiency colleagues" Amos, Daniel, Mary and Vivian has been a pleasure.

Thanks to Markus and Tinoush for joint organization and diverse assistance. The office time shared with Markus has been characterized by diverting conversations and varied debates. Both of us will look back on this time.

Further thanks are due for my colleagues from the agricultural policy – Carsten, Sebastian and Thelma – with whom working also was always a pleasure. Carsten's support in organizing the excursions to Brussels has been of great value.

I'm grateful to Jurij for his reliable data investigations and nice discussions.

I'm much obliged to JProf. Dr. Meike Wollni for taking over the second supervision as well as to Prof. Dr. Ludwig Theuvsen for joining the board of examiners.

Ultimately, I'm indebted to my family who have always encouraged me and I would like to dedicate my work to them. Special thanks apply to my parents and my wife Kerstin.

# 1 Introduction

The understanding of production technology is one of the basic interests of agricultural economists. While the field of research and the related questions are broad, technology remains an underlying element. Although the need for production analysis can differ with the specific research question, empirical evidence of technology and its development provides a reference to the researcher.

Productivity is a fundamental aspect of technology and economic activity. From a firm's perspective, productivity is required for being profitable. From an aggregate perspective, it ensures the efficient utilization of resources and it is a driver of market competition. Färe et al. (2008, p.522) call it "one of the most intuitive and familiar measures of performance at all levels ... [and a] ... key economic indicator".

Full potential of technology cannot be realized if production activities are inefficient. The notion of technical efficiency (TE) implies that deviations from standard assumptions of economic theory and the underlying technology ("best practice") are present in real world applications. Assessment of TE is relevant because it is a precondition for productivity and economic efficiency.<sup>1</sup>

The corresponding model of production is supposed to provide a proper representation of technology. Measurement of productivity change requires panel data that is preferential to cross-sectional data because there is more information on a single decision making unit (DMU). However, in the context of efficiency analysis, an important issue of identification arises. Panel data allow the isolation of time-invariant effects that capture individual characteristics (specific to a single DMU) and that are not connected to TE. These effects are assumed to be persistent while inefficiency can vary over time.

---

<sup>1</sup>Throughout the thesis, (technical) inefficiency is associated with a shortfall in physical output. Here, the relevant output-oriented TE score corresponds to the measure of Debreu who defined the "coefficient of resource utilization" (Debreu, 1951, p.285) to take values between zero and one. A related measure is the reciprocal (output-oriented TE as measured by Farrell (1957)) that indicates the potential expansion of output (with inputs fixed).

Interestingly, while inefficiency is a crucial issue in our applications, for Debreu (1951, p.285) it is "...not, by its very nature, the main concern of the economist". He rather focuses on market failures, such as monopolistic power, causing economic shortfall to an economy.

This distinction is of relevance because its neglect can have consequences for the magnitude and spread of TE scores. Accordingly, the potential improvements may be overestimated if inefficiency is confounded with unobserved heterogeneity.

This subject matters in the field of Stochastic Frontier Analysis (SFA), a methodology frequently applied to agriculture (e.g. Battese (1992), Bravo-Ureta et al. (2007)) and other sectors (e.g. Fried et al. (2008)). The majority of studies do not distinguish between inefficiency and unobserved heterogeneity but the number of alternative studies have increased since Greene (2005) proposed a ‘true’ fixed effects SF model. However, Greene’s approach does not provide a satisfying solution because the model’s error variance cannot be estimated consistently. An innovation by Chen et al. (2014) solves this problem and renders estimation of the fixed effects SF model possible.

This thesis comprises applications that are of interest in agricultural economics. Besides agricultural production at the firm-level as well as at the country-level, data of European meat firms is also analysed. The ultimate objective is to learn about (total factor) productivity and (technical) efficiency and to better understand the developments in these sectors. The results can also serve as a reference in the context of other studies addressing performance at the industry-level (e.g. Timmer et al. (2012)) or those using production data of the Food and Agriculture Organization. Furthermore, the discussion aims at evaluating the role of unobserved heterogeneity in the applications. The relevance or interpretation of individual effects and their consequences for TE differ with the underlying framework.

The structure of the thesis is as follows. The next section presents an overview of the studies. The first paper (Section 3) investigates productivity change in European meat enterprises. Potential productivity spillovers within the European meat chain are analysed in Section 4. Agricultural productivity on a global scale is the subject of the third paper laid out in Section 5. Section 6 consolidates the key findings of the specific applications and critically discusses the underlying assumptions as well as the role of unobserved heterogeneity. Appendix A provides a guide to the econometric methods and



their recent implementation in computer software.

## References

- Battese, G. E. (1992). Frontier production functions and technical efficiency: a survey of empirical applications in agricultural economics. *Agricultural Economics* : 185–208.
- Bravo-Ureta, B. E., Solís, D., Moreira López, V. H., Maripani, J. F., Thiam, A. and Rivas, T. (2007). Technical efficiency in farming: a meta-regression analysis. *Journal of Productivity Analysis* 27: 57–72, doi: 10.1007/s11123-006-0025-3.
- Chen, Y.-Y., Schmidt, P. and Wang, H.-J. (2014). Consistent estimation of the fixed effects stochastic frontier model. *Journal of Econometrics* 181: 65–76, doi:10.1016/j.jeconom.2013.05.009.
- Debreu, G. (1951). The Coefficient of Resource Utilization. *Econometrica* 19: 273–292.
- Färe, R., Grosskopf, S. and Margaritis, D. (2008). Efficiency and Productivity: Malmquist and More. In Fried, H. O., Lovell, K. C. and Schmidt, S. S. (eds), *The Measurement of Productive Efficiency and Productivity Growth*. Oxford University Press.
- Farrell, M. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society* 120: 253–290.
- Fried, H. O., Lovell, K. C. and Schmidt, S. S. (2008). Efficiency and Productivity. In Fried, H. O., Lovell, K. C. and Schmidt, S. S. (eds), *The Measurement of Productive Efficiency and Productivity Growth*. Oxford University Press.
- Greene, W. H. (2005). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics* 126: 269–303, doi:10.1016/j.jeconom.2004.05.003.
- Timmer, M. P., Inklaar, R., O’Mahony, M. and Van Ark, B. (2012). *Economic growth in Europe: a comparative industry perspective*. Cambridge University Press.

## 2 Overview of research papers

### **Firm heterogeneity and divergent patterns of productivity change in European slaughtering and meat processing companies**

The European meat industry is highly export-oriented, and characterised by dynamic firm growth as well as structural change. However, at an aggregate level total factor productivity change (TFPC) seems to be modest. Although relevant at all stages of the meat supply chain, efficiency and productivity studies for slaughterhouses and meat processors are rarely found.

The most appropriate scale for gaining insights into the dynamics at this stage of the supply chain is the firm level. However, the variables that describe the production process can be highly aggregated, as it is the case with this data set. The data are lacking additional information on firm characteristics. Accordingly, an empirical challenge is caused by unobserved heterogeneity that conflicts with the standard stochastic frontier (SF) model.

In view of this lack of information, we apply a SF model that allows to distinguish technical inefficiency from individual fixed effects. The estimation procedure is based on a novel approach that prevents results from being affected by the incidental parameters problem. Besides the standard decomposition of TFPC change, we propose a meaningful economic interpretation of the estimated firm effects.

We find that the panel SF model outperforms the standard model. Our results reveal that firms are working efficiently but their growth paths are highly divergent. Since we consider the industry classification of the meat firms, we can conclude that the productivity of firms engaged in slaughtering tends to stagnate. Conversely, meat processing firms are progressing over the observation period. Firms within this group can profit from higher levels of basic productivity that we associate with the estimated effects.

## **Productivity growth and potential intersectoral spillovers in the European meat supply chain**

Meat production in the European Union is not only characterised by the overall export orientation, but also by the increasing relevance of supply chains. Stronger integration between firms and closer interactions between agriculture and the meat industry are the consequences. Moreover, there are regions in Europe that possess high livestock production or meat manufacturing, or both.

The respective specialisations in production give rise to the investigation of meat clusters and spillovers. Productivity spillovers possibly exist within as well as between sectors of the meat supply chain. However, empirical evidence is rare. Available studies mostly focus on small geographic areas and specific products, respectively.

A more comprehensive analysis requires data on agriculture, the meat industry as well as geographical information. Our data sets can provide information on production activities and regional relationships. The analysis accounts for technical inefficiency that is distinguished from unobserved heterogeneity. We obtain firm-specific TFPC rates, and aggregate these rates into group measures that take the economic relevance of the individuals into account. The aggregated TFPC rates are used in a supply chain model in order to investigate intersectoral productivity spillovers.

Our results show that the output of the livestock sector could be considerably expanded. While there is progress in both sectors, the findings on technical change in agriculture are ambiguous. The regional rates of productivity growth are similar, on average, but less dispersed in case of the meat industry. We find indications for clustering within sectors, whereas the results on intersectoral relationships are vague.

## **Global TFP change in agriculture: consistent frontier estimation with country effects and time-varying inefficiency**

Agriculture is associated with challenges of global reach, such as food security or environmental protection. Accordingly, there is an area of conflict due to

high expectations and opposing objectives. Anyway, growth in TFP is a key aspect in meeting these challenges.

In this context, the notion of technical inefficiency is important because its assessment allows to draw conclusions on unused production potentials. However, the available country-level data are highly aggregated, and efficiency studies do not account for country characteristics by default. Other studies simply ignore inefficiency.

We investigate potential insights which might be due to the differentiation between unobserved country characteristics and technical inefficiency. The data which we analyse are specific because many countries as well as the input variable feed are included. Because of data limitations, this input has been predominantly neglected before. We decompose TFPC into three components and check for their respective relevance in different geographical regions.

The consideration of feed in the world production frontier highlights that TFPC is inevitably overestimated when inputs are neglected. Our results suggest that the level and the variation of predicted TE scores are heavily affected if the analysis accounts for individual effects. Accordingly, studies that ignore these unobserved attributes must result in biased TE measures, and will overestimate production potentials. We weight the results according to the countries' economic relevance and find TC to be the main source of TFPC. However, the other components are important at a regional level as well. The findings confirm increasing TFPC rates over time and indicate a positive relationship between TC and R&D efforts in OECD countries. Databases on agricultural R&D must be improved.

# 3 Firm heterogeneity and divergent patterns of productivity change in European slaughtering and meat processing companies

*Jonathan Holtkamp, Bernhard Brümmer*

## 3.1 Introduction

The food industry is an important sector of total manufacturing in the European Union (EU) and the meat industry is one of the major segments within the food industry. The meat industry became export-oriented as the EU turned from a net importer to a net exporter of pork and poultry. Per capita consumption is high but stagnant while there is an increasing demand in many other regions of the world. Given this situation, efficiency and productivity have implications for the industry as well as for the whole meat chain. For meat firms, higher productivity enhances their competitiveness in the world market. Agricultural suppliers depend on the demand of the industry. In particular, those who are characterized by specialization in livestock farming don't have much flexibility to switch to other production alternatives.

Literature on the European meat industry usually addresses consumer preferences, quality issues or strategic business aspects. Available efficiency and productivity studies mostly focus on the agricultural sector in Europe whereas corresponding studies for the meat industry use data from the United States (Ball and Chambers (1982); Lambert (1994); Macdonald and Ollinger (2005); Ollinger et al. (2005)). However, there are empirical results that indicate poor development in Europe. A recent study of Spanish meat firms (Kapelko et al., 2012) found negative growth in total factor productivity (TFP) in that sector at large. Another report by USDA (Fuglie et al., 2011) analysed TFP growth for the euro area and other developed countries over the period 1980 to 2006. It showed that TFP growth has been far larger in agriculture and total manufacturing than in food manufacturing.

Firm level data provides the most appropriate scale for gaining insights

into the dynamics of the European meat industry. A corresponding analysis may face challenges from firm heterogeneity that conflicts with the standard stochastic frontier (SF) model. Potential firm heterogeneity arises from two aspects. The first aspect refers directly to the structure of the industry and the firms. Market concentration varies between countries as well as between sub-sectors. Meat firms differ with regard to the range of products as well as their internal organization, i.e. firms are (vertically) integrated in the respective supply chain but to a different extent (Wijnands et al. (2008); Trienekens et al. (2009)). The second aspect relates to our data. The data contain almost no covariates that describe relevant firm characteristics and we are not able to adequately determine corresponding differences between firms. In view of this lack of information, we use a fixed effects approach to account for unobserved heterogeneity.

Our study is innovative because data on European meat firms has hardly been analysed in an efficiency and productivity context. We employ a recent fixed effects model (Chen et al., 2014) where estimation is free of incidental parameters. Furthermore, we seek to interpret the estimated firm effects in an economically meaningful manner and explore their relationship to individual productivity growth.

The structure of the paper is as follows. In the next two sections, we introduce the econometric foundations and economic considerations regarding the individual effects. We describe our data and specification in section 3.4 and present the empirical results in section 3.5. Based on these findings, we draw our conclusions which are laid out in section 3.6.

## **3.2 Econometric foundations**

### **Unobserved heterogeneity**

Firms are assumed to be identical units in a basic framework of efficiency analysis. This perspective conflicts with the intuition that there are usually some systematic differences between firms. The consideration of heterogeneity has induced several extensions of the standard SF model (see Greene (2008) for a survey).

This issue is particularly relevant if there is a missing variables problem, i.e. firm heterogeneity is not reflected in the models variables because data is aggregated and/or additional descriptive firm information is lacking. Consequently, this kind of heterogeneity is said to be unobservable and can be interpreted as unmeasured firm characteristics. Available panel data provide additional information on each individual and enable the isolation of time-invariant effects. Since many firm characteristics are presumably (quasi-)fixed it is plausible to assume that these effects capture unobserved heterogeneity. The respective specification of the SF model for panel data is given by:

$$y_{it} = \alpha_i + \beta'x_{it} + v_{it} - u_{it} \quad (3.1)$$

On the left hand side,  $y_{it}$  is a vector of the output of firm  $i$  ( $i=1, \dots, N$ ) in period  $t$  ( $t=1, \dots, T$ ). On the right hand side,  $x_{it}$  is a set of inputs that produce output  $y_{it}$ , and the vector  $\beta$  describes technology parameters to be estimated. The well-known composed error term is given by  $\epsilon_{it} = v_{it} - u_{it}$ , where  $v_{it}$  represents statistical noise and the non-negative component  $u_{it}$  represents inefficiency. In the frequently used normal-half normal model, the distributional assumptions are  $v_{it} \sim N(0, \sigma_v^2)$  and  $u_{it} \sim N^+(0, \sigma_u^2)$ . Unobserved heterogeneity in terms of time-invariant effects is captured by the vector (of incidental parameters)  $\alpha_i$ , i.e. the number of individual intercepts is equal to the number of firms. Replacing  $\alpha_i$  by a common intercept  $\alpha$  yields the so called “pooled model” where the panel structure of the data is irrelevant.

Regarding the model in (3.1), an identification problem arises because it is essential to disentangle firm effects from the error component  $u_{it}$ . An additional challenge results from the specification of time-varying inefficiency. A constant level of inefficiency would be unlikely to prevail over several years for a number of reasons (e.g. because of learning).

Estimation of the model given in (3.1) proves to be a challenge. This specification is known as the “true’ fixed effects” (TFE) model since Greene (2005) proposed to simply include the dummies into the classical SF model. The model is frequently applied but suffers from the incidental parameters problem. Its likelihood function contains the firm effects and the number of

parameters increases with the sample size. As a consequence, the estimated error variance, and hence, the parameter  $\sigma_u^2$  is inconsistent.

Alternative estimation strategies have been proposed. For example, another MLE approach by Wang and Ho (2010) employs within-transformation. This procedure is well-known in the panel data literature:

$$\tilde{z}_{it} = z_{it} - \bar{z}_i \quad (3.2)$$

For each panel  $i$  and any corresponding variable ( $z$ ), the individual mean ( $\bar{z}_i$ ) is subtracted from the observed value in period  $t$  ( $z_{it}$ ). Accordingly, the model can be represented by using deviations from means ( $\tilde{z}_{it}$ ). The resulting formulation is free of the  $\alpha_i$ :

$$\tilde{y}_{it} = \beta' \tilde{x}_{it} + \tilde{v}_{it} - \tilde{u}_{it} \quad (3.3)$$

While this strategy eliminates the  $\alpha_i$ , it does not directly provide a solution for the specification of time-varying inefficiency. For this purpose, Wang and Ho (2010) use a scaling function,  $u_{it} = u_i^* \times h_{it}$ , and a multivariate normal distribution for the noise component. With regard to the TFE specification, this solution is a compromise because  $u_i^*$  is a persistent component. An additional cost of this approach lies in the need to have time-varying covariates included in the scaling function  $h_{it}$ .

### Consistent fixed effects model

Consistent estimation of the fixed effects SF model given in equation (3.1) is demonstrated by Chen et al. (2014). Their solution is also based on deviations from means so that the transformed model looks like equation (3.3). The approach is characterized by the following features. First, within-transformation removes the incidental parameters. Second, the relevant likelihood function is derived from the  $T-1$  deviations. This procedure achieves an implicit correction of the error variance<sup>2</sup>. Third, the model is based on a more general

---

<sup>2</sup>With regards to the degrees of freedom, the correction accounts for the  $N$  individuals:  $df = NT - N - K = N(T - 1) - K$ .



distributional theory that allows this approach to maintain the specification of firm-specific and time-varying inefficiency  $u_{it}$ .

The Closed Skew Normal (CSN) distribution is suitable in the SF context<sup>3</sup>. The normal half-normal SF model has a skew normal distribution due to its parameter  $\lambda$ . While the skew normal distribution is a generalization of the normal distribution, it can be generalized itself by using the CSN distribution. Accordingly, the distribution of the composed error may be written as:

$$\epsilon_{it} \sim CSN_{1,1}(0, \sigma^2, -\frac{\lambda}{\sigma}, 0, 1) \quad (3.4)$$

The density of a  $CSN_{p,q}$ -distribution includes a  $p$ -dimensional pdf and a  $q$ -dimensional cdf of a normal distribution. The five associated parameters describe location, scale and skewness, as well as the mean vector and covariance matrix in the cdf. With panel data, the  $T$ -dimensional vector  $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{iT})'$  is distributed as:

$$\epsilon_i \sim CSN_{T,T}(0_T, \sigma^2 I_T, -\frac{\lambda}{\sigma} I_T, 0_T, I_T) \quad (3.5)$$

where  $I$  is the identity matrix. This vector is partitioned into linear combinations: its mean  $\bar{\epsilon}_i$  and its first  $T-1$  deviations  $\tilde{\epsilon}_i^* = (\tilde{\epsilon}_{i1}, \dots, \tilde{\epsilon}_{i,T-1})'$ . The latter are used to derive the so called “within MLE”. The respective likelihood function includes the conventional SF parameters  $\beta$ ,  $\lambda$  and  $\sigma^2$ , and is not subject to the incidental parameters problem. In order to assess technical efficiency, the point estimator of Battese and Coelli (1988) can be used. Therefore, the composed error has to be recovered:

$$\epsilon_{it} = y_{it} - \hat{y}_{it} = y_{it} - \hat{\beta}' x_{it} - \hat{\alpha}_i \quad (3.6)$$

Here, the way  $\hat{\alpha}_i$  is calculated is labelled as the mean-adjusted estimate by Chen et al. (2014):

$$\hat{\alpha}_i^M = \bar{y}_i - \hat{\beta}' \bar{x}_i + \sqrt{\frac{2}{\pi}} \hat{\sigma}_u \quad (3.7)$$

---

<sup>3</sup>The CSN distribution is not only used by Chen et al. (2014). For example, see Brorsen and Kim (2013) for a plain introduction to the CSN distribution in the SF context.

### 3.3 Economic considerations

#### Interpretation of alpha

In efficiency analysis, the application of fixed effects models is primarily motivated by solving the identification problem between unobserved heterogeneity and inefficiency. Although the estimated firm effects comprise individual information their economic interpretation is usually neglected. Firm heterogeneity –as specified in this conceptual framework– captures unmeasured attributes and constitutes an individual shift parameter. The latter might provide an appropriate linkage to the actual level of productivity. All other things being equal, a higher estimate for the fixed effect will be equivalent to a higher productivity level. In this sense, the estimated  $\hat{\alpha}_i$  reflects the “basic productivity” of firm  $i$ . It is not related to inefficiency but its magnitude presumably has implications for change of total factor productivity.

Our reasoning is similar to the idea of convergence that is extensively explored in the (productivity) literature. This concept focuses on initial productivity levels. Convergence prevails if firms (or regions or countries) that possess low initial productivity levels achieve higher growth rates than firms that initially showed higher productivity levels, i.e. the former catch up. The opposite phenomenon is called divergence. Potential reasons for convergence are higher returns to capital, lower cost of technology adoption, lower institutional constraints and learning from the mistakes of competitors (Ruttan, 2001, p.16). Corresponding studies mostly focus on geographical and sectoral contexts (e.g. Gutierrez (2000); Bernardini Papalia and Bertarelli (2009)) but use micro-level data less frequently (e.g. Mugera et al. (2012)).

An analytical framework that does not account for firm heterogeneity inevitably yields a biased measure of initial TFP. Hence, the use of a fixed effects approach has two advantages. Firstly, it accounts for unobserved heterogeneity and secondly, the estimated firm effects are interpretable as individual basic productivity. The corresponding interpretation is analogous to the convergence perspective but no additional measures are required.

The fixed effects assumption suggests correlation between the regressors and  $\alpha_i$  where the direction is not clear. However, estimation is free of the  $\alpha_i$ s

and the calculation of TFP change (TFPC) is based on changes. Conversely, calculation of the  $\alpha_i$ s is based on (mean) levels of regressors. Thus, the fixed effects assumption is unlikely to affect our interpretation. In our study, we explore how the individual firm effect is related to productivity change.

### 3.4 Data and specification

#### Data

The study uses accounting data on slaughtering and meat processing companies in Europe which are obtained from the Amadeus database (Bureau van Dijk, 2011). The 398 firms in the sample are classified according to their major production activities using NACE codes.<sup>4</sup> This industry classification distinguishes between slaughtering (NACE 1011), poultry slaughtering (NACE 1012) and meat processing firms (NACE 1013). Eurostat (2008, p.21) uses this scheme, accounting for activities that “share a common process for producing goods or services, using similar technologies”. We use data with NACE codes 1011 and 1013. Table 3.1 shows the composition of the sample that is observed over the period 2002 to 2009. Each firm has at least 50 employees and is observed for at least 4 years (on average for 6.5 years).

Table 3.1: Number of firms and observations according to country and classification.

Country	Slaughtering		Processing		Sum	
	Firms	Obs.	Firms	Obs.	Firms	Obs.
Austria	1	5			1	5
Belgium	6	44	9	70	15	114
Finland	2	12	10	71	12	83
France	47	330	41	247	88	577
Germany	4	19	25	118	29	137
Italy	24	171	47	322	71	493
Portugal	15	65	17	74	32	139
Spain	53	384	77	557	130	941
Sweden	8	52	12	80	20	132
Sum	160	1082	238	1539	398	2621

The companies are located in nine of the “old” member states of the European Union (EU); the majority (75%) in France, Italy and Spain. Pro-

<sup>4</sup>Nomenclature statistique des activités économiques dans la Communauté européenne.

cessing companies are more frequently observed than slaughtering firms. The sample covers the most important countries that are major producers in the European meat industry. However, Germany is represented sparsely and the United Kingdom is missing. In the meat industry, medium-sized and large firms (50-249 employees and more than 250 employees) generally account for higher revenue shares than small firms. In 2008, Eurostat (2011) reported the existence of roughly 1700 corresponding meat firms for the countries given in Table 3.1. Therefore, the sample covers approximately a quarter of the firms in the respective size categories.

Table 3.2 shows descriptive statistics of average firm data at constant 2002 prices.<sup>5</sup> Output is measured in terms of revenue while the inputs are measured as labour costs, cost of materials and depreciation. Average revenue is roughly 45 million euros per firm which seems to be a typical figure in the industry since it is close to official statistics.<sup>6</sup> The standard deviations (given in parentheses) demonstrate that the scale of operations is widespread among firms. Furthermore, there are differences according to industry classification. Slaughtering firms tend to be the larger units and have the higher intensity of material. In contrast, processing firms are comparatively labour- and capital-intensive. Revenue growth indicates expansion that is in line with the observed development of meat production and exports.

The data set is relatively large but the variables are highly aggregated. Additional information on firm characteristics which could be economically relevant is scarce (e.g. output composition, firm organization).

## Specification

The data set probably suffers from missing variables, and hence, it seems reasonable to apply models which take unobserved effects into account. Efficiency analysis requires assumptions on the time path of inefficiency. Since our panel has a maximum length of ( $T=$ ) 8 years and we are interested in

---

<sup>5</sup>We use indices of industrial producer prices, agricultural output prices and labour costs in manufacturing to deflate the data (Eurostat, 2011).

<sup>6</sup>Eurostat (2011) reports numbers of firms as well as sectoral revenue according to firm size. A rough calculation gives a mean value of 50 million euros per firm (medium-sized and large) for the nine countries (range 18-69 million euros at 2002 prices).

Table 3.2: Descriptive statistics of firm-specific means.

	Slaughtering (firms = 160)	Processing (firms = 238)
Revenue	52499 (52792)	42114 (61463)
Labour costs	3668 (3994)	4560 (6219)
Material costs	41780 (42143)	25309 (37101)
Depreciation	736 (760)	1306 (2384)

Note: values in thousands of euros; standard deviations in parentheses.

analysing developments, the assumption of time-varying inefficiency is essential. Therefore, we do not use the FE model of Schmidt and Sickles (1984) that allows only a very restrictive interpretation.<sup>7</sup> In fact, we apply the two models described in section 3.2 because the CFE model allows inefficiency to vary over time. The pooled model does not account for firm effects but it is a classical starting point in empirical analysis and may provide a reference for assessing the potential heterogeneity bias.

Pooling of data and estimation of a common frontier are suitable if all meat firms share the same technology. In our case, this approach would not be appropriate because we have a priori information about main activities in terms of industry classification. Furthermore, descriptive statistics indicate differences in technology between groups. Since an LR test rejects pooling all firms into one sample, we split the panel into two subsets (according to NACE classification) and estimate a single group frontier for each subset.<sup>8</sup>

The preferred functional form is the translog specification. We test for (and can reject) the Cobb-Douglas form for each of the two groups.<sup>9</sup> To allow for non-neutral technical change we include a time trend as an additional

<sup>7</sup>It could be used as a benchmark regarding the technology parameters. This was the intention of Abdulai and Tietje (2007) who applied several models to a panel of dairy farms.

<sup>8</sup> $H_0$  : Pooling all firms is appropriate. Reject with test statistic  $\chi^* = 454 > \chi^2_{((2-1)*16; \alpha=0.01)} = 32$

<sup>9</sup> $H_0$  : Coefficients of square and cross terms = 0. Reject with test statistics  $\chi^* = 238; 162 > \chi^2_{(10; \alpha=0.01)} = 23.2$

input. The resulting specification is:

$$\begin{aligned} \ln y_{it} = & \alpha_i + \sum_{j=1}^3 \beta_j \ln x_{jit} + \frac{1}{2} \sum_{j=1}^3 \sum_{l=1}^3 \beta_{jl} \ln x_{jit} \ln x_{lit} \\ & + \beta_t t + \frac{1}{2} \beta_{tt} t^2 + \sum_{j=1}^3 \beta_{tj} \ln x_{jit} t + v_{it} - u_{it} \end{aligned} \quad (3.8)$$

where  $\alpha_i$  reduces to  $\alpha_0$  in case of the cross-sectional model.

### 3.5 Results and discussion

#### Model selection

Since we use an effects framework in dealing with unobserved heterogeneity, we have to decide whether to choose a random or fixed effects approach. If there are correlations between individual effects and explanatory variables, a random effects specification is not appropriate. At first, we estimate a translog production function in the corresponding FGLS and LSDV specifications. In order to test the respective null hypothesis<sup>10</sup>, we conduct a Hausman test. As depicted in Table 3.3 the null hypothesis is rejected for each group, and hence, a fixed effects approach is required.

Table 3.3: Model selection, Hausman tests and Vuong tests for each group.

Group	Hausman	CFE model		Pooled model		Vuong
		ln L	AIC	ln L	AIC	
Slaughtering	51.25***	1889	-3.46	1328	-2.42	6.5***
Processing	147.18***	2301	-2.97	1590	-2.04	12.5***

Note: Critical values: Hausman  $\chi_{14; \alpha=0.01}^2 = 29.1$ ; Vuong  $N_{\alpha=0.01} = 2.3$   
 In case of the CFE model, the values are based on the within-likelihood.

A more sophisticated model should possess an increased explanatory power. Accordingly, the CFE model is expected to be superior to the pooled model. Since the likelihood functions differ, it is not possible to use a standard LR test in evaluating the models. We follow Lai and Huang (2010) in using a Vuong test for the assessment of two competing models (model families) in the SF

<sup>10</sup> $H_0$  : There is no correlation between unobserved effects and explanatory variables.

framework. The model that shows the smaller value of AIC is chosen as the baseline model and the null hypothesis states that both models are equivalent. Rejection means that the baseline model is closer to the true data generating process, i.e. it is superior to the alternative model. The LR based test statistic of Vuong follows a standard normal distribution, i.e.  $T_{LR}^V \sim N(0, 1)$ . Results are shown in Table 3.3. For each group, the CFE model shows the smaller value of AIC and the null hypothesis of the Vuong test is rejected. Therefore, we focus on the CFE model in our analysis.

### Coefficients

Estimation results of the two group frontiers are shown in Table 3.4. Since the variables are normalized at their respective sample means, the estimated coefficients can be directly interpreted as output elasticities. The frontiers satisfy monotonicity because these elasticities are positive and significant.<sup>11</sup> The estimated structure of production corresponds to the cost structure of meat enterprises since material is the most important input, followed by labour and capital. At the sample mean, the corresponding scale elasticity indicates decreasing returns to scale for processors (0.93), whereas slaughtering firms are close to constant returns to scale.<sup>12</sup>

The annual average rate of technical change differs drastically between groups. In fact, there is no progress for slaughtering firms whereas meat processors are progressing at a substantial rate that is slightly increasing over time. In case of the meat industry, it seems plausible to expect technical change that is factor-saving in labour because the industry tends to automate processes and reduce manual labour. However, we do not find this result; as technical change is virtually zero for the first group, the cross terms are less relevant. The fact that technical change is almost neutral for processors underlines differences in technology. Among other things, processing firms are likely to need more comparatively skilled workers.

<sup>11</sup>The positive estimates of the square terms suggest that there could be problems with respect to the usual regularity conditions.

<sup>12</sup>However, the hypothesis that the function exhibits constant returns to scale can be rejected.  $H_0 : \beta_1 = \beta_2 = \beta_3 = 1$ ; and  $\beta_{jl} = \beta_{lj} = 0 \quad \forall j, l$ ;  $\chi^* = 36 > \chi_{(4; \alpha=0.01)}^2 = 13.3$ .

Table 3.4: Estimation results of the CFE model.

	Slaughtering			Processing		
	Coeff.		Std.error	Coeff.		Std.error
log labour cost	0.153	***	0.010	0.173	***	0.012
log material cost	0.823	***	0.010	0.737	***	0.011
log depreciation	0.021	***	0.005	0.021	***	0.007
time trend	0.000		0.001	0.022	***	0.001
$0.5 \times (\log \text{labour cost})^2$	0.079	***	0.012	0.070	***	0.013
$0.5 \times (\log \text{material cost})^2$	0.169	***	0.013	0.119	***	0.017
$0.5 \times (\log \text{depreciation})^2$	0.020	***	0.004	0.027	***	0.008
$0.5 \times (\text{time trend})^2$	0.000		0.000	0.002	**	0.001
log labour cost $\times$ log material cost	-0.089	***	0.010	-0.037	***	0.012
log labour cost $\times$ log depreciation	0.022	***	0.008	-0.011		0.010
log material cost $\times$ log depreciation	-0.046	***	0.007	-0.048	***	0.009
log labour cost $\times$ time trend	0.005	***	0.001	0.003	*	0.002
log material cost $\times$ time trend	-0.002	**	0.001	-0.000		0.001
log depreciation $\times$ time trend	-0.003	***	0.001	-0.001		0.001
$\log \sigma_u^2$	-5.743	***	0.088	-5.580	***	0.118
$\log \sigma_v^2$	-8.109	***	0.207	-6.832	***	0.125
$\sigma_u$			0.057			0.061
$\sigma_v$			0.017			0.033
$\lambda$			3.263			1.870
Log likelihood ( $\ln L_W$ )			1889			2301
AIC			-3.46			-2.97
Observations			1082			1539

Note: estimation output of the pooled model is documented in the appendix.



## Technical efficiency

The descriptive statistics of technical efficiency (TE) scores are presented in Table 3.5. Both groups operate at high levels of efficiency. The range of point estimates are very similar suggesting that inefficiency is of less importance in this industry. Comparing TE scores within groups but between models we find that for slaughtering firms no inefficiency is present when using the pooled model. The share of white noise is smaller under the panel model that is able to identify some inefficiency. In the case of the processing firms, the conclusion drawn from the two models is the same but the CFE model indicates a slightly higher range of TE scores.

Table 3.5: Descriptive statistics of technical efficiency scores.

	Slaughtering		Processing	
	Pooled	CFE	Pooled	CFE
Mean	–	0.957	0.955	0.954
Std.dev	–	0.026	0.015	0.024
Min	–	0.72	0.79	0.71
Max	–	1.00	0.98	0.99

## Decomposition of productivity change

Analysis of TFPC and its components helps in the understanding of developments in European meat chains. Increasing productivity enhances competitiveness of the export-oriented industry as it accompanies lower average costs and higher profitability. The respective components characterize the nature of TFPC and reveal sources for potential improvements. We decompose TFPC into technical efficiency change (TEC), technical change (TC) and a scale component (SC) according to formula (8.2.6) of Kumbhakar and Lovell (2000, p.284).

Each component as well as TFPC is calculated as an index for every firm. We take the geometric mean over all observations in year  $t$  to obtain an index for each year. Results are shown in Table 3.6.

The estimates of TE possess moderate variances indicating a stable situa-

tion. This result is reflected in the two indices of TEC that remain virtually constant over the observation period. However, productivity is affected by the divergent rates of TC as well as the different scale elasticities. The scale elasticity of slaughtering firms is close to one so that the respective SC index essentially does not change. As TEC index and TC index are also constant the situation in this sub-sector appears to be static. In contrast, meat processors experience productivity growth that is driven by TC. Growth is slightly decelerated by the SC due to decreasing returns to scale and increasing input quantities.

Table 3.6: Decomposition of TFP change according to group frontiers.

Year	Slaughtering				Processing			
	TEC <sub>i</sub>	TC <sub>i</sub>	SC <sub>i</sub>	TFPC <sub>i</sub>	TEC <sub>i</sub>	TC <sub>i</sub>	SC <sub>i</sub>	TFPC <sub>i</sub>
2002	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2003	0.995	1.001	0.999	0.995	1.007	1.017	0.998	1.022
2004	0.996	1.001	0.998	0.995	1.001	1.034	0.991	1.025
2005	1.004	1.002	0.998	1.003	0.998	1.050	0.990	1.037
2006	1.011	1.003	0.999	1.012	1.003	1.067	0.988	1.057
2007	0.997	1.004	0.998	0.998	1.008	1.081	0.990	1.078
2008	0.982	1.005	0.998	0.985	0.997	1.106	0.987	1.087
2009	1.000	1.006	1.000	1.003	0.991	1.132	0.986	1.106

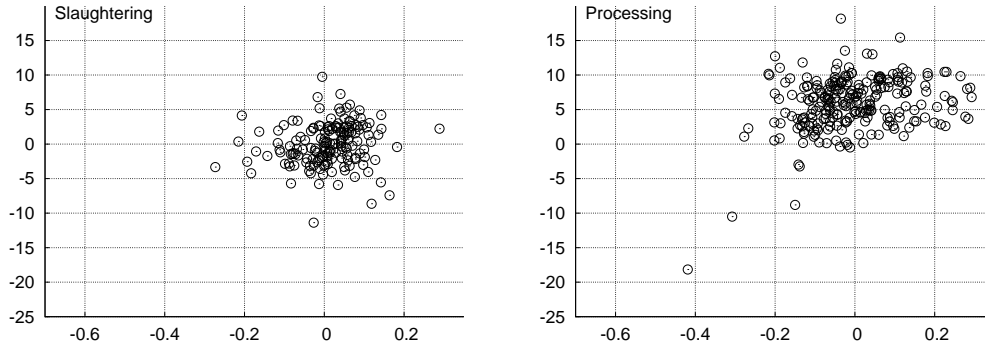
### Productivity change and basic productivity

In section 3.3, we argued that the fixed firm effects are interpretable as measures of basic productivity. In this section, we want to analyse how basic productivity relates to individual productivity growth. We use Figure 3.1 to visualize the corresponding (partial) perspective where the estimated  $\hat{\alpha}_i$ s are plotted on the horizontal axis. The vertical axis depicts the geometric mean of the firm-individual TFPC index. This number gives the average progress of each firm (in percentage terms) given its observation period  $T_i$ .

Figure 3.1 highlights the findings of the TFPC decomposition. It is more specific because individual developments are visible. The slaughtering group (left panel) is not able to keep up with the processing group (right panel). There are some slaughterers that achieve progress but the observations are scattered around zero. In the group of meat processors, more variation in

TFPC indices as well as in estimated  $\hat{\alpha}_i$ s is found. Almost all firms show positive productivity growth but there are also a few firms that are on a negative path.

The partial perspective of Figure 3.1 suggests that there is no interdependence of basic productivity and TFP growth in slaughtering. By contrast, it suggests a positive relationship in the second group, i.e. there seems to be divergence as higher  $\hat{\alpha}_i$ s accompany higher TFPC indices. Additionally, another interesting aspect appears from this simplified illustration. In the right panel, the variance of productivity change tends to decrease as the level of  $\hat{\alpha}_i$  increases. This observation suggests that higher basic productivity could, in a way, lead to stabilization of individual TFP growth.



Note: Vertical axis: geometric mean of firm-specific TFP change index. Horizontal axis: estimates of firm effects (basic productivity).

Figure 3.1: Individual mean of TFP change index and firm effects  $\hat{\alpha}_i$ .

The partial relationship of basic productivity and TFPC index can change if we control for additional factors. We use an OLS regression that includes variables possibly related to productivity growth (Table 3.7). The respective variables are firm size (measured in output quantity) and firm age as well as country dummies and interactions.

There are some country effects in both groups pointing out differences in TFPC. These country effects are the major source of variation for the slaughtering group. Firm size tends to be positively related to productivity change in slaughtering, but the coefficient is insignificant. In this group, age has no effect and the coefficient of basic productivity is negative. The situation is reversed in the case of the processing group. The coefficient of basic productivity as well as age are positively related to TFPC whereas firm

size has no effect. However, the benefits of basic productivity and age seem limited to some extent as the squared term and cross-term suggest decreasing effects.

Table 3.7: Results of OLS regression.

	Slaughtering		Processing		
	Coeff.	Std.error	Coeff.	Std.error	
constant	-5.97**	2.88	constant	5.02	5.25
firm effect	-7.48*	4.39	firm effect	14.82**	6.48
firm size	0.40	0.27	firm size	0.09	0.51
age	0.00	0.02	age	0.03**	0.02
			firm effect <sup>2</sup>	-60.97***	20.42
			firm effect $\times$ age	-0.29**	0.14
Belgium	3.75***	0.80	Belgium	2.36**	0.95
Germany	4.28**	1.86	Germany	-1.78**	0.83
Finland	5.35***	1.64	Finland	3.20***	0.90
France	3.51***	0.63	Italy	1.19*	0.71
Italy	2.63***	0.86	Sweden	-3.78***	0.81
Dep. var. (mean)		-0.09		5.88	
R-squared		0.25		0.33	
Number of firms		160		238	

Note: The dependent variable is the individual geometric mean of each firm's TFP change index. Country names indicate country dummies (with Spain as reference) where insignificant dummies are omitted. White-corrected standard errors are used.

## Discussion

Our results show that the two sub-sectors are characterized by both similar and dissimilar attributes. The observed meat firms handle their production technology confidently as they operate on a high and constant level of efficiency. Hence, meat firms have to be very competitive within the industry because inefficient units drop out quickly. The respective growth paths of TFP are divergent because we find two directions. Slaughtering firms stick to their productivity level, whereas the processing sector shows an active development.

The production environment of the slaughtering industry is static due to the absence of technical progress and inefficiency. This sub-sector is a pioneer of industrial manufacturing and characterized by mature technology where automation of processes has a long history (Perren, 2006). It seems that

even firm size is no advantage in this situation. Accordingly, potentials for improvement are likely to be restricted. Additional challenges arise not only from the increasing importance of regulations including hygiene standards (which apply to the food industry at large), but also from recent issues of animal welfare.

The processing sector faces incentives and the potential of process innovation driven by changing consumption patterns. Like other food items, meat products are increasingly consumed as convenience food. Meat processing was predominantly a business of small-scaled butcheries and private households in the past. This activity has successively shifted to the industry level and this development is likely to be a major reason for TC in that sub-sector. Convenience food often includes poultry meat that is increasingly consumed. This trend is possibly an additional factor that spurs the development of this production activity. As vertical integration tends to be rather strong in the poultry sector, organizational structures are likely to be important as well.

The measure of basic productivity that we derive directly from the individual effects is related to TFPC but its relevance differs between the sub-sectors. In general, this measure captures attributes or surrounding conditions that are not observable. However, higher levels of basic productivity will originate from certain advantages or spillover effects. In case of the meat industry, potential determinants include organizational structure (e.g. vertical integration), specialization (e.g. niches, delicacies) and a favourable environment (e.g. intensive animal husbandry, industry clusters).

Our results show that meat processing firms benefit from higher levels of basic productivity whereas it is rather irrelevant in slaughtering. For slaughterers, the respective coefficient suggests a catch up effect at best. Given their static situation, this interpretation is not very meaningful. Accordingly, the potential of the slaughtering industry appears to be exhausted and hence basic productivity is essentially of no advantage. Technology is presumably rather homogeneous in this segment. A high number of standard operations can be a potential explanation. Although standard operations prevail in any industrial manufacturing, the respective complexity might be comparatively

low in slaughtering.

The variance of estimated firm effects is an additional indication of similar conditions within this sector and results in similar levels of basic productivity (see Figure 3.1). In contrast, the variance of estimated firm effects is larger in meat processing and suggests that the spread of unobserved covariates that affect basic productivity is higher in this segment.

Processing firms are presumably more diverse in terms of their potential to adjust organization and product design. Firm size is of minor relevance in this context. In fact, within this group, the successful utilisation of potential is reflected in higher basic productivity that in turn contributes to higher TFP growth. Meat processing activities in particular facilitate the development of specific products. Therefore, meat processors are able to gain experience and develop skills which provide technical advances for those firms. Results show that experience pays off as older firms tend to be more productive.

### **3.6 Conclusions**

The meat industry is one of the major sectors of the European food industry. However, applied studies considering efficiency and productivity of slaughtering and meat processing companies are rarely found. We use firm level data to gain insights on the dynamics at this stage of the meat supply chain. As panel data are available but descriptive firm attributes are scarce, we take unobserved heterogeneity into account.

In this framework, we apply the recent fixed effects SF model of Chen et al. (2014). This model specification allows for unmeasured firm characteristics and estimation is free of incidental parameters. We know about the firms' main specializations and correspondingly estimate two stochastic group frontiers. In the observation period 2002 to 2009, inefficiency is a minor issue in the meat industry because average scores of TE reach more than 0.95 in each group. In fact, there is no change in TE but technical progress is the most important driver of TFPC. While in the processing segment there are positive growth rates of TC and TFPC (2.2% and 1.8% on average), firms in the slaughtering segment remain in a static situation. Hence, growth paths

of productivity differ considerably between production activities. This fact illustrates that an average measure for the whole industry can be misleading. Furthermore, we treat the estimated firm effects not only as ancillary parameters but explore their interpretation as measures of basic productivity. We find that basic productivity and TFPC are positively related in the sub-sector of meat processing.

Naturally, our study rests on assumptions that inevitably involve some limitations. Production activities are distinguished by a priori information that allows us to derive group frontiers. Although we account for unmeasured attributes, the approach is restricted to such a degree as potential heterogeneity in structural parameters within the groups is ignored. This issue might be less relevant in slaughtering where our results indicate that conditions are rather homogeneous. However, we can expect that processing firms have more production alternatives so that we find a more heterogeneous pattern there. In general, we cannot exclude the possibility that additional activities performed on the sidelines influence the technology of an individual firm. Considering these results and limitations, our study shows that the European meat industry is a (technically) efficient sector. This finding is certainly related to a comparatively mature technology as well as to high intra-sectoral competition.

Relevant insights are due to the consideration of sub-sectors. Analysing the meat industry as a whole, productivity growth appears to be small or even stagnant. In this regard, our results are similar to those of Fuglie et al. (2011) or Kapelko et al. (2012). That perspective hides the divergent developments that we find and that are consistent with observed trends in that industry. Potential in conventional slaughtering is obviously exhausted whereas in the processing segment there are prospects for development. This conclusion is consistent with the concentration of enterprises in the European slaughtering sector and increasing vertical integration (downstream), i.e. there are incentives to engage in meat processing activities.

Our interpretation of basic productivity appears convincing in that we find similar measures within the sub-sector of slaughtering where firms pre-

sumably are rather replaceable. The differences within meat processing are much larger, and it is plausible that unmeasured factors are more relevant and can have positive effects on basic productivity and TFP growth (e.g. specialization, marketing, spillover effects).

With regards to future research, the exploitation of available spatial information would appear to be a promising approach for potential identification of industrial clusters or spillover effects between agriculture and industry. An approach that allows for more flexibility in technology parameters would be another natural extension of our study.

## References

- Abdulai, A. and Tietje, H. (2007). Estimating technical efficiency under unobserved heterogeneity with stochastic frontier models: application to northern German dairy farms. *European Review of Agricultural Economics* 34: 393–416.
- Ball, E. V. and Chambers, R. G. (1982). An Economic Analysis of Technology in the Meat Products Industry. *American Journal of Agricultural Economics* 64: 699–709.
- Battese, G. E. and Coelli, T. J. (1988). Prediction of Firm-Level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data. *Journal of Econometrics* 38: 387–399.
- Bernardini Papalia, R. and Bertarelli, S. (2009). Decomposing productivity patterns in a conditional convergence framework. *Journal of Productivity Analysis* 31: 57–75.
- Brorsen, B. W. and Kim, T. (2013). Data aggregation in stochastic frontier models: The closed skew normal distribution. *Journal of Productivity Analysis* 39: 27–34.
- Bureau van Dijk (2011). AMADEUS Database. Bureau van Dijk Electronic Publishing.
- Chen, Y.-Y., Schmidt, P. and Wang, H.-J. (2014). Consistent estimation of the fixed effects stochastic frontier model. *Journal of Econometrics* 181: 65–76.
- Eurostat (2008). NACE Rev. 2 – Statistical classification of economic activities in the European community. Tech. rep., Eurostat, Methodologies and Working papers.



- Eurostat (2011). Structural business statistics. [http://epp.eurostat.ec.europa.eu/portal/page/portal/european\\_business/data/database](http://epp.eurostat.ec.europa.eu/portal/page/portal/european_business/data/database).
- Fuglie, K. O., Heisey, P. W., King, J. L., Pray, C. E., Day-Rubenstein, K., Schimmelpfennig, D., Wang, S. L. and Karmarkar-Deshmukh, R. (2011). Research Investments and Market Structure in the Food Processing , Agricultural Input , and Biofuel Industries Worldwide. Tech. rep., USDA.
- Greene, W. H. (2005). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics* 126: 269–303.
- Greene, W. H. (2008). The Econometric Approach to Efficiency Analysis. In Fried, H. O., Lovell, K. C. and Schmidt, S. S. (eds), *The Measurement of Productive Efficiency and Productivity Growth*. Oxford University Press.
- Gutierrez, L. (2000). Convergence in US and EU agriculture. *European Review of Agricultural Economics* 27: 187–206.
- Kapelko, M., Oude Lansink, A. and Stefanou, S. E. (2012). Dynamic Productivity Growth in the Spanish Meat Industry. European Association of Agricultural Economists Seminar, Prague.
- Kumbhakar, S. C. and Lovell, K. C. (2000). *Stochastic Frontier Analysis*. Cambridge University Press.
- Lai, H.-p. and Huang, C. J. (2010). Likelihood ratio tests for model selection of stochastic frontier models. *Journal of Productivity Analysis* 34: 3–13.
- Lambert, D. K. (1994). Technological Change in Meat and Poultry-Packing and Processing. *Journal of agricultural and applied economics* 26: 591–604.
- Macdonald, J. M. and Ollinger, M. E. (2005). Technology, Labor wars, and Producer Dynamics: explaining Consolidation in Beefpacking. *American Journal of Agricultural Economics* 87: 1020–1033.
- Mugera, A. W., Langemeier, M. R. and Featherstone, A. M. (2012). Labor productivity convergence in the Kansas farm sector: a three-stage procedure using data envelopment analysis and semiparametric regression analysis. *Journal of Productivity Analysis* 38: 63–79.
- Ollinger, M. E., Macdonald, J. M. and Madison, M. (2005). Technological Change and Economies of Scale in U.S. Poultry Processing. *American Journal of Agricultural Economics* 87: 116–129.
- Perren, R. (2006). *Trade, taste and technology: the development of the international meat industry since 1840*. Ashgate.
- Ruttan, V. W. (2001). *Technology, growth, and development: an induced innovation perspective*. Oxford University Press, New York.

- Schmidt, P. and Sickles, R. C. (1984). Production Frontiers and Panel Data. *Journal of Business & Economic Statistics* 2: 367–374.
- Trienekens, J., Petersen, B., Wognum, N. and Brinkmann, D. (2009). *European pork chains: Diversity and quality challenges in consumer-oriented production and distribution*. Wageningen Academic Publishers.
- Wang, H.-J. and Ho, C.-W. (2010). Estimating fixed-effect panel stochastic frontier models by model transformation. *Journal of Econometrics* 157: 286–296.
- Wijnands, J., Bremmers, H. J., Van der Meulen, B. M. and Poppe, K. J. (2008). An Economic and Legal Assessment of the EU Food Industry’s Competitiveness. *Agribusiness* 24: 417–439.

## Appendix

Table 3.8: Estimation results of the Pooled SF model.

	Slaughtering			Processing		
	Coeff.		Std.error	Coeff.		Std.error
constant	0.005		0.055	0.052	***	0.009
log labour cost	0.144	***	0.006	0.217	***	0.006
log material cost	0.834	***	0.005	0.693	***	0.006
log depreciation	0.032	***	0.005	0.101	***	0.005
time trend	-0.001		0.001	0.017	***	0.001
$0.5 \times (\log \text{labour cost})^2$	0.076	***	0.014	0.105	***	0.012
$0.5 \times (\log \text{material cost})^2$	0.138	***	0.010	0.116	***	0.011
$0.5 \times (\log \text{depreciation})^2$	0.028	***	0.006	0.034	***	0.007
$0.5 \times (\text{time trend})^2$	0.000		0.001	0.001		0.001
log labour cost $\times$ log material cost	-0.100	***	0.009	-0.114	***	0.009
log labour cost $\times$ log depreciation	0.014	*	0.009	0.006		0.007
log material cost $\times$ log depreciation	-0.044	***	0.007	-0.024	***	0.007
log labour cost $\times$ time trend	0.008	***	0.002	0.007	***	0.002
log material cost $\times$ time trend	0.000		0.002	0.002		0.002
log depreciation $\times$ time trend	-0.005	**	0.002	-0.006	***	0.002
$\log \sigma_u^2$	-15.402	***	304.4	-5.653	***	0.354
$\log \sigma_v^2$	-5.292	***	0.043	-5.091	***	0.079
$\sigma_u$		0.000			0.059	
$\sigma_v$		0.071			0.078	
$\lambda$		0.006			0.755	
Log likelihood		1328			2301	
AIC		-2.42			-2.97	
Observations		1082			1539	

# 4 Productivity growth and potential intersectoral spillovers in the European meat supply chain

*Jonathan Holtkamp, Bernhard Brümmer*

## 4.1 Introduction

The European Union (EU) is one of the major producers of meat in the world and a net exporter of pork and poultry. The production processes are increasingly characterised by supply chain features. However, there is much heterogeneity in the structural characteristics of both livestock production and the meat industry itself. For both sectors, regional clustering exists in several European countries. In the case of agriculture, regions with a high livestock density are located in Northwest Germany, Belgium, Northwest France, Eastern Spain and others. Clustering of meat firms prevails in regions like Belgium, Northern France and Northern Italy (at least in terms of number of firms, according to Eurostat (2014b)). Of course, there are overlaps, i.e. where there is a high livestock density and a high concentration of meat processing in the same region or in neighbouring regions.

Livestock production and meat processing are major sub-sectors within their respective branches of agriculture and the food industry. Their productivity growth is relevant because they use a lot of their sector's resources and it is a driver of their development (e.g. with respect to structural change). While the existence of clusters is obvious, it is less clear if there are related spillovers that result in an increased productivity growth within these clusters. This effect or similar consequences are found in the literature. For example, a review of aspects of spatial dynamics in dairy production is provided by Mosnier and Wieck (2010). Another recent study investigates the relevance of spatial clustering in organic farming (Schmidtner et al., 2012). Some studies are more closely related to our approach, including Roe et al. (2002), Larue et al. (2011) as well as Diez-Vial (2011).

The economic relevance of clustering and potential positive spillovers becomes manifest in individual benefits at the firm-level and overall economic gains due to productivity growth. However, the studies available typically address a smaller geographical region; partially, considering only very specific products. Although there is no clear-cut approach for analysing such a relationship, a broader perspective is desirable.

Our study evaluates growth in total factor productivity (TFP) and the relevance of its components for livestock farming and the meat industry. As efficiency change is a potential driver of TFP change (TFPC) we consider technical efficiency (TE). In both sectors, unobserved heterogeneity is likely to be an important aspect. We thus consider and distinguish farm (firm) effects from TE. We investigate potential spillovers at a larger scale, i.e. without restriction on a single region. The analysis doesn't only address the situation within one sector but rather tries to consider intersectoral relationships between livestock production and meat processing.

As separate data sets are available, production technologies are represented by means of a stochastic distance function (agriculture) and a stochastic production frontier (meat industry). We proceed as follows: Section 4.2 lays out the methods for modelling the production technologies, attributing more attention to agriculture. The two data sets and the corresponding specifications are explained in Section 4.3. Section 4.4 comprises the concept of chain productivity and the discussion of required aggregations. Results are presented in Section 4.5. The last section sums up and critically discusses our findings.

## 4.2 Model of agricultural production

### Stochastic production frontier with farm effects

Production analysis of agricultural panel data should address inefficiency as well as unobserved heterogeneity.<sup>13</sup> The corresponding formulation of the

---

<sup>13</sup>Relevant aspects also apply to the meat industry. For the sake of brevity, this section focuses on agricultural production.

stochastic production frontier is given by:

$$y_{it} = \alpha_i + \beta'x_{it} + v_{it} - u_{it} \quad (4.1)$$

The dependent variable is output,  $y$ , that is produced by a set of inputs,  $x$ . Their relationship is characterised by the technology parameters  $\beta$  (to be estimated). Farms are indexed by  $i = (1, 2, \dots, N)$  and time periods are indexed by  $t = (1, 2, \dots, T)$ . The time-invariant  $\alpha_i$  is an individual farm effect and the two remaining components represent the composed error term  $v_{it} - u_{it} = \epsilon_{it}$ . These two components are assumed to be independent of each other and their distributions are assumed to follow the normal-half normal model, i.e.  $v_{it} \sim iid N(0, \sigma_v^2)$  and  $u_{it} \sim iid N^+(0, \sigma_u^2)$ .

Several strategies for the estimation of this model have been proposed, starting with Greene (2005). The approach that is adopted here is a methodological extension by Chen et al. (2014). Their solution is based on the insight that the distribution of the composed error is a special case of the closed skew normal (CSN) distribution, i.e.  $\epsilon_{it} \sim CSN_{1,1}(0, \sigma^2, -\frac{\lambda}{\sigma}, 0, 1)$ . After within-transformation is applied to equation (4.1), the properties of the CSN distribution allow to derive a (“within-”) likelihood function that is based only on the transformed residuals. Therefore, the individual effects are eliminated and the estimation involves only the “conventional” parameters  $\beta$ ,  $\sigma_v^2$  and  $\sigma_u^2$ . The effects are relevant after estimation because the prediction of TE scores requires the calculation of  $\epsilon$  that depends on the effects,<sup>14</sup> given by:

$$\hat{\alpha}_i = \bar{y}_i - \hat{\beta}'\bar{x}_i + \sqrt{\frac{2}{\pi}} \hat{\sigma}_u \quad (4.2)$$

### Estimation of multiple output production

Agricultural production is a typical example for a multiple output production. The respective activities, even for very specialised livestock farms, are usually connected to the use of land. Livestock production inevitably produces manure that is used as organic fertilizer. Consequently, a livestock farm is likely

---

<sup>14</sup>For a more thorough description, see appendix A.

to possess at least some land area to dispose of its manure.<sup>15</sup> Many farms produce a portion of their feed input and/or field crops for sale.

The analysis includes farms that produce livestock as well as crops. The empirical model has to take this structure into account. Hence, a distance function approach is used. Following Kumbhakar and Lovell (2000, p.28ff) the output-oriented distance function can be written as:

$$D_O(x, y) = \min \left\{ \mu : \frac{y}{\mu} \in P(x) \right\} \quad (4.3)$$

where in this case,  $y$  represents the two kinds of output and  $x$  represents the observed input variables. The value of the distance function lies in the unit interval.<sup>16</sup>

Estimation of the distance function in the context of equation (4.1) requires some modification. For an efficient observation, the distance function can be expressed as (ignoring the noise component for now):

$$D_O(x, y e^u; \beta) = 1 = TE \quad (4.4)$$

The inefficiency component enters this equation with a positive sign. Accordingly, the (inefficient) observation  $y$  is scaled onto the frontier by means of  $e^u$  where the function value equals one.<sup>17</sup> Because  $\frac{1}{TE} = e^u$  it holds that:

$$- \ln(TE) = \ln(e^u) = -\ln(D_O(x, y; \beta)) = u \quad (4.5)$$

Adding a noise term, the stochastic distance function for an technical efficient observation is:

$$1 = D_O(x, y; \beta) \exp(u - v) \quad (4.6)$$

The above equations highlight that the value of the distance function is not a dependent variable (as  $u$  is not observed prior to estimation). In order to obtain an estimable model, we follow Brümmer et al. (2002) in using the

---

<sup>15</sup>Although manure can be sold, a farm usually requires some land to be classified as an agricultural enterprise.

<sup>16</sup>Its value corresponds to the *Debreu technical efficiency measure* used here. The relationship  $-\ln(D_O(\cdot)) = u$  holds irrespective of the TE definition.

<sup>17</sup>The underlying rationale is the distance function's property of linear homogeneity.

“ratio model”<sup>18</sup>, i.e. applying the normalisation by means of  $\frac{1}{y_m}$ :

$$\frac{1}{y_m} = D_O(x, \frac{y}{y_m}; \beta) \exp(u - v) \quad (4.7)$$

By taking logarithms and multiplying by minus one, the expression takes the form of a stochastic production frontier:

$$\ln(y_m) = -\ln D_O(x, \frac{y}{y_m}; \beta) + v - u \quad (4.8)$$

### 4.3 Data and specification

#### Data on agricultural production

This study uses data on European farms that are active in livestock production. The data come from the Farm Accountancy Data Network (FADN) (European Commission, 2010). This data set covers 14 countries<sup>19</sup> and covers the period between 1995 and 2008. Most livestock farms are not fully specialised. Data selection thus accounts for farms that produce livestock output as well as crop output and that document input use in the respective categories. The sample only includes farms (N=3674) that are observed for several subsequent periods (on average 6.4 years) to ensure a sound though unbalanced panel structure.

Farms are selected by activity. The type of farming can be distinguished by means of a two-digit code. This sample comprises farms producing cattle (type 42), pig and poultry (type 50) or both kinds of livestock (type 72). Table 4.1 provides an overview of available variables that are given on a yearly basis. Outputs are aggregate volumes of the respective livestock and crop activities. These numbers as well as four of the input variables are measured in monetary terms.<sup>20</sup> Inputs measured in physical terms are labour (in hours worked), land (total utilised agricultural area) and livestock (in live-

<sup>18</sup>The issue of potential endogeneity is discussed in Brümmer et al. (2002) as well as in Kumbhakar and Lovell (2000).

<sup>19</sup>Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal, Spain, Sweden, United Kingdom.

<sup>20</sup>Deflation of monetary variables has been carried out using various price indices from Eurostat.

Table 4.1: Descriptive statistics on livestock farms.

	mean	std.dev	min	max
Livestock output	194,100	410,700	1.0	14,557,000
Crop output	20,095	35,872	1.7	1,177,200
Capital	31,886	38,300	0.8	999,260
Labour (hours)	4,208	3,635	160	114,660
Land (hectares)	58	62	0.1	842
Livestock (units)	246	467	0.9	28,675
Livestock cost	98,455	197,520	0.8	7,621,700
Crop cost	8,199	9,771	1.1	191,830
Other direct cost	23,562	46,745	3.4	2,315,400

Note: monetary values in euros.

stock units). Capital is composed of current costs of machinery and buildings (e.g. small repairs, insurance) plus depreciation. Livestock costs include feed and other livestock-related expenses. Crop cost summarise costs of seeds, fertilizers, crop protection and other crop-related expenses. Direct cost include energy (0.40 of total, on average), other direct inputs (0.36) and services (0.24). The statistics confirm that livestock production is the most important activity. Livestock output has a share of 0.89 at the median of the sample. The variables' standard deviations reveal distributions that are skewed to the left. Some (minimum) values suggest extreme (low) input use. These values can be correct or result from typos or documentation of approximated inputs. It is not trivial to decide whether farms should be excluded based on these statistics. As there is no clear guideline, no criterion is defined and the focus is concentrated on the panel structure. Although statistics on (total) output seem reasonable, the data set is probably subject to incomplete measurement and hence accounting for unobserved heterogeneity is relevant.

### Data on meat industry firms

This fact is even more important in case of the industry data (Bureau van Dijk, 2011). Production activities of meat firms are documented only by aggregate output and three input variables. Output is measured as revenue and the inputs are labour costs, material costs and depreciation. There is information on activities by means of industry classification. These categories subdivide



firms into slaughtering, poultry slaughtering and meat processing firms (a related description is provided in Section 3.4). The firms are either medium-sized or large, i.e. having at least 50 employees. Nine European countries are covered by the data.<sup>21</sup> The observation period includes the years 2002 to 2009.

Table 4.2: Descriptive statistics of meat firms.

	mean	std.dev	min	max
Revenue	44,827	54,914	1,604	548,020
Labour costs	4,157	5,298	38	63,268
Material costs	31,040	38,309	958	314,100
Depreciation	1,010	1,736	7	22,572

Note: values in thousands of euros. N = 473 firms are observed over T = 6.5 years on average.

## Specification

Information on firm activities can generally be used to define group frontiers. This strategy is probably suitable in case of the meat industry. For agricultural production, a clear categorisation is more difficult. Farms often produce multiple outputs and are diversified, i.e. the range of activities is more dispersed in comparison to industrial enterprises. Furthermore, in these specific data, there are some farms which change their classification between the years. Pooling farms thus seems reasonable. In order to maintain comparability and because the objective is to derive regional measures, the same approach is applied to the meat firms.<sup>22</sup>

Agricultural production is represented by a multiple output multiple input technology. Equation (4.9) describes the translog specification of production:<sup>23</sup>

<sup>21</sup>Austria, Belgium, Finland, France, Germany, Italy, Portugal, Spain, Sweden.

<sup>22</sup>Splitting up firm data into group frontiers rather goes into the opposite direction.

<sup>23</sup>A random effects specification is rejected with (Hausman) test statistic  $\chi^* = 2520 > \chi^2_{54; \alpha=0.01} = 81$ .

$$\begin{aligned}
-\ln y_{1it} &= \alpha_i + \alpha_1 \ln \frac{y_{2it}}{y_{1it}} + \frac{1}{2} \alpha_{11} \left( \ln \frac{y_{2it}}{y_{1it}} \right)^2 \\
&+ \sum_{j=1}^7 \beta_j \ln x_{jit} + \frac{1}{2} \sum_{j=1}^7 \sum_{l=1}^7 \beta_{jl} \ln x_{jit} \ln x_{lit} \\
&+ \sum_{j=1}^7 \gamma_{2k} \ln \frac{y_{2it}}{y_{1it}} \ln x_{jit} \\
&+ \beta_t t + \frac{1}{2} \beta_{tt} t^2 + \delta_{ty_2} \frac{y_{2it}}{y_{1it}} t + \sum_{j=1}^7 \beta_{tj} \ln x_{jit} t - v_{it} + u_{it}
\end{aligned} \tag{4.9}$$

This is the log form of equation (4.7), whereas the actual estimation equation corresponds to (4.8), i.e. the expression is multiplied by minus one. The two outputs are livestock output ( $y_1$ ) and crop output ( $y_2$ ). The inputs are capital ( $x_1$ ), labour ( $x_2$ ), land ( $x_3$ ), livestock ( $x_4$ ), livestock cost ( $x_5$ ), crop cost ( $x_6$ ) and other direct cost ( $x_7$ ). A time trend ( $t$ ) is included to allow for non-neutral shifts of the production frontier (technical change).

Industrial meat processing is also modelled by means of a translog specification including a time trend.<sup>24</sup> However, there is only an aggregate output ( $y$ ) that is produced by three inputs  $j$  ( $1 =$  labour costs,  $2 =$  material costs,  $3 =$  depreciation). The formal expression corresponds to equation 3.8 in Section 3.4.

## 4.4 Model of meat chain productivity

### Conceptual considerations

Supply chains increasingly shape meat production in Europe. Organisational structures and clustering of related firms could lead to productivity spillovers between firms or clusters within a sector but also between different sectors. An empirical approach that aims to identify such relationships requires some conceptual considerations. Furthermore, the options of deriving an approach from our production analysis at the firm-level have to be explained.

<sup>24</sup>The Cobb-Douglas specification is rejected with test statistic  $\chi^* = 330 > \chi^2_{(10; \alpha=0.01)} = 23.2$ . A random effects specification is rejected with (Hausman) test statistic  $\chi^* = 152 > \chi^2_{14; \alpha=0.01} = 29.1$ .

From a theoretical viewpoint, industry performance is likely to be a driver of agricultural productivity. There are fewer but larger firms within the industry and hence development of technology goes faster through the aggregate sector. Industry capacity is more concentrated, i.e. it is the bottleneck from primary production towards the retail level. Therefore, the essential characteristics of the raw product are defined at the industry level (or indirectly by retailers and consumer demand).<sup>25</sup>

The potential existence of spillover effects is not necessarily restricted to geographical areas. Ideally, the analysis should focus on economic distance. Information technology simplifies administrative procedures and the flow of animals and materials is not bounded by borders. For meat firms, precise localisation is possible by means of geographic coordinates. However, the position of livestock farms is only documented at the regional level. The analysis is thus restricted to the regional level.

Although the consideration of economic distance is more desirable there are potential factors causing spillover effects that can be associated with physical distance. Concentration of firms can be due to natural conditions as well as infrastructure (e.g. harbours) and related savings of transportation cost and time. A sufficient number of specialised livestock farms can lead to the development of specific agricultural consulting and extension services. Similarly, the availability of specialised inputs such as feed, veterinary care, equipment and loans can be enhanced. Even if many such inputs are transportable, communication and personal relationships matter. Proximity simplifies communication and exchange of experience. It possibly enables easier quality control and visiting of production sites as well as quicker reaction to unexpected events. Visibility of competitors can promote overall motivation and a favourable environment can encourage entrepreneurs to invest. The establishment of local brands (e.g. protected geographical indication) can be another driver of performance.

---

<sup>25</sup>Technological requirements are certainly relevant. There are other potential factors such as origin or organic production.

## Productivity decomposition and regional aggregation

Growth of TFP is calculated from the estimated technology. Firm-specific numbers are derived (within each of the two sectors) before aggregating the results into regional measures.

Following Kumbhakar and Lovell (2000), the individual rate of TFPC is calculated as:

$$TFPC_{it} = [-(u_{it} - u_{i,t-1})] + \frac{\partial \ln \hat{y}_{mit}}{\partial t} + \left[ (RTS_{it} - 1) \times \sum_j \left( \frac{\epsilon_{jit}}{RTS_{it}} \right) (\ln x_{jit} - \ln x_{ji,t-1}) \right] \quad (4.10)$$

where the output elasticity of input  $j$  is  $\epsilon_{jit} = \frac{\partial \ln \hat{y}_{mit}}{\partial \ln x_{jit}}$  and the measure of returns to scale (RTS) is  $RTS_{it} = \sum_j \epsilon_{jit}$ . The rate of TFPC is the sum of change in TE (TEC), technical change (TC) and a scale component (SC).<sup>26</sup>

Since a regional measure of performance is desired there is a need to aggregate the firm-specific results. Färe and Zelenyuk (2003) refer to Farrell's idea of "structural efficiency of an industry" (Farrell, 1957) when discussing the aggregation of TE scores over economic units. In essence, they conclude that the aggregate measure should account for the relevance of the economic unit via its scale of operation. They suggest weighting by output (in their case, revenue) shares.

Furthermore, they show that there is no consistent method of aggregation in case of multiple output production. This finding means that the results and interpretation of the aggregate group technology can be ambiguous. Nevertheless, they advocate a weighting scheme for the case of multiple outputs that is independent of prices. A firm's weight is given by the sum of its output shares (observed firm output in total group output) normalised by the number of activities (output categories). Zelenyuk (2006) refers to the same scheme and provides an intertemporal extension of the aggregation approach considering the Malmquist TFPC index.

Accordingly, in our case, it would be ideal to use firm-specific indices<sup>27</sup>

<sup>26</sup>Results are averaged over years  $t$  and  $t-1$  to obtain the TC rate of year  $t$ . Output elasticities and returns to scale are also averaged over two subsequent years.

<sup>27</sup>These are chained indices, i.e. the index ( $I$ ) of firm  $i$  in period  $t$  is given by  $I_{it} =$

and to weight these indices per period  $t$  on a regional basis (region  $r$ ). For productivity growth:

$$TFPCI_t^r = \sum_i TFPCI_{it}^r \times \left[ \frac{1}{2} \left( \frac{y_{1it}^r}{\sum_i y_{1it}^r} + \frac{y_{2it}^r}{\sum_i y_{2it}^r} \right) \right] \quad (4.11)$$

The regional TFPC index (TFPI) is a weighted sum of the firm-specific indices (firms located in that region) where the individual weight is constructed as in Färe and Zelenyuk (2003). Weighting requires only the observed outputs in the respective period  $t$  which is desirable with respect to aggregation properties. Regional TFPC from period  $s$  to period  $t$  can be calculated as:

$$TFPC_{s \rightarrow t}^r = \frac{TFPCI_t^r}{TFPCI_s^r} - 1 \quad (4.12)$$

However, our data set is unbalanced. The firm-specific indices have no common base year and results from equation (4.11) are unfeasible. Furthermore, it is desirable to consider single periods because spillovers require time to pass through the chain. Unfortunately, the regional distribution of our observations differs over time, leading to different regional matchings. In order to simplify the approach, a more pragmatic weighting is applied. Regional TFPC is calculated as:

$$TFPC^r = \sum_i \sum_t TFPC_{it}^r \times \left[ \frac{1}{2} \left( \frac{y_{1it}^r}{\sum_i \sum_t y_{1it}^r} + \frac{y_{2it}^r}{\sum_i \sum_t y_{2it}^r} \right) \right] \quad (4.13)$$

This measure comprises all periods in which a region is observed. Individual TFPC rates are weighted according to the relevance of a firm (in terms of output quantity). A firm's weight is given by its output share in period  $t$  relative to total regional output. In case of meat firms, a normalisation of weights as in equation (4.13) is not required.

This scheme involves the decision on which outputs to choose, i.e. to choose either output in period  $t$  or the average output of periods  $s$  and  $t$ . Zelenyuk (2011, p.195) advocates to “rest on the weights in one of the periods” when aggregating rates of single units into rates of higher-level units. Otherwise  $I_{i,t-1} \times (1 + a_{it})$  where  $a$  is the growth rate.

aggregation consistency will fail. Accordingly, we rest on output in period  $t$  when using equation (4.13).

### Specification supply chain model

We argue that the meat industry's productivity growth is likely to be a driver of agricultural TFPC. Furthermore, spillovers within agriculture are possible and hence regional TFPC in agriculture is a function of industrial productivity and the productivity of neighbouring observations (as well as of other unknown covariates).

The livestock farms and meat firms are located in  $R$  regions ( $r = 1, \dots, z, \dots, R$ ) that are defined within the FADN framework, predominantly corresponding to level 2 of the NUTS classification.<sup>28</sup> Neighbourhood is considered by means of a spatial weights matrix  $W$  (LeSage, 1999). This is a row-normalised  $R \times R$  matrix where non-zero elements indicate neighbourhood. In our case, the neighbourhood structure is insensitive to the choice of Rook contiguity (regions with a common side) versus Queen contiguity (regions sharing a common side or vertex).

The model structure corresponds to a so-called spatial Durbin model (LeSage and Pace, 2009) that includes a spatial lag of the dependent variable and a spatial lag of an explanatory variable. In fact, the spatial lag is a weighted average of neighbouring observations where the weights are determined by neighbourhood structure. The relationship between regions  $r$  and  $z$  is described by the respective element of the matrix  $W$ . For a single region, using the aggregate measures derived above, the model can be written as:

$$TFPC_{agr}^r = \lambda \sum_{z=1}^R W_{rz} TFPC_{agr}^z + \theta \sum_{z=1}^R W_{rz} TFPC_{ind}^z + \beta TFPC_{ind}^r + \eta + \nu^r \quad (4.14)$$

According to equation (4.14), agricultural ( $agr$ ) TFPC in a region is determined by agricultural TFPC of its neighbours ( $W TFPC_{agr}$ ) as well as in-

---

<sup>28</sup>Nomenclature des unités territoriales statistiques (Eurostat, 2014a).

dustrial (*ind*) productivity growth of its neighbours ( $WTFPC_{ind}$ ) and in its own region. The parameter  $\eta$  is a common constant and  $\nu^r$  is a random error term distributed as  $\nu^r \sim N(0, \sigma^2)$ .

## 4.5 Results

### Results for livestock production

The first part of this section focuses on the results of the agricultural distance function. Results for meat firms are presented in the appendix and a more detailed description of the industry technology is provided in Section 3.5.

The first order coefficients of the distance function are presented in Table 4.3. These point estimates can be interpreted as elasticities at the sample mean because the variables are normalised at their respective means. The crop output elasticity suggests this activity to be more important than livestock production. For livestock farms, this result seems implausible. However, a closer look at the distribution of the elasticities puts the result into perspective.<sup>29</sup>

Table 4.3: Point estimates of agricultural frontier.

	Coeff.	t-value
Crop output	-0.530	-81.2
Capital	0.064	6.2
Labour	0.043	2.9
Land	0.249	17.6
Livestock	0.030	2.5
Livestock cost	0.339	38.9
Crop cost	0.114	13.9
Other direct cost	0.048	4.6
Trend	0.008	5.7
RTS	0.887	

Note: N = 3674, avg. T = 6.4. Inputs are given in logarithms. The complete estimation output of the distance function is documented in the appendix, table 4.7.

<sup>29</sup>This result is partly caused by a large outlier in the normalised variable. The mean of estimated elasticities is 0.33 with a standard deviation of 0.14. The standard deviation of observed crop shares is 0.13 and the correlation with the estimated elasticities is 0.82.

The first order coefficients are positive and have the expected signs (reversed for illustration). With given inputs, an increase of crop output accompanies decreasing livestock output. Signs related to the inputs are positive. The most important input is livestock cost which is consistent with the main activity of farms. Land and crop cost are the next most relevant production factors, whereas the remaining coefficients are of similar magnitude. Accordingly, the structure of agricultural production seems to be plausibly represented. The positive time trend indicates technical progress.

Predictions of TE scores reveal that there is a large scope for improvement (Table 4.4). Output could be expanded by more than 20 percent, on average (with input levels fixed). The notable standard deviation demonstrates that performance differs a lot between farms. There is no remarkable difference between farm types, only the standard deviation of TE for cattle farms is slightly bigger than for the other farm types (0.09 versus 0.07).<sup>30</sup>

Table 4.4: Descriptive statistics of technical efficiency scores.

Mean	Std.dev	Min.	Max.
0.82	0.08	0.25	0.98

Note: observations = 23,386.

As Figure 4.1 shows, average efficiency is changing over years. Weighting of TE scores (analogous to equation (4.11) but for the full sample) highlights the relevance of larger farms whose efficiency leads to a smoothing of the overall average. Smaller farms tend to have lower and rather decreasing TE scores.

The empirical model specification allows for the isolation of farm effects that can contribute to the interpretation of results. Discrimination with respect to farm type shows that the effects of cattle farms diverge from the other types (Figure 4.2). These tend to be smaller in magnitude and are more dispersed. When interpreting the farm effects as basic productivity (as is suggested in Section 3.3) cattle farms are likely to be less productive and

---

<sup>30</sup>The differentiation of farm types can possibly expand the scope of interpretation. As the classification can differ over years, the most frequent category (mode) for each farm is applied.



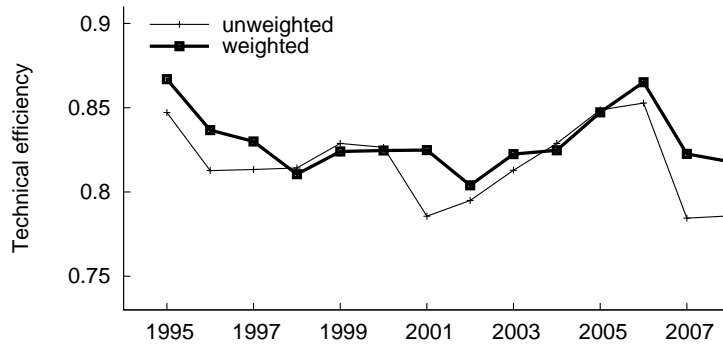


Figure 4.1: Technical efficiency over the observation period.

face more diverse surrounding conditions. Relating basic productivity (farm effects) to the individual average TFPC index shows that there is some indication for a positive relationship.<sup>31</sup>

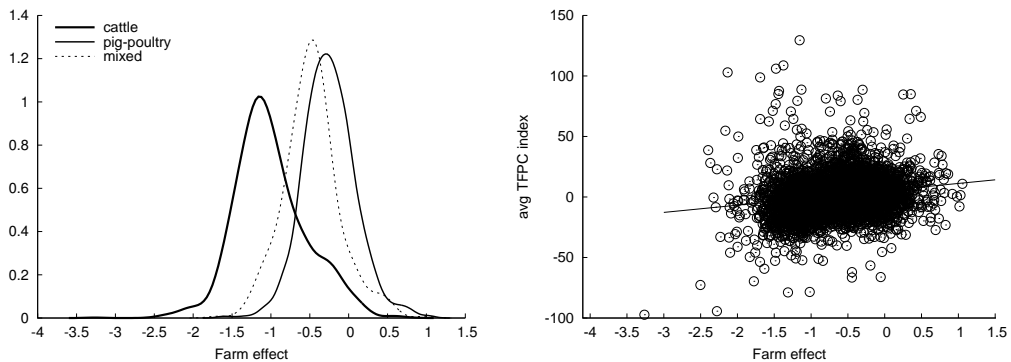


Figure 4.2: Distributions of farm effects according to farm type (left panel) and individual mean of TFP change index versus farm effects (right panel).

The individual average rate of TFPC can take extreme values (ranging from -0.66 to 0.30) and the overall distribution tends towards a negative average rate (with a median of zero). However, the standard deviation is a reasonable 0.06.

<sup>31</sup>A farm's average TFPC index (depicted on the vertical axes) is the geometric mean over the individual period (without base year, the average years are 5.4). Here, the standard deviation is 18 percent (or  $1.18 = 1.063^{2.7}$ ).

## Sectoral results

In order to gain insights into the dynamics of the two sectors, livestock farming and meat industry, Table 4.5 presents the weighted decomposition of TFPC. Both sectors experience productivity growth but the relevance of the sources of the growth is different.

The meat industry is characterised by a high and virtually constant level of TE (0.95 on average). Accordingly, the index of TEC is almost constant. The same result is found for the index of scale-related changes, whereas TC is the key driver of TFPC. These findings are related to the size of slaughtering firms that tend to have the larger output and hence receive a higher weight in the index. Slaughtering firms are close to constant returns to scale and the development in this sub-sector is stagnant as is shown in Section A. Comparing this result with the geometric index (see appendix) highlights that there are (smaller) firms that experience higher technical progress.

Table 4.5: Productivity decomposition for agriculture and meat industry using weighted indices.

Year	Agriculture				Meat industry			
	TECi	TCi	SCi	TFPCi	TECi	TCi	SCi	TFPCi
1995	1.000	1.000	1.000	1.000				
1996	0.973	1.052	1.007	1.032				
1997	0.973	1.099	1.012	1.084				
1998	0.958	1.140	1.020	1.115				
1999	0.976	1.174	1.026	1.176				
2000	0.988	1.199	1.030	1.220				
2001	0.998	1.216	1.035	1.256				
2002	0.988	1.224	1.038	1.255	1.000	1.000	1.000	1.000
2003	1.032	1.222	1.044	1.317	1.003	1.007	0.999	1.009
2004	1.067	1.213	1.046	1.354	1.002	1.014	0.998	1.014
2005	1.106	1.196	1.054	1.395	1.004	1.023	0.997	1.023
2006	1.145	1.172	1.058	1.421	1.012	1.032	0.997	1.040
2007	1.087	1.141	1.059	1.313	1.014	1.043	0.997	1.053
2008	1.119	1.102	1.060	1.308	1.004	1.055	0.996	1.053
2009					1.003	1.067	0.996	1.064

Note: the weighted indices are derived by first weighting the TFPC rates according to years, and then, constructing a chained index for the full sample.

Conversely, in the case of agriculture, this comparison shows that larger

farms dominate the development of productivity. The relevant TFPC components for these farms are changes in efficiency and scale. The index of TC shows a peak in the year 2002, and subsequently declines, suggesting technical regress. There is no obvious explanation for this result. It is, though, one reason for the positive aggregate TEC since observations come closer to the downward shifting frontier.

The results clearly indicate spatial dependence within each of the two sectors. Figure 4.3 depicts weighted aggregate productivity growth for 61 collectively observed regions, i.e. where livestock farms as well as meat firms are observed. The TFPC rate of region  $r$ , on the horizontal axis, is plotted against the rate of neighbouring regions  $z$  to  $R$ , on the vertical axis (spatial lag). For both sectors, Moran's I test statistic confirms spatial clustering.

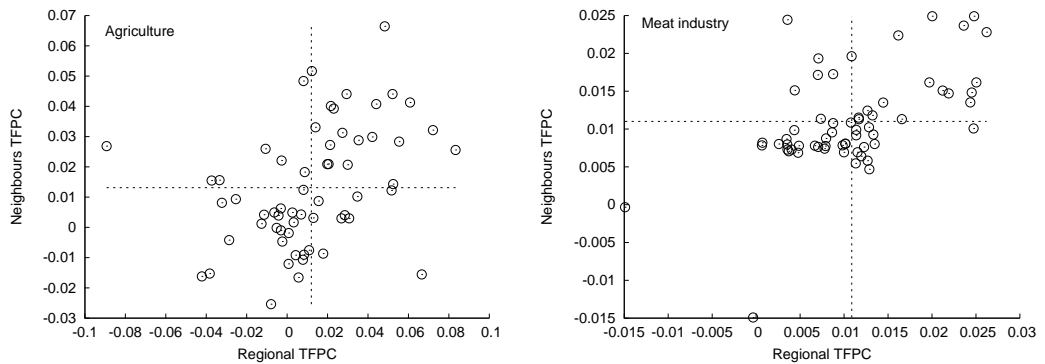


Figure 4.3: Regional clustering of TFPC in agriculture and meat industry.

The average regional rate is about one percent in agriculture as well as in the meat industry (0.012 and 0.01, respectively). However, variation is much larger in case of livestock production. There are numerous regions whose values are close to zero or even negative, whereas there are virtually only positive rates in the meat industry (except for two outliers).

### Results of supply chain model

Next, the perspective is expanded to the potential intersectoral relationships in productivity growth. The observed 61 regions have 3.7 neighbours on average (ranging from 1 to 8). In Section 4.4, we argued that agricultural TFPC is likely to be a function of the meat industry's TFPC. Model A in Table 4.6 corresponds to the proposed specification of the supply chain model.

Table 4.6: Estimation results of the supply chain model.

	Model A		Model B		Model C	
	Coeff.	z-value	Coeff.	z-value	Coeff.	z-value
constant	0.01	1.37	-0.04	-1.40	0.06	1.41
$WTFPC_{agr}$	0.33	2.48	0.32	2.47	0.30	2.31
$WTFPC_{ind}$	-0.17	-0.24	-0.22	-0.35	-0.33	-0.57
$TFPC_{ind}$	-0.12	-0.21	0.27	0.51	0.56	1.12
firm concentration			-0.01	-2.20	-0.01	-3.42
employees per firm			-0.01	-2.11	0.00	-0.89
share crop area			-0.02	-3.34	-0.02	-3.30
livestock density			0.01	2.04	0.01	2.98
$W$ firm concentr.					0.01	2.80
$W$ livestock density					-0.02	-3.04
log likelihood	129		137		144	
AIC	-248		-257		-265	
pseudo $R^2$	0.09		0.31		0.43	
resid. autocor.	rejected		rejected		rejected	

Note: the dependent variable is the weighted regional average of TFPC in livestock production. Notation  $W$  indicates spatially lagged variables. Variables (except for TFPC) are in logarithms.

We find a positive spatial lag for agricultural TFPC, i.e. regions with higher productivity growth in livestock production tend to have neighbours who also have a high TFPC in this activity. This is the finding that is already shown in Figure 4.3. There is no evidence for a connection to the industry's TFPC. The two coefficients representing industrial TFPC in the own region and in neighbouring regions, respectively, are not significant.

The partial visualisation of the variables (Figure 4.4) reveals only scattered patterns. In the centred graph, observations are divided into four groups. The quadrants are obtained by using the respective sample means of agricultural and industrial TFPC, respectively. This grouping is arbitrary to some extent but can provide an idea of how the results are geographically distributed. Regions with agricultural TFPC higher than average (marked with  $\circ$  and  $\square$ ) are predominantly located in Southern Europe and Eastern France. Corresponding observations for the meat industry (marked with  $+$  and  $\square$ ) are more scattered and located e.g. in Germany, Italy and Finland (see appendix for the map). Low TFPC in livestock production as well as in meat processing (marked with  $\triangle$ ) is predominantly found in France. The graph at the bottom of Figure 4.4 suggests a weak negative relationship between TFPC in

agriculture and the neighbouring industry (as does Model A). However, this result is not significant.

The general result of Model A is stable when we control for four other regional characteristics,<sup>32</sup> as demonstrated in Model B. Firm concentration refers to the regional density of food processing enterprises (local units per total area). This variable is a kind of proxy for an industry structure dominated by small firms. It is negatively correlated (-0.41) with the second variable, employees per food firm. The two remaining agricultural characteristics, share in crop production and livestock density (measured in livestock units), relate to total agricultural area.

An LR test suggests to include the spatial lags of firm concentration and livestock density. The resulting model is Model C. Regions with a higher share in crop production show lower TFPC rates, which is plausible, since the focus is on livestock production. The positive effect of agricultural clustering also shows up in the coefficient of (own) livestock density. In this regard, the negative result for livestock density of neighbours is ambiguous. Similarly, there is no clear pattern for the industry-related characteristics. The coefficient for (own) firm concentration is negative, whereas it is positive for concentration in neighbouring regions. In Model C, the parameter of industrial TFPC in the own region, though insignificant, improves.

---

<sup>32</sup>The variables result from averages over the period 2000 to 2007 (Eurostat, 2014b).

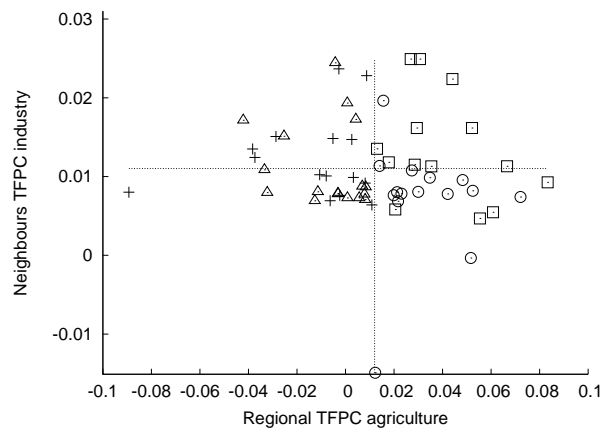
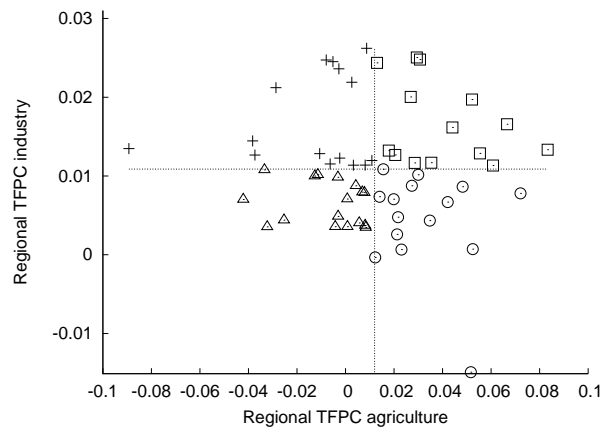


Figure 4.4: Intersectoral relationship of TFPC rates at regional level. Agriculture versus meat industry of own region (centre), and versus meat industry of neighbouring regions (bottom).

## 4.6 Discussion

### Livestock production

The efficiency analysis of livestock production suggests that there is much unused potential within this sector. Large farms are the technological leaders as demonstrated by the weighted statistics of TE. Cattle production is the activity with the worst performance. Cattle farms tend to have a lower basic productivity (as interpreted from the individual effects). In our sample, their TFP growth is lower and more dispersed. These findings are consistent with the characteristics of livestock production in Europe. The production of pig and poultry is predominantly based on indoor housing of animals. This system facilitates management and is rather associated with intensive production activities. Conversely, the range of production systems is much broader in the cattle segment, varying from low-intensity grassland systems to high-intensity fattening. Furthermore, the feed conversion ratio is worse for cattle. Lower productivity results in lower competitiveness. In this regard, our results on the performance of cattle (livestock) production coincide with the EU's trade patterns. There is an excess supply of pig and poultry being exported, whereas the EU is by now a net importer of beef.

The results suggest that unused potentials could be reduced (reduction of technical inefficiency and improvements in TFP) by producing less cattle and shifting livestock production to larger farms. However, there can be regions where alternatives to grazing are limited and where non-market goods (landscape conservation) are produced by means of cattle husbandry. Naturally, beef production is a by-product of the European dairy sector.

Higher basic productivity (farm effect) is positively related to farm size (e.g. the correlation with log of livestock output is 0.70). Smaller farms are less productive but there can be various reasons for staying in the sector. Nevertheless, structural change is likely to be a driver of overall agricultural TFPC.

## Sectoral results

The average regional TFPC rates don't differ much between livestock production and meat processing. While there is progress in both sectors, the variation and characteristics of their development differ. The meat industry is a rather steady branch where larger firms are less dynamic. There is a high level of TE since there are no such exit barriers as in agriculture. Smaller firms and processing enterprises profit from TC that is virtually the only source of productivity growth.

Conversely, all TFPC components contribute to the agricultural development. The interpretation of agricultural TC and TEC is ambiguous though. Negative TC isn't actually plausible. This result can be caused by restrictions to farms (e.g. political measures or malfunctioning factor markets). Another potential reason is an inappropriate deflation of the monetary variables, but this is unlikely to be the only factor, given the observed magnitude. However, the stable level of weighted TE scores results in an increase of the TFPC index (effective TEC).

## Supply chain model

Our analysis suggests that productivity spillovers within each of the sectors are more likely than between the sectors. The negative result with regards to the intersectoral spillovers can have several causes related to the model, data or other factors.

The supply chain model uses regional averages over several periods. It seems plausible that existing spillovers could be reflected in simple regional measures, i.e. if average TFPC in the meat industry is high, it is also high in livestock production. This approach disregards the fact that spillovers need time to pass through the chain (from an unspecified period  $s$  to some period  $t$ ). Ideally, the model would include a lag of industrial TFPC but this not applicable to our data.<sup>33</sup>

The measurement of the regional average TFPC rate has two additional

---

<sup>33</sup>Estimation of the supply chain model restricting the agricultural data to the years 2003 to 2008 doesn't change the interpretation. Data of single years (2003 to 2008) doesn't provide additional insights.



problems. Firstly, the actual situation can be misspecified because there are too few observations representing a specific region. This issue applies to the agricultural as well as to the industry data. The second problem relates to the concept of the region as an aggregate economic unit. Our aggregation scheme maybe performs badly. As pointed out by Färe and Zelenyuk (2003), there is no consistent approach for multiple outputs. However, relevant dynamics are probably not averaged out because this is prevented by using weights.<sup>34</sup> The issue of economic distance is not adequately addressed, though. In Section 4.4, potential regional determinants are discussed but a region is not necessarily an economic cluster. For example, the use of detailed trade (material) flows should represent economic relationships much better.<sup>35</sup> These flows (and their relevance) are neither restricted to the own region nor to neighbouring regions. In fact, the economic relationships can stretch across countries.

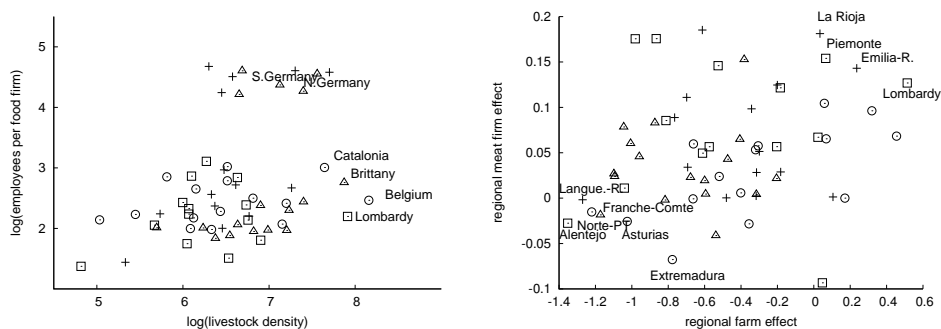


Figure 4.5: Livestock density versus employees per food firm at the regional level (left panel). Weighted farm effects versus weighted meat firm effect (right panel).

Figure 4.5 highlights the mixed results that we find. There are some cases where a very high regional livestock density accompanies higher agricultural TFPC (e.g. Belgium, Catalonia, Lombardy). However, we also find a high density and lower TFPC, as e.g. in Germany. The German industry structure is characterised by larger food firms (“upper cloud” in the left panel) but

<sup>34</sup>The applied weights (using output in the same period) seem reasonable. Using weights with averaged output results in regional measures that are more similar to the simple (unweighted) means. Therefore, the applied regional measures seem appropriate since the unweighted means should not be used. Estimation of the supply chain model using alternative weights does not change its interpretation.

<sup>35</sup>Even more precise geographic information can be useful. The data on meat firm locations are exact. Estimation of a fixed effects spatial autoregressive model (as a translog production function using distance-based weights) reveals spatial dependence among the meat firms in our sample.

this attribute doesn't provide an advantage as indicated by the estimation results.<sup>36</sup>

The positive coefficient of (own) livestock density and the coefficient of (own) industrial TFPC that tends to be positive are eventually indications for some intersectoral spillovers within single regions. However, this is not clear from the supply chain model. The right panel of Figure 4.5 provides a crossplot of the weighted regional farm effects and firm effects, respectively. A kind of aggregated basic productivity could possibly indicate intersectoral clusters by means of a positive relationship between farm effects and firm effects. The highest effects in both sectors are found for regions in Northern Italy as well as for La Rioja in Spain. As the overall pattern remains rather vague, this perspective can also not provide much evidence for intersectoral spillovers.

Besides the shortcomings of the supply chain model discussed above, the underlying dynamics of spillover effects are probably too subtle for the aggregate analysis. There are two related aspects. Firstly, the farms in the sample might be "too diversified", i.e. spillovers are maybe only relevant for the very specialised livestock producers. Secondly, we have shown that slaughtering firms tend to be less dynamic (see Section 3). However, these firms are the link between agriculture and meat processing. The processors are more innovative but their productivity growth doesn't necessarily spill over to the farms. Furthermore, it is still possible that existing productivity spillovers at the firm-level don't translate to the regional level.

## References

- Brümmer, B., Glauben, T. and Thijssen, G. (2002). Decomposition of productivity growth using distance functions: the case of dairy farms in three European countries. *American Journal of Agricultural Economics* 84: 628–644.
- Bureau van Dijk (2011). AMADEUS Database. Bureau van Dijk Electronic Publishing.
- Chen, Y.-Y., Schmidt, P. and Wang, H.-J. (2014). Consistent estimation of

---

<sup>36</sup>The food firm classification can only serve as a proxy for the structure of meat firms.

- the fixed effects stochastic frontier model. *Journal of Econometrics* 181: 65–76.
- Diez-Vial, I. (2011). Geographical cluster and performance: The case of Iberian ham. *Food Policy* 36: 517–525.
- European Commission (2010). Farm Accountancy Data Network. <http://ec.europa.eu/agriculture/rica/index.cfm>.
- Eurostat (2014a). Nomenclature of territorial units for statistics. [http://epp.eurostat.ec.europa.eu/statistics\\_explained/](http://epp.eurostat.ec.europa.eu/statistics_explained/).
- Eurostat (2014b). Regional statistics by NUTS classification. [http://epp.eurostat.ec.europa.eu/portal/page/portal/region\\_cities/regional\\_statistics/](http://epp.eurostat.ec.europa.eu/portal/page/portal/region_cities/regional_statistics/).
- Färe, R. and Zelenyuk, V. (2003). On aggregate Farrell efficiencies. *European Journal of Operational Research* 146: 615–620.
- Farrell, M. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society* 120: 253–290.
- Greene, W. H. (2005). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics* 126: 269–303.
- Kumbhakar, S. C. and Lovell, C. A. K. (2000). *Stochastic Frontier Analysis*. Cambridge University Press.
- Larue, S., Abildtrup, J. and Schmitt, B. (2011). Positive and Negative Agglomeration Externalities: Arbitration in the Pig Sector. *Spatial Economic Analysis* 6: 167–183.
- LeSage, J. P. (1999). The theory and practice of spatial econometrics. University of Toledo. Toledo, Ohio.
- LeSage, J. P. and Pace, R. K. (2009). *Introduction to Spatial Econometrics*. Chapman & Hall/CRC Press.
- Mosnier, C. and Wieck, C. (2010). Determinants of spatial dynamics of dairy production: a review. Institute for Food and Resource Economics, University of Bonn.
- Roe, B., Irwin, E. G. and Sharp, J. S. (2002). Pigs in Space: Modeling the Spatial Structure of Hog Production in Traditional and Nontraditional Production Regions. *American Journal of Agricultural Economics* 84: 259–278.
- Schmidtner, E., Lippert, C., Engler, B., Haring, a. M., Aurbacher, J. and Dabbert, S. (2012). Spatial distribution of organic farming in Germany: does neighbourhood matter? *European Review of Agricultural Economics* 39: 661–683.

Zelenyuk, V. (2006). Aggregation of Malmquist productivity indexes. *European Journal of Operational Research* 174: 1076–1086.

Zelenyuk, V. (2011). Aggregation of economic growth rates and of its sources. *European Journal of Operational Research* 212: 190–198.

## Appendix

Table 4.7: Estimates of agricultural distance function.

	Coefficient	t-value
crop output	0.530	81.2
capital	-0.064	-6.23
labour	-0.043	-2.87
land	-0.249	-17.6
livestock	-0.030	-2.47
livestock cost	-0.339	-38.9
crop cost	-0.114	-13.9
direct cost	-0.048	-4.61
trend	-0.008	-5.73
0.5×crop output <sup>2</sup>	-0.125	-62.7
0.5×capital <sup>2</sup>	0.013	3.45
0.5×labour <sup>2</sup>	0.046	2.37
0.5×land <sup>2</sup>	-0.027	-3.36
0.5×livestock <sup>2</sup>	-0.003	-0.395
0.5×livestock cost <sup>2</sup>	0.066	25.4
0.5×crop cost <sup>2</sup>	0.021	5.65
0.5×direct cost <sup>2</sup>	0.022	3.86
0.5×trend <sup>2</sup>	-0.007	-27.2
crop output×capital	-0.007	-2.06
crop output×labour	0.003	0.473
crop output×land	0.043	13.8
crop output×livestock	-0.021	-5.15
crop output×livestock cost	-0.008	-3.28
crop output×crop cost	0.010	3.51
crop output×direct cost	-0.006	-1.62
crop output×trend	0.004	6.78
capital×labour	-0.035	-3.78
capital×land	0.010	1.99
capital×livestock	-0.008	-1.4
capital×livestock cost	0.001	0.302
capital×crop cost	0.000	0.087
capital×direct cost	-0.001	-0.142
capital×trend	0.002	1.78
labour×land	0.018	1.88
labour×livestock	-0.010	-0.999
labour×livestock cost	0.001	0.176
labour×crop cost	-0.025	-3.34
labour×direct cost	0.020	2.09
labour×trend	0.012	7.61
land×livestock	0.009	1.41
land×livestock cost	0.009	2.25
land×crop cost	-0.001	-0.219
land×direct cost	-0.007	-1.5
land×trend	-0.008	-9.36
livestock×livestock cost	-0.022	-6.04
livestock×crop cost	0.003	0.568
livestock×direct cost	-0.004	-0.717
livestock×trend	0.001	0.761
livestock cost×crop cost	-0.012	-3.61
livestock cost×direct cost	-0.024	-6.99
livestock cost×trend	0.001	1.72
crop cost×direct cost	0.009	2.03
crop cost×trend	-0.001	-1.3
direct cost×trend	-0.005	-4.47
$\ln \sigma_v^2$	-3.746	-101
$\ln \sigma_u^2$	-2.606	-69.3

Note: N = 3674, avg. T = 6.4. Inputs are given in logarithms. Crop output is a normalised variable,  $\frac{y_2}{y_1}$ , where  $y_1$  is livestock output and  $y_2$  is crop output.

Table 4.8: Estimates of production frontier for meat industry.

	Coefficient	t-value
labour	0.184	23
material	0.756	99.8
capital	0.028	6.29
trend	0.011	19.2
$0.5 \times \text{labour}^2$	0.073	7.88
$0.5 \times \text{material}^2$	0.117	11.7
$0.5 \times \text{capital}^2$	0.014	3.27
$0.5 \times \text{trend}^2$	0.001	2.17
labour $\times$ material	-0.061	-7.4
labour $\times$ captial	0.001	0.23
labour $\times$ trend	0.005	5.27
material $\times$ captial	-0.033	-6.71
material $\times$ trend	-0.008	-10.6
capital $\times$ trend	0.002	2.55
$\ln \sigma_v^2$	-6.852	-73.2
$\ln \sigma_u^2$	-5.601	-63.9

Note: N = 473, avg. T = 6.5. Inputs are given in logarithms.

Table 4.9: Descriptive statistics of technical efficiency scores (meat industry).

Mean	Std.dev	Min.	Max.
0.95	0.02	0.71	0.99

Note: observations = 3076.

Table 4.10: Productivity decomposition for agriculture.

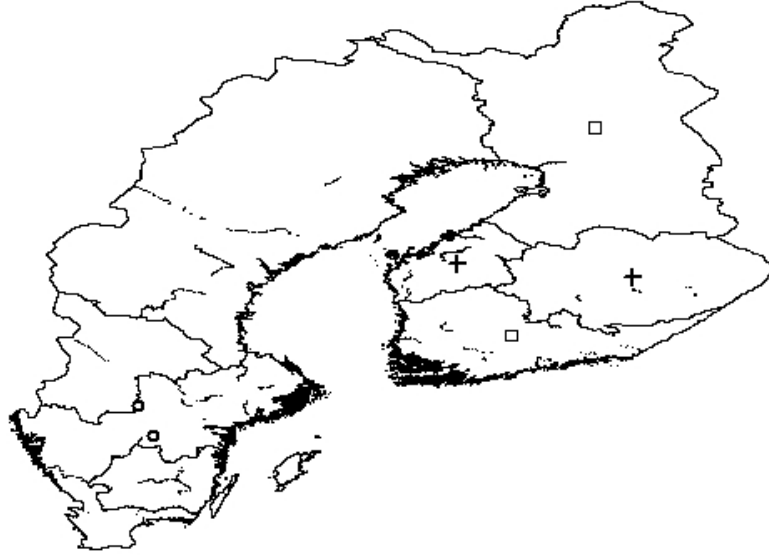
Year	<u>weighted mean</u>				<u>geometric mean</u>			
	TECi	TCi	SCi	TFPCi	TECi	TCi	SCi	TFPCi
1995	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1996	0.973	1.052	1.007	1.032	0.959	1.052	1.004	1.015
1997	0.973	1.099	1.012	1.084	0.961	1.098	1.008	1.066
1998	0.958	1.140	1.020	1.115	0.963	1.139	1.011	1.111
1999	0.976	1.174	1.026	1.176	0.977	1.172	1.015	1.164
2000	0.988	1.199	1.030	1.220	0.975	1.198	1.015	1.187
2001	0.998	1.216	1.035	1.256	0.920	1.215	1.017	1.139
2002	0.988	1.224	1.038	1.255	0.934	1.222	1.017	1.164
2003	1.032	1.222	1.044	1.317	0.959	1.218	1.022	1.196
2004	1.067	1.213	1.046	1.354	0.988	1.208	1.023	1.223
2005	1.106	1.196	1.054	1.395	1.013	1.190	1.024	1.237
2006	1.145	1.172	1.058	1.421	1.021	1.164	1.025	1.221
2007	1.087	1.141	1.059	1.313	0.926	1.130	1.025	1.072
2008	1.119	1.102	1.060	1.308	0.930	1.089	1.023	1.036

Note: the weighted indices are derived by first weighting the TFPC rates according to years, and then, constructing a chained index for the full sample.

Table 4.11: Productivity decomposition for meat industry.

Year	<u>weighted mean</u>				<u>geometric mean</u>			
	TECi	TCi	SCi	TFPCi	TECi	TCi	SCi	TFPCi
2002	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2003	1.003	1.007	0.999	1.009	1.004	1.008	0.999	1.011
2004	1.002	1.014	0.998	1.014	1.002	1.017	0.996	1.015
2005	1.004	1.023	0.997	1.023	1.002	1.028	0.996	1.025
2006	1.012	1.032	0.997	1.040	1.010	1.039	0.995	1.043
2007	1.014	1.043	0.997	1.053	1.011	1.052	0.994	1.055
2008	1.004	1.055	0.996	1.053	0.999	1.065	0.993	1.054
2009	1.003	1.067	0.996	1.064	1.000	1.080	0.994	1.070

Note: the weighted indices are derived by first weighting the TFPC rates according to years, and then, constructing a chained index for the full sample.



Note: Sweden is divided into three FADN regions, i.e. within the four geographical regions two have an identical code of identification.

Figure 4.6: Map of Sweden and Finland.



# 5 Global TFP change in agriculture: consistent frontier estimation with country effects and time-varying inefficiency

*Jonathan Holtkamp, Bernhard Brümmer*

## 5.1 Introduction

Agriculture is an important sector of many countries in the world. It is supposed to meet a variety of challenges, i.e. provide solutions to food security, poverty reduction, provision of natural resources or environmental protection. The conflict of goals is unavoidable. However, productivity growth is one key aspect of addressing these issues as it either increases output or renders input savings possible.

The analysis of productivity growth at a global scale is restricted by data availability. Furthermore, the commonly used data of the Food and Agriculture Organization (FAO) are highly aggregated. There is no unambiguous approach to model production technology and investigate change in total factor productivity (TFP) and its sources.

This paper aims at conclusions from the available data and the determination of important TFP change (TFPC) sources. These issues have often been addressed in the literature. Table 5.1 depicts a small selection of studies that analyse agricultural productivity using aggregated country-level data.

Table 5.1: Selection of productivity studies.

Authors	Year	Countries	Period	Data
Coelli and Rao (2005)	2005	93	1980-2000	FAO
Ludena et al. (2007)	2007	116	1961-2001	FAO
O'Donnell et al. (2008)	2008	97	1986-1990	FAO
Headey et al. (2010)	2010	88	1970-2001	FAO
Fuglie (2012)	2012	172	1961-2010	FAO
Mundlak et al. (2012)	2012	30	1972-2000	World Bank
Butzer et al. (2012)	2012	30	1970-2000	World Bank
Eberhardt and Teal (2013)	2013	128	1961-2002	FAO

However, the topic is of persistent relevance and not all studies consider the panel nature of the data or potential technical inefficiency. The paper provides an update of TFPC sources using a specific data set which includes many countries. We apply a novel specification of a panel Stochastic Frontier (SF) model that allows for time-varying inefficiency as well as for the isolation of individual country effects. The results that we present are weighted, in order to account for the countries' relevance in world agricultural production. Furthermore, we aim at the evaluation of agricultural R&D efforts in the context of TFP growth.

## 5.2 Estimation of the production frontier

The SF model (Aigner et al., 1977) is often used to model agricultural production. The notion of an existing best practice technology is the underlying rationale of this framework. Inefficient producers are those not operating on the best practice frontier. Predictions of technical efficiency (TE) allow to draw conclusions about potential improvements.

The standard (“pooled”) SF model disregards the structure of panel data, i.e. the model interprets the data as a cross-section. Various extensions have been proposed to utilise the gain of information associated with panel data. However, the specification of time-varying inefficiency is challenging. Early models provided only very restrictive patterns of efficiency change or assumed persistent inefficiency. From an economic perspective, a constant level of inefficiency is unlikely to prevail over a longer period. Producers would either improve through learning or drop out of the market. Given a production function with one output and multiple inputs, the SF model in fixed effects panel notation can be written as:

$$y_{it} = \alpha_i + \beta'x_{it} + v_{it} - u_{it} \tag{5.1}$$

Greene (2005) labelled this model as the “true” fixed effects SF model. The notation implies observations on  $N$  individuals ( $i=1,2,\dots,N$ ) over  $T$  periods ( $t=1,2,\dots,T$ ). A set of inputs  $x$  is transformed into output denoted by  $y$ . In-

puts and output are usually provided in the form of the natural logarithm. The functional relationship is described by the vector of technology parameters  $\beta$ . Panel data allows for the isolation of individual (fixed) effects captured in the  $N$ -dimensional vector  $\alpha$ . The feature of the SF model is the composed error term, commonly written as  $\epsilon = v - u$ . For each observation, the component  $v_{it}$  is associated with white noise and assumed to be distributed as  $v_{it} \sim N(0, \sigma_v^2)$ . Inefficiency is represented by  $u_{it}$  that is a non-negative (one-sided) error component. In case of the here applied normal half-normal SF model, it is assumed to be distributed as  $u_{it} \sim N^+(0, \sigma_u^2)$ . The model in equation (5.1) is typically estimated by using the  $N$  dummy variables. This approach implies one additional parameter to be estimated per additional individual, and hence it results in biased error variances (incidental parameters problem).

A methodological extension is provided by Chen et al. (2014). They apply within-transformation to equation (5.1). This firm-specific procedure is a well-known strategy to eliminate the fixed effects.<sup>37</sup> Then, the model is rewritten in deviations from means:

$$\tilde{y}_{it} = \beta' \tilde{x}_{it} + \tilde{v}_{it} - \tilde{u}_{it} \quad (5.2)$$

While the transformation is straightforward, two other issues are more involved. Firstly, the correction of the error variance. Secondly, estimation of  $\sigma_u^2$  from the transformed error term. The solution is based on the Closed Skew Normal (CSN) distribution. The CSN distribution is a generalisation of the (skew) normal distribution and it is thus applicable to the SF model, whose composed error term has a skew normal distribution. In CSN-notation the composed error is distributed as:

$$\epsilon_{it} \sim CSN_{1,1}(0, \sigma^2, -\frac{\lambda}{\sigma}, 0, 1) \quad (5.3)$$

The density of a  $CSN_{p,q}$ -distribution includes a  $p$ -dimensional pdf and a  $q$ -

---

<sup>37</sup>The firm-specific mean of each variable is subtracted from the observation in period  $t$  in order to obtain transformed values (e.g. for arbitrary  $z$ ):  $\tilde{z}_{it} = z_{it} - \bar{z}_i$ .

dimensional cdf of a normal distribution. The basic SF specification considers only the first two parameters of the distribution: the mean (parameter) and the scale (parameter). Here, three additional parameters are explicitly required to formulate the likelihood function: those referring to the skewness as well as to the mean vector and to covariance matrix in the cdf. With panel data, there are  $T$  observations on each individual resulting in the vector  $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{iT})'$ . This vector is distributed as:

$$\epsilon_i \sim CSN_{T,T}(0_T, \sigma^2 I_T, -\frac{\lambda}{\sigma} I_T, 0_T, I_T) \quad (5.4)$$

where  $I$  is the identity matrix. The properties of CSN-distributed random variables allow for partitioning  $\epsilon_i$  into its mean  $\bar{\epsilon}_i$  and into the vector of its first  $T-1$  deviations  $\tilde{\epsilon}_i^* = (\tilde{\epsilon}_{i1}, \dots, \tilde{\epsilon}_{i,T-1})'$ . This  $T-1$  vector is used to derive the (“within”) likelihood function and ensures an implicit correction of the error variance.<sup>38</sup> Furthermore, the likelihood function is parameterised in terms of  $\beta$ ,  $\sigma^2 = \sigma_v^2 + \sigma_u^2$  and  $\lambda = \sigma_u/\sigma_v$ . This conventional parameterisation is free of incidental parameters and allows for time-varying inefficiency.

Once the stochastic frontier is estimated, assessment of TE requires to recover the composed error:

$$\epsilon_{it} = y_{it} - \hat{y}_{it} = y_{it} - \hat{\beta}' x_{it} - \hat{\alpha}_i \quad (5.5)$$

Here, the way used to calculate  $\hat{\alpha}_i$  is labelled as the mean-adjusted estimate by Chen et al. (2014):

$$\hat{\alpha}_i^M = \bar{y}_i - \hat{\beta}' \bar{x}_i + \sqrt{\frac{2}{\pi}} \hat{\sigma}_u \quad (5.6)$$

Predictions of TE are obtained, according to Battese and Coelli (1988), as:

$$TE_{it} = E(\exp(-u_{it})|\epsilon_{it}).$$

---

<sup>38</sup>The distribution of the  $T-1$  vector is more involved, see Chen et al. (2014) for the details. With regards to the degrees of freedom, the correction accounts for the  $N$  individuals:  $df = NT - N - K = N(T - 1) - K$ .

## 5.3 Data and specification

### Data and variable description

The agricultural production data are provided by the FAO. The data set used here is prepared by Fuglie (2012) and available from the USDA (2014). It covers 172 countries over the period 1961 to 2010, resulting in a balanced panel with 8600 observations. However, data for the former USSR are missing from 1961 to 1964 (60 observations). The countries are sorted into nine regions where the categorisation follows criteria of the FAO. Singapore is dropped from the sample due to extreme values (50 observations). The composition of the resulting data set is sketched in Table 5.2.

Table 5.2: Regions and corresponding observations; average land use.

Region	Countries	Observations	Million hectares
South Africa	1	50	16
Sub-Saharan Africa	47	2350	140
Latin America	29	1450	159
North America	2	100	282
Asia	23	1150	744
Europe	26	1300	165
West Asia & North Africa	19	950	102
Oceania	9	450	90
Former USSR	15	690	276
Sum	171	8490	1975

The regional group sizes are very different, in particular, with respect to the land area. North America is the most extreme region with only two countries in this group, but it comprises a high share of the land area. This example demonstrates that there are unfavourable imbalances between observations with respect to aggregation. The regions Asia, North America and former USSR account for two third of the agricultural area. On average, the sum of the agricultural area is 1975 million hectares. Given a global land area of roughly 15000 million hectares, the share of the agricultural area is roughly 13 percent. This share has increased over the observation period. In 1961 the agricultural area sums up to 1700 million hectares (11 percent) whereas the sum is 2200 million hectares (15 percent) in 2010.

The data set includes seven variables that allow to model production technology with a single output and multiple inputs. The input variable feed is not available for all observations and was not included in the analysis of Fuglie (2012). It is also neglected in most of the other studies using corresponding data. Units of measurement and a short description of the variables are listed in Table 5.3. Output is measured in monetary terms (constant prices) whereas all other variables are measured in physical quantities. The variables are not (all) original FAO numbers but are partially adjusted.

Table 5.3: Available variables and respective descriptions.

Variable	Unit	Description
Output	1000 \$	including 189 commodities; measured in international 2005 \$
Land	1000 ha	quality-adjusted rainfed cropland equivalents
Labour	1000 workers	active adults in agriculture; estimates for 1961-1979
Livestock	1000 LSU	measured in cattle equivalents (livestock units)
Machinery	no. of tractors	stock of farm machinery; measured in 40-CV tractor equivalents
Fertiliser	1000 kg	measured in N-fertiliser equivalents
Feed	1000 Mcal	animal feed from crop and crop processing residues

Adjustments apply to land (management / production type), animal species, machinery (size / power) and fertiliser (cost share of nutrients). These values are supposed to be superior to the uncorrected aggregates. For more details, see the article by Fuglie (2012) as well as the description within his data file.

Farm machinery are supposed to serve as a proxy for the capital input. In the current data set, a refined version of the original variable is used. However, it is the most contested variable of the FAO data in the literature (for example, see Butzer et al. (2012) who declare this variable to be biased and set up an own data set with different kinds of capital stock, but for only 30 countries).

Variable description in terms of absolute values is impractical due to the country level aggregation. Figure 5.1 shows ratios of the form input quantity

used to produce one unit of output (1000 dollars). A higher ratio indicates low performance because more input is required (partial perspective). Furthermore, a high (low) ratio points out high (low) factor intensity in production. The reciprocal of the ratio is a measure of average input productivity. Hence, there are improvements in productivity if the ratio decreases over time. Since the numbers presented in Figure 5.1 are averaged over regions, they can hide opposing trends within a region.

Land and livestock productivity increase over the observation period. This development is strong for the input land. At the beginning of the period there is a bigger spread between regions than at the end, i.e. there has been some convergence with respect to land productivity. For most of the other variables the developments are mixed. The ratio of feed to output increases virtually in all regions. This development highlights intensification in livestock production which corresponds to increasing meat consumption at the global level. Europe has the highest ratio in 2010 and is a large livestock producer as well as net exporter in pork and poultry. The highest ratios of labour to output are found for Asia and Sub-Saharan Africa. Asia has experienced the highest increase in labour productivity. Sub-Saharan Africa shows the lowest dynamics in all variables. This region is an illustrative example for the different factor intensities. It is characterised by a high usage of labour and livestock compared to the usage of fertiliser, machinery and feed. The dissolution of the former USSR is quite obvious from the drop in fertiliser to output around 1990. This region shows a spike in the land to output ratio in the in the 1990s (also for machinery). A similar observation is described by Swinnen et al. (2012) who find a U-shape pattern in productivity change for this region. Europe and North America show opponent trends in the machinery ratio. This pattern seems surprising as one would expect different levels but similar trends. The developments of variables suggest gains in productivity (land) as well as increasing intensification (feed). Differences between regions are visible from the levels of the ratios.

## Specification

We estimate two production frontiers including an aggregate output and  $K$  inputs. A time trend enters as an additional input. The preferred functional form is the translog specification:<sup>39</sup>

$$\begin{aligned} \ln y_{it} = & \alpha_i + \sum_{j=1}^K \beta_j \ln x_{jit} + \frac{1}{2} \sum_{j=1}^K \sum_{l=1}^K \beta_{jl} \ln x_{jit} \ln x_{lit} \\ & + \beta_t t + \frac{1}{2} \beta_{tt} t^2 + \sum_{j=1}^K \beta_{tj} \ln x_{jit} t + v_{it} - u_{it} \end{aligned} \quad (5.7)$$

The inputs are labour ( $x_1$ ), land ( $x_2$ ), machinery ( $x_3$ ), fertiliser ( $x_4$ ) and livestock ( $x_5$ ). The inclusion of the input variable feed ( $x_6$ ) into the second production frontier results in a loss of observations.<sup>40</sup>

---

<sup>39</sup>The Cobb-Douglas specification is rejected with test statistics  $\chi^* = 3074 > \chi^2_{(21; \alpha=0.01)} = 39$  and  $\chi^* = 3409 > \chi^2_{(28; \alpha=0.01)} = 48$ . The random effects specification is rejected with (Hausman) test statistics  $\chi^* = 355 > \chi^2_{27; \alpha=0.01} = 47$  and  $\chi^* = 207 > \chi^2_{35; \alpha=0.01} = 57$ .

<sup>40</sup>Affected countries are: Congo, Equatorial Guinea, Somalia, Réunion, Puerto Rico, French Guiana, Taiwan, Afghanistan, Bhutan, Bahrain, Iraq, Oman, Qatar, Papua New Guinea. Furthermore, observations of the former USSR start from 1992, those of Djibouti from 1979.



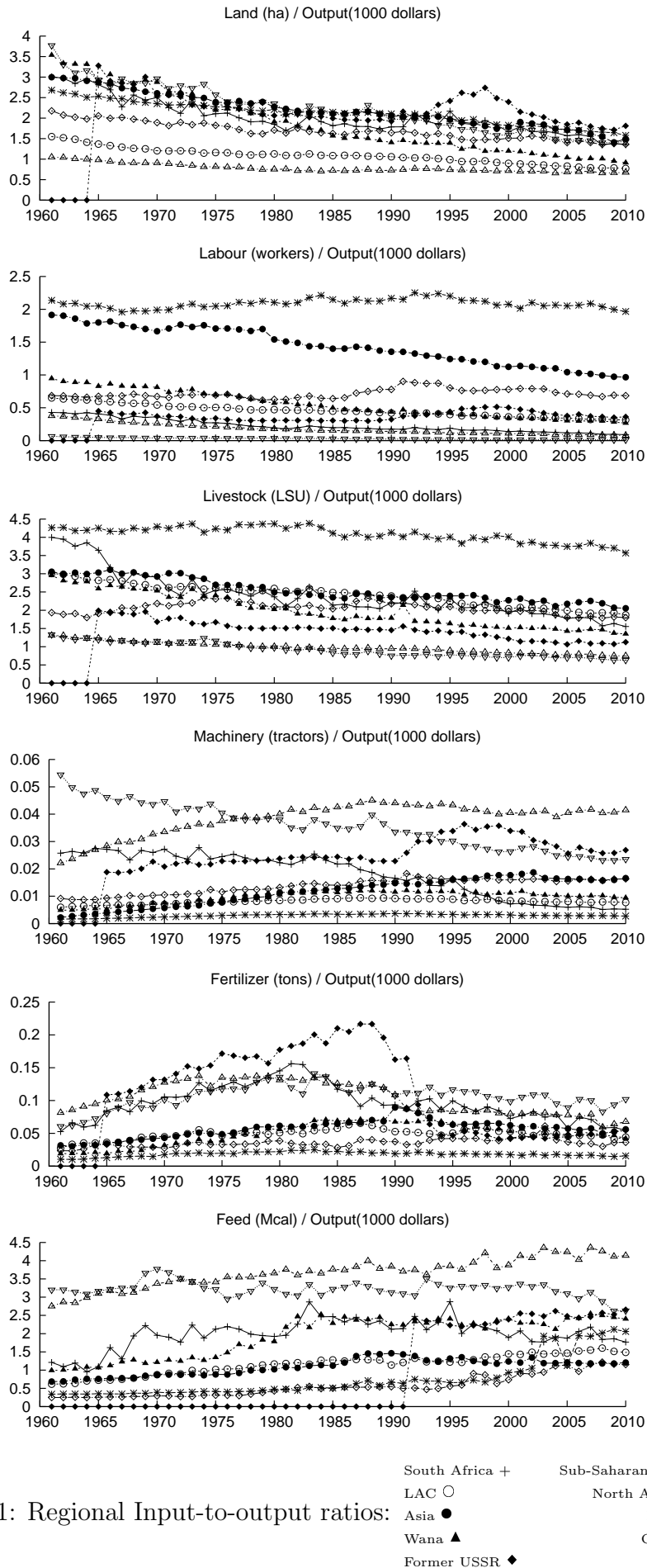


Figure 5.1: Regional Input-to-output ratios:

## 5.4 Agricultural R&D and productivity

There is public investment in agricultural R&D because politics aim to address challenges that are associated with agriculture. These investments are supposed to generate new technological possibilities that in turn should result in productivity gains. Other public efforts, such as education and agricultural extension, rather aim at the full utilisation of the production potentials. There are also private activities in agricultural R&D that are accordingly motivated by private interests.

The activities in R&D will be most profitable if they are goal-oriented. However, it is not trivial to trace back empirically how budget efforts translate into higher rates of TFPC. Difficulties relate to the mechanism itself as well as to its pattern over time. Furthermore, investments and outcomes in different sectors (such as public and private) can interact.

There seems to be scientific consensus about the profitability of agricultural R&D, as a meta-study by Alston et al. (2000) suggests. Mullen (2007) claims that under certain assumptions positive marginal returns to the investments are possible. However, the mechanics are involved, as e.g. demonstrated by Thirtle et al. (2008), for the case of the United Kingdom.

As we analyse agricultural TFPC and some of its components, our study can possibly contribute to the empirical evidence. The comparison of our results with data on agricultural R&D could provide some insights into the relationship of investment and TC ("new possibilities") as well as TEC ("learning"), in the context of this specific data set. The scope of this analysis is restricted to the potential productivity gains due to R&D and the identification of the relevant TFPC components. The assessment of the profitability is much more involved as it requires monetary evaluations and the estimation of supply responses (Alston, 2010).

Data on agricultural R&D is scarce. As our analysis comprises many countries we use three different data sources. The first data set was compiled by Alston et al. (1999, p.61). It covers 22 OECD countries<sup>41</sup> and provides information on related public and private expenditure (measured in constant

---

<sup>41</sup>Organisation for Economic Co-operation and Development.

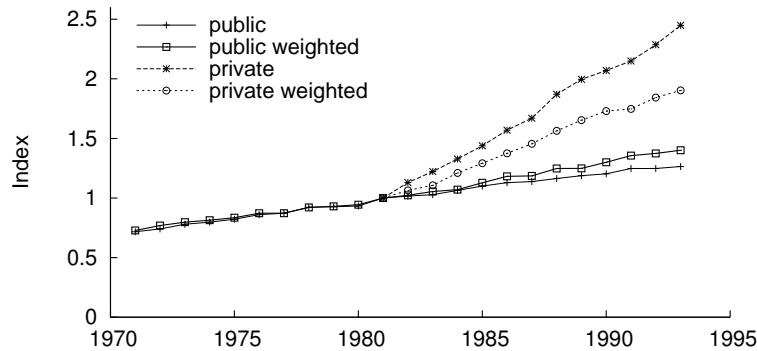


Figure 5.2: Indices of public and private R&D expenditure in 22 OECD countries (1981=1.00; own construction using data by Alston et al. (1999)).

international dollars). The years covered are 1971 to 1993 (public) and 1981 to 1993 (private). Figure 5.2 depicts the corresponding indices. The annual increase in public expenditure is 2.6 percent on average. The share-weighted index suggests that countries with higher expenditure shares spent even more on agricultural R&D (the weighted average rate is 0.03). Private expenditure increases faster (the weighted average rate is 0.055), but in this case, the effort is stronger in case of some smaller countries because the unweighted average rate is about 7.8 percent.

The second data set is directly provided by the OECD (2014). It covers the years 1981 to 2010 and comprises 33 countries of our data set (including some non-OECD members). Although there is a differentiation of expenditure according to sector (e.g. government, higher education), the series are very fragmented.

Unfortunately, the same pattern applies to the third data source provided by the Agricultural Science and Technology Indicators initiative (ASTI, 2014) for the same period. There is a similar categorisation of spendings for 62 developing and transition countries that are included in our data. However, the actual coverage over the years is much lower than the suggested 30 years.

## 5.5 Results

### Coefficients and TE scores

Table 5.4 presents the estimated point elasticities of the two models, without and with the input variable feed. The estimates have the expected signs and are significant. In the first model, the variables with the largest coefficients are livestock, land and labour. The magnitude is lower for fertiliser that is a more flexible input, whereas the others are rather stocks. Machinery shows the lowest coefficient suggesting that the labour to capital ratio is high, i.e. if machinery is interpreted as a proxy for capital, production tends to be rather labour-intensive.

Table 5.4: Point estimates of production frontiers.

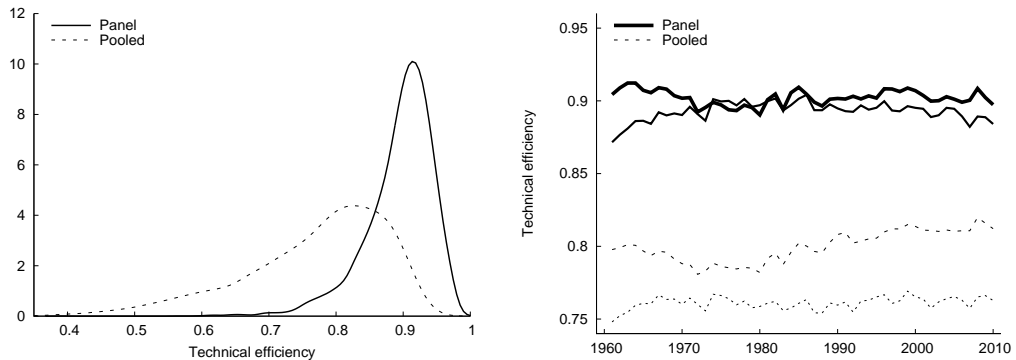
	<u>without feed</u>		<u>with feed</u>	
	Coeff.	t-value	Coeff.	t-value
labour	0.196	17.5	0.198	16.3
land	0.372	19.5	0.321	16.8
machinery	0.062	11	0.035	6
fertiliser	0.129	26.7	0.109	20.5
livestock	0.399	32.9	0.265	19.3
feed			0.157	18.3
trend	0.011	8.4	0.009	26.4
RTS	1.158		1.085	
mean TE	0.882		0.894	
std.dev TE	0.057		0.051	
countries	171		157	
obs.	8490		7362	

Note: inputs are given in logarithms. The complete estimation outputs are documented in the appendix.

Consideration of feed as in the second model results in a decrease of virtually all coefficients. The change is strong for livestock and land but the relative change is especially strong for machinery. The findings for livestock and feed seem intuitive as these variables are closely related. The coefficient of feed shows a distinct magnitude. However, the scale elasticity still indicates increasing returns to scale (IRS), but the value is less extreme than in the first model.

Another important aspect is related to the time trend. Its coefficient suggests technical progress in both models. It drops, though, in the second model where feed is included. This is a plausible finding since disregard of inputs (input growth) can result in an overestimation of technical progress (and TFPC, respectively). As the second model appears to be more reliable, we base our analysis on this model.

Results point out the existence of inefficiencies. Predicted TE is 0.89 on average with a standard deviation of 0.05. The density plots in Figure 5.3 highlight the relevance of considering the panel structure of the data. Disregard of this structure (as immanent to the standard pooled SF model (dashed lines)) results in an overestimation of inefficiency. This consequence applies to the mean TE scores but, more importantly, also to their variances. Accordingly, much of the variation that we can attribute to individual country-effects, is soaked up into TE scores when treating the data as a cross-section.



Note: the solid lines refer to the results of the panel SF model presented in Section 5.2. The dashed lines refer to the TE scores of the standard pooled SF model. Right panel: thicker lines refer to weighted scores; the others are geometric averages.

Figure 5.3: Densities of TE scores (left panel) and respective sample means over the observation period (right panel).

Average TE scores and their development over time are depicted in the right panel of Figure 5.3. The geometric means indicate that TE is changing within some countries. However, the overall average is rather stable when weighting scores by output (for remarks on weighting, see description of equation 5.9). This is a plausible finding since country-level data does not change quickly. The countries that dominate the overall technology at the world frontier and the respective TE index, preserve the efficiency level. An exception is the development from 1961 to the beginning of the 1970s that can eventu-

ally provide an indication for some inconsistencies in the data. For example, Coelli and Rao (2005) omitted this period because of missing data on labour; Headey et al. (2010, p.4) do not trust in the variables and decide to use only data “since 1970, to include only observations with the least error”.

The much bigger divergence of weighted and unweighted averages under the pooled SF model is also a consequence of neglecting unobserved heterogeneity.

### TFP change and components

The decomposition of TFP change follows Kumbhakar and Lovell (2000) and allows to draw conclusions about the nature of productivity growth. The three different sources are change in TE (TEC), technical change (TC) and a scale component (SC). The rate of TFP change (TFPC) is calculated as:

$$TFPC_{it} = [-(u_{it} - u_{it-1})] + \frac{\partial \ln \hat{y}_{it}}{\partial t} + \left[ (RTS_{it} - 1) \times \sum_j \left( \frac{\epsilon_{jit}}{RTS_{it}} \right) (\ln x_{jit} - \ln x_{jit-1}) \right] \quad (5.8)$$

The first term on the right hand side is TEC. The second term is TC that is the partial derivative of the estimated production frontier with respect to the time trend. The last term is the scale component that is supposed to provide a positive average contribution due to the estimated IRS.<sup>42</sup>

The respective rates are used to calculate indices for each country. As pointed out by Färe and Zelenyuk (2003) as well as Zelenyuk (2006) (geometric) means can be inappropriate when presenting economic measures of groups. In our case, the economic units are countries and the groups are geographical regions. We account for the countries’ relevance in overall agricultural production via their output shares. The index of region  $r$  in year  $t$  is obtained as:

$$Index_t^r = \sum_i Index_{it}^r \times \frac{y_{it}^r}{\sum_i y_{it}^r} \quad (5.9)$$

where *Index* corresponds to either the TEC, TC, SC or TFPC index. The

---

<sup>42</sup>Individual results are averaged over years  $t$  and  $t-1$  to obtain the TC rate of year  $t$ . Individual output elasticities and RTS are also averaged over two subsequent years.

group measure is a weighted average of the individual indices. The respective weights are given by the share of country  $i$ 's production in total regional production in year  $t$ .

Results of the TFPC decomposition are presented in Table 5.5. There is a distinct progress in TFP over the observation period. The TFPC index shows a value of 1.79 in 2010 corresponding to an average growth rate of 1.2 percent. TEC does not contribute to TFPC as discussed on the basis of weighted TE scores. The main driver of productivity growth is TC, but scale-related contributions are also relevant. However, the three sources of TFPC are of different relevance among regions as highlighted by the group measures.

Table 5.5: Weighted indices of TFP change and its sources in 2010 (1961 = 1.00).

	TECi	TCi	SCi	TFPCi	avg. TFPC
World	1.00	1.69	1.07	1.79	1.2%
South Africa	1.12	1.54	1.01	1.67	1.0%
Sub-Saharan Africa	0.98	1.24	1.11	1.30	0.5%
Latin America	1.00	1.52	1.07	1.59	1.0%
North America	0.93	2.11	1.00	1.93	1.4%
Asia	1.02	1.71	1.11	1.94	1.4%
Europe	0.97	1.70	0.99	1.61	1.0%
West Asia & North Africa	1.13	1.50	1.07	1.74	1.1%
Oceania	0.94	1.57	1.00	1.45	0.8%
USSR (1992=1.00)	0.99	1.32	0.98	1.26	1.3%

Notably, the growth paths of South Africa as well as West Asia & North Africa are characterised by positive TEC. This contribution can be interpreted as a catch-up of these countries' (unobserved) group frontiers towards the world frontier, i.e. there has been TC within these regions, increasing their production possibilities. Conversely, negative TEC is found for Oceania and North America, which is an unexpected result. North America is the region with the highest TC index. Accordingly, the finding suggests that North America cannot keep up with its own technological progress. While there is a comparable pattern of TC among most of the groups, Sub-Saharan Africa is lagging behind. This region thus shows the lowest progress. However, it can profit from rather high scale-related changes, as can Asia, Latin America and

West Asia & North Africa.

The highest indices of TFPC are found for North America and Asia. Since these regions are important in terms of land area and correspondingly in output, their growth pushes the weighed index of the world frontier. The results from the geometric means are quite different (see appendix), highlighting divergent developments in individual countries. The unweighted average TFPC for the world frontier is only 0.5 percent.

The discussion on agricultural productivity and its growth often addresses the magnitude and the historical pattern of its rate. The concern is about the future path of productivity growth and the ability of the agricultural sector to meet various challenges. A potential slowdown was rejected by Fuglie (2008) as well as by Fuglie (2012), using this data but a growth accounting approach. Our results confirm this notion as demonstrated by Table 5.6.

Table 5.6: Average rates of TFP change according to decades (in %).

	60s	70s	80s	90s	00s
World	0.5	0.8	1.2	1.7	1.7
South Africa	0.4	1.2	1.1	0.9	1.9
Sub-Saharan Africa	0.2	-0.1	0.6	1.3	0.7
Latin America	0.2	0.2	1.1	1.4	1.7
North America	0.8	1.3	1.3	1.7	1.7
Asia	0.5	0.7	1.6	2.0	1.9
Europe	0.6	0.9	1.0	1.2	1.1
West Asia & North Africa	0.6	0.6	1.2	1.7	1.6
Oceania	0.4	1.0	0.5	1.4	0.5
USSR				0.8	1.6

All average growth rates are positive (with only one exception). North America shows a very stable development, whereas Asia speeded up since the 1970s. The growth path of Latin America is also characterised by a sustained increase in the TFPC rate. The finding of an increasing rate seems to be a general tendency. The accumulation of knowledge stocks that generate returns (like compounded interest), is a possible reason for this observation.

### **TFP change and R&D**

Accordingly, R&D expenditure should have an effect on TFPC. We use the data on 22 OECD countries that provide a complete series of agricultural



R&D effort, and compare it with the derived TFPC results. Since the data set is not large, we apply a very basic approach using three models with three different lags (5, 10 and 15 years) to account for the (unknown) time that is required for the diffusion of research outcomes. The R&D data is compared to rates of TFPC, TC and TEC, using three models for each measure (Table 5.7).

Table 5.7: Regression of TFPC rates on lagged R&D expenditure for 22 OECD countries.

<i>public expenditure</i>						
	TFPC		TC		TEC	
	Coeff.	t-value	Coeff.	t-value	Coeff.	t-value
constant	-0.04	-0.09	0.21	5.98	-0.28	-0.66
$\ln R\&D_{t-5}$	0.20	2.36	0.17	22.96	0.04	0.47
$adj.R^2$	0.01		0.51		0.00	
constant	0.26	0.60	0.36	10.19	-0.11	-0.24
$\ln R\&D_{t-10}$	0.15	1.69	0.17	22.84	-0.01	-0.13
$adj.R^2$	0.00		0.51		0.00	
constant	0.34	0.81	0.51	14.30	-0.17	-0.40
$\ln R\&D_{t-15}$	0.15	1.77	0.17	22.46	0.00	-0.04
$adj.R^2$	0.00		0.50		0.00	
<i>private expenditure</i>						
	TFPC		TC		TEC	
	Coeff.	t-value	Coeff.	t-value	Coeff.	t-value
constant	0.30	0.77	0.70	24.19	-0.43	-1.10
$\ln R\&D_{t-5}$	0.15	1.80	0.11	16.99	0.05	0.64
$adj.R^2$	0.01		0.50		0.00	
constant	0.64	1.72	0.84	28.32	-0.23	-0.60
$\ln R\&D_{t-10}$	0.08	0.97	0.11	16.30	-0.02	-0.18
$adj.R^2$	0.00		0.48		0.00	
constant	0.72	1.88	1.00	33.76	-0.31	-0.79
$\ln R\&D_{t-15}$	0.11	1.36	0.10	15.81	0.03	0.33
$adj.R^2$	0.00		0.47		0.00	

Note: public exp., 1971-1993, n=506. Private exp., 1981-1993, n=286.

We find no evidence for a relationship between R&D expenditure and TEC. However, public as well as private spendings are positively related to TC and TFPC. Technical progress drives the level of TFP with an increasing rate, as was shown above and is also apparent from the estimated intercepts. While this is a continuous process, TFPC rates in single years are rather affected by TEC that is more volatile, but tends to be zero on average. Therefore, the results for TFPC are less significant than for TC. R&D expenditure is certainly an important factor for TC in OECD countries. Higher rates of TC are associated with higher budget efforts. This relationship seems to be stable

as suggested by the different lags.

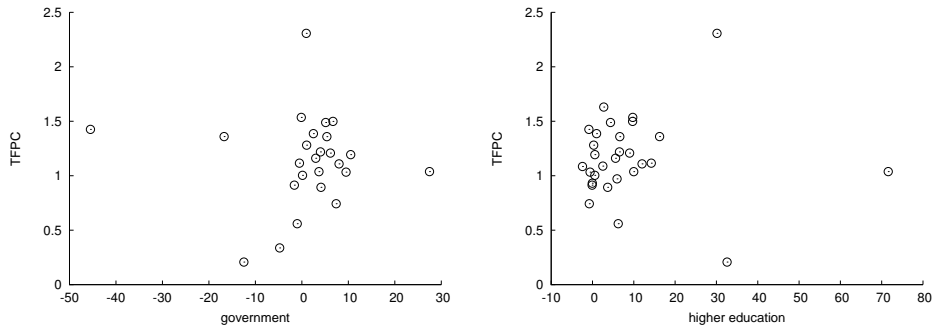


Figure 5.4: Average rates of TFPC and spendings on R&D in the government sector (left panel) as well as in the sector of higher education (right panel), for several OECD and transition countries (in %, 1981-2010; own construction using data by OECD (2014)).

The time series of the other two data sets are fragmented, and for some countries, there are only a few numbers available. Therefore, we calculate average changes where possible, and compare these with the average TFPC rates over the period 1981 to 2010. Most data are available for the sectors government and higher education. There is no apparent pattern as indicated by the data; even countries with no change in the budget show substantial rates of TFPC. There are no other findings in case of TC or TEC.

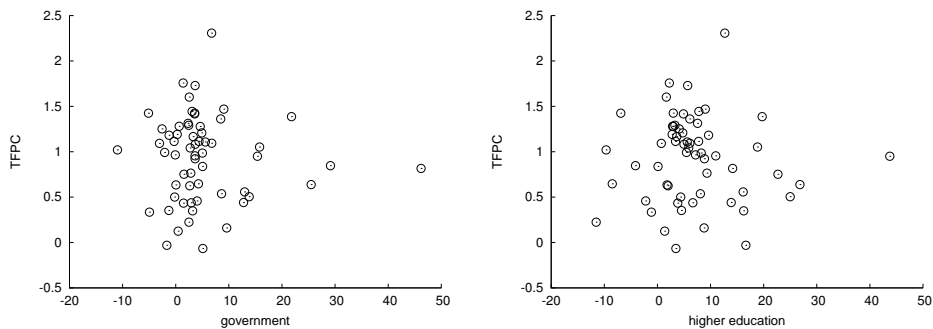


Figure 5.5: Average rates of TFPC and spendings on R&D in the government sector (left panel) as well as in the sector of higher education (right panel), for a set of developing and transition countries (in %, 1981-2010; own construction using data by ASTI (2014)).

## 5.6 Discussion

This study uses data on global agricultural production that are regularly analysed (at least partially). Many challenges that are associated with agriculture

are characterised by their international relevance (e.g. food security, climate change). Hence, modelling of the world frontier is basically meaningful. We use a specific data set that includes many countries as well as adjusted variables that are different from the original FAO numbers (Fuglie, 2012). The estimated production frontier uses feed as an input variable that has been neglected by other studies.

We find the consideration of feed to be important as it results in a decreasing rate of estimated TC (from 0.7 percent to 0.5 percent, on average), and hence in a lower rate of TFPC. This is consistent with the concept of TFP that considers all inputs in the production process. In fact, the data are unlikely to be complete, but this finding stresses the potential relevance of omitted variables.

The consideration of TE is straightforward in the context of the global agricultural frontier because it allows for the assessment of unused potentials. However, our results highlight that the standard SF model is an inappropriate instrument for this purpose. The individual effects can apparently absorb a lot of the unobserved country characteristics.

The reliability of the data is still problematic. Variables are partially estimated, as e.g. labour. The documentation of country-level data can possibly omit important variables, or even overestimate other inputs when recorded several times at different stages in the production process. The precision of variables such as labour or capital is not clear. There is certainly no correction for quality differences in the data when using the number of workers in agriculture. Similarly, the machinery variable is most likely only an incomplete proxy for the capital input. Production can also be misspecified because the aggregation of outputs is inappropriate. This relates to the corresponding indices or the number of outputs considered. Important dynamics, e.g. productivity gains in single outputs, can be averaged out. Furthermore, the correct differentiation between outputs and inputs (e.g. livestock and feed) is not assured.

With regards to the estimated technology parameters, the ratio of labour to machinery (capital) is the most striking result. The coefficients apparently

show an imbalance of these variables. Labour is probably overestimated since the statistics assign too many people to be active in agriculture. Conversely, the use of capital is likely to be underestimated. The frontier thus tends to represent a labour-intensive production technology. This structure might not fit the reality of many regions and is a possible explanation for the negative TEC of North America, Oceania and Europe. Since North America is the technological leader at the world frontier, it is rather implausible that the country does not keep up with its own development.

The stable level of (weighted) TE at the world frontier implies that there is a persistent level of inefficiency. Potential determinants at the country-level are malfunctioning factor markets (also property rights, e.g. land rights) or the structure of production, including education and subsistence agriculture. However, these determinants can change, what is reflected in the TEC index of certain regions. Structural change can also be inferred from increasing SC indices. A main source of the overall persistent inefficiency is most likely the weather. The model's noise component cannot fully account for weather effects. Other factors relate to environmental conditions including effects such as land degradation (e.g. soil erosion, desertification).

We find increasing rates of productivity which is in line with literature findings (e.g. Fuglie (2008), Fuglie (2012)). Although the real magnitude of these rates is unknown, the relative distribution is likely to be robust. The investigation of agricultural R&D expenditure in the context of agricultural productivity requires more data. Related aspects are data quality as well as comparability for international frameworks. How innovations diffuse into the production process is subject to ongoing research. In their meta-study, Alston et al. (2000) report lag lengths of more than 30 years and various specifications for the lag structures.

In case of OECD countries, we find some evidence for a positive relationship between R&D effort and TFP growth. However, these comparisons cannot reveal potential spillovers between institutions (e.g. public and private activities) or countries. Agricultural production in one country can profit from research efforts in another country. We find that R&D expenditure is

stronger related to TC. This is a plausible result because research is supposed to provide the basis for new production possibilities. By contrast, TEC is less related and can rather be associated with agricultural extension or similar activities. The corresponding resources are probably documented in other budget plannings than the statistics on R&D. The use of average rates as for the two data sets of OECD (2014) and ASTI (2014) is too simplistic.

Future research will require improved data on agricultural R&D. Restricted access to technology is probably an relevant aspect. The panel SF model provides a more reliable assessment of TE. To some extent, the model can possibly account for restricted access to the overall technology via the effects (similar to meta-technology-ratios). However, future research should also address the time series properties of the data, as pointed out by Eberhardt and Teal (2013).

## References

- Aigner, D., Lovell, C. A. K. and Schmidt, P. (1977). Formulation and Estimation of Stochastic Frontier Production Function Models. *Journal of Applied Econometrics* 6: 21–37.
- Alston, J. M. (2010). The Benefits from Agricultural Research and Development, Innovation, and Productivity Growth. OECD Food, Agriculture and Fisheries Papers, No. 31.
- Alston, J. M., Marra, M. C., Pardey, P. G. and Wyatt, T. J. (2000). Research returns redux: a meta-analysis of the returns to agricultural R&D. *The Australian Journal of Agricultural and Resource Economics* 44: 185–215.
- Alston, J. M., Pardey, P. G. and Smith, V. H. (1999). *Paying for Agricultural Productivity*. The Johns Hopkins University Press, Baltimore and London.
- ASTI (2014). Data for CGIAR Centers. <http://www.asti.cgiar.org/cgiar-data>, Agricultural Science and Technology Indicators initiative led by the International Food Policy Research Institute.
- Battese, G. E. and Coelli, T. J. (1988). Prediction of Firm-Level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data. *Journal of Econometrics* 38: 387–399.
- Butzer, R., Mundlak, Y. and Larson, D. F. (2012). Measures of Fixed Capital in Agriculture. In Fuglie, K. O., Wang, S. L. and Ball, V. E. (eds), *Productivity Growth in Agriculture: An International Perspective*. CABI.

- Chen, Y.-Y., Schmidt, P. and Wang, H.-J. (2014). Consistent estimation of the fixed effects stochastic frontier model. *Journal of Econometrics* .
- Coelli, T. J. and Rao, D. S. P. (2005). Total factor productivity growth in agriculture: a Malmquist index analysis of 93 countries, 1980-2000. *Agricultural Economics* 32: 115–134.
- Eberhardt, M. and Teal, F. (2013). No Mangoes in the Tundra: Spatial Heterogeneity in Agricultural Productivity Analysis. *Oxford Bulletin of Economics and Statistics* 75: 914–939.
- Färe, R. and Zelenyuk, V. (2003). On aggregate Farrell efficiencies. *European Journal of Operational Research* 146: 615–620.
- Fuglie, K. O. (2008). Is a slowdown in agricultural productivity growth contributing to the rise in commodity prices? *Agricultural Economics* 39: 431–441.
- Fuglie, K. O. (2012). Productivity Growth and Technology Capital in the Global Agricultural Economy. In Fuglie, K. O., Wang, S. L. and Ball, V. E. (eds), *Productivity Growth in Agriculture: An International Perspective*. CABI.
- Greene, W. H. (2005). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics* 126: 269–303.
- Headey, D., Alauddin, M. and Rao, D. S. P. (2010). Explaining agricultural productivity growth: an international perspective. *Agricultural Economics* 41: 1–14.
- Kumbhakar, S. C. and Lovell, C. A. K. (2000). *Stochastic Frontier Analysis*. Cambridge University Press.
- Ludena, C. E., Hertel, T. W., Preckel, P. V., Foster, K. and Nin, A. (2007). Productivity growth and convergence in crop, ruminant, and nonruminant production: measurement and forecasts. *Agricultural Economics* 37: 1–17.
- Mullen, J. (2007). Productivity growth and the returns from public investment in R&D in Australian broadacre agriculture. *The Australian Journal of Agricultural and Resource Economics* 51: 359–384.
- Mundlak, Y., Butzer, R. and Larson, D. F. (2012). Heterogeneous technology and panel data: The case of the agricultural production function. *Journal of Development Economics* 99: 139–149.
- O’Donnell, C. J., Rao, D. S. P. and Battese, G. E. (2008). Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. *Empirical Economics* 34: 231–255.

- OECD (2014). Innovation in science, technology and industry: Research and Development Statistics. <http://www.oecd.org/sti/rds>, Organisation for Economic Co-operation and Development.
- Swinnen, J., Van Herck, K. and Vranken, L. (2012). Agricultural Productivity Path in Central and Eastern European Countries and the Former Soviet Union: The Role of Reforms, Initial Conditions and Induced Technological Change. In Fuglie, K. O., Wang, S. L. and Ball, V. E. (eds), *Productivity Growth in Agriculture: An International Perspective*. CABI.
- Thirtle, C., Piesse, J. and Schimmelpfennig, D. (2008). Modeling the length and shape of the R&D lag: an application to UK agricultural productivity. *Agricultural Economics* 39: 73–85.
- USDA (2014). International Agricultural Productivity – Overview. [http://www.ers.usda.gov/data-products/international-agricultural-productivity.aspx#.UuY\\_mvswfGg](http://www.ers.usda.gov/data-products/international-agricultural-productivity.aspx#.UuY_mvswfGg), United States Department of Agriculture – Economic Research Service.
- Zelenyuk, V. (2006). Aggregation of Malmquist productivity indexes. *European Journal of Operational Research* 174: 1076–1086.

## Appendix

Table 5.8: List of countries.

### *Africa, developed*

South Africa

### *Africa, Sub-Saharan*

*Central:* Cameroon, Central African Republic, Congo, Congo DR, Equatorial Guinea, Gabon, Sao Tome and Principe; *Eastern:* Burundi, Kenya, Rwanda, Seychelles, Tanzania, Uganda; *Horn:* Djibouti, Ethiopia, Somalia, Sudan; *Nigeria:* Nigeria; *Sahel:* Burkina Faso, Cape Verde, Chad, Gambia, Mali, Mauritania, Niger, Senegal; *Southern:* Angola, Botswana, Comoros, Lesotho, Madagascar, Malawi, Mauritius, Mozambique, Namibia, Réunion, Swaziland, Zambia, Zimbabwe; *Western:* Benin, Côte d'Ivoire, Ghana, Guinea, Guinea-Bissau, Liberia, Sierra Leone, Togo, Uganda;

### *Latin American countries*

*Caribbean:* Bahamas, Cuba, Dominican Republic, Haiti, Jamaica, Lesser Antilles, Puerto Rico, Trinidad and Tobago; *Central America:* Belize, Costa Rica, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama; *Andean:* Bolivia, Colombia, Ecuador, Peru, Venezuela; *North East:* Brazil, French Guiana, Suriname, Guyana; *Southern Cone:* Argentina, Chile, Paraguay, Uruguay;

### *North America*

Canada, United States

### *Asia*

*Developed:* Japan, Korea-Republic, Taiwan; *North East:* China, Korea-DPR, Mongolia; *South East:* Brunei Darussalam, Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Thailand, Timor Leste, Vietnam; *South:* Afghanistan, Bangladesh, Bhutan, India, Nepal, Pakistan, Sri Lanka;

### *Europe*

*Northwest:* Austria, Belgium-Luxembourg, Denmark, Finland, France, Germany, Iceland, Ireland, Netherlands, Norway, Sweden, Switzerland, United Kingdom; *Southern:* Cyprus, Greece, Italy, Malta, Portugal, Spain; *Transition:* Albania, Bulgaria, Czechslovakia-former, Hungary, Poland, Romania, Yugoslavia-former;

### *West Asia and North Africa*

*North Africa:* Algeria, Egypt, Libya, Morocco, Tunisia; *West Asia:* Bahrain, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Oman, Qatar, Saudi Arabia, Syria, Turkey, United Arab Emirates, Yemen;

### *Oceania*

*Developed:* Australia, New Zealand; *Developing:* Fiji, Micronesia, New Caledonia, Papua New Guinea, Polynesia, Solomon Islands, Vanuatu;

### *Former USSR*

*Baltic:* Estonia, Latvia, Lithuania; *Central Asia and the Caucasus:* Armenia, Azerbaijan, Georgia, Kyrgyzstan, Tajikistan, Turkmenistan, Uzbekistan; *East Europe:* Belarus, Kazakhstan, Moldova, Russian Federation, Ukraine;



Table 5.9: Production frontier with five inputs.

	Coefficient	t-value
labour	0.196	17.5
land	0.372	19.5
machinery	0.062	11
fertiliser	0.129	26.7
livestock	0.399	32.9
trend	0.011	38.4
$0.5 \times \text{labour}^2$	0.030	4.38
$0.5 \times \text{land}^2$	0.060	5.03
$0.5 \times \text{machinery}^2$	0.003	1.54
$0.5 \times \text{fertiliser}^2$	0.020	14.6
$0.5 \times \text{livestock}^2$	0.148	18
$0.5 \times \text{trend}^2$	0.000	18.1
labour $\times$ land	0.028	3.66
labour $\times$ machinery	-0.002	-0.572
labour $\times$ fertiliser	-0.016	-7.53
labour $\times$ livestock	-0.041	-7.85
labour $\times$ trend	-0.001	-7.11
land $\times$ machinery	-0.013	-3.37
land $\times$ fertiliser	0.010	3.54
land $\times$ livestock	-0.087	-11
land $\times$ trend	0.001	6.99
machinery $\times$ fertiliser	0.002	1.29
machinery $\times$ livestock	0.003	1.27
machinery $\times$ trend	0.001	8.03
fertiliser $\times$ livestock	-0.010	-5.33
fertiliser $\times$ trend	0.001	8.45
livestock $\times$ trend	-0.001	-4.27
$\ln \sigma_v^2$	-4.685	-75.1
$\ln \sigma_u^2$	-3.618	-52.6

Note: N = 171, avg. T = 49.7. Inputs are given in logarithms.

Table 5.10: Production frontier with six inputs.

	Coefficient	t-value
labour	0.198	16.3
land	0.321	16.8
machinery	0.035	6
fertiliser	0.109	20.5
livestock	0.265	19.3
feed	0.157	18.3
trend	0.009	26.4
$0.5 \times \text{labour}^2$	0.028	3.89
$0.5 \times \text{land}^2$	0.018	1.4
$0.5 \times \text{machinery}^2$	0.004	1.71
$0.5 \times \text{fertiliser}^2$	0.015	9.77
$0.5 \times \text{livestock}^2$	0.132	14.2
$0.5 \times \text{feed}^2$	0.019	5.13
$0.5 \times \text{trend}^2$	0.000	13.1
labour $\times$ land	0.061	7.59
labour $\times$ machinery	0.003	0.826
labour $\times$ fertiliser	-0.010	-4.33
labour $\times$ livestock	-0.041	-7.13
labour $\times$ feed	-0.024	-6.87
labour $\times$ trend	0.000	-1.91
land $\times$ machinery	-0.011	-2.6
land $\times$ fertiliser	0.001	0.262
land $\times$ livestock	-0.096	-10.7
land $\times$ feed	0.019	4.56
land $\times$ trend	0.001	5.48
machinery $\times$ fertiliser	0.002	1.44
machinery $\times$ livestock	-0.003	-0.906
machinery $\times$ feed	-0.005	-1.9
machinery $\times$ trend	0.000	2.39
fertiliser $\times$ livestock	-0.002	-0.985
fertiliser $\times$ feed	0.002	1.39
fertiliser $\times$ trend	0.000	0.199
livestock $\times$ feed	-0.003	-0.777
livestock $\times$ trend	-0.001	-5.87
feed $\times$ trend	0.001	8.98
$\ln \sigma_v^2$	-4.795	-63.5
$\ln \sigma_u^2$	-3.853	-42

Note: N = 157, avg. T = 47. Inputs are given in logarithms.

Table 5.11: Weighted indices of TFP change and its sources in 2010 (1961 = 1.00).

	TECi	TCi	SCi	TFPCi	avg. TFPC
World	1.00	1.69	1.07	1.79	1.2%
South Africa	1.12	1.54	1.01	1.67	1.0%
Sub-Saharan Africa	0.98	1.24	1.11	1.30	0.5%
Latin America	1.00	1.52	1.07	1.59	1.0%
North America	0.93	2.11	1.00	1.93	1.4%
Asia	1.02	1.71	1.11	1.94	1.4%
Europe	0.97	1.70	0.99	1.61	1.0%
West Asia & North Africa	1.13	1.50	1.07	1.74	1.1%
Oceania	0.94	1.57	1.00	1.45	0.8%
USSR (1992=1.00)	0.99	1.32	0.98	1.26	1.3%

Table 5.12: Geometric means of indices in 2010 (1961 = 1.00).

	TECi	TCi	SCi	TFPCi	avg. TFPC
World	1.01	1.24	1.05	1.27	0.5%
South Africa	1.12	1.54	1.01	1.67	1.0%
Sub-Saharan Africa	1.00	1.05	1.09	1.10	0.2%
Latin America	1.04	1.21	1.05	1.29	0.5%
North America	0.96	2.02	1.00	1.90	1.3%
Asia	0.99	1.34	1.09	1.41	0.7%
Europe	0.98	1.53	0.99	1.46	0.8%
West Asia & North Africa	1.16	1.33	1.09	1.44	0.7%
Oceania	0.97	0.99	1.02	0.95	-0.1%
USSR (1992=1.00)	1.00	1.24	0.99	1.20	1.0%

## 6 Discussion

The first part of this section briefly summarises the key findings of the applications. The focus is on the interpretation of the results and their compatibility with the underlying assumptions, i.e. the robustness of estimated technical efficiency (TE) scores and growth in TFP.

The second part addresses the relevance of unobserved heterogeneity in the applications. It refers to the consequences of its consideration and the corresponding economic interpretation, rather than to statistical evaluations. A brief discussion aims at the scope of the interpretation as well as the need for future research.

### 6.1 Findings on efficiency and productivity growth

#### Technical efficiency

The applications comprise firm-level data on meat firms and livestock farms in Europe, as well as agricultural country-level data for many countries of the world. In case of the meat industry, results indicate a very high level of TE. The variation in predicted TE scores is low. The analysis suggests that there are no firms that are unable to handle the respective technology. Correspondingly, firms with a low performance are likely to drop quickly out of competition. The small remaining inefficiencies can be due to some excess capacities (e.g. machinery). However, these resources might be required in order to maintain some flexibility.

Conversely, the average efficiency in livestock farming is much lower and the variation in TE scores much higher. These divergencies can be associated with a more heterogeneous structure in agricultural production, i.e. although the analysis accounts for individual effects, the performance among farms differs a lot. There are several plausible reasons for this situation. Factor markets do not always work efficiently or are subject to restrictions (e.g. the land market). Farmers might not quickly switch to alternative jobs. Furthermore, inefficient production can partly be compensated by subsidies.

If there are technical inefficiencies at the country-level, production cannot

be fully efficient at the global scale. This is reflected in the FAO data analysed in Section 5. However, the aggregated data shows relatively lower variation in TE and a stable overall level at the sample mean (weighted TE scores). This finding is reasonable since country data are naturally less dynamic and most resources tend to be fixed. Countries that dominate the world frontier define the overall technological possibilities. If these countries have control over their technology, there should not be much dynamics in TE. The fact that the remaining inefficiency tends to be persistent, is rather due to the characteristics of agricultural production. For the aggregated data, natural conditions, such as weather, are more likely to be relevant than knowledge or malfunctioning markets. Probably, full potential of production is not utilised because of these natural conditions that are not completely captured by the noise component.

The model framework rests on assumptions about the distributions of the error terms. While the assumption of the normally distributed noise component is usually less contested, there are several alternative specifications for the inefficiency component in case of cross-sectional data. In these models, the basic interpretation of inefficiency tends to be robust, irrespective of the selected distribution. In case of panel data, there are less alternatives for the specification of the inefficiency distribution. The here applied model of Chen et al. (2014) is based on the closed skew normal distribution, and is thus restricted to the normal-half normal case.

However, as with cross-sectional data, panel data models relying on different inefficiency distributions should result in similar interpretations. The more important aspect that is emphasised here, refers to the differentiation between cross-sectional models and panel data models. The corresponding results when using the *same* distributional assumption (half-normal) imply different economic interpretations as will be pursued below.

Although the model specification allows for time-varying inefficiency, the residuals are assumed to be independent between two periods. This assumption disregards the notion of autocorrelation and the fact that “inefficiency” can accompany rational economic behaviour. The model possibly ignores dy-

namics such as intertemporal decisions associated with adjustment cost or investments. For example, as pointed out in the case of meat firms, potential excess capacities could be a meaningful requirement.

Since TE is assessed relative to a best practice frontier, this reference technology has to be a valid benchmark. The assumption of a common technology that provides a comparable basis, is a strong one. Some related aspects are mentioned in Section 6.2.

### **Productivity growth**

Change in TE (TEC), technical change (TC) and a scale component (SC) are the three sources that are assumed to determine the pattern of change in total factor productivity (TFPC).

TEC is virtually irrelevant in case of the meat industry. As the corresponding firms are working almost efficiently, there is not much scope for changes. Conversely, this component is one of the productivity drivers of livestock farms. Notably, larger farms are characterised by increasing TE. However, the sector-wide contribution of TEC is related to the ambiguous finding of negative TC. Increasing efficiency is also found in the context of global agricultural production. This finding does not apply to the overall TE at the world frontier. In fact, TEC is relevant in single countries or regions where it can be associated with a catch-up to the world frontier. This effect requires no inefficient production at the unobserved group (regional) frontier because it can be interpreted as a shift of this group frontier towards the meta (world) frontier.

Technical progress tends to be the main driver of TFPC in our applications. This finding is in line with the literature and theoretically plausible, since TC historically has been viewed as the only source of productivity growth. However, we find some peculiarities related to the data as well as to the interpretation. Negative TC is found for European livestock production. Political measures or market events can cause some restrictions to agricultural production. However, the magnitude is implausible. Furthermore, the interpretation of a comprehensive deterioration of production possibilities seems

to be beside the point. In the case of slaughtering firms, there is almost no TC as well as no productivity growth. This finding highlights the relevance of the differentiation between sub-sectors for the interpretation of results.

The SC is the minimal source of TFPC, which is understandable. The overall scale of an enterprise tends to change at a slow rate. If there are mergers that can be realised in the short term, the observed firms are likely to drop from the data. We analyse large meat firms and find no evidence of relevant scale changes. The industry is already higher concentrated and less dynamic in comparison to agriculture where structural change is faster. This conclusion also applies to the country-level. The SC and the underlying structural changes are partially important for global TFPC, e.g. in case of the regions Asia and Sub-Saharan Africa.

The applied decomposition of TFPC is one approach that can be contrasted with alternative concepts. We assume away any components other than TEC, TC and SC. For example, the association of technical inefficiency with malfunctioning factor markets implies the notion of allocative inefficiency and corresponding changes.

Extensions of the productivity decomposition can address the mixtures of outputs and inputs, as for example proposed by O'Donnell (2012). The ignorance of such effects is mainly due to data limitations (not only in these applications). We estimate a distance function for livestock farms that at least accounts for two output categories. The respective results are more plausible than those of the corresponding production function with respect to TE and TFPC (both measures tend to be higher).

However, these concepts require the correct measurement of outputs and inputs that is ultimately dependent on the data sources. We use proxies for capital that should be represented by means of service flows from capital stock into the production process. The construction of alternative measures usually requires additional assumptions that must be justified. We assume that the proxies are sufficient for modelling the capital input.

Related considerations apply to other variables that are measured in monetary terms. An implicit correction for quality can be an advantage of these

variables. However, the disadvantage is manifest in the choice of appropriate indices for deflation. One aspect refers to the relevant categories of outputs and inputs, i.e. the decision for a correct price index. The availability of the respective index is another issue. Apparently, many databases can only provide approximate information.<sup>43</sup>

With regards to productivity analysis, this issue is relevant for the magnitude of the rates as well as for the composition of aggregate outputs. The relative performance of firms might be unaffected. For example, the finding of divergent growth paths in sub-sectors of the meat industry is robust to deflation. In the case of global agricultural production, the aggregate output can be biased due to an inappropriate construction of the variable. The negative TC that we find for livestock farms can, at least partially, be affected by indices that do not represent all economic changes adequately (e.g. number of countries, length of the period, introduction of the euro).

## 6.2 Consideration of unobserved heterogeneity

The standard Stochastic Frontier (SF) model of Aigner et al. (1977) is unable to account for unobserved heterogeneity. However, the original objective of efficiency analysis is to make inferences on inefficiency. Our applications suggest that the panel model of Chen et al. (2014) should be a standard tool, whenever panel data are available. This recommendation is not based on the view that the model is the “true” one. It can provide valuable insights, though, because economic interpretations derived from the standard (pooled) model can be misleading. This section compares some results from the two models, assuming a normal-half normal specification. The description does not draw on statistical arguments but rather focuses on the interpretation of the results.

---

<sup>43</sup>For example, Eurostat uses weights to account for the relevance of EU member states when constructing producer price indices. On occasion, aggregated indices for very different groups are identical, as in the case of the producer price index for meat products. In this case, the index of the EU-15 is identical to the index of the EU-27.



## Coefficients

As expected, most coefficients differ when comparing the models. Figures 6.1 and 6.2 provide selected examples from the firm-level data of meat firms and livestock farms. The changes of estimated elasticities can be drastic; particularly, in case of some “suspect” variables. Capital is such a variable, in general and in case of meat firms. The results suggest that the effect of capital on output is overestimated in the pooled model, because larger elasticities are positively related to the individual effects that are obtained from the panel model. Correspondingly, there is an output response that is rather caused by basic productivity than by capital. However, there are more negative elasticities under the panel model. The respective estimate of labour is apparently more efficient.

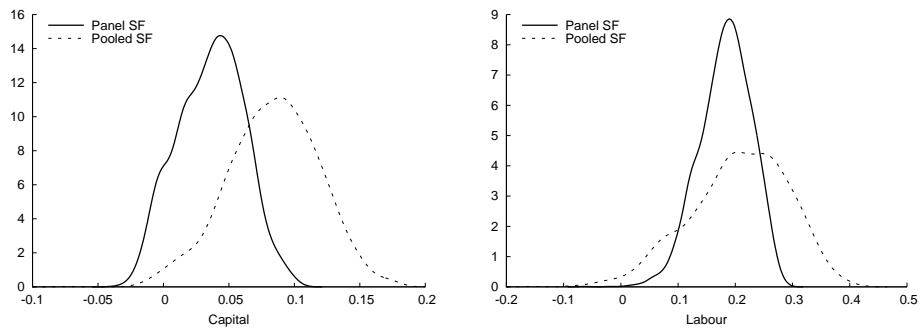


Figure 6.1: Selected distributions of calculated elasticities for meat firms, capital (left panel) and labour (right panel).

A drastic change is found for the elasticities of land in the context of the distance function (livestock farms). Land and labour are special inputs in agricultural production. Land is immobile by definition; labour is supposed to be rather inflexible (in the short and medium term). As illustrated in the left panel of Figure 6.2, monotonicity is violated for most of the observations when using the pooled model. Conversely, the alternative estimator shows almost exclusively positive values (only 0.2 percent are negative). The positive elasticities are required for consistency with economic theory, and are meaningful in the case of land because it is an important input that can represent a high cost share (in particular, for livestock farms with low land endowment).

For labour, the panel model suggests a much lower elasticity (but also

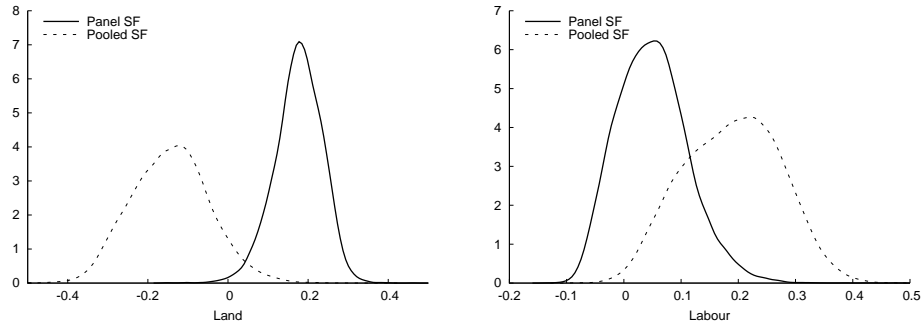


Figure 6.2: Selected distributions of calculated elasticities for livestock farms, land (left panel) and labour (right panel).

shows more negative elasticities). This is an intuitive result because the cost share labour is rather low, due to structural change that results in a lower labour-intensity in agricultural production.

In case of the panel model, the findings on estimated scale elasticities would deserve further investigation in future research. Apparently, there is a stronger trend to decreasing returns to scale for firm-level data, while for country-level data, we find distinct increasing returns to scale. The individual effects absorb production potential, i.e. there are unmeasured factors, resulting in output, that we attribute to basic productivity. Correspondingly, the output elasticities of the inputs are lower. This is a possible interpretation at the firm-level. However, there is no clear pattern in case of the country-level data.

### Technical efficiency

Different coefficients change the reference technology which has consequences for the predicted inefficiencies. In general, TE scores are expected to increase because the panel model decomposes the residual into one more component (composed error plus individual effect). We find this result for agriculture at the firm-level as well as at the country-level.

Table 6.1 presents the descriptive statistics of TE scores for the three studies. The general trend is an increase in average efficiency and a decrease in the respective variation. The changes are particularly strong for the data on international agriculture.

The finding for the meat industry is different. There is virtually no change

for meat processors, but instead of an TE increase, only the panel model suggests inefficiency in case of the slaughtering firms.

Table 6.1: Comparison of TE scores in the applications: pooled SF model and panel SF model, with half-normal distribution of the inefficiency component.

	<u>Pooled SF</u>		<u>Panel SF</u>		correlation
	mean	std.dev	mean	std.dev	
<i>Section 3</i>					
slaughtering	–	–	0.957	0.026	
processing	0.955	0.015	0.954	0.024	0.57
<i>Section 4</i>					
livestock farms	0.773	0.093	0.818	0.082	0.62
<i>Section 5</i>					
South Africa	0.712	0.042	0.899	0.033	0.80
Sub-Sahara Africa	0.749	0.123	0.895	0.050	0.40
Latin America	0.818	0.066	0.892	0.055	0.50
North America	0.822	0.022	0.901	0.024	0.39
Asia	0.758	0.099	0.897	0.041	0.49
Europe	0.756	0.098	0.898	0.038	0.39
W. Asia & N. Africa	0.766	0.120	0.888	0.071	0.41
Oceania	0.808	0.122	0.887	0.065	0.41
Former USSR	0.745	0.067	0.898	0.039	0.46

In general, the results of the efficiency analysis indicate that inefficiency is not an important problem in the meat industry. Furthermore, the potential expansion of agricultural output can be considerably lower than suggested by the standard model.

### **Individual effects and group technology**

The individual effects represent a kind of heterogeneity that is said to be unobservable. Accordingly, the interpretation of these effects must inevitably remain speculative to some extent. In case of the country-level data, there is definitely no straightforward interpretation. The consideration of unobserved heterogeneity is very meaningful, though.

In case of the firm-level data, the individual effects provide additional scope of interpretation when treated as basic productivity. Furthermore, their distribution allows to assess the relevance of the spread in firm characteristics. In our studies, slaughtering firms as well as pig and poultry producers show

a lower variation, whereas it is higher for meat processing firms and cattle farms. The conditions and products in the slaughtering business are clearly defined. Similarly, pig and poultry are predominantly produced under well-specified conditions. Accordingly, these enterprises can be interpreted as more “homogeneous”, i.e. each unit tends to be rather replaceable. This is different for meat processors because they can specialise or diversify their skills and products. It is also different for cattle production because the range of breeds, systems and surrounding conditions is larger than in the other segments.

However, these considerations can possibly cast doubt on the assumption of a common technology. If the technology of these firms (or farms) is not comparable, then their efficiency scores don’t provide any information value. Such a strict view would consequently result in no analysis being conducted. This lack of information is undesirable from an economic perspective, i.e. this basic assumption is usually required.

The notion of technological heterogeneity remains important, though. Two major aspects refer to this kind of heterogeneity: restricted access to the meta-technology, and individually different technology parameters, e.g. as resulting from varying input qualities. Other models will require additional information or alternative assumptions. However, our results suggest that the effects are able to capture a lot of the individual characteristics, that are in turn not confounded with technical inefficiency. Moreover, factors, such as managerial ability, are most likely better modelled using the fixed effects framework than using flexible technology parameters.

A strict view on economic theory suggests that there should be no inefficiency in production. Accordingly, the analysis would address a non-existent problem, i.e. it would be a “fruitless exercise” (Greene, 2008, p.100). The literature on efficiency analysis and our interpretation suggest that inefficiency matters.

In our framework, inefficiency is assumed to vary over time. Theoretically, this assumption is absolutely required in panel data. The assumption on time-invariant unobserved heterogeneity is a more “philosophical issue[s]” (Chen et al., 2014, p.65). Since many unobserved characteristics are fixed

for a certain period, this assumption should be acceptable for the firm-level data (but it might be different for the long series of agricultural data). The acceptance of persistent technical inefficiency within this framework (Colombi et al., 2014), seems more problematic. Firms that show a constant level of inefficiency are likely to drop out of competition. However, such a component could provide indications of market failures or restricted access to technology. Therefore, this approach should also be considered in future research.

Furthermore, from a methodological perspective, a more flexible parameterisation of the composed error term, in the context of the recent model, is clearly desirable. Our results suggest that future research should also address the consequences of variable aggregation in productivity analysis. The necessity of improving (sectoral) data quality and comparability is another important insight.

## References

- Aigner, D., Lovell, C. A. K. and Schmidt, P. (1977). Formulation and Estimation of Stochastic Frontier Production Function Models. *Journal of Applied Econometrics* 6: 21–37.
- Chen, Y.-Y., Schmidt, P. and Wang, H.-J. (2014). Consistent estimation of the fixed effects stochastic frontier model. *Journal of Econometrics* 181: 65–76.
- Colombi, R., Kumbhakar, S. C., Martini, G. and Vittadini, G. (2014). Closed-Skew Normality in Stochastic Frontiers with Individual Effects and Long/Short-Run Efficiency. *Journal of Productivity Analysis* .
- Greene, W. H. (2008). The Econometric Approach to Efficiency Analysis. In Fried, H. O., Lovell, K. C. and Schmidt, S. S. (eds), *The Measurement of Productive Efficiency and Productivity Growth*. Oxford University Press.
- O’Donnell, C. J. (2012). An aggregate quantity framework for measuring and decomposing productivity change. *Journal of Productivity Analysis* 38: 255–272.

# A Stochastic Frontier Analysis using SFAMB for Ox

*Jonathan Holtkamp, Bernhard Brümmer*

## A.1 Introduction

**SFAMB** (Stochastic Frontier Analysis using ModelBase) is a package for estimating stochastic frontier production (as well as cost, distance, and profit) functions. This version includes different specifications for cross sectional data as well as four models for panel data. **SFAMB** is a class written in **Ox** (Doornik, 2009) and is used by writing programs that use an object of this class.

The console versions of **Ox** are free for research and educational purposes. **Ox Console** uses **OxEdit** to run programs. The commercial version of the programming language, **Ox Professional**, uses the graphical user environment **OxMetrics** instead.

The structure of the paper is as follows<sup>44</sup>. In the next section, we briefly introduce the available estimators and their corresponding econometric foundations. Data organization and usage of the code are explained in Section A.3. Section A.4 is a technical documentation of related member functions. We present practical examples using real world data in Section A.5.

## A.2 Stochastic frontier production function estimation

**SFAMB** provides frontier models of Aigner et al. (1977); and Meeusen and van den Broeck (1977), respectively, with extensions; Schmidt and Sickles (1984); Greene (2005); Wang and Ho (2010) as well as Chen et al. (2014). The available estimators are:

There are several other software packages that incorporate (some of) the estimators listed in Table A.1. **LIMDEP** (Econometric Software, Inc., 2014) and **Stata** (StataCorp LP, 2014) are comprehensive commercial packages that im-

---

<sup>44</sup>Structure and format follow the documentation “Panel Data estimation using DPD for **Ox**” (Doornik et al., 2012).

	SetMethod	Ineff.distribution	Example
SFA - cross section	POOLED	$\mu$ and/or $\sigma_u$	hbest1.ox
Least squares with dummies	LSDV		hbest2.ox
SFA - with dummies	TFE		hbest2.ox
SFA - within-transformation	WT	$\mu$ or $\sigma_u^2$	hbest3.ox
SFA - consistent fixed effects	CFE		hbest2.ox

Table A.1: Available estimators.

plement frontier techniques in their standard distributions. In case of **Stata**, there are additional third-party add-ons such as those of Wang (2012) or Belotti et al. (2012).

Hughes (2008) has written two free packages called **sfa\_hetmod**, and **sfa\_mod**, that can be used with **gretl** (Cottrell and Lucchetti, 2014). Both include variations of the standard model where the first one allows for heteroscedasticity.

The recent package **spfrontier** (Pavlyuk, 2014) deals with (the specific family of) spatial Stochastic Frontier models. It is implemented in **R** (**R** Core Team, 2014) and allows for various specifications.

The first program to implement frontier techniques was **Frontier** (Coelli, 1996). Later, the original code was transferred to **R** by Coelli and Henningsen (2013). This package provides extensions of the standard model, namely, the so-called “error components specification” (Battese and Coelli, 1992) and the more frequently used model of Battese and Coelli (1995). Its functionality is augmented by some additional options (e.g., for calculating marginal effects).

Similarly, **SFAMB** offers specific member functions that can be extended by the user. To date, it is the only package including the **CFE** model.

## The Stochastic Frontier model

This section is intended as a short, concise introduction to Stochastic Frontier Analysis (SFA) techniques. A more detailed introduction can be found in Coelli et al. (2005). More advanced material is covered in Kumbhakar and Lovell (2000). The basic problem in efficiency analysis lies in the estimation of an unobservable frontier (production, distance or cost) function from observable input and output data, together with price data when necessary.

Standard estimation techniques like OLS are inappropriate in this setting since they aim at the identification of average relationships, which are not in the focus of an efficiency model.

The basic approach was simultaneously developed by Aigner et al. (1977), and Meeusen and van den Broeck (1977). An exposition in terms of the production function highlights its most important characteristics. The basic production function model is given by:

$$y_i = \alpha + \beta^\top x_i + v_i - u_i \tag{A.1}$$

On the left hand side,  $y_i$  is the output (or some transformation of the output) of observation  $i$  ( $i=1\dots N$ ). On the right hand side,  $x_i$  is a matrix of inputs that produce output  $y_i$ , and the vector  $\beta$  describes technology parameters to be estimated. The most commonly used transformation of the variables is the natural logarithm. The crucial part of this formulation is the composed error term given by  $\epsilon_i = v_i - u_i$ , where  $v_i$  represents statistical noise and  $u_i$  represents inefficiency. Estimation is possible by means of Maximum Likelihood Estimation (MLE) where distributional assumptions concerning the error components are required. The noise component is a conventional two-sided error, distributed as  $v_i \sim N(0, \sigma_v^2)$ . The inefficiency component is a non-negative disturbance that can be modelled using several distributions. However, the *truncated normal* and *half normal* distributions are most frequently used and are implemented in **SFAMB**. Accordingly, the random variable  $u_i$  is distributed as  $u_i \sim N^+(\mu, \sigma_u^2)$ . If  $\mu$  equals zero the model is labelled as the *normal-half normal* SF model; *normal-truncated normal* SF model otherwise.

The independence assumption for the inefficiency distribution in the basic SFA model can be changed by introducing covariates into the distribution, thereby accounting for differences in inefficiency between individuals. The corresponding covariates are often labelled as Z-variables. These can be used to model either the location parameter or scale parameter of the underlying distribution or both, cf. Alvarez et al. (2006). An useful overview is given by Lai and Huang (2010) who summarize and categorize several well-known



models. A model describing  $\mu$  by means of an exponential form is labelled as the KGMHLBC model<sup>45</sup>,  $u_i \sim N^+(\mu \exp(\theta^\top z_i), \sigma_u^2)$ . If  $\mu$  is set to zero and the scale is modelled using an exponential form it is the RSCFG model<sup>46</sup>,  $u_i \sim N^+(0, \exp(2 \delta^\top z_i))$ . Combination leads to  $u_i \sim N^+(\mu_i = \mu \exp(\theta^\top z_i), \sigma_{u,i}^2 = \exp(2 \delta^\top z_i))$ , labelled by Lai and Huang (2010) as a generalized exponential mean model. *Note:* In **SFAMB**, the respective parameter modelled in the **POOLED** model is (the natural logarithm of)  $\sigma_{u,i}$ , not  $\sigma_{u,i}^2$ !<sup>47</sup>

In addition to standard results, the estimation output of the **POOLED** model provides three other results:

**gamma** is given by  $\gamma = \sigma_u^2 / \sigma^2 = \sigma_u^2 / (\sigma_v^2 + \sigma_u^2)$

where (in case of the half-normal specification)  $\sigma_u^2 = \frac{1}{n} \sum_i \exp(2 \delta^\top z_i)$ .

**VAR(u)/VAR(total)** describes the “correct” variance decomposition of the composed error (recall that given  $u$  is a one-sided disturbance,  $\sigma_u^2$  is not the variance  $\text{var}[u]$  of the one-sided error). The share of the variance of  $u$  in the total variance of the composed error is given by  $\text{var}[u] / \text{var}[\epsilon] = [(\pi - 2) / \pi] \sigma_u^2 / [(\pi - 2) / \pi] \sigma_u^2 + \sigma_v^2$ , cf. (Greene, 2008, p.118).

**Test of one-sided err** provides a likelihood ratio test statistic for the presence of inefficiency, i.e., for the null hypothesis  $H_0: \gamma=0$ . The critical value cannot be taken from a conventional  $\chi^2$ -table, see Kodde and Palm (1986).

A point estimator of inefficiency is given by  $E(u_i | \epsilon_i)$ , see Jondrow et al. (1982). If the dependent variable is in logarithms, a more appropriate estimator is the point estimator of technical efficiency  $TE_i = E(\exp(-u_i) | \epsilon_i)$ , see Battese and Coelli (1988).

---

<sup>45</sup>Kumbhakar et al. (1991); Huang and Liu (1994); Battese and Coelli (1995);

<sup>46</sup>Reifschneider and Stevenson (1991); Caudill et al. (1995);

<sup>47</sup>While  $\sigma_u^2$  is often used, the original formulation of CFG involved  $\sigma_u$ .

## Unobserved heterogeneity

With panel data, additional information on each individual is available. Each cross section  $i$  is observed over a certain period of time  $T_i$  ( $t=1\dots T_i$ ):

$$y_{it} = \alpha_i + \beta^\top x_{it} + v_{it} \quad (\text{A.2})$$

This formulation differs from equation (A.1) in that it involves time dimension  $t$ , only one error component (two-sided) and an individual intercept  $\alpha_i$ . The model is estimated by OLS, and hence,  $v_{it} \sim N(0, \sigma_v^2)$ . Its virtue lies in the identification of the  $N$  time-invariant individual (“fixed”) effects. These effects may capture unmeasured attributes, and hence, this approach is one way to deal with (unobserved) heterogeneity. The model has different names in the literature; one that is commonly used is “Least squares with dummy variables” (LSDV). Instead of estimating all  $N$  dummies, the usual approach is to employ a transformation: for each panel  $i$ , the respective variables (e.g.,  $x_{it}$ ) are transformed by subtracting the individual mean (out of  $T_i$ ) from the observation in period  $t$ , i.e.,  $\tilde{x}_{it} = x_{it} - \bar{x}_i$ . This procedure (within-transformation) removes the individual effects (because  $\tilde{\alpha}_i = \alpha_i - \alpha_i = 0$ ) and estimation works only with deviations from means, i.e., with the transformed variables. Estimates of the individual effects are calculated as:

$$\hat{\alpha}_i = \bar{y}_i - \hat{\beta}^\top \bar{x}_i \quad (\text{A.3})$$

Schmidt and Sickles (1984) use the model in a frontier context. They interpret the individual with the highest intercept as 100% technically efficient. The inefficiency of the remaining groups is assessed by  $u_i = \max(\hat{\alpha}) - \hat{\alpha}_i$ ; efficiency estimates are time-invariant and are given by  $TE_i = E(\exp(-u_i))$ . Estimation output of the LSDV model differs from the other models to some extent:

`sigma_e` describes  $\sigma_v$  that is the square root of the corrected estimate of the error variance  $\sigma_v^2 = \frac{SSR}{N(T-1)-K}$ . This estimate is also used to compute the standard errors.

`AIC1 (all obs)` is given by  $AIC1 = -2 \ln L + 2(K + 1)$ ; it uses the likeli-

hood function  $\ln L = -\frac{NT}{2}\ln(2\pi) - \frac{NT}{2}\ln(\sigma^2) - \frac{\sum_i \sum_t \tilde{v}_{it}^2}{2\sigma^2}$  with the uncorrected  $\sigma^2 = \frac{SSR}{NT}$ .

AIC2 uses a different formula for the criterion,  $AIC2 = \ln(\frac{SSR}{NT}) + (2 \frac{K+N}{NT})$ ; that does not need the likelihood function and considers the number of individuals in the penalty term.

## Unobserved heterogeneity in SFA

### Dummy variables - TFE model

The approach outlined above does not distinguish between inefficiency and unobserved heterogeneity because there is no one-sided error component. The respective (“true”) specification of the SF model for panel data is given by:

$$y_{it} = \alpha_i + \beta^\top x_{it} + v_{it} - u_{it} \quad (\text{A.4})$$

This model was proposed by Greene (2005) and is known as the “true fixed effects” (TFE) frontier model. Estimation involves all  $N$  individual effects, and hence, the model suffers from the incidental parameters problem. In micro panels ( $T$  fixed),  $\sigma^2$  is inconsistent as the sample size increases.

The point estimators for inefficiency and technical efficiency are the same as for the POOLED model. Output of the TFE model provides `lambda`, given by  $\lambda = \sigma_u/\sigma_v$ .

### Elimination of dummies - WT model

To overcome the incidental parameters problem Wang and Ho (2010) propose an extension that is based on deviations from means<sup>48</sup>:

$$\tilde{y}_{it} = \beta^\top \tilde{x}_{it} + \tilde{v}_{it} - \tilde{u}_{it} \quad (\text{A.5})$$

This within-transformation (WT) model is estimated by MLE. The transformed noise component is distributed as multivariate normal, i.e.,  $\tilde{v}_{it} \sim MN(0, \Pi)$ . However, simple transformation of the one-sided error component would result in an unknown distribution. Therefore, time-varying inefficiency is specified

---

<sup>48</sup>In addition they demonstrate how the model can be estimated by first-differencing.

as  $u_{it} = u_i^* \times h_{it}$ . The persistent part  $u_i^*$  (inefficiency) is assumed to follow a half-normal or truncated-normal distribution, i.e.,  $u_i^* \sim N^+(\mu, \sigma_u^2)$ , where  $\mu$  is equal to zero in case of a half-normal distribution. The scaling function  $h_{it} = f(\delta^\top z_{it})$  includes firm- and time-specific variables ( $z_{it}$ ) that might affect the inefficiency distribution. The vector  $\delta$  describes the corresponding parameters and the function takes an exponential form, i.e.,  $f(\delta^\top z_{it}) = \exp(\delta^\top z_{it})$ . The use of within-transformation does not affect the component  $u_i^*$  but the function value of  $h_{it}$  is transformed:  $\tilde{u}_{it} = u_i^* \times \tilde{h}_{it}$ . Wang and Ho (2010) present the conditional expectation of  $u_{it}$  in their equation (30); efficiency estimates are given by  $TE_{it} = E(\exp(-u_{it}|\tilde{\epsilon}_{it}))$ . Estimation output of the WT model additionally provides:

`lambda` is given by  $\lambda = \sigma_u/\sigma_v$ ; where  $\ln(\sigma_u^2) = c_u$ ; truncated-normal:  $\sigma_u^2 = \exp(c_u)$ ; half-normal:  $\sigma_u^2 = \frac{1}{NT} \sum_i \sum_t \exp(c_u + \delta^\top z_{it})$ ;

Although the individual effects are not directly estimated with the WT model, they can be recovered, assuming that  $\bar{v}_i = 0$ :

$$\hat{\alpha}_i = \bar{y}_i - \hat{\beta}^\top \bar{x}_i + \bar{u}_i \quad (\text{A.6})$$

### Consistent estimation with time-varying inefficiency - CFE model

Consistent estimation of the fixed effects SF model given in equation (A.4) is demonstrated by Chen et al. (2014). Their solution is also based on deviations from means so that the transformed model looks like equation (A.5). However, the respective likelihood function is derived only from the first  $T-1$  deviations, i.e., from  $\tilde{\epsilon}_i^* = (\tilde{\epsilon}_{i1}, \dots, \tilde{\epsilon}_{i,T-1})^\top$ . This procedure has two advantages. First, within-transformation removes the incidental parameters. Second, an implicit correction of the error variance is achieved by means of the first  $T-1$  deviations.<sup>49</sup> Wang and Ho (2010) use a multivariate normal distribution to model  $v_{it}$  but have to accept a persistent basic inefficiency component  $u_i^*$ . The current model is based on a more general distributional theory and allows for firm-specific and time-varying inefficiency  $u_{it}$ .

<sup>49</sup>With regards to the degrees of freedom, the correction accounts for the  $N$  individuals:  $df = NT - N - K = N(T - 1) - K$ .

The composed error,  $\epsilon = v - u$ , has a skewed distribution (to the left) due to the non-negativeness of  $u$ . Accordingly, the standard (half-normal) SF model has a skew normal distribution, with skewness parameter  $\lambda$  and density:

$$f(\epsilon) = \frac{2}{\sigma} \phi\left(\frac{\epsilon}{\sigma}\right) \Phi\left(-\lambda \frac{\epsilon}{\sigma}\right) \quad (\text{A.7})$$

While the skew normal distribution is a generalization of the normal distribution, it can be generalized itself by using the closed skew normal (CSN) distribution.<sup>50</sup> The composed error has a CSN distribution; what is written as:

$$\epsilon_{it} \sim CSN_{1,1}(0, \sigma^2, -\frac{\lambda}{\sigma}, 0, 1) \quad (\text{A.8})$$

The density of a  $CSN_{p,q}$ -distribution includes a  $p$ -dimensional pdf and a  $q$ -dimensional cdf of a normal distribution. The five associated parameters describe location, scale and skewness, as well as the mean vector and covariance matrix in the cdf. With panel data, the  $T$ -dimensional vector  $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{iT})^\top$  is distributed as:

$$\epsilon_i \sim CSN_{T,T}(0_T, \sigma^2 I_T, -\frac{\lambda}{\sigma} I_T, 0_T, I_T) \quad (\text{A.9})$$

where  $I$  is the identity matrix. Chen et al. (2014) partition the vector  $\epsilon_i$  into linear combinations: its mean  $\bar{\epsilon}_i$  and its first  $T-1$  deviations  $\tilde{\epsilon}_i^*$ . The CSN distribution is “closed under linear combinations” (p.10). The density and respective log likelihood function for the model are derived from  $\tilde{\epsilon}_i^*$ . Accordingly, the likelihood function is free of incidental parameters and the parameters to be estimated are  $\beta, \lambda$  and  $\sigma^2$  –as in the basic SF model.  $\bar{\epsilon}_i$  and  $\tilde{\epsilon}_i^*$  are not independent, unless  $\lambda = 0$ . If  $\lambda = 0$  the model is the fixed effects model with normal error.

In order to obtain the inefficiency index, the composed error has to be recovered:

$$\epsilon_{it} = y_{it} - \hat{y}_{it} = y_{it} - \hat{\beta}^\top x_{it} - \hat{\alpha}_i \quad (\text{A.10})$$

---

<sup>50</sup>Chen et al. (2014) explain how the SF model is related to the CSN distribution and present the required properties of CSN distributed random variables. Another plain introduction to the CSN distribution in the SF context is provided by Brorsen and Kim (2013).

There are two ways to calculate  $\hat{\alpha}_i$ . The one used here is labelled as the mean-adjusted estimate by Chen et al. (2014):

$$\hat{\alpha}_i^M = \bar{y}_i - \hat{\beta}^\top \bar{x}_i + \sqrt{\frac{2}{\pi}} \hat{\sigma}_u \quad (\text{A.11})$$

The point estimators for inefficiency and technical efficiency are the same as for the POOLED model. The output provides `lambda`, given by  $\lambda = \sigma_u/\sigma_v$ .

### A.3 Data organisation and model formulation

#### Data organisation

Different data file formats can be read directly into a **SFAMB** object (.xls, .dta,...), for details see the **Ox** manual (Doornik, 2009).

The data have to be organized in columns where the first row holds the variable name. Each row refers to the same time period. Missing values are also called NaN (Not a Number) in **Ox**. In case of panel data **SFAMB** needs to recognize the structure of the data. In the data file specify the variable names of the individuals and time periods. The data have to be stacked by individual ( $i = 1, 2 \dots N$ ) and within individuals by time period ( $t = 1, 2 \dots T_i$ ). The panel may be unbalanced.

Example:

id	time	y	x1
31	1	298384	24145
31	2	333522	27725
31	3	378768	38115
37	1	62473	3401
37	2	212442	12529
37	3	295142	16734
101	1	150037	10752
101	2	158909	10418
101	3	172744	10671

#### Model formulation

The sequence of model formulation is sketched in Figure A.1.

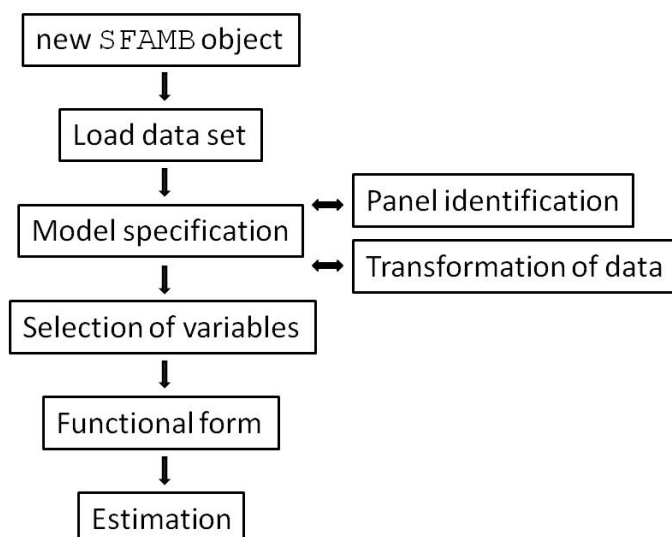


Figure A.1: Model formulation.

In each input file a new object is created. This object is an instance of the **SFAMB** class and can use the functionality of the class. The function `Load` loads the data file and creates the data base. You choose the model type with `SetMethod`. Five estimators /arguments are available (see Section A.2). In case of panel data, specify the panel structure using `Ident`. If the original data are in levels you can use `PrepData` (or other functions of the `Modelbase` or `Database` classes) for transformation. There are several types of variables that can be selected according to the underlying model<sup>51</sup>. The respective function is called `Select` and works with the variable names.

To formulate the frontier function:

- Use `Select(Y_VAR, {".",...})` to select the dependent variable.
- Use `Select(X_VAR, {".",...})` to select the independent variable(s).

To include variables that affect the distribution of the inefficiency component:

- Use `Select(U_VAR, {".",...})` to select variables that shift the mean of the distribution.
- Use `Select(Z_VAR, {".",...})` to select the variables that are hypothesised as responsible for heteroscedasticity, i.e., those that affect

<sup>51</sup>In case of the panel models, a common constant is not identified. However, you can leave `"Constant"` in the selection because it is ignored automatically.

the scale parameter of the distribution.

`SetTranslog` can be used to choose the functional form of the frontier function. In case of the translog specification, we recommend to normalize the variables by the respective sample means. Estimation of the model is executed via `Estimate`. For more details, see the documentation of member functions in the next section.

## A.4 Class member functions

These functions (user interface) together with the data members and several other functions build up the **SFAMB** class. These other functions are not listed here. The interested user may consult the package's header file and source code file. Note that the class derives from the `Ox Modelbase` class, and hence, all underlying functions may be used<sup>52</sup>.

**Ident** `Ident(const vID, const vPer);`

*No return values*

*Description*

-Only panel data- Identifies the structure of the panel.

`vID` is a NT x 1 vector holding the identifier (integer) of the individual.

`vPer` is a NT x 1 vector holding the identifier (integer) of the time period.

**PrepData** `PrepData(const mSel, iNorm);`

*Return value*

Returns logarithms of the specified variables, either normalized or not.

*Description*

This function expects your data in levels and can do two things: It takes logarithms of your specified variables (if `iNorm = 0`) or it normalizes your data (by the sample mean if `iNorm = 1`) before taking logarithms. The transformed variable should receive a new name.

---

<sup>52</sup>In turn, the `Modelbase` class derives from the `Database` class. Accordingly, the corresponding member functions are available in **SFAMB**.



`mSel` is a  $NT \times k$  matrix holding the respective Y- and X-variables.

`iNorm` is an integer: 0=no normalization; 1=normalization;

**SetTranslog** `SetTranslog(const iT1);`

*No return values*

*Description*

This function expects an integer to control the construction of additional regressors from the selected X-variables.

- A value of zero indicates no further terms to be added, e.g., for a log-linear model, this corresponds to the Cobb-Douglas form.
- A value of one indicates that all square and cross terms of all independent variables should be constructed, e.g., for a log-linear model, this corresponds to the full translog form.
- An integer value of  $k > 1$  indicates that the square and cross terms should be constructed for only the first  $k$  independent variables (useful when the regressor matrix contains dummy variables).

**SetConfidenceLevel** `SetConfidenceLevel(const alpha);`

*No return values*

*Description*

-Only POOLED model- This function expects a double indicating the error probability for the construction of confidence bounds (default 0.05).

**TestGraphicAnalysis** `TestGraphicAnalysis();`

*No return values*

*Description*

Only useful in conjunction with the free `0x` package **GnuDraw** (Bos, 2014) that is an `0x` interface to `gnuplot` (gnuplot, 2014). This function draws two or three graphs, respectively: A histogram of the efficiency point estimates and a boxplot of these estimates. In case of the POOLED model: in addition, a sorted graph depicting the interval estimates for technical efficiency at the specified significance.

**TE** TE();

*Return value*

Returns point estimates of technical efficiency, NT x 1 vector.

*Description*

These predictions are given by the conditional expectation of  $\exp(-u)$  (MLE), see Section A.2 for details.

**TEint** TEint(const dAlpha);

*Return value*

Returns point estimates of technical efficiency as well as lower and upper bounds.

*Description*

-Only POOLED model- This function expects a double indicating the error probability for the construction of confidence bounds (default 0.05), for details see Horrace and Schmidt (1996), for an application Brümmer (2001). It returns a NT x 3 matrix structured as (point estimate-lower bound-upper bound).

**Ineff** Ineff();

*Return value*

Returns point estimates of technical inefficiency, NT x 1 vector.

*Description*

These predictions are given by the conditional expectation of  $u$  (MLE), see Section A.2 for details.

**AiHat** AiHat();

*Return value*

Returns the calculated individual effects  $\hat{\alpha}_i$ , N x 1 vector.

*Description*

-Only panel data- These values can be obtained after estimation, see Section A.2 for the respective formulas.

**Elast** Elast(const sXname);

*Return value*

Returns the calculated output elasticity as well as the respective t-value.

*Description*

-Use with `SetTranslog()`- Only if a translog functional form is used.  
The observation-specific output elasticity of input  $k$  is  $\delta \ln y_i / \delta \ln x_{ki}$ .

`sXname` is the name of the corresponding input variable (string).

**GetResults** `GetResults(const ampar, const ameff, const avfct, const amv);`

*No return values*

*Description*

-So far, only POOLED model- This function can be used to store the results of the estimation procedure for further use. All four arguments should be addresses of variables.

`mpar` consists of a `Npar X 3` matrix, where `Npar` is the number of parameters in the model. The first column contains the coefficient estimates, the second column the standard errors, and the last the appropriate probabilities.

`eff` consists of a `Nobs X 3` matrix, where `Nobs` is the number of total observations. The first column holds the point estimate for technical efficiency, the second and third columns contain the upper and lower bound of the (1-alpha) confidence interval.

`fct` Holds some likelihood function values (OLS and ML), as well as some information on the correct variance decomposition of the composed error term.

`v` Variance-Covariance-Matrix.

**Different functions** to extract data:

*Return value*

Different vectors or matrices.

*Description*

These functions can be used with convenient (Database) functions such as `Save`, `Renew` or `savemat`.

`IDandPer()`; is a  $NT \times 2$  matrix holding the number of the individual (e.g., 1,1,1,2,...N,N) as well as the individual group size  $T_i$ . -Only panel data-

`GetLLFi()`; returns the individual log-likelihood values. It is a  $NT \times 1$  vector for models `POOLED` and `LSDV` but a  $N \times 1$  vector for the other models.

`GetResiduals()`; returns the (composed) residual of the respective observation,  $NT \times 1$  vector.

`GetTldata()`; returns the corresponding vectors of  $Y$ ,  $X$ , square and cross terms of  $X$ . -Use with `SetTranslog()`-

`GetMeans()`; returns the means of  $Y$ - and  $X$ -variables,  $N \times (k+1)$  matrix. -Only panel data-

`GetWithin()`; returns the within-transformed  $Y$ - and  $X$ -variables,  $NT \times (k+1)$  matrix. -Only panel data-

**SetStart** `SetStart(const vStart);`

*No return values*

*Description*

This function expects a column vector of appropriate size containing starting values for the maximum likelihood iteration<sup>53</sup>. If the function is not called at all, OLS values are used in conjunction with a grid search for the SFA specific parameters  $\sigma_u$  and  $\lambda$ .

**SetPrintDetails** `SetPrintDetails(const bool);`

*No return values*

*Description*

-Not for `LSDV` model- Prints starting values, warnings and elapsed time if `bool`  $\neq 0$ .

**DropGroupIf** `DropGroupIf(const mifr);`

*No return values*

*Description*

---

<sup>53</sup>Corresponding to the technology parameters. In case of the `TFE` model, a vector of zeros is used for the alphas.

-Only panel data- Allows to exclude a whole individual from the sample if the condition in one (single) period is met. Call after function `Ident`.

`mifr` is the condition that specifies the observation to be dropped, see the general documentation of `selectifr`.

## A.5 Examples

### Example: `hbest1.ox`

The first example is a generalized exponential mean model (cf. Lai and Huang (2010)) where  $u_i \sim N^+(\mu_i, \sigma_{u,i} = \exp(\delta^\top z_i))$ . The original data are in levels and are transformed using member function `PrepData` to accommodate the translog functional form. The data are a subset of FAO/USDA data prepared by Fuglie (2012) including the regions Sub-Saharan Africa and South Africa.

General usage and details of the `0x` language are explained in Doornik and Ooms (2006). The sample file `hbest1.ox` looks like follows. At the beginning of every program some header files are linked in:

```
#include <oxstd.h>
#include <packages/gnudraw/gnudraw.h>
#import <packages/sfamb/sfamb>
```

The first so-called standard header file ensures that all standard library functions can be used. The second line includes the header file of **GnuDraw** (Bos, 2014), an `0x` interface to `gnuplot` (gnuplot, 2014). If it is not installed or you do not want to use this package, delete this line. However, graphics output will then be disabled in the free **Ox Console** version (in the commercial **OxMetrics** version, graphics would still be available). Alternatively, you can comment it out via `//`:

```
//#include <packages/gnudraw/gnudraw.h>
```

The third line imports the (compiled) source code of the package (you may also use `#include <packages/sfamb/sfamb.ox>`). Every `0x` program is executed by the `main()` function that contains the main loop of `0x`.

```
main(){...
}
```

The next steps outlined follow the structure of Figure A.1. A new object of class `Sfa` has to be declared.

```
decl fob = new Sfa();
```

The data are loaded with a call to the member function `Load`. The argument of `SetMethod` chooses the respective estimator (see Table A.1). Here, the model for cross-sectional data is specified. The function `SetConstant` creates a constant (intercept).

```
fob.Load("USDAafrica.xls");
fob.SetMethod(POOLED);
fob.SetConstant();
```

Data are either used directly or prepared within the code. Here, the output variable, five input variables and a time variable are transformed where logarithms of the mean-normalized inputs (output) are taken.<sup>54</sup> New names are assigned to the prepared variables. These names are used for further instructions. The function `Info` is useful here because it prints summary statistics, thereby, allowing to check the transformed data. The program always stops at an `exit` function (that is why it is commented out here).

```
decl inorm = 1;
fob.Renew(fob.PrepData(fob.GetVar("output"), inorm), "lny");
fob.Renew(fob.PrepData(fob.GetVar("labour"), inorm), "lnlab");
fob.Renew(fob.PrepData(fob.GetVar("land"), inorm), "lnland");
fob.Renew(fob.PrepData(fob.GetVar("machinery"), inorm), "lnmac");
fob.Renew(fob.PrepData(fob.GetVar("fertilizer"), inorm), "lnfert");
fob.Renew(fob.GetVar("time") - meanc(fob.GetVar("time")), "trend");
//fob.Info(); exit(1);
```

Selection of variables is carried out by `Select` where `Y_VAR` is the selection of the dependent variable, `X_VAR` is the selection of the regressors. The function uses the new variable names defined above (if your data file already includes transformed variables you would use the names from within the file). The intercept ("`Constant`") is available because `SetConstant` is called above.

---

<sup>54</sup>Normalization of inputs (and output):  $\ln\left(\frac{x_{jit}}{\bar{x}_j}\right)$ ; normalization of time trend:  $t - \bar{t}$ . `PrepData` is a member function of this package (see Section A.4). Both of the other functions are member functions of the `Database` class (see Doornik and Ooms (2006)).

Within the `Select` function there are arrays with three elements (variable name, start lag, end lag). Here, the lags are set to zero. Note that there must not be a comma before the closing curly brace of `Select`.

```
fob.Select(Y_VAR, {"lny", 0, 0});
```

```
fob.Select(X_VAR, {  
  "Constant", 0, 0,  
  "lnlab", 0, 0,  
  "lnland", 0, 0,  
  "lnmac", 0, 0,  
  "lnfert", 0, 0,  
  "trend", 0, 0  
});
```

The above selections define the production frontier. Additional covariates associated with the underlying inefficiency distribution can be introduced (POOLED and WT model). Covariates used to model the location parameter of the distribution are selected into `U_VAR`. Here, only "Constant" is selected meaning that  $\mu_i \neq 0$  but more variables can be added.

```
fob.Select(U_VAR, {  
  "Constant", 0, 0  
});
```

Likewise, covariates intended to model the scale of the distribution are selected into `Z_VAR`, i.e., these variables parameterize  $\sigma_{u,i}$  (in case of the WT model, it is  $\sigma_{u,i}^2$ ).

```
fob.Select(Z_VAR, {  
  "Constant", 0, 0,  
  "lnlab", 0, 0,  
  "lnland", 0, 0,  
  "lnmac", 0, 0,  
  "lnfert", 0, 0  
});
```

The next three lines allow for different adjustments. `SetSelSample` is required and can be used to choose a subset of the data (here: full sample). `SetPrintSfa` ensures that estimation output is printed. `MaxControl` is an optional function that allows for documentation and adjustments of the maximization procedure.

```
fob.SetSelSample(-1, 1, -1, 1);
fob.SetPrintSfa(TRUE);
MaxControl(1000, 10, TRUE);
```

The functional form of the production frontier is chosen by `SetTranslog` where the options are Cobb-Douglas or translog. Here, a translog form is specified. Estimation of the model is invoked via `Estimate`.

```
fob.SetTranslog(1);
fob.Estimate();
```

A number of results can be obtained after estimation. In the SF context, the efficiency scores ( $TE_i$ ) are of particular interest. Here, the point estimates are extracted, together with the lower and upper bounds of a 95% confidence band. The respective function is `TEint`. The function `Ineff` extracts the point estimates of inefficiency,  $E(u_i|\epsilon_i)$ . These results are labelled and appended to the object using `Renew`. The original database together with the transformed variables and results is saved to file via `Save`.

```
fob.Renew(fob.TEint(0.05), {"TE", "lower", "upper"});
fob.Renew(fob.Ineff(), {"jlms"});
fob.Save("out.xls");
```

There is a graphical functionality involving the package `GnuDraw` that allows for a visual assessment of the efficiency scores. The function `TestGraphicAnalysis` displays the graphics presented in Figure A.2. The setting of the confidence band included can be changed with `SetConfidenceLevel` where an error probability of 0.05 is the default.

```
fob.SetConfidenceLevel(0.05);
fob.TestGraphicAnalysis();
```

The output of this program looks like follows (omitting information on the maximization procedure). Some general information:

```
Sfa package version 1.0, object created on 19-02-2014
Constructing Squares and Cross-Products...done.
-Pooled model-
```



---- Sfa ----

The estimation sample is: 1 - 2400

The dependent variable is: lny

The dataset is: USDAafrica.xls

The transformed variables facilitate the interpretation of the estimated coefficients of the translog functional form. Thus, the first order coefficients listed below can be interpreted as output elasticities at the sample mean. These estimates are positive (except for the machinery input whose estimate, however, is not significant), and hence, meet the requirement of monotonicity. The parameter associated with `trend` indicates the estimated average rate of technical change per year.

	Coefficient	Std.Error	robust-SE	t-value	t-prob
Constant	0.418511	0.01734	0.01604	26.1	0.000
lnlab	0.128542	0.01338	0.01105	11.6	0.000
lnland	0.747665	0.01552	0.01301	57.5	0.000
lnmac	-0.0103591	0.009488	0.008851	-1.17	0.242
lnfert	0.0753081	0.006573	0.006243	12.1	0.000
trend	0.0104214	0.0007006	0.0006763	15.4	0.000

Further, the output shows the coefficients of the squared and cross terms that can be used to calculate the individual output elasticities.

	Coefficient	Std.Error	robust-SE	t-value	t-prob
.5*lnlab <sup>2</sup>	-0.0555308	0.02432	0.02387	-2.33	0.020
.5*lnland <sup>2</sup>	-0.170596	0.02547	0.02843	-6.00	0.000
.5*lnmac <sup>2</sup>	-0.0152330	0.005151	0.004632	-3.29	0.001
.5*lnfert <sup>2</sup>	0.0611979	0.003107	0.003063	20.0	0.000
.5*trend <sup>2</sup>	0.000420185	6.481e-005	6.132e-005	6.85	0.000
lnlab*lnland	0.189014	0.02492	0.02557	7.39	0.000
lnlab*lnmac	-0.125613	0.008138	0.007344	-17.1	0.000
lnlab*lnfert	-0.0294984	0.006109	0.005248	-5.62	0.000
lnlab*trend	-0.000443247	0.0007217	0.0006231	-0.711	0.477
lnland*lnmac	0.137893	0.008829	0.008381	16.5	0.000
lnland*lnfert	-0.0633866	0.006748	0.006383	-9.93	0.000
lnland*trend	-0.000495269	0.0007838	0.0007483	-0.662	0.508
lnmac*lnfert	-0.0135746	0.002997	0.002857	-4.75	0.000
lnmac*trend	0.000810360	0.0002892	0.0002743	2.95	0.003
lnfert*trend	0.000898462	0.0002366	0.0002062	4.36	0.000

After the technology parameters, the estimates of  $\sigma_v$  and  $\sigma_u$  are listed in form of their natural logarithms. The next line refers to the noise component.

	Coefficient	Std.Error	robust-SE	t-value	t-prob
$\ln\{\sigma_u\}$	-2.64680	0.1459	0.1361	-19.4	0.000

Since  $\ln(\sigma_u)$  is parameterised using covariates there are several estimates to look at. The order of coefficients corresponds to the specification  $\ln(\sigma_u) = \delta_0 + \sum_{l=1}^4 \delta_l \times z_l$  where  $l = 1$ (labour),  $2$ (land),  $3$ (machinery),  $4$ (fertilizer); and the  $z$ 's are in logarithms. Higher use of  $z_l$  is associated with a lower level of inefficiency (or higher technical efficiency) if the estimated parameter has a negative sign.

	Coefficient	Std.Error	robust-SE	t-value	t-prob
Constant	-1.04439	0.04104	0.04791	-21.8	0.000
lnlab	0.232693	0.04300	0.05044	4.61	0.000
lnland	-0.146195	0.04176	0.05050	-2.90	0.004
lnmac	-0.00976602	0.01491	0.01671	-0.584	0.559
lnfert	-0.0149101	0.01372	0.01647	-0.905	0.365

Here, the inefficiency distribution is supposed to have a non-zero mean,  $u_i \sim N^+(\mu_i = \theta_0, \sigma_{u,i}^2)$ , i.e., it is a constant ( $\theta_0$ ) common to all individuals. Additional covariates can be introduced. The omission of U.VAR in the model specification leads to  $\mu_i = 0$ , and hence, results in the normal half-normal model. Note that, if specified, this output (here, the third Constant) is always the last Constant term in the list.

	Coefficient	Std.Error	robust-SE	t-value	t-prob
Constant	0.454143	0.02926	0.03249	14.0	0.000

Some additional information is provided, for details see Section A.2.

log-likelihood	-458.928611				
no. of observations	2400	no. of parameters	28		
AIC.T	973.857222	AIC	0.405773842		
mean(lny)	-1.14273	var(lny)	2.98932		
\gamma:	0.9618	VAR(u)/VAR(total)	0.9016		
Test of one-sided err	172.93	mixed Chi <sup>2</sup> !!			

Finally, the graph created by the function TestGraphicAnalysis is depicted in Figure A.2.

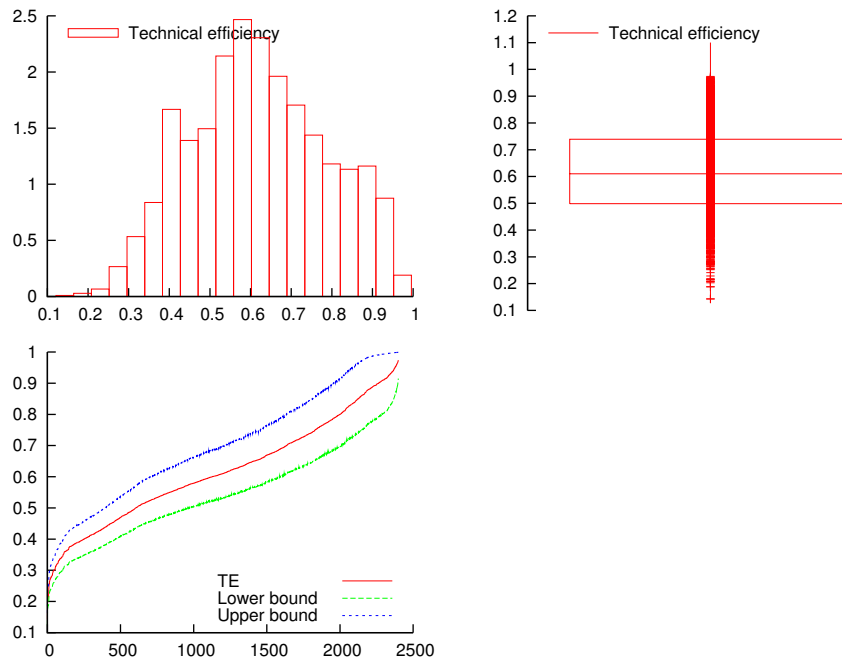


Figure A.2: TE scores of the POOLED model.

### Example: hbest2.ox

In this example, the CFE model of Chen et al. (2014) is specified using again the data set `USDAafrica.xls` and a translog functional form. You can immediately switch to the LSDV or TFE model, respectively, by changing the argument of `SetMethod`. A large part of this example corresponds to the code of the previous subsection. However, as panel data are involved here some things are different.

```
#include <oxstd.h>
#include <packages/gnudraw/gnudraw.h>
#import <packages/sfamb/sfamb>

main(){
  decl fob = new Sfa();
  fob.Load("USDAafrica.xls");
  fob.SetMethod(CFE);
  fob.SetConstant();
}
```

CFE is the estimator selected. Here, the function `SetConstant` does not create a constant because it is not required. However, this line can be kept for convenience. The function `Ident` identifies the panel structure of the

data. The required information includes the variable names of the individuals ("ID") and the period ("time").

```
fob.Ident(fob.GetVar("ID"), fob.GetVar("time"));
```

Data transformation and model specification correspond to the previous example. Note that U\_VAR or Z\_VAR are not available here.

```
decl inorm = 1;
fob.Renew(fob.PrepData(fob.GetVar("output"), inorm), "lny");
fob.Renew(fob.PrepData(fob.GetVar("labour"), inorm), "lnlab");
fob.Renew(fob.PrepData(fob.GetVar("land"), inorm), "lnland");
fob.Renew(fob.PrepData(fob.GetVar("machinery"), inorm), "lnmac");
fob.Renew(fob.PrepData(fob.GetVar("fertilizer"), inorm), "lnfert");
fob.Renew(fob.GetVar("time") - meanc(fob.GetVar("time")), "trend");

fob.Select(Y_VAR, {"lny", 0, 0});

fob.Select(X_VAR, {
"Constant", 0, 0,
"lnlab", 0, 0,
"lnland", 0, 0,
"lnmac", 0, 0,
"lnfert", 0, 0,
"trend", 0, 0});

fob.SetSelSample(-1, 1, -1, 1);
fob.SetPrintSfa(TRUE);
MaxControl(1000,10,TRUE);
fob.SetTranslog(1);
fob.Estimate();
```

For this model, there is no calculation of confidence bounds involved. The efficiency scores can be extracted as point estimates using function TE.

```
fob.Renew(fob.TE(), {"TE"});
fob.Renew(fob.Ineff(), {"jlms"});
fob.TestGraphicAnalysis();
```

The output of this program looks like follows. Additional information on the panel structure is printed.

```
Sfa package version 1.0, object created on 10-02-2014
#groups:   #periods(max):  avg.T-i:
  48.000      50.000      50.000
Constructing Squares and Cross-Products...done.
-CFE model-

---- Sfa ----
The estimation sample is:  1 - 2400
The dependent variable is: lny
The dataset is: USDAafrica.xls
```

A common intercept is not identified, and hence, there is no Constant.

	Coefficient	Std.Error	t-value	t-prob
lnlab	0.00883654	0.03048	0.290	0.772
lnland	0.677192	0.02304	29.4	0.000
lnmac	0.106177	0.009083	11.7	0.000
lnfert	0.0837343	0.007086	11.8	0.000
trend	0.00920993	0.0006800	13.5	0.000
.5*lnlab^2	0.138565	0.02083	6.65	0.000
.5*lnland^2	0.177254	0.02047	8.66	0.000
.5*lnmac^2	0.0121082	0.003350	3.61	0.000
.5*lnfert^2	0.0245012	0.002852	8.59	0.000
.5*trend^2	0.000407978	3.744e-005	10.9	0.000
lnlab*lnland	-0.138300	0.02024	-6.83	0.000
lnlab*lnmac	-0.0247345	0.007611	-3.25	0.001
lnlab*lnfert	0.00218990	0.005678	0.386	0.700
lnlab*trend	-0.000134440	0.0005109	-0.263	0.792
lnland*lnmac	0.0243333	0.008190	2.97	0.003
lnland*lnfert	-0.0319551	0.006178	-5.17	0.000
lnland*trend	0.000212194	0.0004843	0.438	0.661
lnmac*lnfert	0.00379000	0.001959	1.93	0.053
lnmac*trend	0.000346355	0.0001844	1.88	0.060
lnfert*trend	-0.000171510	0.0001308	-1.31	0.190

This model is restricted to the normal half-normal case. Here, the estimates of (the natural logarithms of)  $\sigma_v^2$  and  $\sigma_u^2$  are given.

ln{\sigma_v^2}	-4.94563	0.1464	-33.8	0.000
ln{\sigma_u^2}	-3.44008	0.1125	-30.6	0.000

log-likelihood	1476.81739		
no. of observations	2400	no. of parameters	22
AIC.T	-2909.63478	AIC	-1.21234782
mean(lny)	7.55183e-018	var(lny)	0.127523
lambda	2.123		

The function `TestGraphicAnalysis` is used to create the graph depicted in Figure A.3.

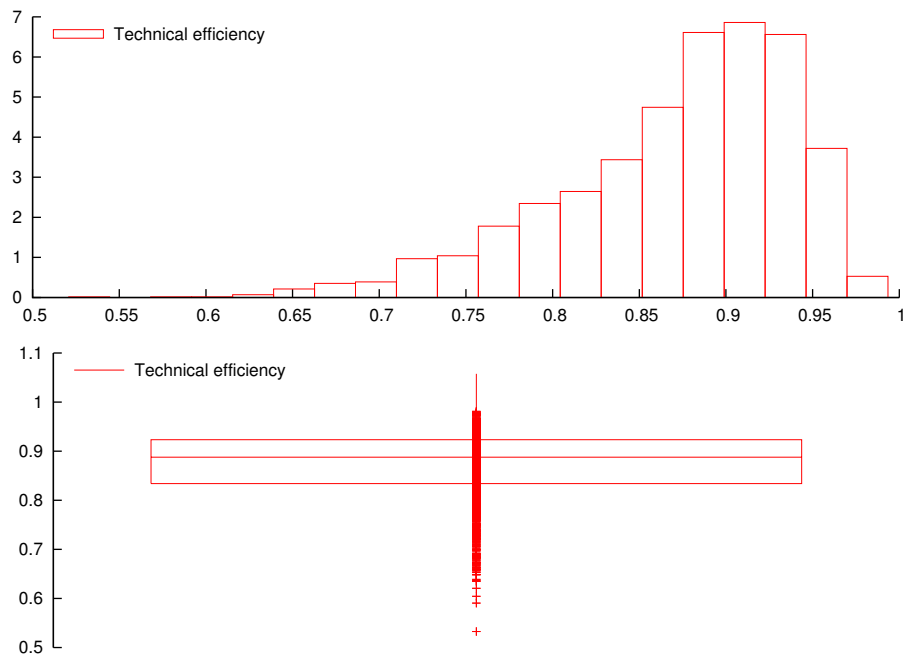


Figure A.3: TE scores of the CFE model.

### Example: member functions `SetTranslog` and `Elast`

The member function `SetTranslog` allows for convenient specification of a translog functional form. In the following, we refer to the current instance of the class as `fob`. Suppose your selection of regressors looks like this:

```
fob.Select(X_VAR, {
"Constant", 0, 0,
"lnx1", 0, 0,
"lnx2", 0, 0,
"lnx3", 0, 0,
"trend", 0, 0});
```

The default specification is Cobb-Douglas, i.e., `SetTranslog(0)`, changing the argument to 1 invokes construction of the respective square and cross

terms of `X_VAR`. In general notation:

$$\ln y_i = \beta_0 + \sum_{j=1}^K \beta_j \ln x_{ji} + \frac{1}{2} \sum_{j=1}^K \sum_{l=1}^K \beta_{jl} \ln x_{ji} \ln x_{li} \quad (\text{A.12})$$

If your selection includes dummies the variables should be ordered like this:

```
fob.Select(X_VAR, {
"Constant", 0, 0,
"lnx1", 0, 0,
"lnx2", 0, 0,
"lnx3", 0, 0,
"trend", 0, 0,
"dummy1", 0, 0,
"dummy2", 0, 0});
```

Specification of a translog form is then possible by means of `SetTranslog(4)` because only the first four regressors are used ("`Constant`" is ignored automatically).

After estimation the member function `Elast` can be used to calculate the output elasticity ( $\epsilon_{ji}$ ) of each input for each observation:

$$\epsilon_{ji} = \beta_j + \sum_{l=1}^K \beta_{jl} \ln x_{li} \quad (\text{A.13})$$

The following example illustrates one possible way the function may be used. Here, results are plotted as histograms (see Figure A.4). Note that indexing starts at 0 in `0x` (`Elast` returns a `NTx2` matrix but only the first column is considered here).

```
decl vEps1 = fob.Elast("lnx1");
decl vEps2 = fob.Elast("lnx2");
decl vEps3 = fob.Elast("lnx3");
decl vEpst = fob.Elast("trend");

DrawDensity(0, vEps1[] [0]', {"eps1"}, 1, 1, 0, 0, 0, 0, 1, 0, 1);
DrawDensity(1, vEps2[] [0]', {"eps2"}, 1, 1, 0, 0, 0, 0, 1, 0, 1);
DrawDensity(2, vEps3[] [0]', {"eps3"}, 1, 1, 0, 0, 0, 0, 1, 0, 1);
DrawDensity(3, vEpst[] [0]', {"epst"}, 1, 1, 0, 0, 0, 0, 1, 0, 1);
ShowDrawWindow();
```

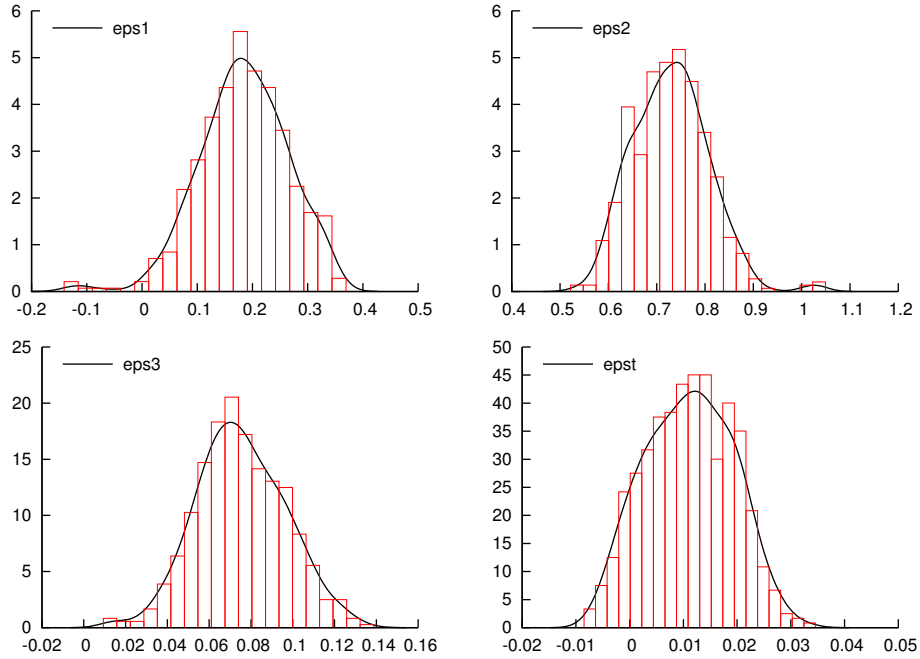


Figure A.4: Histograms of calculated elasticities (by observation).

## A.6 Future developments

The basic version of **SFAMB** dates back to the mid 1990’s where the capability was restricted to cross-sectional data. As it now allows for panel data and the literature on SF methods is broad as well as still growing there is scope for potential extensions. Some related possibilities are mentioned here.

In the model framework of Chen et al. (2014) there are two ways to calculate the individual effects. As an alternative to equation (A.11) the individual “*between estimator of  $\alpha_i$* ” can be used. It could be implemented as an optional function, involving a second maximization. Its availability would allow to compare results and check the consequences for TE scores.

While the current focus of panel methods is on fixed effects estimation, a more comprehensive supplement might involve random effects models. The most recent SF approach using the CSN distribution is presented by Colombi et al. (2014). Its specification is similar to equation (A.4) but the time-invariant part is further decomposed into two residuals (persistent inefficiency and time-invariant unobserved heterogeneity). Greene (2014) introduces computational simplifications and labels the model as the “*Generalized True Random Effects SF model*”.



Since version 7, `0x` supports parallel programming (usage of multiple processors). The implementation of *parallel loops* in the package could lead to a substantial increase in computational speed.

## Acknowledgements

All code is pure `0x` code. However, code of the WT model is partially adapted from `Stata` code by Hung-Jen Wang. We thank Hung-Jen Wang for providing data to check our `CFE` code.

## References

- Aigner, D., Lovell, C. A. K. and Schmidt, P. (1977). Formulation and Estimation of Stochastic Frontier Production Function Models. *Journal of Applied Econometrics* 6: 21–37.
- Alvarez, A., Amsler, C., Orea, L. and Schmidt, P. (2006). Interpreting and Testing the Scaling Property in Models where Inefficiency Depends on Firm Characteristics. *Journal of Productivity Analysis* 25: 201–212, doi:10.1007/s11123-006-7639-3.
- Battese, G. E. and Coelli, T. J. (1988). Prediction of Firm-Level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data. *Journal of Econometrics* 38: 387–399.
- Battese, G. E. and Coelli, T. J. (1992). Frontier production functions, technical efficiency and panel data: With application to paddy farmers in india. *Journal of Productivity Analysis* 3: 153–169.
- Battese, G. E. and Coelli, T. J. (1995). A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data. *Empirical Economics* 20: 325–332.
- Belotti, F., Daidone, S., Ilardi, G. and Atella, V. (2012). Stochastic Frontier Analysis Using `Stata`. CEIS Research Paper 251, Tor Vergata University, CEIS.
- Bos, C. S. (2014). **GnuDraw** – an `0x` Package for Creating `gnuplot` Graphics.
- Brorsen, B. W. and Kim, T. (2013). Data aggregation in stochastic frontier models: The closed skew normal distribution. *Journal of Productivity Analysis* 39: 27–34, doi:10.1007/s11123-012-0274-2.

- Brümmer, B. (2001). Estimating Confidence Intervals for Technical Efficiency: The Case of Private Farms in Slovenia. *European Review of Agricultural Economics* 28: 285–306.
- Caudill, S. B., Ford, J. M. and Gropper, D. M. (1995). Frontier Estimation and Firm-Specific Inefficiency Measures in the Presence of Heteroscedasticity. *Journal of Business & Economic Statistics* 13: 105–111.
- Chen, Y.-Y., Schmidt, P. and Wang, H.-J. (2014). Consistent estimation of the fixed effects stochastic frontier model. *Journal of Econometrics* 181: 65–76, doi:10.1016/j.jeconom.2013.05.009.
- Coelli, T. J. (1996). A Guide to **FRONTIER 4.1**: A Computer Program for Stochastic Frontier Production and Cost Function Estimation. CEPA Working Papers, University of New England.
- Coelli, T. J. and Henningsen, A. (2013). **frontier**: Stochastic Frontier Analysis. R package version 1.1-0.
- Coelli, T. J., Rao, P. D., O’Donnell, C. J. and Battese, G. E. (2005). *An Introduction to Efficiency and Productivity Analysis*. Springer-Verlag, New York.
- Colombi, R., Kumbhakar, S. C., Martini, G. and Vittadini, G. (2014). Closed-Skew Normality in Stochastic Frontiers with Individual Effects and Long/Short-Run Efficiency. *Journal of Productivity Analysis* doi:10.1007/s11123-014-0386-y.
- Cottrell, A. and Lucchetti, R. (2014). **gret1** User’s Guide – Gnu Regression, Econometrics and Time-Series Library.
- Doornik, J. A. (2009). *An Object-Oriented Matrix Language Ox 6*. Timberlake Consultants Press, London.
- Doornik, J. A., Arellano, M. and Bond, S. (2012). Panel Data Estimation Using DPD for Ox. Boston College.
- Doornik, J. A. and Ooms, M. (2006). Introduction to Ox.
- Econometric Software, Inc. (2014). LIMDEP, Version 10.0. ESI, New York.
- Fuglie, K. O. (2012). Productivity Growth and Technology Capital in the Global Agricultural Economy. In Fuglie, K. O., Wang, S. L. and Ball, V. E. (eds), *Productivity Growth in Agriculture: An International Perspective*. CABI.
- gnuplot (2014). **gnuplot 4.6** – An Interactive Plotting Program.

- Greene, W. H. (2005). Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model. *Journal of Econometrics* 126: 269–303, doi:10.1016/j.jeconom.2004.05.003.
- Greene, W. H. (2008). The Econometric Approach to Efficiency Analysis. In Fried, H. O., Lovell, C. A. K. and Schmidt, S. S. (eds), *The Measurement of Productive Efficiency and Productivity Growth*. Oxford University Press.
- Greene, W. H. (2014). True random effects in stochastic frontier models. North American Productivity Workshop VIII, Ottawa, Canada.
- Horrace, W. C. and Schmidt, P. (1996). Confidence Statements for Efficiency Estimates from Stochastic Frontier Models. *Journal of Productivity Analysis* 7: 257–282.
- Huang, C. J. and Liu, J.-T. (1994). Estimation of a Non-Neutral Stochastic Frontier Production Function. *Journal of Productivity Analysis* 5: 171–180, doi:10.1007/BF01073853.
- Hughes, G. (2008). **sfa\_hetmod** and **sfa\_mod**. User-contributed function packages for **gretl**.
- Jondrow, J., Lovell, C. A. K., Materov, I. S. and Schmidt, P. (1982). On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model. *Journal of Econometrics* 19: 233–238.
- Kodde, D. A. and Palm, F. C. (1986). Wald Criteria for Jointly Testing Equality and Inequality Restrictions. *Econometrica* 54: 1243–1248.
- Kumbhakar, S. C., Gosh, S. and McGuckin, J. T. (1991). A Generalized Production Frontier Approach for Estimating Determinants of Inefficiency in U.S. Dairy Farms. *Journal of Business and Economic Statistics* 9: 279–286.
- Kumbhakar, S. C. and Lovell, C. A. K. (2000). *Stochastic Frontier Analysis*. Cambridge University Press, Cambridge.
- Lai, H.-p. and Huang, C. J. (2010). Likelihood Ratio Tests for Model Selection of Stochastic Frontier Models. *Journal of Productivity Analysis* 34: 3–13, doi:10.1007/s11123-009-0160-8.
- Meeusen, W. and Broeck, J. van den (1977). Efficiency Estimation from Cobb-Douglas production functions with composed error. *International Economic Review* 18: 435–444.
- Pavlyuk, D. (2014). **spfrontier**: Spatial Stochastic Frontier Models Estimation. R package version 0.1.10.

- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Reifschneider, D. and Stevenson, R. (1991). Systematic Departures from the Frontier: A Framework for the Analysis of Firm Inefficiency. *International Economic Review* 32: 715–723.
- Schmidt, P. and Sickles, R. C. (1984). Production Frontiers and Panel Data. *Journal of Business & Economic Statistics* 2: 367–374.
- StataCorp LP (2014). *Stata*, Version 13. College Station, Texas, USA.
- Wang, H.-J. (2012). Manual of Hung-Jen Wang’s *Stata* Codes.
- Wang, H.-J. and Ho, C.-W. (2010). Estimating Fixed-Effect Panel Stochastic Frontier Models by Model Transformation. *Journal of Econometrics* 157: 286–296, doi:10.1016/j.jeconom.2009.12.006.