2018

Structural convergence between the dairy sectors of the EU-27 Member States

SINCE THE EASTERN ENLARGEMENT

WINAND DE JONG

Author: Winand de Jong
Registration number: 950906406060
Program: Management, Economics & Consumer Studies
Specialization: Economics, Environment and Governance
Course: MSc Thesis Agricultural Economics and Rural Policy
Course code: AEP-80433
Publication date: 19 February 2018
Place of publication: Wageningen
Supervisor: dr. R. Ihle & dr.ir. R. Jongeneel
Second reader: dr. ir. C. Gardebroek
Educational institute: Wageningen University
Chair group: Agricultural Economics & Rural Policy
Hollandseweg 1 6706 KN Wageningen

WAGENINGEN
UNIVERSITY & RESEARCH

ABSTRACT

In this thesis, we examine structural convergence across the regional dairy sectors in the EU since the Eastern enlargement. Structural disparity existed between the dairy sectors of old and new Member States when those Member States accessed the European Union. Few researchers have addressed the question whether the dairy sectors of the regions in the EU-27 converged towards or diverged from each other. It is of particular political interest whether sectorial cohesion took place after the Eastern Enlargement, given that all regions were subject to the common policy framework of the CAP and convergence is one of the central goals of European integration. This thesis examines whether convergence has taken place across these regions with respect to farm-gate milk prices, productivity and farm income. In this thesis we examine convergence according to the σ-convergence definition; examining the dispersion of the cross-sectional distribution. Using an OLS regression of the coefficient of variation of national farm-gate milk prices of 26 MS, we find that price dispersion decreased till some extent from 2004-2007 between the OMS and NMS that accessed in 2004. For the productivity and income convergence analysis, a balanced panel dataset of 91 aggregated FADN regions (Specialist dairy farms, TF45) has been used. With respect to productivity, Kernel density plots show that there is persistent bimodality in the distribution of biological and mechanical productivity. Markov chains show that the probability to transition from the lowest productivity class (<50% EU-average) to a higher class is very low. With respect to income convergence, it is shown that convergence was the greatest for family farm income. For both productivity and income, we find that convergence was less and slower in the period 2007-2015 compared to the period 2004-2015 (without Bulgaria and Romania). Although some NMS regions have caught-up and we have seen convergence for some variables, the findings of this study support the idea that only very limited convergence has taken place between the dairy sectors of the EU regions. This demonstrates that structural economic convergence between the dairy sectors of the EU regions is not an automatic process. Convergence, as a central goal of both the CAP framework specifically and European integration in general, has taken place, to a limited extent, between the dairy sectors of the EU-27 Member States.

Keywords: convergence, dairy sector, EU, Markov-chain, Kernel estimation, coefficient of variation.
JEL: O47, R12, Q18

Source cover image: (IEG Policy, 2018)

PREFACE

In the past five months, I have worked on this thesis in the AEP-group. I want to thank my supervisors dr. Ihle and dr. Jongeneel. Dr. Ihle for his detailed feedback moments and his reflective words on academic research. Dr. ir. Jongeneel in particular for his outstanding knowledge on the dairy markets and for the ‘sparring’ on the research design of my thesis. I want to thank dr. ir. Gardebroek as the second reader of this thesis. Furthermore, I want to thank my colleagues at AEP for providing a workspace and for their friendly help. Not to forget my fellow students for five pleasant years of study.
Last year when I went on exchange to the University of Helsinki, I followed a most-interesting and challenging course on Discrete Markov Processes. It was a pleasure to come across this fascinating mathematical model again. This particular combination of applied quantitative economics and societal relevant topics has proved to be of particular interest to me. I am grateful that I got the opportunity to be taught in economics for the past five years.

Winand de Jong
19 February 2018, *Wageningen*

“It is in the agricultural sector that the battle for long-term economic development will be won or lost.”(Gunnar Myrdal, Nobel Laureate in economics)

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

LIST OF ABBREVIATIONS

AWU	Annual Working Unit
AIC	Akaike's Information Criterion
CAP	Common Agricultural Policy
CV	Coefficient of variation
DEA	Data Envelopment Analysis
EEC	European Economic Community
EU	European Union
FADN	Farmer Accountancy Data Network
FWU	Family Working Unit
GDP	Gross Domestic Product
LOP	Law of One Price
MMO	Milk Market Observatory
MLE	Maximum Likelihood Estimation
NMS	New Member States that accessed since 2004
NMS(2004)	New Member States that accessed in the year 2004
NMS(2007)	New Member States that accessed in the year 2007
OMS	Old Member States that accessed before 2004
PDO	Protected Designation of Origin
PDI	Protected Geographical Indication
SFA	Stochastic Frontier Analysis
SMP	Skimmed Milk Powder
TF	Type of Farming category in FADN database
TFP	Total Factor Productivity
WMP	Whole Milk Powder

LIST OF SYMBOLS

A	Total Factor Productivity
L	Labour input
K	Capital input
$P_{i,t}$	Price of a country i at time t
CV_t	Coefficient of variation at time t
τ	Change point of the structural break
t	Time period
σ_t	Standard deviation at time t
p_{ij}	Transition probability from state i to state j
$\mathbf{P}$	Transition probability matrix
π	Distribution of the Markov chain
M	Mobility index
h	Kernel density bandwidth
k	Kernel function

1 INTRODUCTION

In this chapter, we introduce the topic of this thesis. First it is explained why this research is relevant today. Followed by the objectives of the research. Then the theoretical framework of this research is described, accompanied by the empirical methods and data. At last, an overview of the content of this thesis is given.

1.1 BACKGROUND AND RELEVANCE OF THE THESIS

In 2004 and 2007, the European Union (EU) welcomed in total twelve New Member States. Those new Member States (NMS), all brought their own agricultural sectors with them to the common market. For instance, Polish dairy farmers therefore suddenly faced the competition of Danish dairy farmers. For particularly this sector, the dairy sector, this enlarged market had a large impact on the farmer. Dairy farmers are bounded to their cows and can hardly substitute between products, and are therefore locked-in their business. Above that, milk products represent the largest share in total agricultural production in the EU (i.e. 13.9%), meaning that it is the largest internal agricultural market (European Commission, 2012).

When those NMS accessed the EU, their dairy production was totally different from the Member States that were already member before 2004, the Old Member States (OMS). An example of this difference is that in 2004, an average Polish dairy farmer in the Mazowsze i Podlasie region had a farm net income of €7.513 while an average German dairy farmer in Sachsen had a farm net income of €87.771 (FADN, 2017a). Moreover, in 2007 an average dairy farmer in Yuzhen Tsentralen (Bulgaria) had a milk yield of 3741.4 kg/cow while a dairy farmer in Denmark had a milk yield of 8209.37 kg/cow. Above that, in the past there has been a process of farm exits and scaling up in the OMS (Van Berkum and Helming, 2006). While today in Bulgaria, Poland and Lithuania still so many small farms exist, that these regions probably still have to face a structural scaling-up process in the future (Van Berkum, 2009). Therefore, dairy farmers from the NMS faced a gap with the OMS; hence, we can speak of disparity between regional dairy sectors.

The period during and after the accession of the NMS can be characterized as turbulent. To begin with, the Common Agricultural Policy (CAP) has undergone multiple reforms and that resulted in a more market-oriented policy. Secondly, the policy of the dairy sector in the European Union changed but also the dairy market itself has changed; this can be illustrated by the severe price fluctuations in the period 2006-2009 (Jongeneel and Van Berkum, 2015). Without price floors and production quota, the farmers are in volatile times, and therefore more vulnerable to price shocks (O'Connor and Keane, 2011). The latest and most fundamental change was the end of the milk quota in 2015. For mostly the north-western EU MS, this quota was a real limitation on the milk production (Versteijlen, 2013). The quota abolishment was preceded by the 'soft landing' policy, meaning that the milk quota was gradually increased from 2009 till the quota abolishment (Jongeneel, 2011). Despite the 'soft landing', the abolishment led to a crisis in the dairy markets and this forced the European Commission to adapt crisis regulations (European Commission, 2016a).

These three developments have led to increased competition. The question then arises to what extent farmers in the different regions were able to cope with all these changes. Increased competition has inevitably led to structural change in the farms, markets and policies on all levels (Zimmermann and Heckelei, 2012; Gocht et al., 2012). More important is the question: how did the dairy farmers in the EU develop compared to each other? For example, is the milk yield per cow on a Bulgarian dairy farm now closer to the one of a Danish dairy farm, or did it even diverge more? Further disparity could lead to a structural problem with respect to the viability of dairy farming in in some regions. If the regions that lag behind do not catch-up because they are not able to adapt to the new market conditions, then disparity can sustain or even increase. As Jansik and Irz (2015a) describe, this could lead to a situation of winners and losers in the European dairy sector, where farmers in some regions cannot compete with the farmers in the most competitive regions. From a European point-of-view, the absence of cohesion could hamper the 'continuation of the integration process' (Kuokannen and Vihinen, 2006, p. 6).

Multiple studies have already examined the effects of changes in the dairy sector at regional or national levels (e.g. Huettel and Jongeneel, 2011; Salou et al., 2017; Tonini and Jongeneel, 2009). Less is known on the effects across EU regions, particularly on whether the dairy sectors of these regions converged towards or diverged from each other. Other scholars have examined whether there was convergence between the agricultural sectors of the EU countries (Jambor et al., 2016; Rezitis, 2010; Alexiadis, 2010) but they do not specifically mention the dairy sector. Only Cechura et al. (2016) provide a detailed study on productivity growth across Member States for the dairy sector.

It is crucial to examine convergence across EU regions, since the policy is made on a European level. Having the CAP 2020 reform in mind, and the upcoming Brexit, an important issue will be the distribution of the CAP money across the regions. In early communication on drafts on the CAP after 2020, it has already been mentioned that the new CAP should: *"…, reduce differences between Member States in CAP support. Even if the wide diversity of relative costs of labour and land as well as the different agronomic potentials across the EU should be acknowledged, all EU farmers face similar challenges."* (European Commission, 2017a, p.16). Bulgaria, who will hold the Presidency of the Council of the European Union in spring 2018, declared via the agricultural minister Porodnov that: a key priority for his country was to move towards greater convergence in support *"to ensure a level playing field on the single market"* (Agrafocus, 2017, p. 23). Hence, the debate on convergence between the agricultural sectors of the EU MS will be an important point for the future CAP.

With that in mind, this thesis gives more insight into the process of convergence by analysing the data for the period since 2004. For that reason, it will be assessed whether there was convergence between the dairy sectors of the EU MS from 2004 onwards with a focus on productivity, prices and farm incomes.

1.2 RESEARCH QUESTIONS AND OBJECTIVES

The main research question of this thesis is:

To what extent has convergence taken place between the dairy sectors of the EU-27 Member States since 2004?

By giving special focus on convergence with respect to productivity, prices and farm income.

This research question will be answered by shedding light on the following three specific research questions:

i) *What disparities existed in the EU dairy sector, and how did they change over time?*

ii) *How can convergence be modelled and measured?*

iii) *What is the empirical evidence on convergence in prices, productivity and farm income of the dairy sector across the MS and regions of the EU?*

1.3 Theoretical framework

The question on convergence can be researched from many different scientific disciplines. As people can have opposing views at the issue of convergence, this thesis uses numerical data to analyse the issue. Moreover, we use an economic approach to assess whether there was convergence or divergence. More specifically, we use of micro- and macro-economic theory to define convergence. Econometrical methods are used to shed light on the empirical evidence of convergence.

Convergence is defined in the dictionary as: "*the act of converging and especially moving toward union or uniformity*" (Merriam-Webster, 2018). With the term 'convergence' in the economic theory scholars refer usually to macroeconomic models like the neo-classical Solow model, in which is shown that poorer regions tend to grow faster than richer regions. Consequently, theory predicts that all regions would in the end be more or less equally rich, since they converge to the same steady state (Bernard and Durlauf, 1996). In the nineties, the neo-classical Solow model was revisited by Barro (1991), who showed that economic growth was also dependent on more factors, such as political stability and human capital. Shortly after that, Fujita, Krugman and Venables (1999) gave a thought-provoking turn to the research on the spatial aspect of economics by the 'New Economic Geography'. Within their theory, they try to explain economic agglomeration in the geographical space with mathematically based models like the core-periphery model (Fujita and Krugman, 2004). However, most relevant framework for this study is the framework of European integration. Within this framework, the term convergence is linked to the term 'cohesion'. Dunford and Smith (2000, p.173) state that: *"Cohesion depends on the degree of equality in the distribution of GDP per head and the extent to which there are processes of catch-up in which less developed countries and region and lower-income groups enjoy faster rates of income growth than more developed areas or richer groups."* Within this framework, convergence can therefore be defined as: 'increased cohesion'. As is mentioned in section 1.1, there was a significant disparity between the dairy sectors of the EU in the beginning. Illustrated in Figure 1.1, the starting situation is disparity, then three processes can take place: convergence, a status-quo or divergence. Cohesion then means that the greater cohesion there is, the less dispersion there is. This thesis examines the disparity, it tests whether convergence has taken place and by that if greater cohesion has occurred.

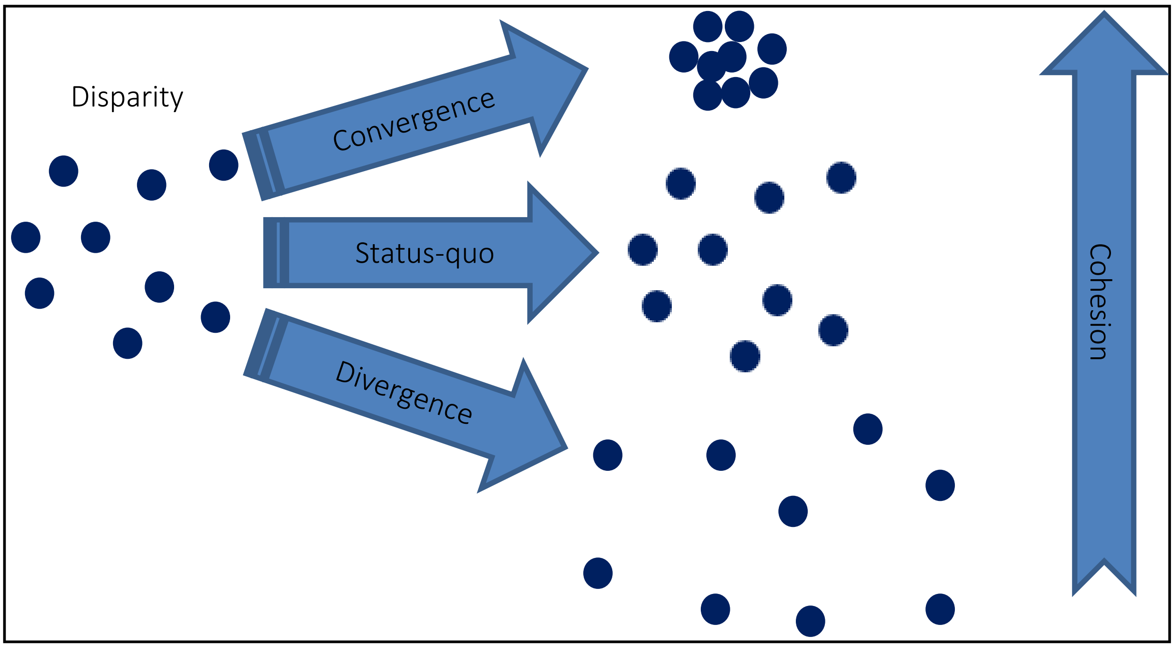


Figure 1.1. Disparity, convergence and cohesion illustrated.
(Source: author)

Kuokkanen and Vihinen (2006) define three levels of convergence: between EU Member States, between EU regions and between EU citizens. In this thesis, the focus is on the first and second level, meaning that convergence across EU MS and EU regions is assessed.
From a theoretical point, this study could contribute to the existing literature in three ways. First, it focuses on convergence on a micro/meso-scale, the dairy sector, while most convergence studies have focussed on the macro-economy. Secondly, it provides an overview of the empirical methods that can be used to test convergence since we measure convergence for three different variables. Lastly, it can contribute to the existing field by quantifying convergence for an atypical part of the economy, namely agriculture. Thereby, it can quantify the concept of convergence and link it to the agricultural policy, in which convergence is an important element.

1.4 METHODOLOGY AND DATA

To apply the question of convergence to the current societal and political debate, we have to link it with the policy on the dairy sector, the CAP. The original objectives of the CAP that are still in effect are: i) stimulate agricultural productivity ii) ensure a fair farm income iii) stabilise markets iv) assure availability of supplies v) ensure reasonable consumer prices (European Union, 2012). The availability of milk supply and the assurance of reasonable consumer prices are not major issues (European Commission, 2016b, 2017b). For that reason, it is assessed whether there was convergence between the Member States with respect to productivity, prices and farm income in the period from 2004 onwards. This specific period entails the eastern and central enlargements of the EU. Croatia is excluded from the study, since it accessed only recently in 2013 which is from the perspective of data collection a time period which is too limited. So this leaves us to the Member States that were member at 1 June 2007, the EU-27.
For each variable (productivity, prices and farm income) we conduct: i) an identification of the concept of convergence ii) a literature review iii) an overview of the empirical methods iv) an explanation of the data and empirical model v) an overview of the results vi) a discussion on the chapter.
The empirical analysis is done with the statistical software programs STATA and R. Data on productivity and farm income is obtained from the Farm Accountancy Data Network (FADN). This data is yearly data aggregated at FADN regional level. Data on individual farm level is not available due to privacy restrictions of the data. The latest validated FADN survey available is from 2015, which is the first year that the quota was abolished. Within the FADN survey, dairy farms are specified by their category in the FADN data: TF45: Specialist dairying (FADN, 2014; 2018a). Milk prices are obtained from the EU Milk Market Observatory. Price data is monthly data aggregated on Member State level. The price analysis entails also the years 2015-2017 because price data for more recent years are available (Milk Market Observatory, 2017a).

1.5 OVERVIEW OF THE CONTENT

In chapter 2, an analysis of the disparity in the dairy sector in the EU is done; starting with the situation in 2004, we look at the disparity between the different regions, how this developed till the end of the quota and what the current situation in the sector is. In chapter 3, the economic theory on economic convergence is described. It is explained what the importance of convergence is on the European level, how convergence is defined in and what other theories had influence on the convergence debate.
The fourth, fifth and sixth chapter are devoted to the convergence with respect to milk prices, productivity and farm income. These empirical chapters start with a definition of convergence, followed by a literature review, an outline of the empirical methods and the data and in the end with the results and conclusive thoughts. Finally in chapter 7, the results are summarized and discussed in the light of the issues that are present in the sector. The chapter gives an indication how the overall convergence developed and what this implies for the future of the sector and the future of the policy on the sector. Furthermore, we critically reflect on the approach, models and tests that were used in this thesis.

2 DISPARITY AND STRUCTURAL CHANGE FROM 2004 TO THE PRESENT

In this chapter an outlook is given on the dairy sectors of the MS throughout the period 2004-2014. We start in 2004 by explaining the disparity that existed between the OMS and NMS. Then a subsection is devoted to the structural change that took place between 2004 and 2014, to illustrate the processes that could be of influence with respect to convergence. The last part gives a short overview of the current situation, by explaining the latest developments. Each time period is divided in four sections: policy, markets, farms and industry. The first three characteristics are particularly relevant since structural change has been going on in the past decade. Industry is also added as a topic, since it has an important share in the vertical integration of markets. For that reason, it is an important factor that can explain the disparity between the regional dairy sectors in the EU.

2.1 STARTING SITUATION IN 2004

2.1.1. Policy

The major tone for the dairy policy in the EU was set just before 2004. During the 2003 CAP reform, the European Commission proposed a set of new policy measures to the European policy for the dairy sector. In June 2003, the ministers of the European countries decided to implement most of these measures (although not as radical as the Commission proposed). The main measure was that intervention prices for butter and Skimmed Milk Powder (SMP), were gradually decreased in the period 2004-2007. However, farmers were given a compensatory payment for this lower price. Besides, it was decided to increase the milk quota with 1.5%, and to sustain the quota till 2014/2015. Furthermore, the payments to dairy farmers were aggregated to one single payment instead of multiple payments (De Bont et al., 2003).

Above the 2003 CAP reform, came the accession of ten new Member States to the European Union. Before their accession, there were quite some issues in relation to the CAP. First of all, there were budgetary concerns since some studies estimated that the costs of the CAP would increase with a third (Bach et al., 2000). Secondly, the old agricultural policies of the NMS should be aligned to the CAP to have a smooth transition. At the Copenhagen Summit, the result of the negotiation was that direct payments were gradually introduced for the NMS. The NMS started off with 25% of the direct payment level of the OMS rate in 2004. Eventually, this would increase till 100% of the rate in 2013 (European Commission, 2002). Moreover, the NMS had to comply with a milk quota that was based on a historic reference production, just like all the old Member States.

2.1.2. Markets

The period just before 2004 was still marked by high intervention prices for dairy related commodities like SMP and WMP. For that reason, the milk prices were quite stable, as can be seen in Figure 2.1. All prices seem to fluctuate in a bandwidth between 25 cents and 40 cents. Apart from some seasonal variation in the price, no large shocks hit the market.

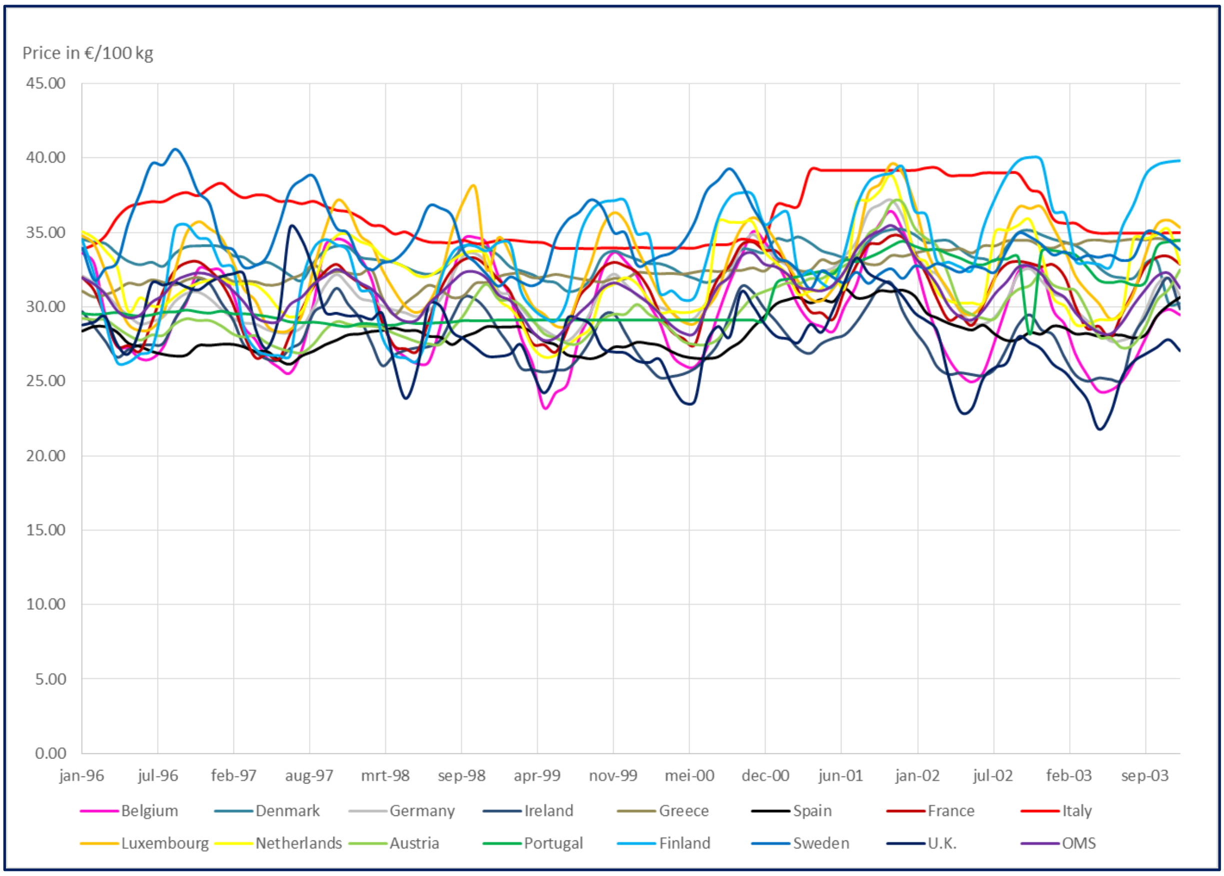


Figure 2.1. Milk price received by dairy farmers in selected EU Member States from 1995-2003 in (€/100 kg)
(Source: author, based on MMO, 2017)

2.1.3. Farms

The general story since the introduction of the milk quota in 1984, is that the number of dairy farms has declined dramatically. In Table 2.1 it can be seen that for most countries that the 2003 index number of dairy farmers is around 30, meaning that the number of dairy farms decreased with 70% in only two decades. That has also resulted in a smaller dairy cow herd in the EU. Nevertheless, maybe the most important development that can be seen in the table is that the number of dairy cows per farm increased. For Italy, the number of dairy cows per farm more than tripled in twenty years (index 354).

Table 2.1. Number of farms with dairy cows and average number of cows per farm in 1983 and 2003

	1983		*2003*		*Index 2003 (1983=100)*		
	Farms	*Cows/Farm*	*Farms*	*Cows/Farm*	*Farms*	*Cows*	*Cows/Farm*
Belgium	48,740	20	16,570	35	*34*	*60*	*175*
Denmark	35,480	28	7,950	75	*22*	*59*	*265*
Germany	396,920	14	121,280	36	*31*	*79*	*258*
France	420,430	17	113,930	36	*27*	*56*	*207*
Ireland	91,440	18	27,000	43	*30*	*69*	*235*
Italy	331,530	8	67,500	28	*20*	*72*	*354*
Luxembourg	2,510	27	1,040	39	*41*	*59*	*143*
Netherlands	63,540	40	25,000	59	*39*	*58*	*147*
UK	57,600	58	28,210	78	*49*	*66*	*134*

(Source: Van Berkum and Helming, 2006, based on Eurostat and LEI data)

2.1.4. Industry

Dairy processors play the key role in the valorisation of milk in the EU. Since milk is a very perishable product, the processor is needed to extend the shelf-life of the milk. Additionally, the processor is able to market the product to various wholesalers in order to bring it to the (international) consumer. Dairy farmers are therefore highly dependent on the processors. The dairy processing- and retailing stage are characterized by a high degree of market concentration. Currently, in the retail sector, dairy products are very often sold at a loss, to attract consumers (Ihle et al., 2017).

2.2 STRUCTURAL CHANGE BETWEEN 2004-2014

2.2.1. Policy

From 2004 onwards, the measures of the 2003 CAP reform came into force. In the period from 2004 onwards, most payments were combined into one Single Farm Payment. To receive payments, farmers had to comply with some ‘cross-compliance’ standards with respect to animal welfare, animal health and environment (Meester et al., 2013). In 2008, some revisions of the CAP were formalized in the ‘2009 Health Check’. An important measure was the liberalisation of the Article 68. This article gives Member States, the possibility to redistribute a part of the income support to other targets (Meester et al., 2013). In 2010, the European Commission (2010a) introduced the ‘milk-package’ in order to increase bargaining power for dairy farmers. This measure would improve the position of the farmer with regard to the dairy processer, by providing collective negotiation. In the years before the quota, the quota has been gradually increased, also known as the ‘soft-landing’ policy. From 2009 onwards, the milk quota of each Member State was increased by 1% each year in order to guarantee a smooth transition to the post-quota era. As can be seen in Figure 2.2, not every country used their quota fully. Romania and Bulgaria, used only half of their quota rights. While other countries like Austria, Denmark and Poland exceeded the quota.

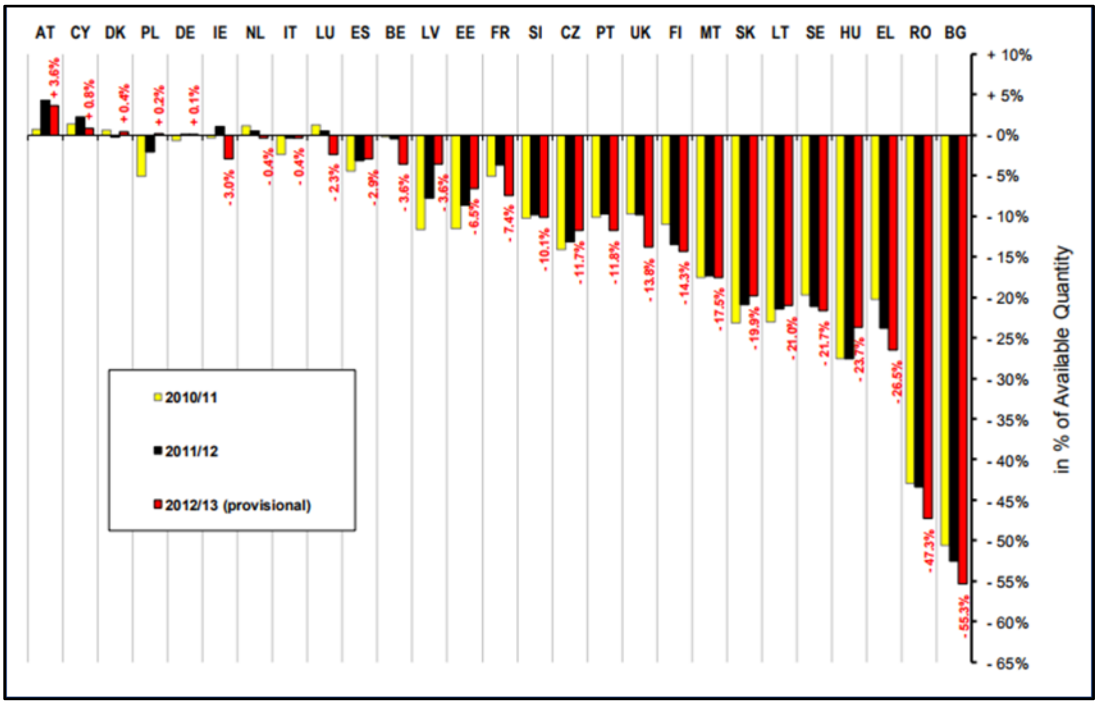


Figure 2.2. EU deliveries overshoot/underuse in % of quota.
(Source: Versteijlen, 2013)

2.2.2. Markets

Due to the lower intervention prices of SMP and WMP, the equivalent milk price also declined. In Figure 2.3, it can be seen that the period 2004-2014 was denoted by a large price peak followed by a large price drop in the years 2008 and 2009. The price drop naturally led to a severe crisis for dairy farmers. When the prices returned to normal, and the crisis disappeared, one thing remained: price volatility. Price volatility in the commodity markets for dairy products is reflected in a volatile farm-gate milk price. A highly unstable milk price is hard for farmers, since it becomes difficult to determine whether investments will pay off in the future. With the 2003 CAP reform, price floors and tariffs were lowered, which means that the EU milk price is now more closely linked to the world market (O'Connor et al., 2015).

Another remarkable sign that can be observed in Figure 2.3 is that in 2007 and 2010, the world market price level exceeded the European milk price (marked yellow). This would generally mean that the EU milk price became competitive on the world market, since no export refunds would be needed to offset the demand. Overall, this picture of competitiveness on the world market is more complicated. Quality of dairy products in the EU is an important factor for exports (e.g. special cheeses or high-quality milk powder). For that reason it does not mean that when EU milk price is above the world milk price that exports are not profitable (Jongeneel, 2011). In fact, the EU has a large trade surplus on milk products. In the past decade, this trade surplus on dairy products has almost doubled. Cheeses, skimmed milk powder (SMP) and whey milk powder (WMP) have been the growing export markets since 2005. The importance of China as a trade partner has only larger and larger. For instance, 33% of EU whey powder export is going to China (Ragonnaud, 2014).

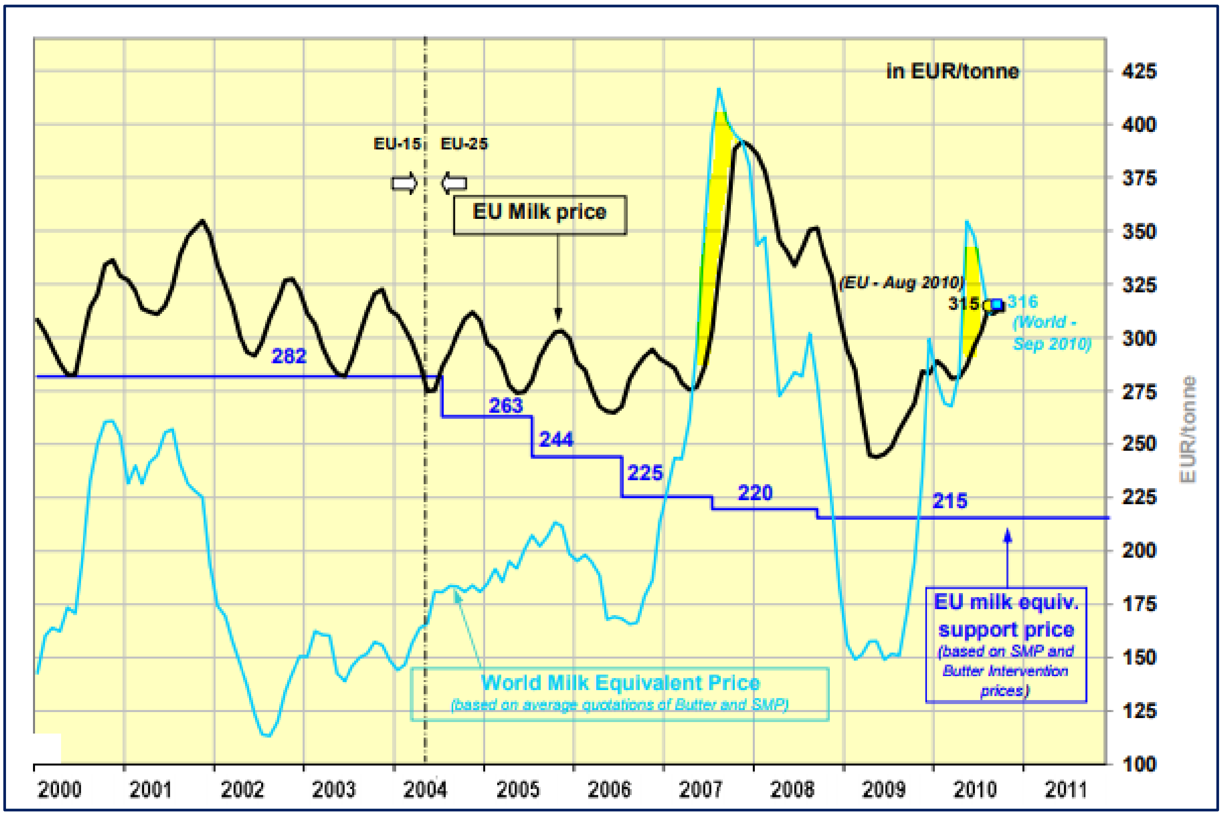


Figure 2.3. Price developments of EU milk price, EU milk equivalent support price and the world milk equivalent price from 2000-2011 in EUR/tonne (Source: European Commission, 2010b).

In Figure 2.4, it is shown how the raw milk prices developed per Member State. Up to 2003, most prices moved more or less within the same bandwidth. By the introduction of the NMS, the price disparity increased within the EU market. In the years that follow, it can be seen that indeed there is a lot of volatility in the national milk prices. The highest price can be observed on the geographically isolated islands of Cyprus and Malta, but also Greece and Finland show an above average price. On the lower bound, countries like Lithuania, Romania and Bulgaria can be found.

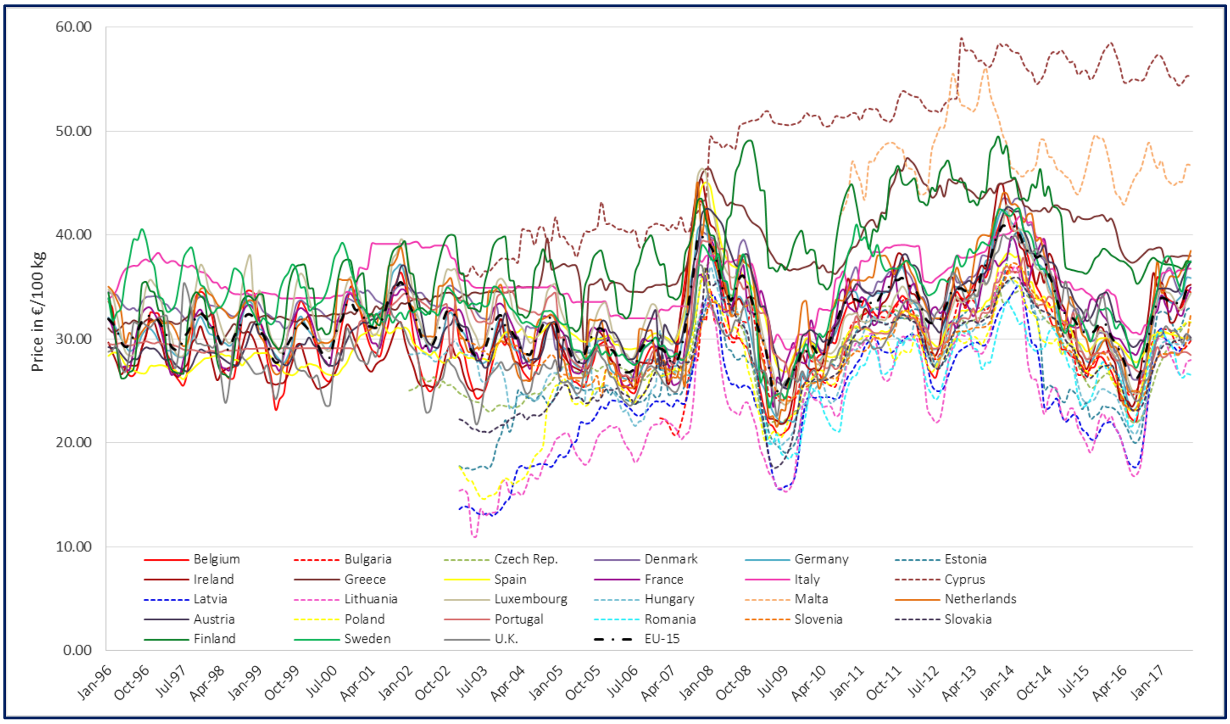


Figure 2.4. Historical EU price series of cow's raw milk in euro/100 kg per EU MS
(Source: author based on MMO, 2017)

2.2.3. Farms

As can be seen in Table 2.2, a few years after accession the dairy sector of the NMS consisted for a substantial part of small-sized farms. For a country like Poland, with 278 thousand dairy cows, the data shows that almost 90% of the dairy farms owned less than 20 dairy cows. Also for Romania, about 99% of the farms owned less than 5 dairy cows. It is obvious that the scale of production is completely different compared to the OMS. On the long term, this means that structural change is unavoidable.

Table 2.2. Average number of dairy cows per farm and share of dairy cow farms per size in 2007

NMS	*Average heads per farm*	*% farms <5 cows*	*% farms 5-19 cows*	*% farms >20 cows*	*Dairy herd (x1000)*
Bulgaria	2.7	91.5	6.8	1.7	336
Cyprus[1]	228.2	5.3 (≤10)	6.1 (11-50)	88.6 (>50)	56
Czech Republic	165	4.3 (≤10)	13.7 (11-50)	82 (>50)	423
Estonia	14.5	65 (1-2)	25 (3-19)	10	104
Hungary[1]	19.8	56 (1-2)	33 (3-10)	11 (>10)	321
Latvia	4.6	84.4	11.7	3.9	179
Lithuania	3.3	90.5	8.3	1.2	396
Malta	50.6	5 (1-2)	29.5 (3-29)	65.5 (≥30)	7.5
Poland	4.2	88.4 (<10)	10.2 (10-30)	1.4 (>30)	2787
Romania	1.6	98.7	0.7 (5-10)	0.6 (>10)	1700
Slovakia	183	23 (≤10)	10 (11-50)	67 (>50)	181
Slovenia	6.5	60	33.5	6.5	124

Note: 1) 2005 data used (Source: Van Berkum, 2009)

In Table 2.3, it can be seen that during the period 2004-2015, the overall image is that the number of cows per dairy farm has increased. However, the growth is not the same in every country. As is reflected in the index number, Romania has not seen an increase in the number of dairy cows per farm, nor has Malta. On the contrary, the number of dairy cows per farm has doubled in Bulgaria, although in absolute terms they are still at the lower spectrum. If we look at the OMS- and NMS averages, we see that they do not differ that much in 2004, partly due to the high number of cows/farm in Slovakia. In the period that follows, it is clear that on average the OMS have a higher scaling-up process than the NMS that accessed in 2004. This is quite remarkable since we would assume that NMS have higher returns to scale when they scale up their farms.

Table 2.3. Number of dairy cows per specialist dairy farm (TF:45) aggregated on MS level from 2004-2015

Country	*2004*	*2005*	*2006*	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*	*2014*	*2015*	*Index 2015 (2004=100)*
BEL	48	48	48	52	53	58	59	59	66	66	68	70	*144*
BGR	-	-	-	8	9	14	13	13	14	15	16	17	*209[1]*
CZE	79	82	83	87	88	101	92	98	87	86	90	117	*148*
DAN	91	97	96	117	117	144	150	144	170	168	166	165	*182*
DEU	44	44	45	47	47	53	53	54	60	61	62	63	*144*
ELL	27	-	39	-	-	-	-	-	51	49	-	-	-
ESP	29	30	32	34	36	45	43	43	50	50	50	53	*179*
EST	49	50	51	58	58	73	58	61	68	74	75	73	*149*
FRA	43	43	44	46	47	50	52	53	56	57	58	59	*138*
HUN	30	31	32	36	41	41	41	41	43	40	42	44	*143*
IRE	51	52	52	56	57	63	64	65	67	67	67	72	*141*
ITA	45	46	48	43	42	47	46	48	53	54	54	53	*118*
LTU	8	8	7	11	10	12	12	11	10	9	10	10	*118*
LUX	42	43	43	46	46	53	53	55	61	64	65	68	*163*
LVA	13	13	13	15	15	15	15	14	16	16	16	16	*127*
MLT	63	64	62	71	71	64	61	57	59	58	59	57	*90*
NED	65	66	68	72	73	81	82	82	92	94	94	91	*139*
OST	14	14	14	15	15	17	16	16	17	18	18	18	*128*
POL	12	12	12	13	13	14	14	15	15	16	16	17	*136*
POR	23	23	23	25	25	26	27	28	31	31	31	32	*140*
ROU	-	-	-	4	5	5	5	5	4	4	4	4	*95[1]*
SUO	21	21	22	25	25	29	30	31	34	33	35	34	*162*
SVE	43	45	46	52	54	65	58	59	76	77	76	81	*189*
SVK	167	196	187	190	201	197	184	206	198	205	220	190	*114*
SVN	13	14	14	12	12	17	17	17	19	16	16	16	*123*
UKI	95	97	99	115	111	119	121	121	127	132	133	136	*143*
OMS	45	48	48	53	53	61	61	61	67	68	70	71	*157*
NMS (2004)	48	52	51	55	57	59	55	58	57	58	60	60	*124*
NMS (2007)	-	-	-	6	7	9	9	9	9	10	10	10	*170[1]*

Note: Rounded to integers. 1)2007=100
(Source: author, based on FADN, 2017b)

In Figure 2.5 it can be seen that there is also a clear difference with respect to labour income. The Netherlands, Italy and some parts of the UK, Spain and Germany show labour incomes on dairy farms that are higher than the regional average income. More remarkable is that the Nordic regions show relative lower labour incomes compared to their regional average income.

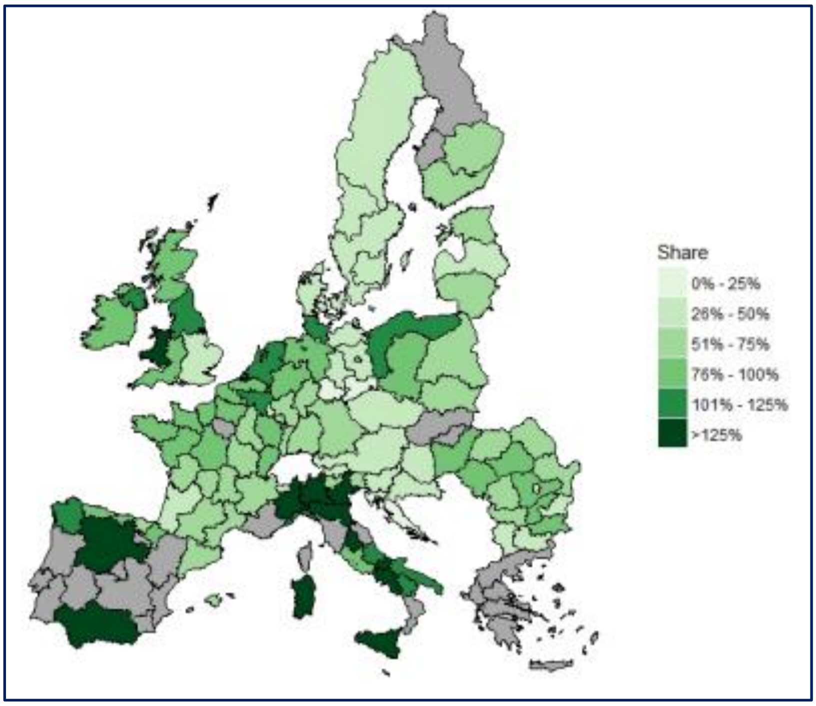


Figure 2.5. Labour income (in AWU/year) of specialist dairying farms (averaged over 2011-2013) in comparison to regional average income (GDP/capita) in 2014.
Note: Regions marked grey had no adequate data. (Source: Ihle et al., 2017, p. 84)

At last, there is a substantial differentiation with respect to the specialization of livestock farms in the EU. As shown in Figure 2.6, we see in some countries that specialist dairy farms hold the largest share of the livestock in a region. In France, Romania, Poland, we see that specialist dairy farms have a minor share in the total amount of cattle. This means that in several regions, mixed livestock/crop farms or fattening farms (for beef production) account for a considerable share in the total cattle. So, in the Netherlands, Germany, Denmark, Bulgaria and some parts of Italy, we see that that specialist dairy arms account for the majority of cattle in the country.

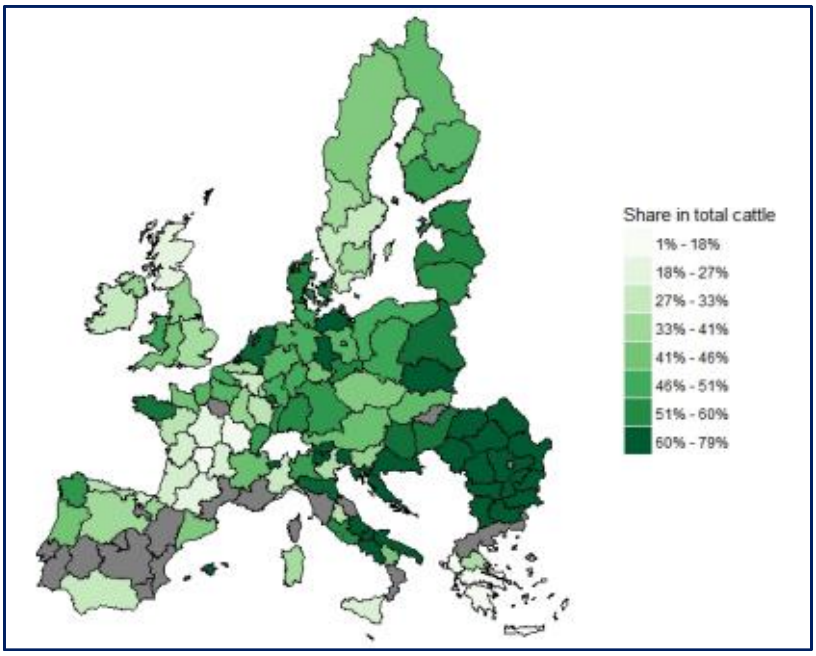


Figure 2.6. The distribution of the share of total cattle kept by specialist dairying farms (TF:45) in total cattle number per FADN region in 2013.
Note: Regions marked grey had no adequate data.
Source: Ihle et al.., 2017, p. 133)

As can be seen in Figure 2.7 some regions have higher number of specialist dairy farms than others. Still, on the eastern border of the EU, Romania, Poland and Lithuania are countries with a lot of active dairy farmers.

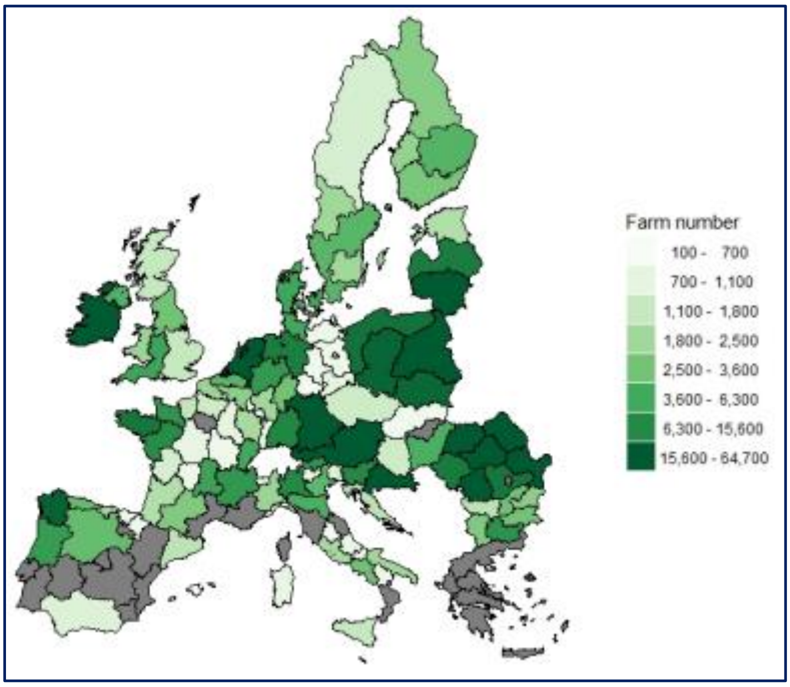


Figure 2.7. The number of specialist dairying farms (TF:45) per FADN region in 2010.
Note: Regions marked grey had no adequate data.
(Source: Ihle et al., 2017, p. 69)

2.2.4. Industry

In the EU, we see that the cooperative structure is a very important organizational form. In Table 2.4, it can be seen that in the Nordic countries, the Netherlands and Ireland more than 90% of the dairy market belongs to cooperatives. So, in these countries, the farmers collectively own the processing company. This means that they have a stronger bargaining position with respect to the farm-gate milk price.

Table 2.4. Cooperative market shares per Member State in eight agricultural sectors in 2010

Member State	*Cooperative market share in %*	*Member State*	*Cooperative market share in %*
Austria	95	Latvia	33
Belgium	66	Lithuania	25
Bulgaria	n.d.	Luxemburg	n.d.
Cyprus	10	Malta	91
Czech	66	Netherlands	90
Denmark	96	Poland	72
Estonia	35	Portugal	70
Finland	97	Slovakia	25
France	55	Slovenia	80
Germany	65	Spain	40
Greece	0	Sweden	100
Hungary	31	UK	35
Ireland	99		
Italy	42	*EU-average*	*57*

(Source: Bijman et al., 2012)

In Table 2.5 it can be seen, that concentration of the dairy industry in a country differs per Member State. Again, the Nordic countries and the Netherlands stand out due to the market share of the biggest processor. Among those processors are Arla and FrieslandCampina, who are in fact one of the largest milk processors worldwide. Besides, we have seen some large mergers in the dairy industry: Lactalis and Parmalet, Friesland Foods and Campina (FrieslandCampina), Humana Milchunion and Nordmilch (Deutsche Milk Kontor). On the other hand, in Bulgaria and Romania still have a low concentration in the dairy industry with many small-scale processing companies (Ihle et al. 2017).

Table 2.5. Share of national milk delivery by processors.

Share of national milk delivery	*Date of data*	*Biggest processor*	*Sum of second to ninth biggest processors*	*Rest (processors or not)*	*Name of biggest processor*	*Comment*
Austria	2010	Approx. 40%	Approx. 40%	Less than 20%	Berglandmilch	-
Belgium	2008	-	-	-	Milcobel Belgomilk	-
Bulgaria	2011	-	-	-	Poliday-2 Ltd Karlovo	-
Cyprus	2011	-	-	-	Vivartia Cyprus	-
Czech Republic	2011	Over 60%	-	Less than 40%	Madeta a.s.	-
Denmark	2011	Approx. 90%	-	Approx. 10%	Arla Foods	-
Estonia	2010	Approx. 25%	Over 50%	Approx. 25%	TERE AS	-
Finland	2010	Approx. 85%	Approx. 15% (*)	0%	Valio Oy	(*) Only five processors are responsible for 100% of national Valio Oy milk delivery
France	2010	Approx. 25%	Over 50%	Approx. 25%	Lactalis	-
Germany	2010	Approx. 25%	Approx. 40%	Approx. 40%	DMK	-
Greece	2011	-	-	-	Vivartia SA	-
Hungary	2007	-	-	-	Sole-Mizo Zrt	-
Ireland	2011	Approx. 25%	Approx. 60%	Less than 15%	Glanbia	-
Italy	2010	-	-	-	Parmalat (*)	(*) In terms of turnover
Latvia	2011	Approx. 15%	Approx. 45%	Less than 40%	Rigas piena kombinats	-
Lithuania	2010	Approx. 30%	Approx. 60% (*)	Approx. 10%	SC Rokiskio suris	(*) sum of the second to the fourth processors
Luxembourg	2011	Approx. 45%	Approx. 55% (*)	0%	Luxlait	(*) Only five processors are responsible for 100% of national milk delivery
Netherlands	2010	Approx. 75%	Approx. 25% (*)	0%	Friesland Campina	(*) Only seven processors are responsible for 100% of national milk delivery
Poland	2010	Approx. 15%	-	Approx. 85%	SM Mlekpol (*)	(*) In terms of turnover
Portugal	2010	Approx. 25%	Approx. 60%	Approx. 15%	Agros	-
Romania	2011	-	-	-	SC Friesland Romania SA	-
Slovakia	2011	Approx. 15%	Approx. 35%	Approx. 50%	Rajo a.s., Bratislava	-
Slovenia	2011	Approx. 55%	Approx. 45% (*)	Approx. 1%	Ljubljanske mlekarne	(*) sum of the second to the seventh processor
Spain	2009	-	-	-	Danone S.A.	(*) In terms of turnover
Sweden	2011	Approx. 65%	Approx. 35%	0%	Arla Foods Sverige	-
UK	2011	Approx. 15%	Approx. 65%	Approx. 20%	Dairy Crest	-

(Source: Ernst and Young, 2013, based on IFCN-data)

2.3 The present situation

2.3.1. Policy

After 2014, the capstone of the European dairy policy was removed. Since 1984, the milk quota had put a limit on the production volume. Since 1 April 2015, the quota has been ended and now there is no official limit on milk production. Unfortunately, this led to a crisis because many farmers could not cope with the low prices that hit the market after the abolishment. For that reason, crisis measures were taken to assist farmers during this period. The Commission launched two support packages of €500 million each to support the farmers (European Commission, 2015;2016a). Article 222 was applied, which meant that producer organisations and cooperatives could make voluntary agreements about their production. Moreover, temporary state aid up to €15.000 per farmer per year was allowed.

In the Netherlands, the abolition of the quota led to a substantial increase of the dairy herd. In fact, this caused that the Dutch farmers were producing more phosphate than was allowed. The Ministry of Economic Affairs therefore had to come up with a 'phosphate reduction scheme'. This policy obliges farmers to reduce their herd to their reference herd in 2015, minus 4% (Van Dam, 2017). In fact, this policy created again a limit to milk production, but now established by environmental problems.

2.3.2. Markets

The abolition of the milk quota was the initiator of a sharp price decrease in the dairy market. Despite the soft-landing policy, excess supply led to a crisis that affected the majority of EU dairy farmers. In the years before the quota abolition, milk prices were at a quite high level. Droughts in Oceania led to low production in Australia and New Zealand, which led to a high world market price (Polet and Kuypers, 2017). European farmers faced high farm-gate prices, which made it possible for them to invest and expand regarding the lifting of the quota. However, a Russian import ban and a hampering export to China led to an excess supply that pushed the EU milk prices to record lows (Polet, 2015). In the meantime, prices recovered till some extent. Nevertheless, in Figure 2.8 it can be seen that there is still about 400.000 ton of SMP in storage. This will obviously have a depressing effect on the price in the short future, since they have to get rid of these stocks at some point.

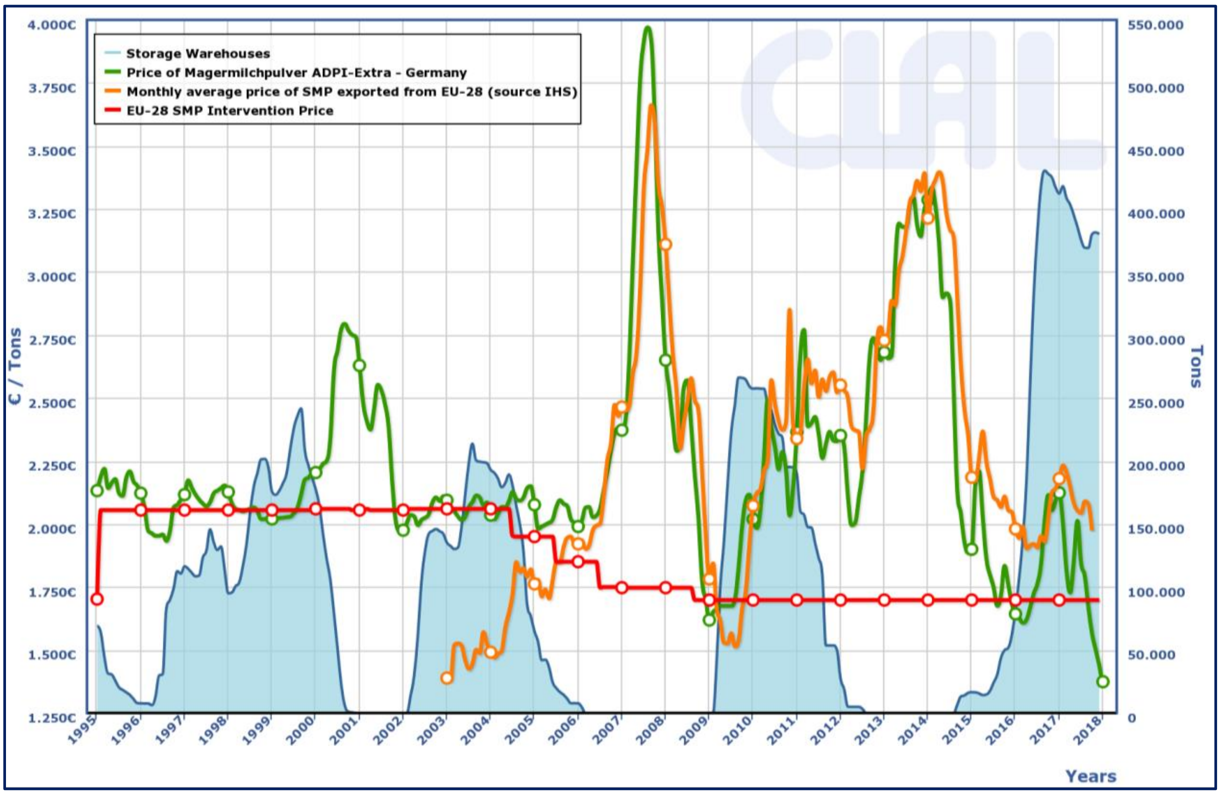


Figure 2.8. EU-28 comparative overview for SMP stocks (Public and Private), SMP market price in Germany and SMP intervention price (1.698 €/Ton to Dec 2017)
(Source: CLAL, 2017)

2.3.3. Farms

The post-quota period caused many problems for the farmers. Adequate data on this period is still lacking. Nevertheless, the general picture shows that 2015 and 2016 were problematic years for a considerable part of the farmers. Several cases are known where farmers were forced to stop because their cost price was simply too high. Also the front-runner countries like the Netherlands and Germany show examples of severe problems. In the Netherlands, a government regulation to financially support farmers to quit their business had to close after one day, due to the large number of farmers that applied for the regulation (Van Ammelrooy, 2017).

2.3.4. Industry

Also for the dairy industry, the abolition of the milk quota gave some problems. For example, FrieslandCampina offered farmers a premium when they did not raise their milk delivery (Smit, 2015). Due to the increase in production of farmers, the cooperative did not have enough capacity to process all the milk (which is in fact a key statute of most dairy cooperatives). More recently in Germany, Deutsche Milk Kontor announced that they will close several processing plants due to the fact that many farmers have given up their business (Deutsche Milk Kontor, 2017). One would expect that the abolition of the quota would only benefit the processors. Hence no limit on the main input of the processor would in generally be beneficial. However, the constant milk flow to the processing plant and the cooperative structure of the processing industry make it in reality complex for processors.

3 ECONOMIC THEORY ON CONVERGENCE

First it is explained why convergence is so important within the framework of the EU. In the section that follows, we go deeper into the theory of economic convergence; the definition, the new growth models. Followed by a discussion on the theory that is linked to the convergence debate. At last, it is defined how convergence is linked to this thesis.

3.1 THE IMPORTANCE OF CONVERGENCE IN THE EU

With the growth of the European Union, the EU became also more heterogeneous. Since no country is the same, each country brought also a different economy to the European Union. The common ideal of all members is of course to progress by cooperating with other countries. The 2nd article of the Maastricht *Treaty on European Union* (European Union, 1992) states this goal specifically:

*"To promote throughout the community a harmonious, balanced and sustainable development of economic activities, a high level of employment and of social protection, equality between men and women, sustainable and non-inflationary growth, a high degree of competitiveness and **convergence** of economic performance, a high level of protection and improvement of the quality of the environment, the raising of the standard of living and quality of life, and **economic and social cohesion** and solidarity among Member States."*

As can be seen in Article 2, convergence of economic performance and economic and social cohesion were from the beginning important objectives. Since policy-makers understood that convergence and cohesion were not naturally processes, they designed several structural polices to achieve this. These policies were translated into funds that backed this policy, and examples of these funds are: European Regional Development Fund (ERDF), European Social Fund (ESF), Cohesion Fund (CF), European Agricultural Fund for Rural Development (EAFRD), European Maritime and Fisheries Fund (EMFF) and Instrument for Pre-Accession Assistance (IPA) (European Commission, 2014). In this thesis we do not test the effectiveness of these funds on convergence, since it is hard to identify the impact of these funds on specifically the dairy sector.

Since the beginning of the EEC in 1957, agriculture has always been a prominent area that was tackled at the European level. In the EEC-treaty, agriculture was dealt with detail and became therefore a frontrunner in economic European integration (Meester et al., 2013). The Common Agricultural Policy therefore had an important influence on the development of the regions. The CAP in itself has also an important effect on convergence and cohesion between the regions. If we look at the objectives of the CAP (European Union, 2012):

- *(i)* *To increase agricultural productivity by promoting technical progress and by ensuring the rational development of agricultural production and the optimum utilisation of the factors of production, in particular labour;*
- *(ii)* *Thus to ensure a fair standard of living for the agricultural community, in particular by increasing the individual earnings of persons engaged in agriculture;*
- *(iii)* *To stabilise markets;*
- *(iv)* *To assure the availability of supplies;*
- *(v)* *To ensure that supplies reach consumers at reasonable prices.*

If we look at the objectives, the first two objectives are directly aimed at improving the economic performance of farmers. The CAP should aim at enhancing the productivity and income of farmers in regions that lag behind. The third objective is in place to ensure a stable level playing field in the European agricultural market. The last two objectives aim at improving the situation of consumers;

ensuring enough food at a fair price. With regard to cohesion of food consumption, people in Romania and Lithuania still spend around 25% of their total expenditure to food, while in Ireland and Austria this is lower than 10% (Eurostat, 2015). Apparently, there is also disparity in food consumption between the MS. As mentioned in chapter 1, we do not go deeper into the policy objectives (iv) and (v). For the future it is expected that in Less Developed Areas, the employment in agriculture will decrease further, due to the low productivity. If regions want to keep up employment in agriculture, then those regions should aim to upgrade their agriculture to higher quality segments. In the past decade, we have seen a trend within the EU of narrowed interregional disparity with respect to GDP per head (European Commission, 2017c). In fact this is what is desirable for the agricultural sector, the catching up of the areas that have agricultural sectors that are lagging behind.

3.2 CONVERGENCE IN THE ECONOMIC THEORY

3.2.1. The New Growth Theory

Convergence has already for a long time drawn the interest of economists. Whether poorer regions or nations could come closer to the richer ones is a crucial question for society. Most convergence studies are covered within the so-called Neo-classical Growth Theory. The first economic model that deals with convergence was developed by Robert Solow (1956). This neoclassical model has been the workhorse of economists to study convergence. The main equation of this model is:

$$Y = F(K, L) \tag{3.1}$$

In which the output of an economy, Y is a function of capital, K and labour, L. One assumption of the model was that an economy had decreasing returns to scale, which eventually led to the convergence hypothesis. Countries that were closer to the steady state had less economic growth then countries that were further away from the steady state level. In the past decades, this model has been further developed and it was used to assess convergence. Although this model became frequently applied in the literature, it has at the same time received as much critical feedback in the literature. Starting with Baumol (1986), who shows that there is proof for convergence amongst industrialized countries, but that this cannot be extended to the non-industrialized countries. Shortly after this paper, Barro and Sala-I-Martin (1992), state that across-country convergence is only found when accounting for differences between the steady-state characteristics between countries. Mankiw et al. (1992) contribute to the debate by stating that human capital was the missing link in the Solow model. So there is no consensus on a single model for convergence. This also resulted in a wide variety of empirical models. Islam (2003) distinguished four general approaches to test convergence: cross-section, panel, time series and distributional. The Solow model was mainly tested with the cross-sectional approach. Quah (1993) published an influential paper in which he criticizes this common method in convergence tests: regressing cross sections of average growth rates on initial level incomes. According to Quah this method is misleading for the given hypothesis, since the method suffers from Galton's fallacy; meaning there is regression towards the mean. Quah (1997) proposes non-parametric methods to examine whether there is convergence. In this method he studies the dynamics of cross-sectional income distribution with kernel density estimates. The author shows examples of twin-peak distributions, in which so-called convergence clubs exist. These clubs or groups of countries converge within their club, but they do not converge with other groups of countries.

All these disputes in the literature have led to various dichotomies in the study of convergence, which are explained in the next section.

3.2.2. The economic definitions of convergence

The wide spread of theoretical and empirical approaches to convergence have also led to plurality in the definition of convergence. Islam (2003) provides an excellent overview of the scientific debate on economic convergence. In this overview he shows that economists have different perceptions of convergence, as can be summarized in the summation below (Islam 2003, p.312)

(a) Convergence within an economy vs. convergence across economies;

(b) Convergence in terms of growth rate vs. convergence in terms of income level;

(c) σ-convergence vs. β-convergence;

(d) Unconditional (absolute) convergence vs. conditional convergence;

(e) Global convergence vs. local or club-convergence;

(f) Income-convergence vs. TFP (total factor productivity)-convergence;

(g) Deterministic convergence vs. stochastic convergence.

(a) The first duality shows that there are two economic scales to measure convergence. One is the measurement of convergence within an economy and the other is to measure convergence across economies. Originally, the Solow model was constructed to show a stable dynamic equilibrium for a single economy. Eventually, the model was mostly used for assessing across-economy equilibria (Islam, 2003).

(b) Secondly, the term can be used to define convergence in terms of growth rate or in terms of income level. Convergence in terms of the growth rate is that all countries gain equally from technological progress, thus in steady state all countries will have the same growth rate. Convergence in terms of income level is that if we assume that the aggregate growth function of each country is identical it means that the steady state of income level will be the same in every country (Islam, 2003).

(c) The definitions of σ-convergence vs. β-convergence are common used terms in convergence analysis. β-convergence means that poorer regions are grower faster than richer regions, and therefore poor regions will catch up with the rich regions (Monfort, 2008). This is related to the initial/income vs. growth rate regressions, which were used to study this phenomenon, a negative correlation between initial income per capita and the growth rate of income per capita would proof this type of convergence. As mentioned earlier, Quah (1993) showed that it could suffer from *reversion to the mean* (Islam, 2003). Still the study of β-convergence with these regressions is popular in the academic literature.

σ-convergence is a measure of the standard deviation or more generally: the degree of dispersion across income levels or income growth rates. If we find that the degree of dispersion reduces over time, we speak of σ-convergence. In Figure 3.1, σ-convergence vs. β-convergence is illustrated. In the left panel it can be seen, that σ-convergence implies that the dispersion of for example GDP/capita between countries becomes smaller over time. In the right-panel β-convergence is shown, it can be seen that when initial GDP/capita is lower, that the growth rates of GDP/capita are higher.

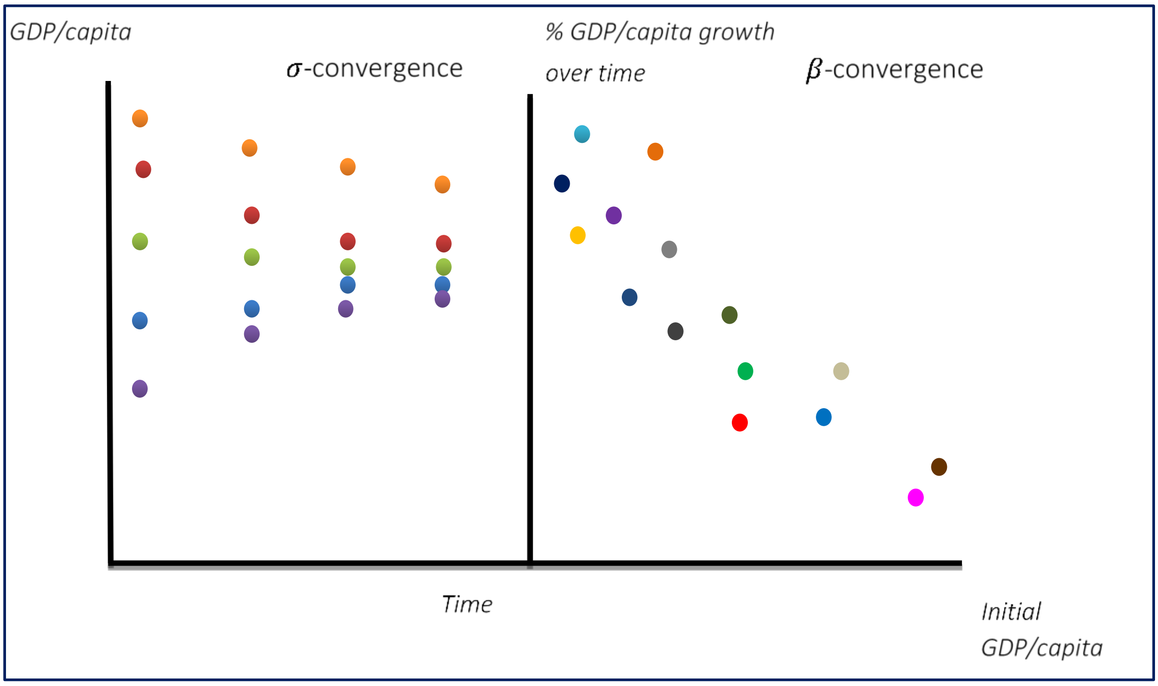


Figure 3.1. Illustration of σ-convergence vs. β-convergence
Note: Each coloured dot represents a region.
(Source: author)

(d) Another dichotomy is the unconditional vs. conditional convergence. In a simple Solow model with a Cobb-Douglas production function we can define (Islam, 2003):

$$Y_t = K_t^{\alpha}(A_t L_t)^{\beta} \tag{3.2}$$

Where Y_t is the output, K_t the capital, A_t the total factor productivity and L_t the labour. The steady state output per capita is solved in the Solow model, and gives the following equation (Islam, 2003):

$$y^* = A_0 e^{gt}\left(\frac{s}{n+g+\delta}\right)^{\frac{\alpha}{1-\alpha}} \tag{3.3}$$

Where A_0 is the initial TFP, s is the savings rate, n is the population growth, g the economic growth and δ the depreciation rate. Now the case with unconditional convergence is that we assume that these variables are the same for all countries. However, if we make take into account for differences in these variables between countries, we speak of conditional convergence.

(e) With the global convergence idea, it is meant that every country should move to one universal level steady state. This idea was criticized by many scholars since there are obvious structural differences between countries. Club convergence, does allow for multiple steady states. Perhaps some clubs of countries who share the same initial wealth and economic structure may reach convergence but only within this club. As an example, one could think of the initial EEC-countries who had quite a similar economy and also shared a common market and an administrative institution (Islam, 2003).

(f) Income convergence can be a result of two processes: technological catch-up or capital deepening. Most studies focussed on the second one, but a few also on the first one. These studies used Total Factor Productivity (TFP) as a measure for technology to see whether the technological levels of countries came closer to each other. This led to a side-branch of the convergence literature that was devoted to the convergence of TFP (Islam, 2003).

(g) The last dichotomy of the convergence definition is the stochastic vs. deterministic definition. In convergence studies that use the time series approach, it is common to index the data to one reference country. Then a unit root test is applied to test for convergence. In this test one can chose to include a deterministic or a stochastic trend, which resulted in these two definitions (Islam, 2003).

3.3 ECONOMIC THEORY THAT LINKS TO CONVERGENCE

Convergence is not an economic phenomenon that is independent of any other process in the economy. As many economic phenomena it is linked to various other processes in the economy. Naturally, convergence is dependent on the economic performance of the units of a country. Logically, for a catching-up process, some kind of growth of the units within a country is required.

To explain these processes in the economy, we shift more to the framework of economic geography.

Johann Heinrich Von Thünen (1783-1850) used a model with an isolated area with one city and a homogenous area around the city. With his model Von Thünen tried to explain the use of agricultural land. According to the model, land use becomes less intensive when the distance to the market becomes larger (Heijman and Schipper, 2010). Transportation costs differ by product, and the most valuable and delicate products have the highest costs, they are therefore produced close to the city. Fresh vegetables and milk are therefore located close to the city, since these products are highly perishable. While cereals, fuel wood and beef are located further away, since they can be stored longer and they are easier to transport. According to this simplistic model, differentiated transport costs determine where production is located. In real life, this model is too simplistic since it assumes an isolated area, with homogenous land and only one city. Nevertheless, the effect of transport costs on the location of a firm is of great importance.

Closely linked to previous theory, is the one of David Ricardo. His idea of 'comparative advantage' shows that a region specializes in an activity in which it has a relative price advantage. If this concept would be perfectly true, it is unlikely that convergence across regions with respect to one sector will happen. It would mean that certain regions that have the comparative advantage in dairying simply specialize more in dairy farming than other regions. As Lafougère (2012) argues dairy farms will more and more concentrate around the coastal areas of the Baltic sea, the North Sea and the Atlantic Ocean. Since these areas have more pastures, they are able to produce for lower costs. While on the other hand, Central European regions could more and more shift to crop production, since they have more suitable conditions for this type of farming (Jansik et al., 2014).

At last, the more recent theory of the New Economic Geography can be applied to this case. The new part of this theory was that general equilibrium models were applied in location theory (Krugman and Fajita, 2004). This theory tries also to explain why some regions flourish and others not. The main argument is that economic agglomeration arises because of increasing returns to scale. Keeping in mind transportation costs, the most suitable location for a business is close to the market. The labour force will move to a location where employment and consumption is nearby. When those workers concentrate near the firms, this will attract more firms since all these workers consume. This causes and upward spiral of success, also called centripetal forces. On the other hand, you have forces that work against this upward spiral: centrifugal forces. Concentration of firms and workers can lead to higher housing prices, land rents and traffic congestions. This will in fact create an opposite stimulus. The balance of centripetal and centrifugal forces shapes the agglomerations in the economy (Schmutzler, 1999). If we would apply this to the dairy sector, we could also identify certain centripetal and centrifugal forces. For example, a clear centripetal form is the concentration of the dairy processing industry. If many farms cluster together, it will attract the processing industry. If processors deliver high quality products, then farmers can profit from a higher milk price. This will strengthen this dairy cluster, and farmer can invest more in improving production which can result in an upward spiral in the cluster. On the other hand, if farmers cluster more and more together, land rents will go up and environmental pressure occurs. As we have seen in section 2.3.3., the Netherlands, as a very competitive dairy country, has for example increasing difficulty to manage their manure surplus.

3.4 From theory on convergence to the applied case of the dairy sector

As we have seen in the wide spread of theories and empirical methods, it is not straightforward how to approach convergence in this case. As we have seen in section 3.2, there is no uniform definition of convergence and it is therefore not surprisingly that there is not one single model that is used to test for convergence. By definition, this thesis focuses on across regions/countries convergence, since we want to study how disparities evolved between regions.

Overall, we could say that the definition of β-convergence is difficult to apply for the dairy sector. In consideration of the volatile market conditions in the dairy sector, milk prices and farm income, and in lesser extent productivity vary from year to year. As we discussed in chapter 2.2.2, since 2007 the dairy market has known several crisis periods. The accession years make it also problematic to decide on a starting year and an end year. For that reason, initial income vs. income growth regressions do not make sense, because it would mainly be troubled by the yearly market circumstances. Above that, in the academic debate Friedman (1992) and Quah (1993) have pointed at the structural weaknesses in the β-convergence definition. Additionally, if we would condition this regression on national characteristics, it is quite arbitrary to determine what variables to pick. The development path of a low productive might depend largely on a lack of financial resources, while for a high productive region it might be the lack of agricultural land. Examining β-convergence is therefore a highly arbitrary approach. The approach of σ-convergence seems therefore more suitable; examining the dispersion of the cross-sectional distribution over time. If we look back at Figure 1.1, this definition seems to fit better in our research approach. Our starting situation is disparity between dairy sectors, and the particular interest is to find how this disparity developed over time. In this thesis we examine convergence for several variables. By nature, these variables are different in their behaviour. Prices for example are a signal of the market and fluctuate heavily, while productivity is a rather stable concept which is determined at the farm-level. Additionally, for productivity and income ongoing growth is desirable, while for prices it is not desirable to increase endlessly. In the coming chapters, we continue the study by empirically assessing convergence. For every variable in this study, it is explicitly mentioned what definition is used and which empirical approach goes with this definition. Practically, this means that several empirical models are used that fit the applied case of the dairy sector the best. It demonstrates whether these variables show the same extent of convergence and whether the empirical measures yield the same conclusions on convergence.

4 PRICE CONVERGENCE

In this chapter we test for convergence in milk prices. First, it is explained how the concept of convergence can be applied to prices. After that, an overview of the academic literature on price convergence is given. Then, the choice of empirical measures is clarified and the empirical measures themselves are explained. Then we explain the empirical model and show the results.

4.1 THE CONCEPT OF PRICE CONVERGENCE

As we have seen in the previous chapter, there is not one definition of convergence. So for analysing convergence with respect to the milk price, we have to choose what definition to use. It in this case not of interest whether the price level is high or low, since prices are a signal of demand and supply. As we have seen in chapter 2.2.2, milk prices have been fluctuating heavily in the past. Assessing β-convergence based on initial price levels and growth rates of prices is therefore less suitable, since the measure would largely depend on the market situation in the corresponding years. σ-convergence is therefore a more useful definition, since it measures how the price dispersion developed over time. For that reason, we stick to the definition of σ-convergence in this chapter. In particular, we are interested to find whether the farm-gate milk prices of the NMS and OMS came closer to each other. As was seen in Figure 2.4 in section 2.2.2, the milk prices of the NMS were mostly lower than the OMS. If convergence would occur, i.e. dispersion would decrease, then we would expect that the prices of the OMS and the NMS came closer to each other. So that would mean that the sum of the distances between the several milk prices would decrease over time.

If eventually all EU milk prices converged, then we would expect that when a market is fully integrated that the 'Law of One Price' (LOP) holds. The LOP means that the price of a specific good must be the same everywhere due to arbitrage, thereby taking into account transportation and transaction costs (Fackler and Goodwin, 2001). This law can formally be written as (Richardson, 1978):

$$P_i = \beta_0 P_j^{\beta_1} E_{ij}^{\beta_2} T_{ij}^{\beta_3} X_{ij}^{\beta_4} \tag{4.1}$$

Where P_i is the price of the commodity in country i, respective j. E_{ij} is the exchange rate between the countries i and j, T_{ij} is a variable that accounts for the transaction costs between country and X_{ij} is a random variable that accounts for reasons why prices may differ between the countries, the βs are parameters. If $\beta_0, \beta_1, \beta_2, \beta_3 = 1$ and $\beta_4 = 0$, we would expect that there is perfect arbitrage over the commodity, so that the LOP holds. As we have one single market in the EU, we would assume that the market is fully integrated. All trade barriers are removed so every MS can freely trade with any other MS. Is it however realistic that all EU milk prices would converge and that in the end the LOP would hold? Perhaps not, since a condition for the LOP to hold is that the market has many consumers and many producers that do not have power to influence the price (Dreger et al., 2007). As was pointed out in chapter 2, there exist structural differences between the regional dairy markets in the EU that can keep up the market segmentation.

As can be seen in Figure 4.1, there exist several components that explain the pricing of milk. This starts at farm-level, where the quality of the milk in terms of protein- and fat content determines the price the farmer receives for his milk. Next to that, if a farmer has a higher hygienic standard this will be reflected in the price he receives. Chavs and Kim (2001) have provided an interesting work on the how the underlying product components of dairy products are reflected in the dairy price with the help of hedonic pricing models. Due to the high aggregation of milk prices in our dataset it is hard to retrieve the underlying components of the milk price, so therefore we do not go deeper into the topic of hedonic pricing. Also the CAP influences the farm-gate price. As discussed in 2.2.2, the EU intervention price helped to sustain a certain price level in the EU in some periods. Above that, the milk quota limited the production of certain countries meaning that there was no full competition. The trade barriers imposed by the EU on dairy imports, have also hindered price transmission from the world- to the EU market. In the processing stage, it is also dependent whether the processor is an Investor Owned Firm (IOF) or a cooperative. Cooperative farmers have a stronger bargaining position with respect to the price they receive, since they are the owners of the company. As reported in section 2.2.4, there is a substantial difference in the cooperative market share between EU MS. At last, the portfolio of products that a processor produces influences farm-gate price. The processor tries to maximize profit given the composition of each product and the costs of the ingredients. Some of these products are highly tradable, like WMP or SMP, but for fresh dairy products like yoghurts it might be harder to trade them internationally, due to the shorter shelf-life. We might have to take into account whether the goods are tradable or non-tradable (Officer, 1986) because it is harder to arbitrage over non-traded goods. The diversity of dairy products is also an important factor to take into account. The end-products like yoghurts, cheeses, butters are so diversified in terms of quality and other characteristics that it will be difficult to find evidence for convergence. Not to forget, the regional speciality products that hold a Protected Designation of Origin (PDO) or Protected Geographical Indication (PGI) who create a niche market for their product. It is therefore more likely that we find price convergence in the input or intermediate products. Furthermore, some countries have strong seasonal patterns in the milk price, with Ireland as a well-known example (Bergmann et al., 2015). Considering all these factors that affect the farm-gate milk price, it is not realistic to assume full convergence of prices.

Dreger et al. (2007) identifies two main processes that could lead to price convergence in an internal market. First, due to increased competition there will be downward pressure on the mark-up of all firms and therefore prices will decrease. Secondly, low income countries will have an upward pressure on the price level due to their catching up process in the transition period.

We are interested to find whether the dispersion of prices became larger or smaller over time. Due to the pricing mechanisms and the structural differences in the dairy sector it is expected that only limited convergence has taken place.

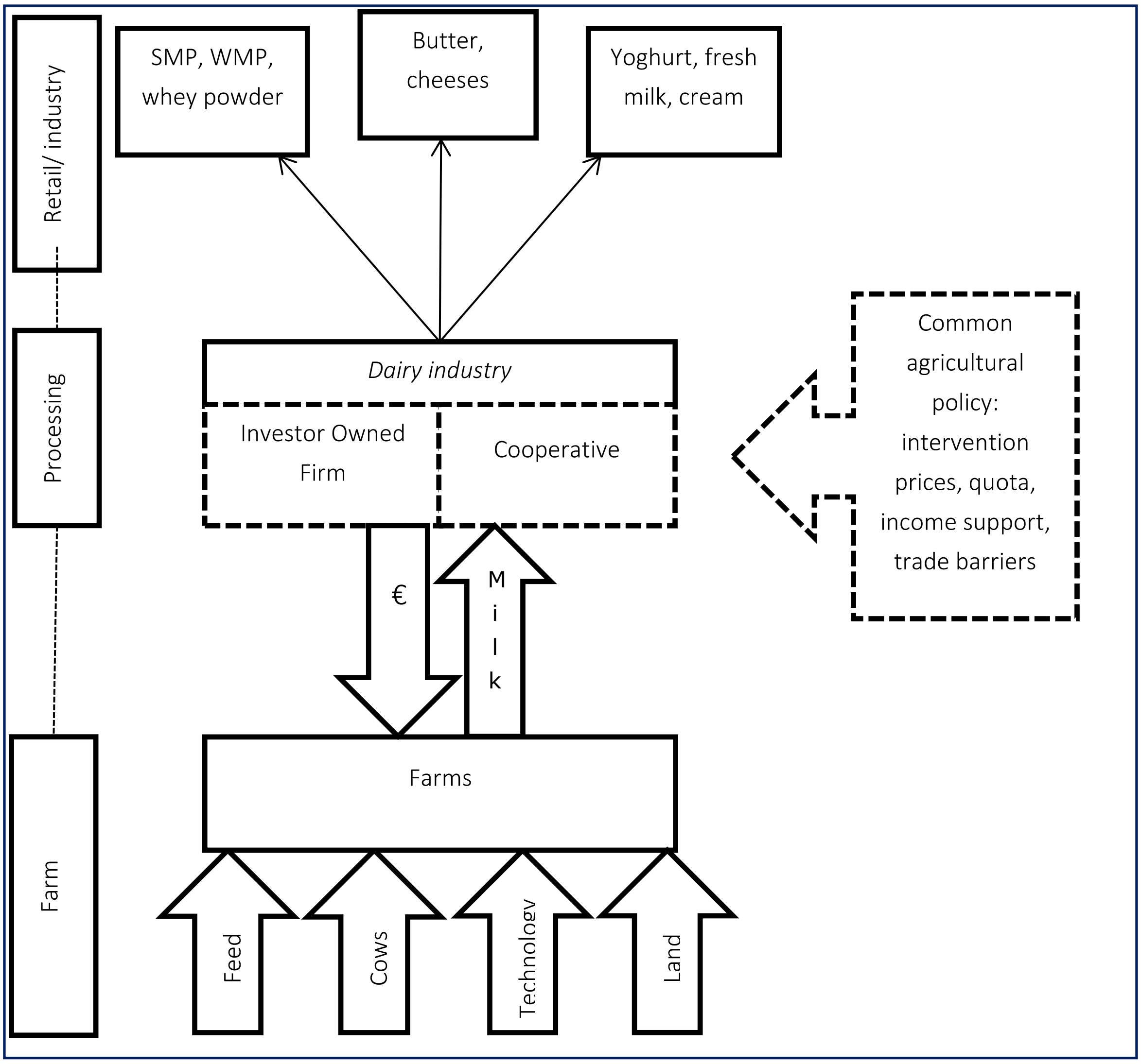


Figure 4.1. Illustration of the potential factors that influence the farm-gate milk price.
(Source: author)

4.2 LITERATURE REVIEW

Studies on convergence of dairy commodity prices do not exist. Most studies on dairy prices focus on price transmission analysis between farm-gate milk prices and retail dairy product prices[1]. In this way, it can be identified whether the retail sector softens price fluctuations or not. However, this is a measure of vertical relationship in the supply chain. There are no studies that have examined price convergence of dairy commodities in the EU so far. Although there are no studies found on dairy commodities, there are some scholars who have studied price convergence across the EU for several goods. Parsley and Wei (1996) studied the convergence of U.S. prices towards 'Purchasing Power Parity', a law that is closely tied to the LOP. They find that between 1975 and 1992, commodity prices across for tradable goods 48 cities in the US converged fast to price parity. Sosvilla-Rivero and Gil-Parejo (2004) studied price convergence for a set of consumer-price indices across 12 EU countries. The authors use an Augmented Dickey-Fuller regression to test for convergence. They find evidence for convergence of traded goods, however for non-traded goods they did not. Particularly for countries that had stable bilateral exchange rates, the speed of convergence was higher. Sosvilla-Rivero and Gil-Parejo (2008) also use the coefficient

[1] For more information on price transmission in the dairy industry see the working paper of Bonnet et al. (2015)

of variation (CV) to examine price convergence in the EU car market and they find clear evidence for price convergence. Goldberg and Verboven (2005) estimate convergence to the LOP, also for the European car market. In their study they use a hedonic pricing mechanism to estimate convergence for quality adjusted prices. They find strong evidence for convergence between car prices across EU MS. Rogers (2007), finds that price dispersion for traded goods declined remarkably in the period 1990-2004. According this study, the price dispersion in the EU was now close to the price dispersion in the USA. Especially, the price dispersion for the 11 EMU-members in 1998 was not significantly different from price dispersion in the USA. This means that the creation of an internal market has at least resulted in a low level of price dispersion, which is comparable to a large federal state. Fischer (2012) estimate price convergence for a specific good market: washing machines. In this study it is emphasized that using aggregated relative price level measures may not be accurate because they are not homogenous and comparable. For that reason, they use washing machines, since these products are traded on a large scale, are comparable due to detailed characteristics and brands, and are non-perishable. They find no evidence for price convergence in the EU washing machine market, also not when distinguishing between clubs of countries. Given the highly tradable character of the washing machines market, they doubt whether there is evidence for price convergence on an aggregate level.

4.3 EMPIRICAL METHODS

4.3.1. Overview of empirical methods

As can be seen in the previous section, many studies have examined price convergence. However, the theoretical framework and empirical approaches were different for the studies. The absence of a strong single theoretical framework for price convergence has resulted in multiple empirical measures of convergence. First there is a large share of studies that assesses price convergence with the help of panel unit root analysis. Parsley and Wei (1996), Sosvilla-Rivero and Gil-Parejo (2004), Goldberg and Verboven (2005), Fischer (2012) all used the panel unit root approach. This method uses panel data to test whether prices converge to each other. This method is common-used in price convergence analysis, and it is also a common-used method in income convergence analysis (Islam, 2003). Although it must be mentioned that it is not a perfect measure for convergence. As Dreger et al. (2007) point out that although while price differences are stationary, price dispersion can still increase. However, this method can still show whether price differentials between countries have the tendency to go back to zero.

Next to that the CV is used by Sosvilla-Rivera and Gil-Parejo (2008), this measure can easily calculate the price dispersion over time, since it is the standard deviation over the mean. It can be used to conduct convergence tests, based on OLS regressions.

Rogers (2007) uses an F-test to test whether price dispersion increases over time. Basically, he calculated the variance of several price indexes between cities. After that, he uses an F-test to test whether the variance of period $t+1$ is smaller than the variance of period t. In this way, it can be tricky to assess convergence with only one product, since market circumstances in two periods can be quite different.

Fischer (2012) uses also an alternative approach by using the convergence test of Philips and Sul (2007) to test for convergence. This is a recently developed test which uses a simple regression based on a one-sides t-test for convergence. It can be used with panel data, and it does not rely on assumptions of stationarity. Still, there exists only a small amount of studies on this empirical measure of convergence.

Table 4.1. Empirical methods of price convergence

Measure	*Visual/Quantitative*	*Main characteristics*
Unit root analysis	Quantitative	Convergence in the short-run dynamics. Uses the panel-data aspect in price series.
Coefficient of variation	Quantitative /Visual	Unit-free measure, based on a cross-section of data in a specific year. Regression tests can be used to view the long-run development of this variation measure.
Philips-Sul method	Quantitative	Novel approach in convergence tests. Uses the panel-data aspect in price series. Able to identify convergence clusters.
Rogers F-test	Quantitative	Based on an F-test of two moments in time.

(Source: author)

In this chapter we use the coefficient of variation as a measure for convergence, because it is a straightforward measure which can show the long-run development of price dispersion and it is considered adequate for our purpose. Unit root analysis is not suitable since this measure makes use of the short-term price dynamics to test for convergence, while our particular interest lies in the long-run. Philips-Sul method has the disadvantage that the empirical applications are limited, which makes it hard to cross-check results. The Rogers F-test is a too sensitive measure since it only uses two points in time.

4.4 THE EMPIRICAL MODEL AND DATA

4.4.1. Farm –gate milk prices and data

We start out with an analysis of the raw milk price in € per 100kg, as provided by the Milk Market Observatory (MMO) (2017a). In this analysis, data of the MMO is used to assess price convergence. The MMO is a European institution that tracks the price developments with respect to milk products. Price data is available only at a national level, so in this part our focus is on convergence between the MS. The prices are monthly data which goes back to 1977. In this analysis we use only the time series after 1996, because since that year all OMS had price data available. The NMS that accessed in 2004, had price data since 2003, Hungary and Czech Republic even since 2002. Bulgaria has data from 2007 onwards, Romania from 2009 onwards. Since we divide the group in several groups of MS, we exclude Bulgaria (2007, 2008), Hungary (2002) and Czech Republic (2002) from our analysis. This loss of information is a pity, but it improves the consistency of the analysis. Since Malta had only data from 2011 onwards, and Croatia from 2013 onwards, we exclude these two countries from the analysis.
In Table 4.2, it can be seen for which groups and time frames the CV is calculated. First, we calculate the CV for three groups (OMS, NMS (2004), NMS (2004+2007)) and visually inspect the CV over time.
Subsequently, for four specific groups the CV is calculated and a convergence test is applied to these groups. It is chosen to make a distinction between large and small producing countries to see whether there is a difference in the convergence process. It could be that the largest producing countries have more complete mechanisms to transmit price shocks. For example, for large countries a demand shock like the Russian trade ban can be moderated because they have multiple export destinations to which they can shift their export. The selection between large and small producing countries is based on the total collection of cow's milk per MS in 2005 (Eurostat, 2017). Countries with a higher than median collection of cow's milk are considered as large, the others small. A remarkable statistic that can be observed in Table 4.2 is the group of the 12 largest producing countries consist of 11 OMS and Poland as the only NMS. This is in line with the findings of Ihle et al. (2017, p.61), who find that the OMS account for 86% of the total milk delivered to dairies in the EU.

Table 4.2. Summary table of estimation of CV

Group of countries	*Countries*	*Time period*	*Number of countries*	*CV calculation*	*Convergence test*
OMS	BE, EL, PT, ES, LU, FR, DK, DE, IT, NL, FI, AT, SE, IE, UK	1996-2017	15	X	
NMS (2004)	CZ, EE, CY, LV, LT, HU, SI, SK, PL	2003-2017	9	X	
NMS (2004+2007)	CZ, EE, CY, LV, LT, HU, SI, SK, PL +BG, RO	2009-2017	11	X	
OMS+NMS(2004)	BE, EL, PT, ES, LU, FR, DK, DE, IT, NL, FI, AT, SE, IE, UK + CZ, EE, CY, LV, LT, HU, SI, SK, PL	2003-2017	24	X	X
OMS+NMS (2004+2007)	BE, EL, PT, ES, LU, FR, DK, DE, IT, NL, FI, AT, SE, IE, UK + CZ, EE, CY, LV, LT, HU, SI, SK, PL+ BG, RO	2009-2017	26	X	X
12 largest producers	BE, ES, FR, DK, DE, IT, NL, AT, SE, IE, UK, PL	2003-2017	12	X	X
12 smallest producers	EL, PT, LU, FI, CZ, EE, CY, LV, LT, HU, SI, SK,	2003-2017	12	X	X

(Source: author)

4.4.2. The Coefficient of variation

The CV is a measure that can estimate price dispersion. It is defined as the standard deviation divided by the mean at time t (Monfort, 2008). Where the standard deviation σ_t is defined as:

$$\sigma_t = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_{i,t} - \overline{P}_{i,t})} \tag{4.2}$$

With $P_{i,t}$, as the price in euros of country i in month t.

The CV can then be defined as:

$$CV_t = \frac{\sigma_{i,t}}{\overline{P}_{i,t}} \quad (4.3)$$

In which $\overline{P}_{i,t}$ is the average price of all n countries. In Figure 4.2, it can be seen that when the distance between the individual prices and the mean price become smaller ($P_{i,t} - \overline{P}_{i,t}$), the CV decreases. Hence as the individual prices come closer to the mean price, the dispersion decreases. When the dispersion decreases we can speak of σ-convergence.

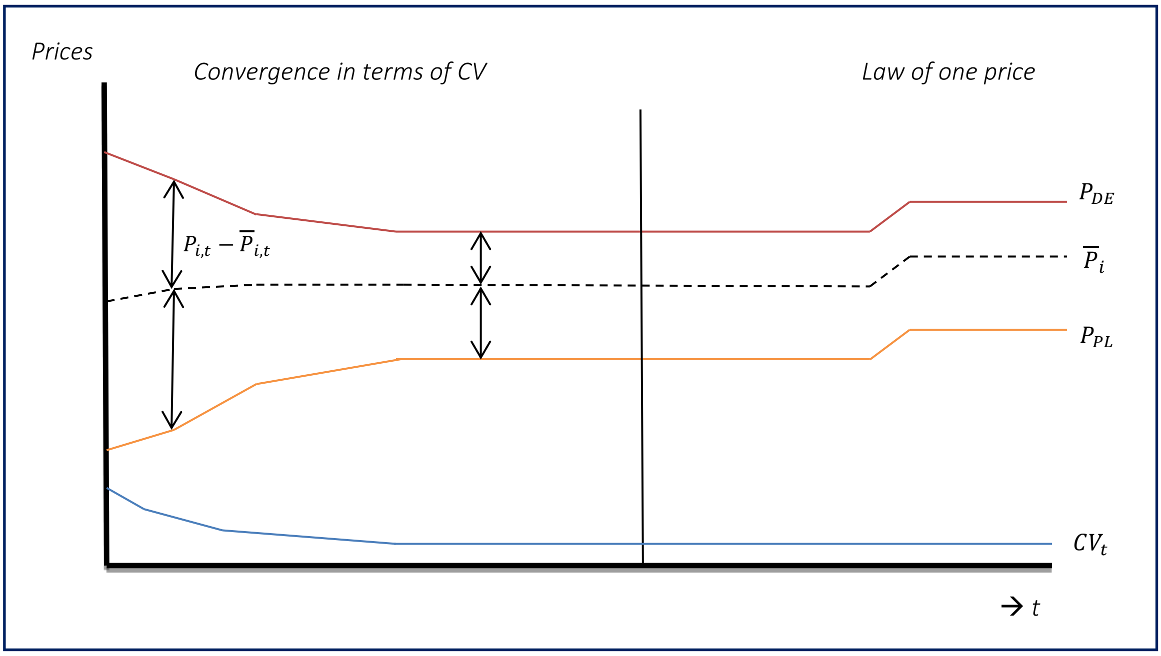


Figure 4.2. Illustration of convergence within the coefficient of variation.
Note: the red line represents an example of the German milk price, the black dashed line represents an example of an average EU milk price, the yellow line represents an example of the Polish milk price and the blue line represents the CV. The arrow indicates the distance between the national milk price and the average milk price, which refers to part of the equation 4.2.
(Source: author)

4.4.3. The convergence test

To measure whether the dispersion has decreased over time we use an OLS-regression to test for σ-convergence as defined in Gil-Pareja and Sosvilla-Rivero (2008). This is a simple regression of CV_t verses a linear time trend t with a constant α and the error term ε_t:

$$CV_t = \alpha + \sigma t + \varepsilon_t \quad (4.4)$$

To test for convergence, a simple one-sided Wald test is conducted on the parameter σ. If the null-hypothesis holds, σ is not significantly different from zero which means that a process of convergence is not present. The alternative hypothesis is that σ is negative, since this would imply a decreasing dispersion over time, hence convergence. Formally:

$$H_0\!: \sigma = 0, \text{no convergence vs. } H_a\!: \sigma < 0, \text{convergence} \quad (4.4)$$

4.4.4. Structural break test

In chapter 2, we have seen that there existed periods of severe price volatility in the past decades. The empirical model can therefore be influenced by the volatility. Hence, volatility can interfere the continuous process of convergence. Therefore, a simple structural break test is applied, to test whether there is a structural break present in the series. We use the method that is developed by Andrews (1993). This test uses a Wald-test to test for a structural break in the parameter. So it tests whether there is parameter stability:

$$H_0 = \beta_0 = \beta_t \text{ for all } t > 1 \tag{4.4}$$

If the null-hypothesis is rejected, there is a structural break present in the series. The alternative hypothesis is that there is structural break point τ. So, suppose there is one single break point in the sample T, then the alternative is written as (Andrews, 1993, p.823):

$$H_{1T}(\tau) = \beta_t = \begin{cases} \beta_1(\tau) \text{for } t = 1, \ldots, T\tau \\ \beta_2(\tau) \text{ for } t = T\tau + 1, \ldots \end{cases} \tag{4.4}$$

It is found that for every equation, that the null-hypothesis of no structural break was rejected at 1%-level[2]. For that reason, we include a dummy variable (D) and an interaction variable $(D \cdot t)$ to account for the structural break:

$$CV_t = \alpha + \sigma t + \gamma D + \beta(D \cdot t) + \varepsilon_t \tag{4.6}$$

Additionally, from visual inspection of the price data we have seen that the price drops in 2009,2015 and 2016 cause the main rise in the CV. From the Milk Market Observatory (2017b) we obtain data about the monthly public intervention in the SMP market, we take this data as a proxy for 'extraordinary' market circumstances. For the months SMP is bought up from the market, we apply the structural break test. In this case for the period March 2009-October 2009 and the period March 2015-December 2016. By running the OLS regression with, the null-hypothesis of no structural break was rejected for both periods for every regression, at 1%-level. To account for this structural break, two dummies $(D_{1,2})$ and two interaction variables $(D_{1,2} \cdot t)$ are added for each period. Formally this model is written as:

$$CV_t = \alpha + \sigma t + \gamma_1 D_1 + \beta_1(D_1 \cdot t) + \gamma_2 D_2 + \beta_2(D_2 \cdot t) + \varepsilon_t \tag{4.6}$$

Additionally, we have applied a log-linear and linear-log functional form. In general the log-linear model showed lower AIC and R^2 values. The linear-log model has shown similar values for the AIC and R^2, but the interpretation of the model is difficult and the dummy- and interaction variables are hardly significant. In this chapter we show therefore only the results of the linear model. The results for the other functional forms can be found in Appendix II.

[2] Results of the structural break test can be found in Appendix II

4.5 Results

4.5.1. Coefficient of variation

In Figure 4.3, the CV for the three different MS sub-groups can be found. In this graph we can identify four interesting patterns in the dispersion. The first pattern is that between 2003 and 2007, the prices of the 9 NMS (Malta excl.) that entered in 2004, came closer to the OMS. At the end of 2007, the difference between the dispersion of OMS and the dispersion of OMS+NMS(2004) is very small (see grey area). It confirms the pattern that was visible in Figure 2.4 in section 2.2.2, in which it could be seen that there was an upward price pattern of the milk price of several NMS between 2004 and 2007.
Secondly, we can see that the inclusion of Bulgaria and Romania did not change the dispersion so much. As third, it can be seen that when there is a price decline in the EU milk market then the dispersion increases (marked yellow). Above that, the gap between the dispersion of OMS and the dispersion of OMS+NMS grows larger when there is a price decline in the market. The last pattern that can be observed is that the increased volatility in the market after 2009, has also affected the price dispersion in the market. Until 2008 price dispersion is quite stable, except for some seasonal variation. After 2008, the dispersion fluctuates also more, probably to non-symmetrical price changes in the national milk prices.

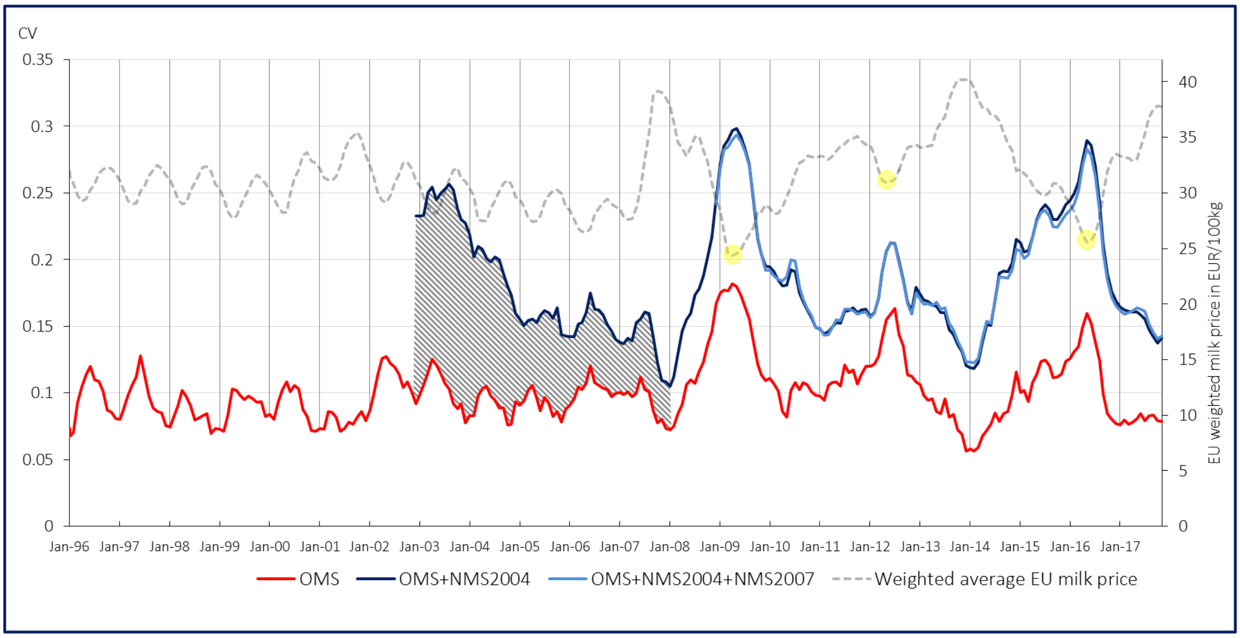


Figure 4.3. Coefficient of variation for 3 MS subgroups between 1996 and 2017
Note: The CV is represented on the left vertical axis. On the right vertical axis the weighted EU raw milk price can be found, this is measured in €/100kg.
(Source: author, based on data from MMO, 2017a)

In Figure 4.4, the CV for the largest and smallest group of dairy countries are shown. It can be seen that there is a wide gap between the group of largest and smallest countries. Again, the CV increases with the large price drops in 2009 and 2015/2016. Remarkable is that the CV of the largest producing countries is quite stable over time. The CV of the largest producing countries is even lower than the CV of the OMS. The prices of the largest producing countries are moving together, resulting in a more or less stable CV. An explanation for this phenomenon could be that there is considerable price transmission between these large markets. Since these countries are major players on the EU- and world market, the price transmission is more complete than for the smaller producing countries. In the middle of 2014, the Russian trade ban came into place, which resulted in much more price dispersion amongst the smallest producers. While the largest producing countries only have a minor increase in the dispersion.

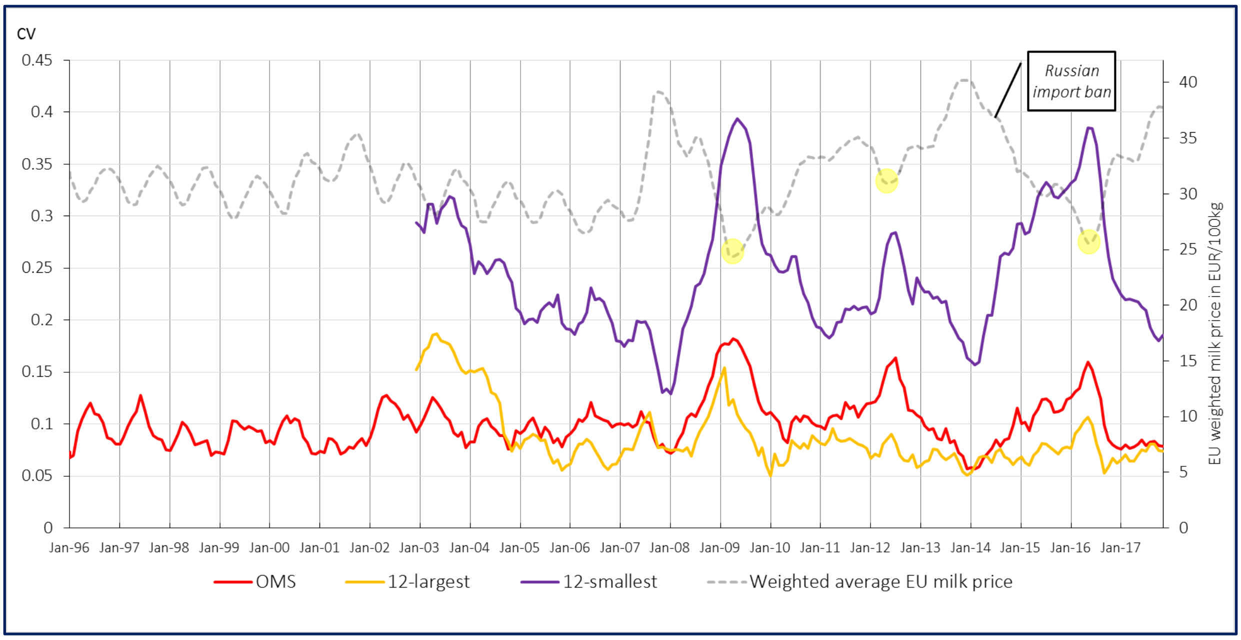


Figure 4.4. Coefficient of variation for 3 subgroups between 1996 and 2017
Note: The CV is represented on the left vertical axis. On the right vertical axis, the weighted EU raw milk price can be found, this is measured in €/100kg.
(Source: author, based on data from MMO, 2017a)

4.5.3. Regression results

In Table 4.3 the results of the linear regression with one dummy variable for the structural break are shown. For all four groups the variable t is significant at 1%-level. From one-sided t-tests[3], we find that the null-hypothesis of $\sigma \geq 0$ was rejected at 1%-level. So, we can expect this coefficient to be negative for all groups of countries, meaning that *ceteris paribus* the CV has decreased over time, hence dispersion decreased, which means that price convergence has occurred. The group of the 12 largest countries have the lowest coefficient for t, which would mean that this group has faced the strongest price convergence process. The interaction variable $t \cdot D$ is significantly different from zero for 3 out of 4 models at 1%-level. This would mean that indeed the time effect on the CV is different after the structural break. For the OMS+NMS(2004&2007), the interaction variable was not significant. This can be explained by the fact that this variable only has values from 2009 onwards. The dummy variable is only significantly different from zero for the 12 largest countries, which indicates at a clear level shift after the structural break. The constant is significant for all four groups of countries. They show reasonable values, with the lowest value for the 12 largest countries (i.e. 0.19333), in Figure 4.4 it was also seen that this variable shows the lowest CV in general.
The model statistics show clearly that the model for the 12 largest countries seems to fit the best, which is reflected in the highest F- and R^2 values and the lowest AIC value.

3 For the p-values of the one-sided test, see Appendix II

Table 4.3. Regression results linear model with one dummy

Variables	$OMS + NMS(2004)$	$OMS + NMS(2004\&2007)$	$12\ largest$	$12\ smallest$
t	-0.00192***	-0.00157***	-0.00364***	-0.00222***
	(0.00)	(0.00)	(0.00)	(0.00)
D	-0.02837*	0.12888***	-0.10247***	-0.02542
	(0.02)	(0.04)	(0.01)	(0.02)
$t \cdot D$	0.00177***	-0.00026	0.00353***	0.00208***
	(0.00)	(0.00)	(0.00)	(0.00)
$constant$	0.23885***	0.23752***	0.19333***	0.29856***
	(0.01)	(0.01)	(0.01)	(0.01)
n	*180*	*108*	*180*	*180*
F	*24.12875*	*29.56878*	*156.56350*	*22.17043*
R^2	*0.29143*	*0.46032*	*0.72742*	*0.27426*
AIC	*-664.63791*	*-428.37694*	*-968.44545*	*-557.49285*

Note: *** significant at 1%-level, ** significant at 5%-level *significant at 10%-level
(Source: author, based on data from MMO, 2017)

The linear regression for the functional form with the two dummies based on the intervention in the SMP, can be seen in Table 4.4. Again for this model, we find that for all groups of countries, the variable t was significantly different from zero and also significantly negative. It follows that *ceteris paribus* the CV has decreased over time, which means that dispersion has decreased. Again, the lowest coefficient for t is for the group of the 12 largest countries, this hints that this group has known the largest convergence process. Looking at the interaction variables, it is found that for the 12 largest countries there is one interaction variable ($\mathrm{t} \cdot D_1$) that is significant. With respect to the dummy variables only the dummy for the first period (D_1) is significant for two models. The dummies and interactions for the second period of volatility is not significant at all. The constants show again reasonable values; the lowest constant for the 12 largest countries, the highest for the 12 smallest countries.
The reported model statistics show that the group of the 12 largest countries again has the lowest AIC values. The highest R^2 is found for the group of OMS+NMS (2004&2007). For three out of four groups of MS, the AIC values for the model with two dummies are lower than for the model with only one dummy. On the other hand, the model with two dummies has shown only a few significant variables.

Table 4.4. Regression results for the linear model with two dummies based on SMP public intervention

Variables	$OMS + NMS(2004)$	$OMS + NMS(2004\&2007)$	$12\ Largest$	$12\ Smallest$
t	-0.00020***	-0.00029***	--0.00039***	-0.00015**
	(0.00)	(0.00)	(0.00)	(0.00)
D_1	0.49370	0.12045***	0.67445**	0.36418
	(0.39)	(0.03)	(0.30)	(0.51)
D_2	0.21152	0.14580	0.03902	0.34296
	(0.24)	(0.10)	(0.18)	(0.31)
$t \cdot D_1$	-0.00491	-0.00394	-0.00832**	-0.00278
	(0.01)	(0.00)	(0.00)	(0.01)
$t \cdot D_2$	-0.00081	-0.00082	-0.00010	-0.00144
	(0.00)	(0.00)	(0.00)	(0.00)
$constant$	0.18950***	0.18682***	0.11899***	0.23947***
	(0.01)	(0.01)	(0.00)	(0.01)
n	*180*	*108*	*180*	*180*
F	*32.48266*	*40.13285*	*22.36224*	*35.83601*
R^2	*0.48278*	*0.66299*	*0.39121*	*0.50733*
AIC	*-717.29906*	*-475.23141*	*-819.80462*	*-623.21737*

Note: *** significant at 1%-level, ** significant at 5%-level *significant at 10%-level
(Source: author, based on data from MMO, 2017)

4.6 DISCUSSION

The regression results show us that there is a significant downward time trend of the CV. Hence, this would mean that the dispersion decreased over time and that price convergence has occurred. However, the results should be interpreted with caution. Plots of the residuals versus the fitted values (see Appendix II) suggest that there might be non-linearity in the model. If we inspect the residuals and the CV together over time, it can also be seen that the residuals increase when there are large peaks in the CV. With OLS regression, according to the Gauss-Markov theorem says that the OLS estimators are the BLUE (best linear unbiased estimators) when the properties of linearity, unbiasedness and minimum variance hold (Gujarati, 1992). It is very unlikely, given these patterns in the residuals that this theorem holds. Additional tests for heteroscedasticity and serial autocorrelation confirm this suspicion. The Durbin-Watson tests with the null-hypothesis of no autocorrelation was rejected for all cases[4]. The Breusch-Pagan test with the null-hypothesis of constant variance was only rejected for some cases. Still, this means that serious problems with respect to the Gauss-Markov assumption remain if there is structural serial autocorrelation present. Further research should therefore be conducted to find a model that is statistically more reliable. Given the time aspect in the model, it might be wise to also look for time-series models. For example time series beta-regression models, which is a model that is particularly suitable for proportions with a time series aspect (Guolo and Varin, 2014).

We can therefore not proof that convergence has taken place over time. Is it surprising that we cannot find a stable convergence path over time? Perhaps not, as was discussed in 2.2.2, the period from 2007-2017 could be seen as an exceptional period of volatility for the milk market. In Figures 4.3 and 4.4, it could be seen they have led to peaks and drops in the CV. In the residual plots it can also be seen that the residuals peak during this period and also that in high values of the fitted CV patterns are observable. It might well be that these periods therefore cause the serial autocorrelation present in the series.

[4] The outcome of the Durbin-Watson and Breusch-Pagan tests can be found in Appendix II.

Above that, there are simply too many external factors that influence the milk prices as we have seen in Figure 4.1. For example, as was pointed out in section 2.2.4, in some countries the dairy industry is organized as cooperatives and farmers have a stronger bargaining position, while in other countries the dairy industry only consists of IOFs. Since these factors are not clearly observable, and we do not account for these factors, it is extremely difficult to identify the 'true' convergence process. For the group of OMS+NMS(2004+2007), there are only observations since 2009, during the large price drop. It is therefore difficult in what terms this group of convergence has experienced 'true' price convergence. Still, the analysis has shown some interesting patterns over time. Between 2004-2007, under stable market conditions, the dispersion between OMS and NMS(2004) has decreased substantially.

5 PRODUCTIVITY CONVERGENCE

In this chapter we test for convergence in productivity. First, it is explained how the concept of convergence can be applied to productivity. After that, an overview of the academic literature on productivity convergence is given. Then, the choice of empirical measures is clarified and the empirical measures themselves are explained. Then we explain the empirical model and show the results.

5.1 THE CONCEPT OF PRODUCTIVITY CONVERGENCE

About productivity, opposite to prices, we can say that the higher the productivity, the better the economic performance of a company. Increasing the productivity vis-à-vis other firms or regions is a way to strengthen the economic position of firm or region (and to achieve convergence). In the economic literature on convergence, a large share is devoted to the convergence of productivity, more specifically Total Factor Productivity (TFP). As we have seen in chapter 3.2, a significant share of the academic literature on convergence was devoted to the convergence of TFP.
In a simple Cobb-Douglas production like:

$$Y = AL^{\alpha}K^{1-\alpha} \tag{5.1}$$

Here A is the Total Factor Productivity, it is a factor that accounts for the output growth that is not clarified by the input-side; labour (L) and capital (K) (Miller and Upadhyay, 2002). In terms of growth, it is also known as the *Solow residual*, the part of output growth that is not explained by input growth.
In recent theoretical advances, Romer (1990) has endogenised technologic growth in his model, while the original Solow-model assumes that technologic growth it is exogenous.
Productivity is a rather theoretical concept in the neoclassical growth model. In dairy farming however, it is easy to make productivity practical, a classical example is the milk yield per cow. The nature of the development of productivity is quite different between the dairy sector and the general economy.
Theory on technological advance has been developed by agricultural economists like Cochrane. Cochrane (1958) pointed out that new technology in farming is first only adopted by a few farms, the frontrunners. This small group of high-tech farms will reduce their per units costs of production by this new technology. They gain extra income by this technology since output prices do not change much. After some time, the majority of the farmers will adopt this new technology. However, this will result in an increase of total output, subsequently by a decrease in output prices. Furthermore, any extra gain in income is capitalized in income, or let us say additional milk quota, which then leads to an increase in the price of land or the price of milk quota. In the end, the average farmer is back at its prior situation, this is what Cochrane calls the *technological treadmill.* Hayami and Ruttan (1971) propose an agricultural development model in which technical change is endogenous. According to the authors, there is an instant capacity of the agricultural sector to improve the agricultural productivity. The authors argue that there are roughly two ways to improve productivity in agriculture: biological innovations and mechanical innovations.
What can be concluded is that the development path of productivity for agriculture is different from the whole economy. It is unlikely that productivity convergence follows a deterministic pathway. Hence, productivity growth is a combination of farmer innovations on the micro-scale combined with larger spill-overs and interactions on the meso- and macro-scale. In this chapter we use again the definition of σ-convergence. However, the distributional approach to σ-convergence is applied in this case. So it is studied how the distribution of the cross-section changes over time. β-convergence is disregarded

since it requires a strong theoretical framework how productivity changes over time and this is simply not present. Quah (1996a) proposed this *distributional approach*, because σ- and β-convergence (regression) tests show only a partial story that is conditioned on the mean. Also the variance measures linked to the σ-convergence, only show one aspect of the cross-section namely the variance (Islam, 2003) By studying the distribution over time, we can identify changes in the *internal* distribution: are there specific regions that are able to climb on the productivity ladder? Moreover, it can also give an outlook on the *external* distribution; what is the shape of the across distribution?; do we observe twin-peaks (clusters)?

5.2 LITERATURE REVIEW

The domain of productivity studies on the dairy sector is quite well represented in the scientific literature. The literature on productivity is very much linked to efficiency analysis. One of the pioneers of efficiency analysis, Kumbhakar published in 1991 together with other scholars a paper on dairy farm inefficiency. Using a Stochastic Frontier Analysis (SFA) on they estimate the efficiency of a sample of U.S. dairy farms. They find that larger farms are relative more technically and allocatively efficient, that they have lower returns to scale and that they are more efficient given the output price. The authors suggest therefore that larger farms are more profitable and that without the price support system, more farms would grow to a large size.

Brümmer et al. (2003) are the first that estimate Total Factor Productivity (TFP) based on an SFA for European dairy farms in three countries. They find that TFP-growth in Poland and Germany was mainly based on technical change, while in the Netherlands the allocative change was more important for TFP-growth. The authors therefore argue that measures to stimulate productivity could be differentiated, because for some farmers the allocation might hinder productivity growth while for other farmers it is the technology that hampers growth. Jansik et al. (2014) provide a detailed cross-country study for the Nordic and Baltic countries. In their study, they measure the actual competitiveness of the dairy sector with the help of several indicators, including productivity. A methodological outcome of this study was that it is hard to find a good measure to compare cross- country competitiveness. With respect to productivity they also used two measures: partial productivity measures and Total Factor Productivity. The latter involves complete productivity measures based on inputs and outputs, but is methodologically complex, more sensitive to measurement error and it can lead to ambiguous outcomes (Matthews, 2014). Partial productivity measures are easier to calculate and understand, but that is at the same time its weakness, since they provide an oversimplified image. Jansik et al. (2014) found that the initial competitive position did not change much over the years. The dairy sector in Denmark stayed the most productive, while only Estonia showed a remarkable catch-up. Next to that, they found that with respect to the productivity of the processing industry, the Baltic countries were able to catch up with the 'old' Nordic EU members. Baráth and Ferto (2017) provide a study on TFP-convergence between the agricultural sectors of the EU. Their results show that TFP of the agricultural sector has somewhat decreased between 2004 and 2013. In terms of TFP-levels there was little change in ranks between countries. At last they find that the TFP of the different countries are converging, although this convergence is slow. The authors point out that their method is based on the assumption that production inputs and outputs are homogenous and that they have a common production function. For the reason that they aggregate data of many different countries, it is hard to account for all the factors that might undermine this assumption.

At last, the paper of Cechura et al. (2017) gives a very detailed study of convergence of the dairy sectors of the EU. They use a stochastic meta-frontier multiple-output distance function to estimate TFP-growth of dairy farms across 24 EU Member States at NUTS2-level between 2004 and 2011. They find that the north-western part of the Old Member States have the highest TFP growth rates. They find no evidence that the poor performing regions are catching up with the better performing regions. Their results on technical change put forward that dairy farmers in Central and Eastern EU MS have sub-optimal farm sizes. They argue that farms in those regions have problems to structurally change. They think that larger farms may be better in adapting new innovations and technology, since their results indicate that regions with large-scale farming are performing better with respect to technical change.

5.3 EMPIRICAL METHODS

As opposed to the empirical methods of price convergence, measuring productivity involves two empirical choices. First, the question is how to measure productivity. There exist several methods to estimate/calculate productivity. The second step that should be taken is: how to measure convergence? Starting with the first question, there exist generally three ways to estimate productivity: Data Envelopment Analysis (DEA), Stochastic Frontier Analysis (SFA), and partial productivity measures.

The method of DEA is a non-parametric method that can measure productivity of a decision making unit (Ray, 2004). With the DEA method, you can measure Total Factor Productivity on which you can conduct convergence tests. When multilateral comparisons are made, it is crucial to have transitive productivity measures. For example, when country i produces 10% more than country j, and country j produces 30% more than country k, then country i should produce 1.1x1.3=1.43 more than country k. To obtain those transitive measures, you have to calculate input and output price and quantity index numbers for each pair of firms (or regions). The advice in multilateral comparisons is to compute multilateral Törnqvist indices based on the Elteto-Koves-Szulc (EKS) method. Then to find convergence one can use panel unit root tests like Baráth and Fertő (2017) did.

Opposite to DEA, SFA is a parametric technique to measure productivity. This technique uses output distance functions to measure productivity. The flexible trans-log functional form is mostly used to estimate productivity. Then a Törnqvist-Theil Index is used to compute TFP-measures that are suitable for multilateral comparisons (Caves et al., 1982). An advantage of the SFA method is that it is possible to see where the Total Factor Productivity growth consists of. Like Cechura et al. (2017), you can decompose TFP in a scale effect, a technical efficiency effect, a technological change effect and a heterogeneity effect. Especially the last effect would be a very interesting component for the case of the dairy sector.

Partial productivity measures are straightforward to calculate, it is a measure of only two production factors. It is simple, because it only shows the productivity with respect to one input and output, like milk yield per cow. Partial productivity measures will not give a good understanding of the overall productivity, since not all inputs and outputs are taken into account. For that reason, it can be that milk yields per cow are very high, but that the feed per cow is also very high (Jansik and Irz, 2014). Examples of partial productivity measures for dairy farms could be milk yield per cow, workers per cow or cows per hectare.

In this thesis, it is decided to use partial productivity measures. Although DEA and SFA can be good methods to calculate TFP, they require a lot of data. Given the structural differences between the different regions, it is also hard to assume one and the same production function for all the regions. Diewert (2002) makes clear that to accurately measure the TFP, data on the price and quantity of outputs *and* data on the price, quantity and quality of inputs are required. Specifically, the quality aspect of inputs cannot be measured, while there are several reasons to suspect significant differences in quality between countries. Partial productivity measures require fewer assumptions, can enlighten several aspects of dairy farming and are easy to calculate. However, they cannot estimate overall productivity, we decide that it is better to have multiple targeted productivity measures than one TFP measure that is problematic given the required assumptions.

Table 5.1. Empirical methods of price convergence

Empirical method	*Advantages*	*Disadvantages*
Data Envelopment Analysis	-Transitive measures -Multilateral comparisons possible -Multi-output vs. multi-inputs	-Time- consuming -Assumes single production function -Requires a lot of data
Stochastic Frontier Analysis	-Multilateral comparisons possible -Can be decomposed in several effects -Multi-input – multi-output	-Time- and data- consuming -Assumes single production function -Requires a lot of data
Partial productivity measures	-Easy to calculate -Straightforward interpretation -One output vs. one input	-Does not show overall productivity

(Source: author)

The second choice to make is which empirical method should be to test convergence. As we have chosen to measure productivity with partial productivity measures, the convergence test that are related to the TFP are not suitable. Instead we might have to focus on alternative methods to measure convergence. There exist several methods to analyse convergence across regions. None of the measures can give a complete and exact measure of convergence. Each measure has its advantages and drawbacks. Monfort (2008) provides an excellent overview of the measures that are available to assess convergence, as can be seen in Table 5.2. These measures are in particular used for convergence of GDP per capita across regions, but can also be applied to measures of productivity.

Table 5.2. Measures of inequality and convergence

	Measure	*Visual/Quantitative*	*Range*	*Main characteristics*
Beta-convergence	Beta-coefficient	Quantitative	0- ∞	Estimated rather than computed
Sigma-convergence	Coefficient of variation (CV)	Quantitative	0- 1	Sensitive to changes in the mean, in particular when the mean value is near zero
	Gini index	Quantitative	0- 1	Sensitive to changes in inequality around the median/mode
	Atkinson index	Quantitative	0- 1	Weight given to gaps between incomes in lower or upper tail of the distribution parameterised through the "aversion to inequality".
	Theil index	Quantitative	0- ∞	Gives equal weights across the distribution.
	Mean Logarithmic Deviation	Quantitative	0- ∞	Gives more weight to gaps between incomes in the lower tail of the distribution.
Analysis of distribution	Salter graphs	Visual	-	No possibility of statistical inference. Possibility of identifying individual regions.
	Markov chain analysis	Quantitative	0-1	Possibility of statistical inference and of identifying individual regions.
	Kernel estimation	Visual	-	No possibility of statistical inference. No possibility of identifying individual regions.
	Cumulative frequency	Visual	-	No possibility of statistical inference. No possibility of identifying individual regions

(Source: author, based on Monfort, 2008, p. 20)

The particular interest of productivity convergence is to find whether regions with low productivity are catching up with the most productive regions. As defined in section 5.1 the focus in this chapter is on the analysis of the distribution. This approach requires less theoretical assumptions and can still provide useful information about productivity growth. Kernel density plot estimation is a useful tool to see whether bimodality exists in the distribution. It could show whether there exist several 'clubs' or 'groups' of regions that have the same productivity. It can therefore show us the dynamics of the *external distribution.* To also statistically assess convergence, Markov chain analysis is applied. The Markov chain analysis seems a useful empirical model since it has the possibility of statistical inference *and* it can particularly show the dynamics of the *internal* distribution.

5.4 THE EMPIRICAL MODEL AND DATA

5.4.1. Partial productivity measures and data

To measure partial productivity, we first have to choose what inputs versus what inputs we use. Like Jansik and Irz (2015b) two measures for labour productivity are applied: output per dairy cow and cows per worker. These two measures also correspond with the mechanical and biological innovation path as was explained in section 5.1. The first one can reflect innovations in a biological sense, like breeding and genetics or feed input to improve milk yield. The second one can show underlying growth in the mechanical sense, like milking machines, feed robots or tractors. Together they form an identity for labour productivity:

$$\frac{Yield}{Labour} = \frac{Yield}{Cows} \cdot \frac{Cows}{Labour} \tag{5.2}$$

$$Biological\ productivity = \frac{Milk\ yield\ (kg)}{Number\ of\ cows} \tag{5.3}$$

$$Mechanical\ productivity = \frac{Number\ of\ cows}{Labour\ input} \tag{5.4}$$

The FADN database provides several input and output variables from which partial productivity indices can be constructed. The dataset that is used is a dataset from 2004-2015, with year, country, region and TF14 classification. So with this database we have aggregated data on regional FADN-level that allows to only select the specialist dairy farms (TF:45). For the biological productivity measure, we use the FADN variable: SE125 Milk Yield. It is defined as *the average production of milk and milk products (in milk equivalents) per dairy cow*. For the mechanical productivity measure, we divide: SE085 Dairy cows by SE010 Total labour input. SE085 *includes female bovine animals (including female buffaloes) which have calved and are primarily held for milk production for human consumption, cull dairy cows excluded*. SE010 *is defined as total labour input expressed in Annual Working Unit (AWU) (AWU=full-time person equivalent)*. (FADN, 2018b)

STATA 14.1 is used to estimate Kernel density plots. The R-package *markovchain* is used to estimate the transition probability matrices.[5]

5.4.2. Kernel density estimation

Kernel density estimation is a non-parametric technique to show the density of a distribution. Like a histogram it can show how a certain variable is distributed, but then in a smooth way without sub-intervals (Monfort, 2008). The Kernel density estimator of a series X with a specific point x can be defined as (Silverman, 1986, p.4):

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x - X_i}{h}\right) \tag{5.5}$$

n is the number of observations, k the kernel function, and h is the smoothing parameter. Like Fingleton and López-Bazo (2003), Quah (1997), Hansen and Teuber (2011) the Gaussian Kernel function (K) is used in the estimation. The optimal bandwidth h is based on the paper Silverman (1986). If σ-convergence would occur, we would expect that the spread of the distribution becomes smaller over time.

[5] For more info: https://cran.r-project.org/web/packages/markovchain/vignettes/an_introduction_to_markovchain_package.pdf

5.4.3. The Markov chain[6]

The Markov chain is a model that was developed by the Russian mathematician Andrej Markov. The model shows a system of several 'states' in which it is possible to move from one state to the other state over time. An important property of the Markov chain is that it has no memory, i.e. that the future steps in the system from the current do not depend on the past. An example of a Markov chain can be seen in Figure 5.1. This Markov chain has three states, so an entity can only be at one state at one time. However, over time it can move to all states in the system. The special property of the Markov chain in Figure 5.1 it that all states can be reached from all other states within a finite time.

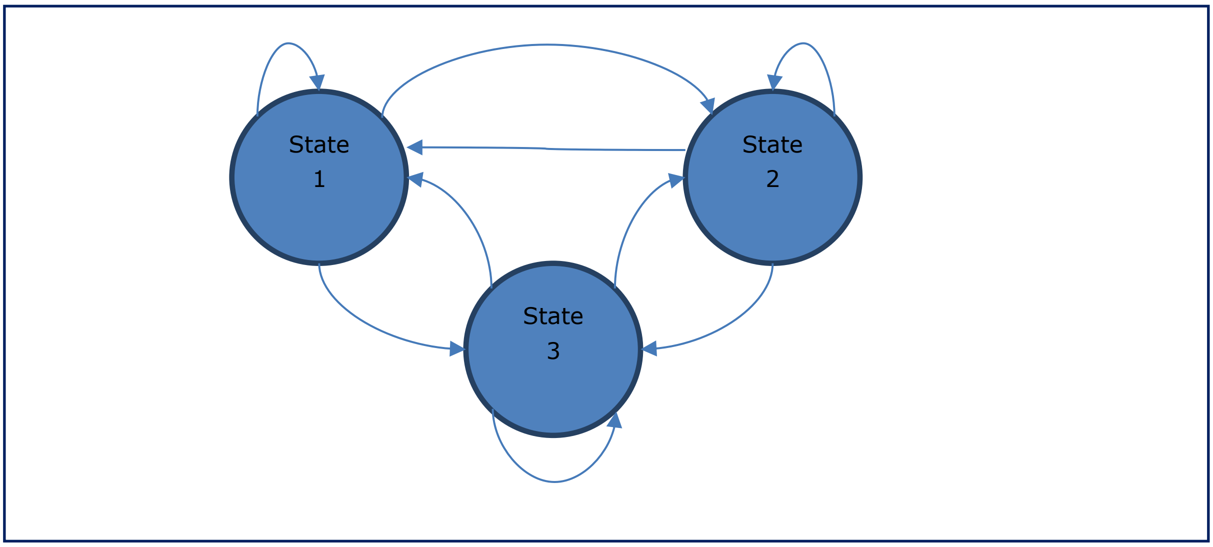


Figure 5.1. An example of a Markov chain with 3 states
(Source: author)

The transition probability matrix $\mathbf{P}$, can show is what the probabilities are to move from one state to the other state. p_{11}, is the probability that being at state one at time t, you will stay in state 1 at time $t+1$. p_{13}, is the probability that you will move from state 1 to state 3, and p_{31} is the probability that from state 3 you will move to state 1.

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \tag{5.6}$$

This matrix $\mathbf{P}$ is our object of study, because our interest lies in the part that tells us what the probability is that entities relocate to certain states. In this case, what the probability is that a region that falls within a certain productivity class, is able to move up our down certain productivity classes.
If there is an initial distribution of for example a population with five classes $\pi^t(\pi_1^t, \pi_2^t, \dots, \pi_5^t)$, then the evolution of the distribution can be described by the transition matrix $\mathbf{P}$:

$$\pi^{t+1} = \mathbf{P} \cdot \pi^t \tag{5.7}$$

To estimate these transition probabilities, data is required that allows us to derive these transition probabilities p_{ij}. Generally, two types of data are distinguished in Markov chain studies: micro- and macro-data. Micro-data is data that can show us the movements of the entities from one state to

[6] For more information on the properties of Markov chains I recommend the reader the first two chapters of the book of Lawler (2006)

another over time (Zimmermann et al., 2009). While macro-data can only show the number of entities in each state at time t, so the individual movements of the entities between the states cannot be observed. In this case, micro-data is available, so we limit ourselves to the estimation procedure of micro-data. Anderson and Goodman (1957) have shown a Maximum Likelihood Estimation (MLE) procedure to estimate the transition probabilities[7]. We define $n_{ij}(t)$ as the number of individuals in state i at $t-1$ and j at time t $(i,j = 1, \dots, m;\ t = 1, \dots, T)$. $n_{i(0),i(1),\dots,i(t)}$ is the number of individuals for which the sequence of states is $i(0), i(1), \dots, i(T)$. We assume the transition probabilities p_{ij} to be stationary (or called homogenous over time), meaning that they do not change over time. To estimate the transition probabilities, we have to imply two restrictions (Anderson and Goodman, 1957, p.92):

$$p_{ij} \geq 0 \text{ and } \sum_{j=1}^{m} p_{ij} = 1 \qquad i = 1,2, \dots, m \tag{5.8}$$

These restrictions in equation 5.8 imply that transition must be non-negative, hence only positive probabilities exist. The second restriction says that the sum of each probability from state $j = 1$ to the states $1, \dots, m$ should add up to one. This makes sense, since the number of movements can never exceed the number of individuals in the system. Then the Maximum Likelihood estimator for p_{ij} can be defined as (Anderson and Goodman, 1957, p.92):

$$\hat{p}_{ij} = n_{ij}/n_i^* = \sum_{t=1}^{T} n_{ij}(t) \Big/ \sum_{k=1}^{m}\sum_{t=1}^{T} n_{ik}(t) = \sum_{t=1}^{T} n_{ij}(t) \Big/ \sum_{t=0}^{T-1} n_i(t) \tag{5.9}$$

This holds for the i-th sample $(i = 1,2, \dots, m)$ that consists of $n_i^* = \Sigma_j n_{ij}$ multinomial trials with the probabilities $p_{ij}(i,j = 1,2, \dots, m)$.
From the transition matrix, several statistics on convergence can be conducted.
The first one is to find a stationary distribution. If the transition matrix $\mathbf{P}$ is stationary (or called homogenous over time) it means that all transition probabilities are equal over time (Anderson and Goodman, 1957, p.92):

$$p_{ij}(t) = p_{ij} \quad \text{for all } t \tag{5.10}$$

If the transition matrix $\mathbf{P}$ is of an *irreducible* and *aperiodic* Markov chain and it is homogenous over time it means that the Markov chain moves towards a steady state. It is then possible to estimate the stationary distribution, which shows the probability distribution in the steady state. This stationary distribution $\bar{\pi}$ is a probability distribution that does not change when time progresses. Formally, we can define this stationary distribution $\bar{\pi}$ (also called invariant probability vector) as Lawler (2006, p.22):

$$\bar{\pi}\mathbf{P} = \bar{\pi} \tag{5.11}$$

This distribution is not a prediction for the future, since the circumstances for one-time period may be completely different from another time period. Rather the stationary distribution characterizes the process of the past period.

[7] For more explanation regarding the statistical inference of Markov chains see: Anderson, T. W., and Goodman, L. A. (1957).

Secondly, the half-life of a transition matrix can be calculated. The half-life is the number of periods it takes to close half of the gap towards the stationary distribution. The half-life can be calculated as (Shorrocks, 1978, p.1021):

$$\text{Half-life} = \frac{-\log 2}{\log|\lambda_2|} \tag{5.12}$$

λ_2 is the second-to-largest eigenvalue of the transition matrix $\mathbf{P}$. The half-life is value between zero and infinity. If it is zero, it means that the stationary distribution has already been reached.

A third measure is the mobility index (M^{OV}), which is an indicator of the degree of mobility in the distribution. So it is a measure that indicates what the overall likelihood of the transition matrix is to remain in a certain state. If there would be no mobility in the distribution, it would mean that all probabilities along the diagonal are equal to one, hence it would be the identity matrix of $\mathbf{P}$. Then the mobility index is equal to zero. . For the case of perfect mobility, we assume a quasi-maximal diagonal for $\mathbf{P}$ ($\text{there exists a positive } \mu_1, \ldots, \mu_n \textit{ such that } \mu_j p_{jj} \geq \mu_i p_{ij} \text{ for all } i, j$) (Shorrocks, 1978, p. 1017). This means that the probability to remain in the same class is not less than the probability to move to another class. So in case of perfect mobility, the trace of the probability matrix is then equal to one, consequently the mobility index will be one. This measure proposed by Shorrocks (1978, p.1017) is defined in Equation 5.13. Where $tr(\mathbf{P})$ is the trace of the transition matrix $\mathbf{P}$ and n is the number of classes.

$$M^{OV} = [n - tr(\mathbf{P})] \cdot [n-1]^{-1} \tag{5.13}$$

Additionally, two mobility measures are calculated to interpret the direction of the mobility: upward or downward mobility. These indicators can be seen as shares of the overall mobility indicator. The sum of the upper triangle of transition probabilities represent the upward mobility ($\sum_i \sum_{j>i} p_{ij}$), while the sum of the lower triangle of transition probabilities represent the downward mobility ($\sum_i \sum_{i>j} p_{ij}$). The mobility of the diagonal element is defined as $\sum_j (1 - p_{jj})$. The upward and downward mobility are 'deflated' by the sum of this diagonal element and therefore $M^U + M^D = 1$.

If there is no downward mobility and no persistence along the diagonal, it would mean that there is perfect upward mobility and M^U would be equal to one. If there is no upward mobility and no persistence along the diagonal, it would mean that there is perfect downward mobility and M^D would be equal to one. The upward mobility can therefore be defined as (Jongeneel and Huettel, 2011, p.513):

$$M^U = \left[\sum_i \sum_{j>i} p_{ij}\right] \cdot \left[\sum_j (1 - p_{jj})\right]^{-1} \tag{5.14}$$

And the downward mobility is defined as (Jongeneel and Huettel, 2011, p.513):

$$M^D = \left[\sum_i \sum_{j<i} p_{ij}\right] \cdot \left[\sum_j (1 - p_{jj})\right]^{-1} \tag{5.15}$$

If a process of convergence would happen, it is expected that there are substantial probabilities to move from the lowest category to higher categories (marked green) in the transition matrix $\mathbf{P}$. On the other hand, the transition probabilities to move from middle categories to lower categories should not be too high (marked pink). If the probabilities to stay at the tails of the distribution (marked red) are much higher than the other probabilities along the diagonal, there is a high risk that the distribution moves towards a twin-peak distribution, so that a persistent structural gap between two groups will exist. If a process of convergence would occur, we would also expect the stationary distribution to have higher shares in the three middle categories. Hence, this would mean that regions are more closely located to the average, so then the dispersion decreasing, meaning that σ-convergence takes place.

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} & p_{15} \\ p_{21} & p_{22} & p_{23} & p_{24} & p_{25} \\ p_{31} & p_{32} & p_{33} & p_{34} & p_{35} \\ p_{41} & p_{42} & p_{43} & p_{44} & p_{45} \\ p_{51} & p_{52} & p_{53} & p_{54} & p_{55} \end{bmatrix} \tag{5.16}$$

For this study we divide the FADN regions into specific productivity classes. To identify these classes, we average the productivity numbers of the regions by the FADN-average. This has two advantages: fluctuations in the whole dairy market are eliminated (like EU crisis years) and it makes it easier to compare the regions with each other. The choice of the number and width of the classes is rather sensitive. The higher the number of classes, the better the density of the distribution is approximated, but it comes with less reliable transition probabilities. The lower the number of classes, the rougher the distribution is division and the less information is abstracted from the distribution (Geppert and Stephan, 2008)
We chose to divide the sample in five classes, with equal steps like Monfort (2008) and Pellegrini (2002) did. In total the sample is sub-divided in five classes as can be seen in Table 5.3. As could be seen, the least productive regions are the 'laggards', the regions that lag behind on a European scale. While at the other end, the 'frontrunners' are a group of regions that are leading in terms of productivity or income.

Table 5.3. Class division and initial distribution for the productivity measures in 2007

Class	***<75% EU-average*** *Laggard*	***75-90% EU-average*** *Low-productive*	***90-105% EU-average*** *Average*	***105-120% EU-average*** *High-productive*	***>120% EU-average*** *Frontrunner*
Biological productivity	14%	13%	27%	23%	22%
Class	***<50% EU-average*** *Laggard*	***50-80% EU-average*** *Low-productive*	***80-110% EU-average*** *Average*	***110-140% EU-average*** *High-productive*	***>140% EU-average*** *Frontrunner*
Mechanical productivity	24%	8%	14%	25%	29%

(Source: author)

In total we include all regions that appear consecutively in the sample. This assures that these regions at least show stable movements over time. Some countries only have data for some years, but we only consider regions that are consecutively present in the dataset. This ensures that the mean value is not influenced by regions that appear only for one or two years. If for example, in 2008 a region shows up hat has a high productivity value, it could move the mean value for the group up, which then leads to movements between classes which are only the result of the region being in the sample.
Since Bulgaria and Romania joined in 2007, these countries also only have data from 2007 onwards. Therefore the empirical model is conducted for two time periods. First for all the regions that have data

from 2004 onwards. Secondly, for all the regions that have data from 2007 onwards. It was chosen to not take into account OMS regions that have consecutive data from for example 2006 onwards. This to make sure that we can clearly see what the effect on convergence is if we add the NMS(2007) regions.

5.5 Results

5.5.1. Biological productivity (Milk yield per cow)

As can be seen in Figure 5.2, there is a considerable variation in the biological productivity. The three dark dashed lines represent the density for the 84 regions between 2004 and 2006. There is no clear pattern visible in the distribution of the productivity for these years. During the period 2007-2015, bimodality appears in the distribution. In 2015, there is a clear peak around 3000 kg, and a second peak at 7500 kg. A second process is observable in the second peak, which is slightly shifting to the right over time.

In terms of convergence this means that we can speak of a status-quo or divergence. If convergence would occur then the left peak should shift more and more to the right over time. What actually happens is that the bimodality in the distribution has grown over time, and that the right peak has moved further away from the left peak. Hence there is a persistent gap between most of the regions with a small group of lower productive regions. Moreover, the spread of the distribution has increased, for the year 2015, we observe the widest spread of all years (see black dashed lines). So dispersion decreased, and we cannot find evidence for convergence in this plot.

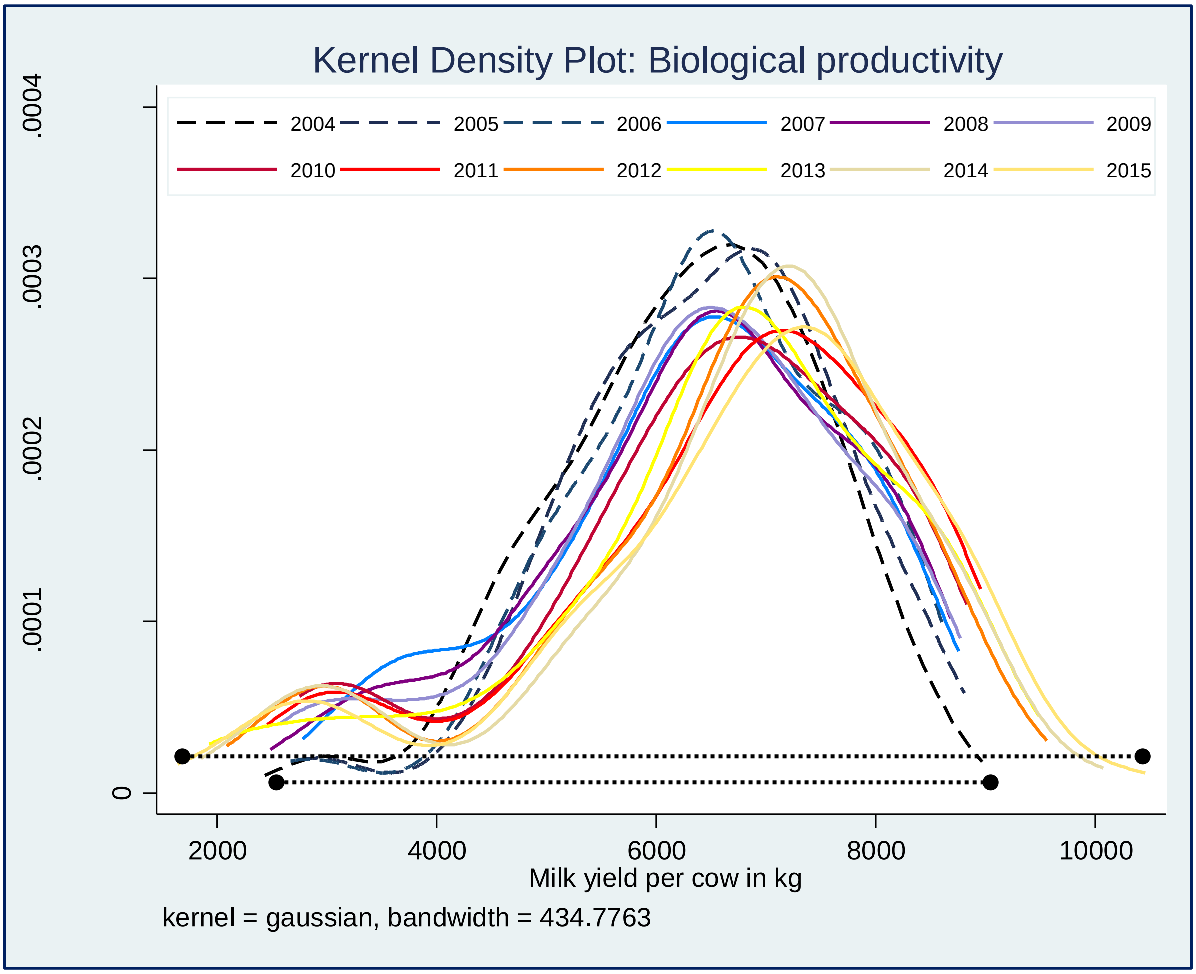


Figure 5.2. Kernel density biological productivity (milk yield per cow), for 91 FADN regions from 2004-2015
Note: the black dashed lines illustrate the difference in spread between the 2004-2007 period and the 2013-2015.
(Source: author, based on FADN data)

Looking at the transition matrix of the regions between 2004 and 2015 in Table 5.4, several things stand out. First the transition probability to remain in the highest productivity is quite high, 0.862 respectively. At the lower productivity classes there seems to be slightly more mobility between the classes. For example, a region is in the lowest productivity class has a probability of 0.167 to move to a higher class. This would imply that convergence is possible, because there is a probability of upward mobility in the lower classes. The upward mobility is also higher than downward mobility. In general, the overall mobility is very low (0.234), which means that the overall probability to move from one class to the other is low. The stationary distribution shows that most of the regions stay in the middle classes.

Table 5.4. Transition probability matrix of biological productivity (milk yield per cow) for 84 FADN regions from 2004-2015

Transition probability matrix						
Initial distribution (2004)	*Class*	*<75*	*75-90*	*90-105*	*105-120*	*>120*
8	*<75*	0.833	0.154	0.000	0.000	0.013
15	*75-90*	0.070	0.790	0.140	0.000	0.000
25	*90-105*	0.006	0.055	0.820	0.113	0.006
25	*105-120*	0.000	0.000	0.150	0.760	0.090
11	*>120*	0.007	0.000	0.007	0.124	0.862
Summary statistics						
Class		*<75*	*75-90*	*90-105*	*105-120*	*>120*
Stationary distribution		0.082	0.146	0.331	0.252	0.188
Half-life		7.877				
Mobility index		0.234				
Upward mobility		0.552				
Downward mobility		0.448				

Note: EU-sample average=100, estimates are rounded to three decimals. Standard errors can be found in Appendix III.
(Source: author, based on FADN data)

If we add the regions that accessed in 2007, the transition probabilities change considerably. In Table 5.5, it can be seen that the probability that being in *class <75* and remaining in *class <75* is now 0.929. This means that adding the NMS (2007) regions, changes the mobility in lower classes. The stationary distribution shows that there is also a larger share in the lowest class (0.109).

It is evident from the two matrices that adding the NMS (2007) regions decreases the overall mobility. The convergence process towards the stationary distribution is also slower, which is reflected in the higher half-life, which is twice as high.

Table 5.5. Transition probability matrix of biological productivity (milk yield per cow) for 91 FADN regions from 2007-2015

Transition probability matrix						
Initial distribution (2007)	*Class*	*<75*	*75-90*	*90-105*	*105-120*	*>120*
13	*<75*	0.929	0.071	0.000	0.000	0.000
12	*75-90*	0.062	0.825	0.103	0.010	0.000
25	*90-105*	0.000	0.054	0.801	0.140	0.005
21	*105-120*	0.000	0.005	0.111	0.794	0.090
20	*>120*	0.000	0.000	0.006	0.108	0.885
Summary statistics						
Class		*<75*	*75-90*	*90-105*	*105-120*	*>120*
Stationary distribution		0.109	0.125	0.235	0.291	0.239
Half-life		15.882				
Mobility index		0.191				
Upward mobility		0.547				
Downward mobility		0.453				

Note: EU-sample average=100, estimates are rounded to three decimals. Standard errors can be found in Appendix III.
(Source: author, based on FADN data)

In Figure 5.3, it can be seen how the class distribution has changed over time. Most regions in the highest productivity class are located in the north-western part of the EU, also in the part where the specialized milk farms are concentrated (Ihle et al. 2017). While the regions with the lowest productivity classes can be found on the eastern border of the EU. As can be seen on the maps, there is only limited mobility between the productivity classes. From the NMS regions, only Slovakia, Czech Republic and Estonia have caught up with the EU-average. Some regions in France and Italy also have improved their productivity.

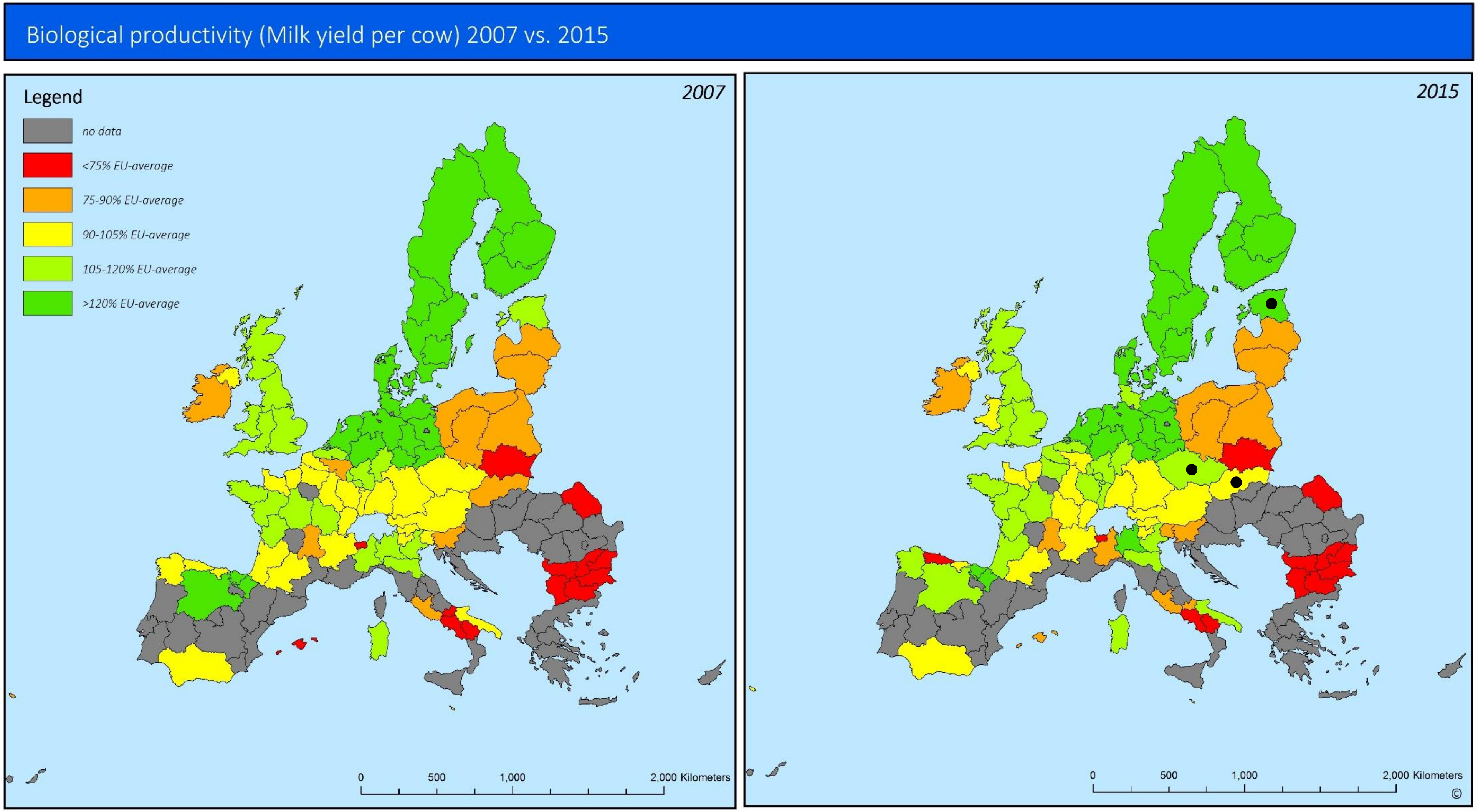


Figure 5.3.Class transition per region for biological productivity (milk yield per cow) for 91 FADN regions from 2007 vs. 2015
Note: EU-sample average=100. The black dot represents NMS regions that have caught up.
(Source: author, based on FADN data)

5.5.2. Mechanical productivity (Cows per worker)

As shown in the Kernel density plot for mechanical productivity (Figure 5.4), again a bimodal or twin-peak distribution can be observed. For the years 2004-2006 this bimodality is already present. In the period 2007-2015 the sharp peaks disappear, resulting in a wider spread. The largest peak has shifted to the right over time (see dashed arrow), while the smallest peak stays at the same position. The spread of the distribution has not substantially changed over time. Since the twin-peak distribution stays in place, there are no signs of convergence present in the distributional dynamics of the Kernel density plot.

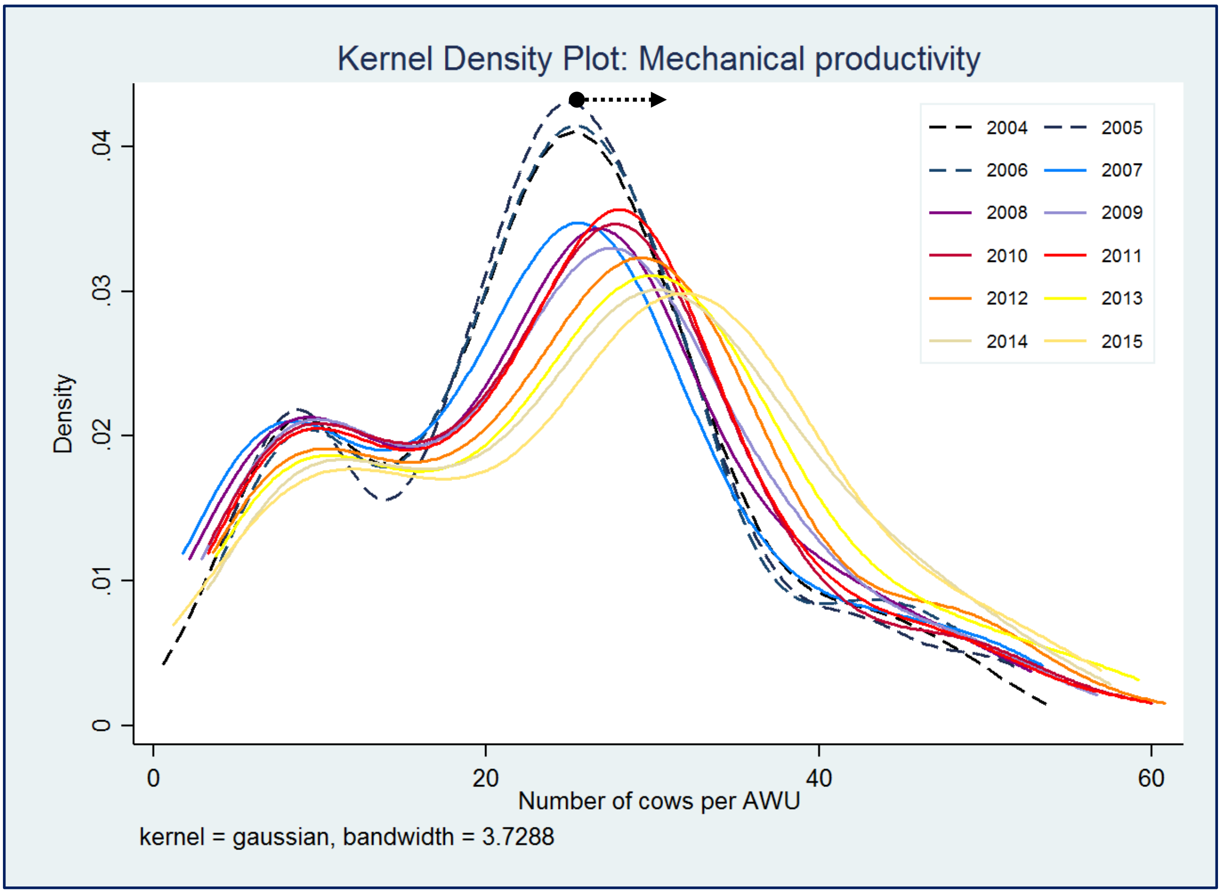


Figure 5.4. Kernel density mechanical productivity (cows per AWU), for 91 FADN regions from 2004-2015
Note: the dashed arrow shows the movement of the right peak to the left over time.
(Source: author, based on FADN data)

In the transition matrix in Table 5.6, it can also be seen that there is a high persistence in the distribution, $\hat{p}_{11}$ and $\hat{p}_{55}$ are both around 0.9, meaning that at both ends of the distribution there is a high probability to stay at both ends. It appears that in the three middle classes there is more mobility. If we compare it to biological productivity, we find that the overall mobility is even lower for mechanical productivity. This means that the likelihood to transition to another class is very low. The half-life is 13 meaning that it takes 13 years to close half of the gap between the initial and the stationary distribution. From the stationary distribution it is remarkable that only 7.9% will end up in the lowest category, given the high probability at the tails ($\hat{p}_{11}$ and $\hat{p}_{55}$).

Table 5.6. Transition probability matrix of mechanical productivity (cows per AWU) for 84 FADN regions from 2004-2015

Transition probability matrix						
Initial distribution (2004)	*Class*	*<50*	*50-80*	*80-110*	*110-140*	*>140*
18	*<50*	0.925	0.075	0.000	0.000	0.000
10	*50-80*	0.039	0.819	0.142	0.000	0.000
24	*80-110*	0.004	0.040	0.839	0.113	0.004
19	*110-140*	0.000	0.009	0.127	0.816	0.047
13	*>140*	0.000	0.000	0.000	0.073	0.927
Summary statistics						
Class		*<50*	*50-80*	*80-110*	*110-140*	*>140*
Stationary distribution		0.079	0.120	0.327	0.279	0.196
Half-life		13.084				
Mobility index		0.168				
Upward mobility		0.565				
Downward mobility		0.435				

Note: EU-sample average=100, estimates are rounded to three decimals. Standard errors can be found in Appendix III.
(Source: author, based on FADN data)

If we add the seven regions that accessed the EU in 2007, there is even more persistence in the distribution, as shown by the lower mobility index. Now it can be seen that a region that has a productivity less than 50% of the EU-average has a probability of 0.956 to remain in that class. This is also reflected in the higher stationary value for the lowest class (i.e. 0.138). The half-life also increased compared to the previous matrix. Like the biological productivity it shows that if we add the seven NMS (2007) regions, that there is more persistence in the distribution and a slower pace towards convergence.

Table 5.7. Transition probability matrix of mechanical productivity (cows per AWU) for 91 FADN regions from 2007-2015

Transition probability matrix						
Initial distribution (2007)	*Class*	*<50*	*50-80*	*80-110*	*110-140*	*>140*
22	*<50*	0.956	0.044	0.000	0.000	0.000
9	*50-80*	0.034	0.862	0.103	0.000	0.000
20	*80-110*	0.007	0.046	0.795	0.139	0.013
23	*110-140*	0.000	0.005	0.103	0.851	0.041
17	*>140*	0.000	0.000	0.007	0.073	0.920
Summary statistics						
Class		*<50*	*50-75*	*75-100*	*100-125*	*>125*
Stationary distribution		0.138	0.132	0.228	0.307	0.195
Half-life		19.866				
Mobility index		0.154				
Upward mobility		0.553				
Downward mobility		0.447				

Note: EU-sample average=100, estimates are rounded to three decimals. Standard errors can be found in Appendix III.
(Source: author, based on FADN data)

In Figure 5.5. there is a clear east-west division visible in the map. Most productive regions are located in north-western Europe (Denmark, UK, Ireland and the Netherlands). At the eastern side of the EU we find the low-productive regions (Poland, Baltic, Slovakia, Czech Republic, Romania and Bulgaria). Over time only limited mobility has taken place. Southern Finland and southern Sweden have made a leap forward. From the NMS regions, only Estonia was able to catch-up. This can perhaps explain the high transition probability to stay in the lowest productivity category ($\hat{p}_{11} = 0.956$), since all NMS remain in the lowest category.

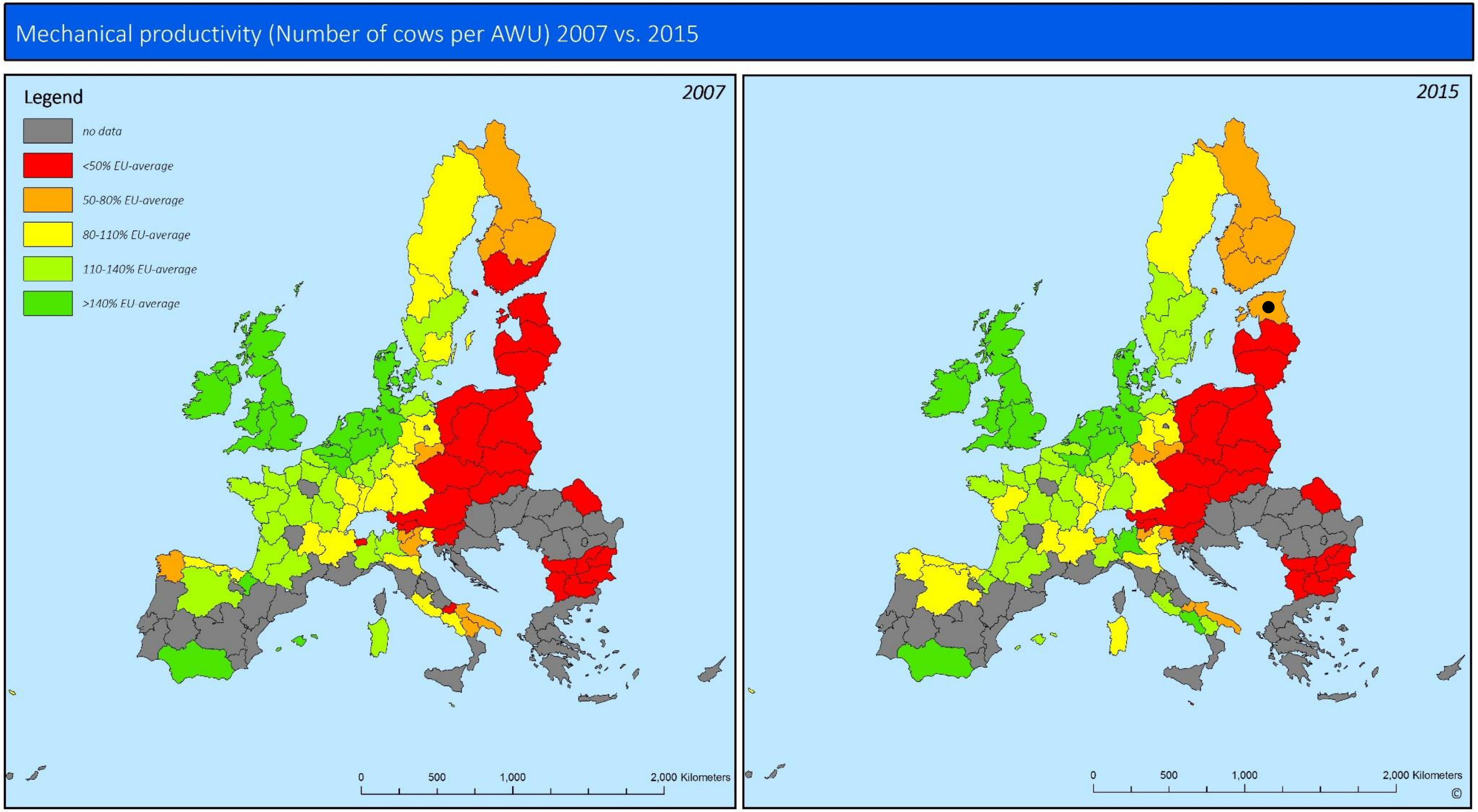


Figure 5.5. Class transition per region for mechanical productivity (Number of cows per AWU) for 91 FADN regions from 2007 vs. 2015
Note: EU-sample average=100. The black dot represents NMS regions that have caught up.
(Source: author, based on FADN data)

5.6 Discussion

Considering the results, there has been limited convergence in the period 2004-2015. For biological productivity there has been a catch-up process for some regions in the NMS (2004). However, in the Kernel density plot it could be seen that there was substantial bimodality in the distribution that only became more present over time. The spread of the distribution became larger over time, meaning that dispersion increased. While at the same time, the smaller peak stayed at the same place, resulting in a wider gap between a large group of productive regions and a smaller group of low-productive regions. For mechanical productivity there seems to be even less evidence for convergence. The high probabilities for $\hat{p}_{11}$ suggest that low-productive regions were not able to catch-up. For mechanical productivity, the distribution also shows bimodality, which remained in place over time. The half-life indicated that it could take 20 years to close half of the gap between the initial distribution and the stationary distribution. A high mobility index for all productivity measures has shown that the overall probability to move from one category to another is quite low. Although in all cases, the upward mobility was somewhat larger than the downward mobility. A possible explanation for this result could be that once a productivity level is reached, it is feasible to hold on to this productivity level. For example, when once an investment in a milking machine is done, it is very likely that this will result in a permanent increase in productivity. All in all, the evidence suggests that the process towards convergence is fragile and slow. There seems to be a persistent gap with respect to productivity between the NMS and the OMS. Our results are in accordance with Cechura et al. (2017), who also do not find evidence that the regions with poorer dairy farms are catching up. Like Jansik and Irz (2015b), we find that Estonia is one of the regions that has shown a catch-up process. For Lithuania, Latvia, Slovenia, Poland, Romania, Bulgaria there is no evidence for a catch-up process with respect to productivity.

The question arises what could be the explanation for the limited convergence that has taken place for productivity across the regions. In Figure 5.6 and 5.7 some additional benchmarks are shown. First of all, to increase productivity a farmer can invest in new technologies. In Figure 5.6, the investment per cow in a region is presented. Back in 2007, investments per cow were the highest in the Nordic countries and Austria, with more than €1000 investment per cow. Particularly in Spain, Italy and the NMS (2007) investments were the lowest. Most NMS (2004) showed average levels of investments per cow. In 2015, the situation has changed completely. Still, in north-western Europe the levels of investment per cow are high. Most NMS now have very low levels of investment, with Lithuania and the Czech Republic as the positive exemptions. In Figure 5.7 the stocking density (number of cows per hectare) for the FADN regions is shown. The more cows per hectare the more intensive the farming system is. In 2007, most intensive farming systems could be found in southern Spain, Italy, the UK, Ireland, the Netherlands and Bulgaria. At the other hand, the most extensive farming systems can be found in the Baltic, Sweden, Czech Republic, Slovakia, eastern Germany and eastern France. In the period 2004-2015, north-western and eastern Germany, Denmark and Wallonia (Belgium) became more intensive.

In general, we can see that the most productive regions (Netherlands, UK and Ireland, Denmark, southern Sweden) have above average investments per cow and above average stocking densities. For the NMS, the Baltic countries Slovakia, Slovenia and the Czech Republic have high investments per cow, this is partly reflected in the biological productivity, but not at all in mechanical productivity. Poland, seems to be a laggard with respect to productivity, and it also belongs to the group with the lowest investments per cow and the lowest stocking density. It is difficult to pinpoint the specific factors that make a dairy sector in a region successful. Still, it can be seen that some regions lag behind in terms of productivity and that also the structural factors have come to a standstill.

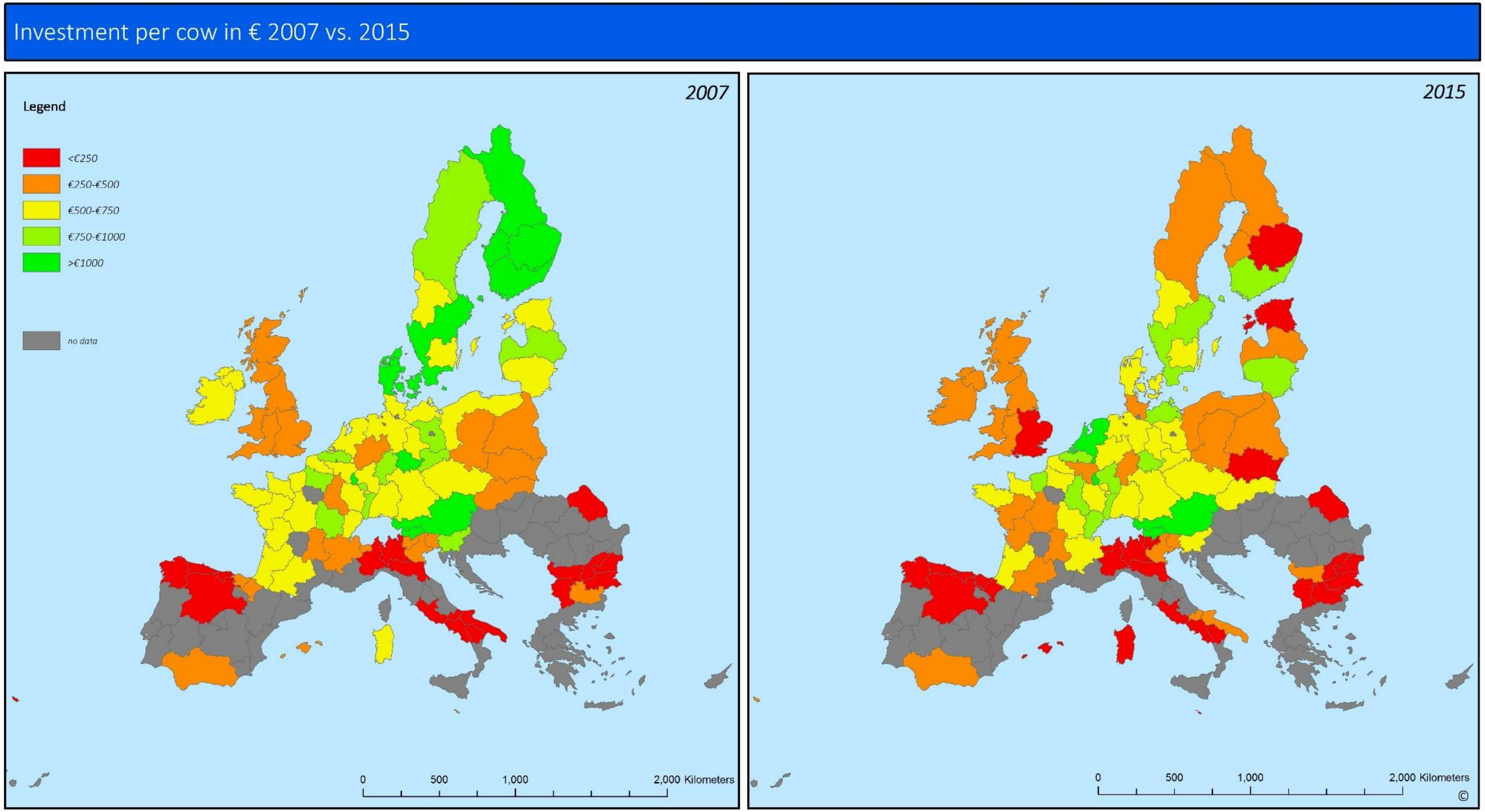


Figure 5.6. Investment per cow in € for 91 FADN regions from 2007 vs. 2015.
(Source: author, based on FADN data)

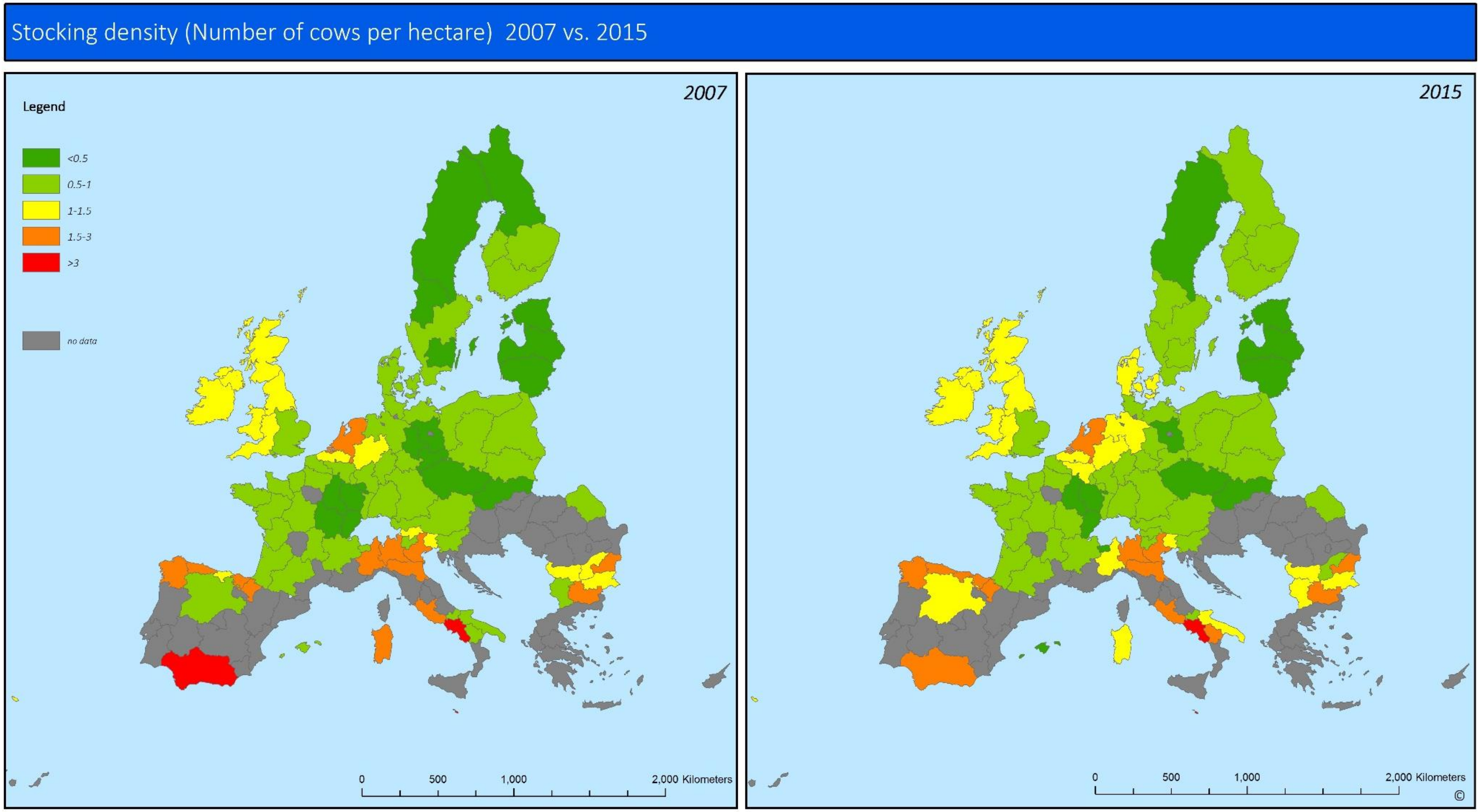


Figure 5.7. Stocking density (cows per hectare) in € for 91 FADN regions from 2007 vs. 2015
(Source: author, based on FADN data)

6 INCOME CONVERGENCE

In this chapter we test for convergence in farm income. First, it is explained how the concept of convergence can be applied to incomes. After that, an overview of the academic literature on income convergence is given. Then, the choice of empirical measures is clarified and the empirical measures themselves are explained. Then we explain the empirical model and show the results.

6.1 THE CONCEPT OF INCOME CONVERGENCE

The study of income convergence is the most importance within the convergence theory. Since it is linked to the topics of economics of development and economic inequality, it is widely studied. As one of the three variables in this study (prices, productivity, income), income is the most important because it is most directly felt by people. It is therefore not surprising that most convergence studies are devoted to income.

In the neo-classical Solow model, it was expected that the income per capita moved towards a steady state. Since this model assumes diminishing returns to the production factors, high-income countries will have lower growth rates than low-income countries. In a way, convergence seems therefore like a natural process, where no policy is required. For the dairy sector we could say that the returns to capital and labour are very high for the NMS. That would naturally result in a higher growth rate for these countries. In the previous chapter, we have seen that productivity convergence only occurred for a small number of regions. It seems not straightforward that regions that are lagging behind automatically catch up. Specifically for income there are also several ways to improve income; cost minimization, scaling up, increasing output, obtaining a higher output price are all ways to improve the income position. It is therefore crucial for income convergence that also some extent of price convergence takes place. Since the milk price determines for a large extent the income of dairy farmers, large differences in the milk price between countries can hamper the convergence of income.

From section 2.2, it is clear that in the NMS still many small-scale farms exist. For example, in Bulgaria still 90% of the farms had five cows or less. For these farms, the only way up seems to be a scaling-up process. For more intensive, high-innovative countries like Denmark or the Netherlands it is questionable how much room there is present for income improvement. Since land resources are limited in these countries, their dairy sectors cannot continue to grow unlimitedly. So the source of growth can be different for the regions. In the end, it is also important for a farm how the income is distributed. In an industrial farm with many employers, income is a different concept than in a family farm that are self-sufficient in terms of labour. Also, the income that a farmer gains from off-farm activities might change the perception of farm income. It is therefore not reasonable to assume a single economic model that can account for farm income growth across the EU regions. Although, there are several causes for income growth it seems unlikely that there is a deterministic model that can account for all the ongoing processes in dairy farming. Again, the σ-convergence definition is used to see how the dispersion changed over time. Likewise, for the empirical part the distributional approach is taken. In order to inspect the cross-section of incomes over time. In this way, convergence can be measured and at the same time we can keep track of the individual income process of the regions.

6.2 LITERATURE REVIEW

Income convergence has been widely studied in literature. However, the focus in those papers is often on the convergence of Gross Domestic Product or Income (GDP/GDI) per capita across countries or regions. Next to the literature on growth theory, there are some specialized papers that studied the income distribution of farmers. Brasili et al. (2006) use a panel of EU regions and US states to test for convergence of agricultural incomes. They find evidence on convergence for EU regions but not for US states. They show that convergence of family farm income is greater than convergence of the net added value per hectare. In this paper two methods are used: a panel unit root analysis and a stochastic kernel. A related field of research are the studies on structural change in agriculture. This type of studies tries to measure structural change within the dairy sector. They apply the Markov chain method to estimate the transition of the (economic) farm size over time. In most studies, transition probabilities are estimated, which indicate what the probability is that a farm transforms over time (e.g. Tonini and Jongeneel, 2009; Ben Arfa et al., 2015). Secondly, some studies add a second step and try to identify what drives these transitions, by estimating what factors influence the transition probabilities (Zimmermann et al., 2009). Soares and Ronco (2000) have studied if growth in agricultural incomes in the EU has led to convergence. They measure convergence for per capita Gross Agricultural Value and per capita Final Agricultural Output. Only for France and Spain they find a pattern of convergence. The 'richer' countries were able to maintain their position or even widen the gap with the 'poorer' countries in terms of agricultural output. Bivand and Brunstad (2003) have studied if agricultural support may have an effect on regional convergence. With a geographically weighted regression they find some evidence that agricultural support might have a negative effect on convergence, although they identify certain weaknesses of this type of regression. Hansen and Teuber (2011) studied the impact of the CAP on regional convergence for some regions in Germany. They use several empirical measures like the CV and Kernel density functions to study convergence. They find that inequality with respect to revenues increased, but that the CAP softened the cross-sectional inequality. Nevertheless, the CAP did not hamper the diverging trend in the period 1991-2004. They measure that the structural differences between the farmers account for most of the inequality between the farmers' revenue.

6.3 EMPIRICAL METHODS

Unlike productivity, income is a clear variable that does not need complicated estimation methods to find the income. Again we need an empirical measure to test for convergence. As could be seen in Table 5.2, there exist several methods to measure convergence. In this chapter, the distributional approach to convergence is chosen. For that reason, it is straightforward to again chose for Kernel density estimation and Markov chain analysis. For the same reasons as the previous chapter: an adequate combination of visual inspection and statistical inference of the dynamics of the distribution.

6.4 THE EMPIRICAL MODEL AND DATA[8]

6.4.1. Measures of income and data

Unlike traditional income measures like GDP/capita or GNI/capita, farm income can be specified in several ways. First it has to be determined how income should be defined. One could choose for a measure like total output of milk in euros, but there is a wide gap between output and the eventual income. As can be seen in Figure 6.2, there is a wide gap between the total output and the farm net income that consists of VAT, subsidies, consumption, depreciation, etc.

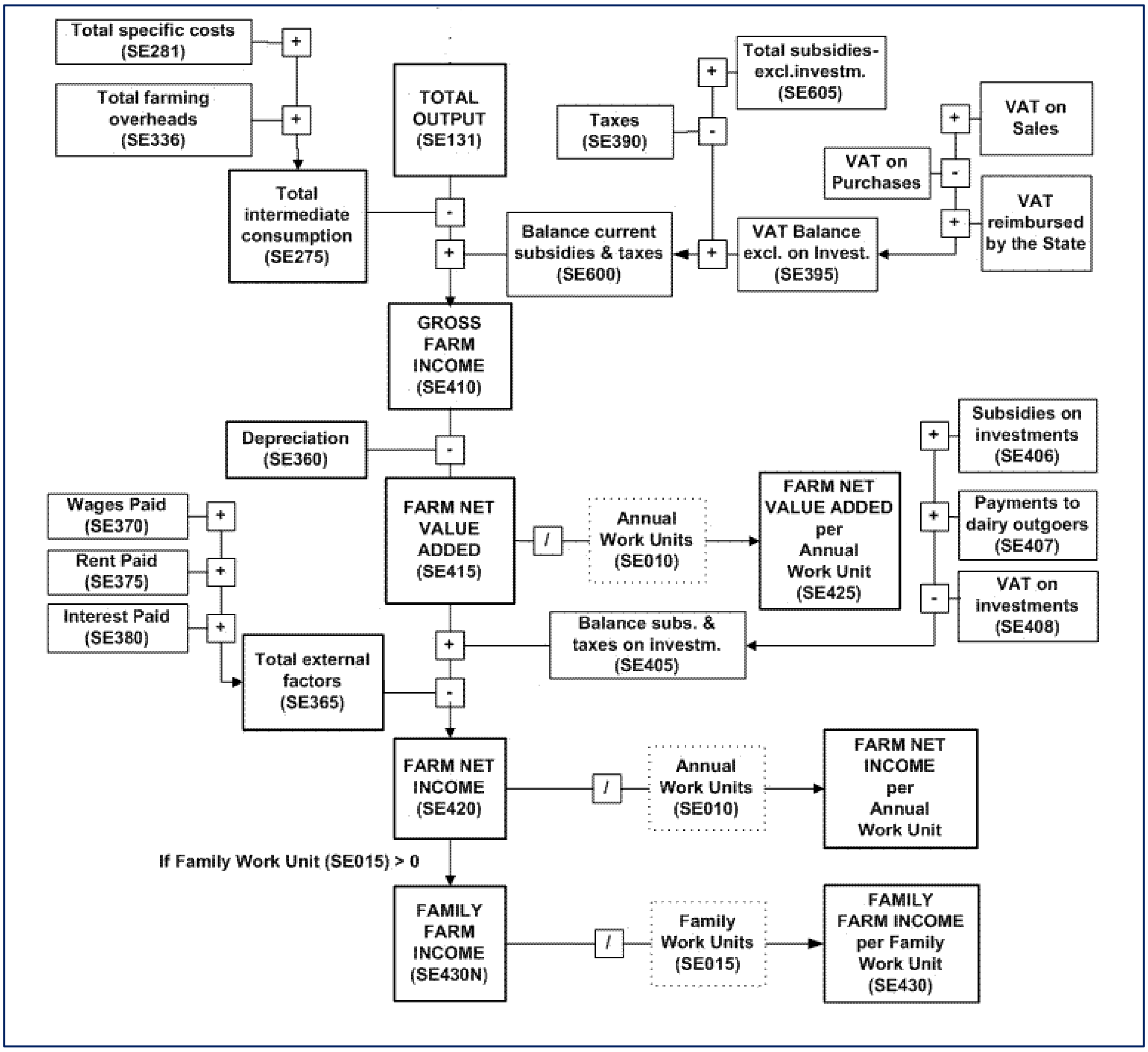


Figure 6.1. Income components of a farming unit in the FADN standard results
(Source: FADN, 2010)

Given the heterogeneity of the EU dairy sector it is wise to choose measures that are comparable across regions. This makes it possible to compare industrial-based farming systems that uses a lot of labour vs. small-scale self-sufficient farming systems.

For that reason, the first income measure is the *labour-adjusted value added*:

$$Labour - adjusted\ value\ added = \frac{Farm\ net\ value\ added}{Labour\ input} \qquad (6.1)$$

[8] To avoid repetition only the main equations of the empirical models are shown, for more background we refer to section 5.4.

In the FADN database the variable SE425 Farm Net Value Added/AWU can be used as the variable for Labour-adjusted net value added. It is defined *as the Farm Net Value Added per Annual Working Unit*. As second income indicator we use the Farm net income instead of the Farm net value added. The farm net value added is corrected for external factor costs and for subsidies and taxes on investments. We define this variable as *labour income*:

$$Labour\ income = \frac{Farm\ net\ income}{Labour\ input} \tag{6.2}$$

In the FADN database the variable can be calculated by dividing SE420 Farm Net Income by SE010 AWU (Annual Working Unit).
As third income indicator we take a closer look at the family aspect in farming. In many regions in the EU farming is still organized as a family business. Usually (unpaid) family labour is a major component in the dairy farm. As a second income measure we therefore propose *family farm income*:

$$Family\ farm\ income = \frac{Farm\ net\ income}{Family\ labour} \tag{6.3}$$

Family farm income can shine a light on the family component present in dairy farming. It is important since a lot of families depend on their farm for their income, so it can measure an important living condition for the families. In the FADN database it is defined as: SE430 Family Farm Income / FWU. The variable is described as *family farm income expressed per family labour unit. Takes into account difference in the family labour force to be remunerated per holding. It is calculated only for the farms with family labour.* (FADN, 2018b)

6.4.2. Kernel density estimation

The Kernel density estimator of a series X with a specific point x can be defined as (Silverman, 1986, p.4):

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x - X_i}{h}\right) \tag{6.4}$$

n is the number of observations, k the Kernel function, and h is the smoothing parameter. Similar as in chapter 5, the Gaussian Kernel function is used, with the optimal bandwidth as described by Silverman (1986). If σ-convergence would occur, we would expect that the spread of the distribution becomes smaller over time.

6.4.2. Markov chain

The Maximum Likelihood estimator for p_{ij} can be defined as (Anderson and Goodman, 1957, p. 92):

$$\hat{p}_{ij} = n_{ij}/n_i^* = \sum_{t=1}^{T} n_{ij}(t) \,/ \sum_{k=1}^{m}\sum_{t=1}^{T} n_{ik}(t) = \sum_{t=1}^{T} n_{ij}(t) \,/ \sum_{t=0}^{T-1} n_i(t) \tag{6.5}$$

This holds for the i-th sample ($i = 1,2,\dots,m$) that consists of $n_i^* = \Sigma_j n_{ij}$ multinomial trials with the probabilities $p_{ij}(i,j = 1,2,\dots,m)$.
From the transition matrix, several statistics on convergence can be conducted.
If the transition matrix $\mathbf{P}$ is of an *irreducible* and *aperiodic* Markov chain and it is homogenous over time it means that the Markov chain moves towards a steady state. It is then possible to estimate the stationary distribution, which shows the probability distribution in the steady state. This stationary distribution $\bar{\pi}$ is a probability distribution that does not change when time progresses. Formally, we can define this stationary distribution $\bar{d}$ (also called invariant probability vector) as Lawler (2006, p.22):

$$\bar{\pi}\mathbf{P} = \bar{\pi} \qquad (6.6)$$

This distribution is not a prediction for the future, since the circumstances for one-time period may be completely different from another time period. Rather the stationary distribution characterizes the process of the past period.
Secondly, the half-life of a transition matrix can be calculated. The half-life is the number of periods it takes to close half of the gap towards the stationary distribution. The half-life can be calculated as (Shorrocks, 1978, p.1021):

$$\text{Half-life} = \frac{-\log 2}{\log|\lambda_2|} \qquad (6.7)$$

λ_2 is the second-to-largest eigenvalue of the transition matrix $\mathbf{P}$. The half-life is value between zero and infinity. If it is zero, it means that the stationary distribution has already been reached.
A third measure is the mobility index (M^{OV}), which is an indicator of the degree of mobility in the distribution. So it is a measure that indicates what the overall likelihood of the transition matrix is to remain in a certain state. If there would be no mobility in the distribution, it would mean that all probabilities along the diagonal are equal to one, hence it would be the identity matrix of $\mathbf{P}$. Then the mobility index is equal to zero. For the case of perfect mobility, we assume a quasi-maximal diagonal for $\mathbf{P}$ (there exists a positive $\mu_1,\dots,\mu_n$ $such\ that$ $\mu_j p_{jj} \geq \mu_i p_{ij}$ for all i,j) (Shorrocks, 1978, p. 1017). This means that the probability to remain in the same class is not less than the probability to move to another class. So in case of perfect mobility, the trace of the probability matrix is then equal to one, consequently the mobility index will be one. This measure proposed by Shorrocks (1978, p.1017) is defined in Equation 5.12. Where $tr(\mathbf{P})$ is the trace of the transition matrix $\mathbf{P}$ and n is the number of classes.

$$M^{OV} = [n - tr(\mathbf{P})] \cdot [n-1]^{-1} \qquad (6.8)$$

Additionally, two mobility measures are calculated to interpret the direction of the mobility: upward or downward mobility. These indicators can be seen as shares of the overall mobility indicator. The sum of the upper triangle of transition probabilities represent the upward mobility ($\sum_i \sum_{j>i} p_{ij}$), while the sum of the lower triangle of transition probabilities represent the downward mobility ($\sum_i \sum_{i>j} p_{ij}$). The mobility of the diagonal element is defined as $\sum_j (1 - p_{jj})$. The upward and downward mobility are 'deflated' by the sum of this diagonal element and therefore $M^U + M^D = 1$.
If there is no downward mobility and no persistence along the diagonal, it would mean that there is perfect upward mobility and M^U would be equal to one. If there is no upward mobility and no

persistence along the diagonal, it would mean that there is perfect downward mobility and M^D would be equal to one.

The upward mobility can therefore be defined as (Jongeneel and Huettel, 2011, p.513):

$$M^U = \left[\sum_i \sum_{j>i} p_{ij}\right] \cdot \left[\sum_j (1 - p_{jj})\right]^{-1} \tag{6.9}$$

And the downward mobility is defined as (Jongeneel and Huettel, 2011, p.513):

$$M^D = \left[\sum_i \sum_{j<i} p_{ij}\right] \cdot \left[\sum_j (1 - p_{jj})\right]^{-1} \tag{6.10}$$

If a process of convergence would happen, it is expected that there is a substantial probability to move from the lowest category to higher categories (marked green) in the transition matrix $\mathbf{P}$. On the other hand, the transition probabilities to move from middle categories to lower categories should not be too high (marked pink). If the probabilities to stay at the tails of the distribution (marked red) are much higher than the other probabilities along the diagonal, there is a high risk that the distribution moves towards a twin-peak distribution, so that a persistent structural gap between two groups will exist.
If a process of convergence would occur, we would also expect the stationary distribution to have high shares in the three middle categories. Hence, this would mean that regions are more closely located to the average, so then the dispersion decreasing, meaning that σ-convergence takes place.

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} & p_{15} \\ p_{21} & p_{22} & p_{23} & p_{24} & p_{25} \\ p_{31} & p_{32} & p_{33} & p_{34} & p_{35} \\ p_{41} & p_{42} & p_{43} & p_{44} & p_{45} \\ p_{51} & p_{52} & p_{53} & p_{54} & p_{55} \end{bmatrix} \tag{6.11}$$

Again, the distribution is split into five separate income classes. In Table 6.1, it can be seen how the 91 regions were distributed over the five classes. As could be seen, the least productive regions are the 'laggards', the regions that lag behind on a European scale. While at the other end, the 'frontrunners' are a group of regions that are leading in terms of productivity or income.

Table 6.1. Class division for the income measures as percentage of the total in 2007

Class	***<50% EU-average*** *Laggard*	***50-80% EU-average*** *Low-productive*	***80-110% EU-average*** *Average*	***110-140% EU-average*** *High-productive*	***>140% EU-average*** *Frontrunner*
Labour adjusted value added	21%	18%	19%	20%	23%
Labour income	21%	21%	19%	16%	23%
Family farm income	20%	23%	22%	11%	24%

(Source: author)

6.5 RESULTS

6.5.1. Labour adjusted value added

As can be seen in Figure 6.2, the Kernel density plots vary much more by year as in the previous chapter. Still, for most years a pattern can be observed. There is a small peak at the beginning of the distribution followed by a very large peak. The distribution is skewed to the right, which can be seen at the densities for the region above €50.000. For the year 2009, the distribution is completely different, since the distribution is shifted to the left. This corresponds to the dairy crisis in that year as was discussed in section 2.2.2, which is reflected in a shift to lower levels of value added. Like in the previous chapter we see a shift of the larger peak to the right (see dashed arrow), while the peak at the left stays at the same point. The spread of the distribution has stayed the same. Overall, there are no signs of convergence or catch-up processes in the distributional dynamics.

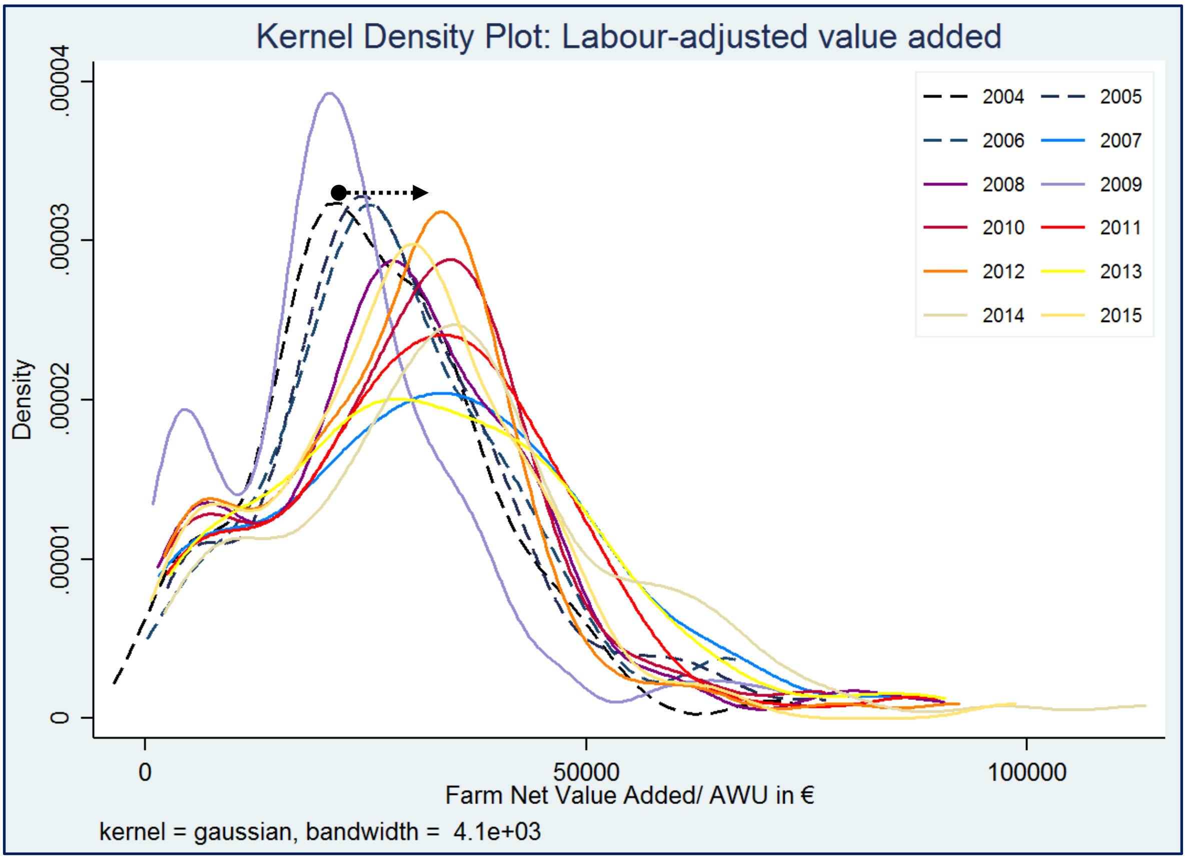


Figure 6.2. Kernel density labour adjusted value added (farm net value added/AWU), for 91 FADN regions from 2004-2015
Note: the dashed arrow shows the movement of the peak to the right.
(Source: author, based on FADN data)

In Table 6.2 the transition matrix of the labour-adjusted added value from 2004 onwards is shown. The $\hat{p}_{11}$ is around 0.889, which means a high probability to stay in the lowest class. In the middle classes, some more mobility is present which is reflected in the non-zero probabilities. Compared to the transition matrices of the productivity variables, there is a lower half-life and higher mobility index. The lower half-life implicates in general that it takes less time to convergence to the stationary distribution. A higher mobility index indicates that the process is less stable which implies that the probability to move from one category to another category is higher overall.

Table 6.2. Transition probability matrix of labour adjusted value added (farm net value added/AWU) for 84 FADN regions from 2004-2015

Transition probability matrix						
Initial distribution (2004)	*Class*	*<50*	*50-80*	*80-110*	*110-140*	*>140*
12	*<50*	0.889	0.111	0.000	0.000	0.000
20	*50-80*	0.070	0.654	0.232	0.038	0.005
19	*80-110*	0.008	0.151	0.660	0.154	0.027
17	*110-140*	0.000	0.021	0.234	0.548	0.197
16	*>140*	0.000	0.000	0.064	0.242	0.694
Summary statistics						
Class		*<50*	*50-80*	*80-110*	*110-140*	*>140*
Stationary distribution		0.141	0.189	0.302	0.206	0.163
Half-life		6.652				
Mobility index		0.389				
Upward mobility		0.492				
Downward mobility		0.508				

Note: EU-sample average=100, estimates are rounded to three decimals. Standard errors can be found in Appendix IV.
(Source: author, based on FADN data)

By adding the seven regions that accessed in 2007, there is more mobility in the matrix. Still, the probability to stay in the lowest class is 0.918, which means that there is only a very small chance to move to a higher class. In the stationary distribution, it is also found that around 16% of the regions ends up in the lowest class.

Table 6.3. Transition probability matrix of labour adjusted value added (farm net value added/AWU) for 91 FADN regions from 2007-2015

Transition probability matrix						
Initial distribution (2007)	*Class*	*<50*	*50-80*	*80-110*	*110-140*	*>140*
19	*<50*	0.918	0.082	0.000	0.000	0.000
16	*50-80*	0.061	0.591	0.313	0.026	0.009
17	*80-110*	0.012	0.182	0.506	0.259	0.041
18	*110-140*	0.000	0.033	0.285	0.464	0.219
21	*>140*	0.000	0.000	0.082	0.212	0.705
Summary statistics						
Class		<50	50-80	80-110	110-140	>140
Stationary distribution		0.160	0.166	0.261	0.213	0.199
Half-life		8.091				
Mobility index		0.454				
Upward mobility		0.522				
Downward mobility		0.478				

Note: EU-sample average=100, estimates are rounded to three decimals. Standard errors can be found in Appendix IV.
(Source: author, based on FADN data)

In Figure 6.3 a map on the transitions between 2007 and 2015 is shown. As can be seen most NMS regions remain in the lowest class. Exceptions are the Czech Republic and Estonia who climbed one class. Next to that, some regions in the UK, western Germany and Spain seem to have faced a decline in the labour-adjusted value added. On the other hand, Sweden seems to have made a leap forward compared to the EU-average.

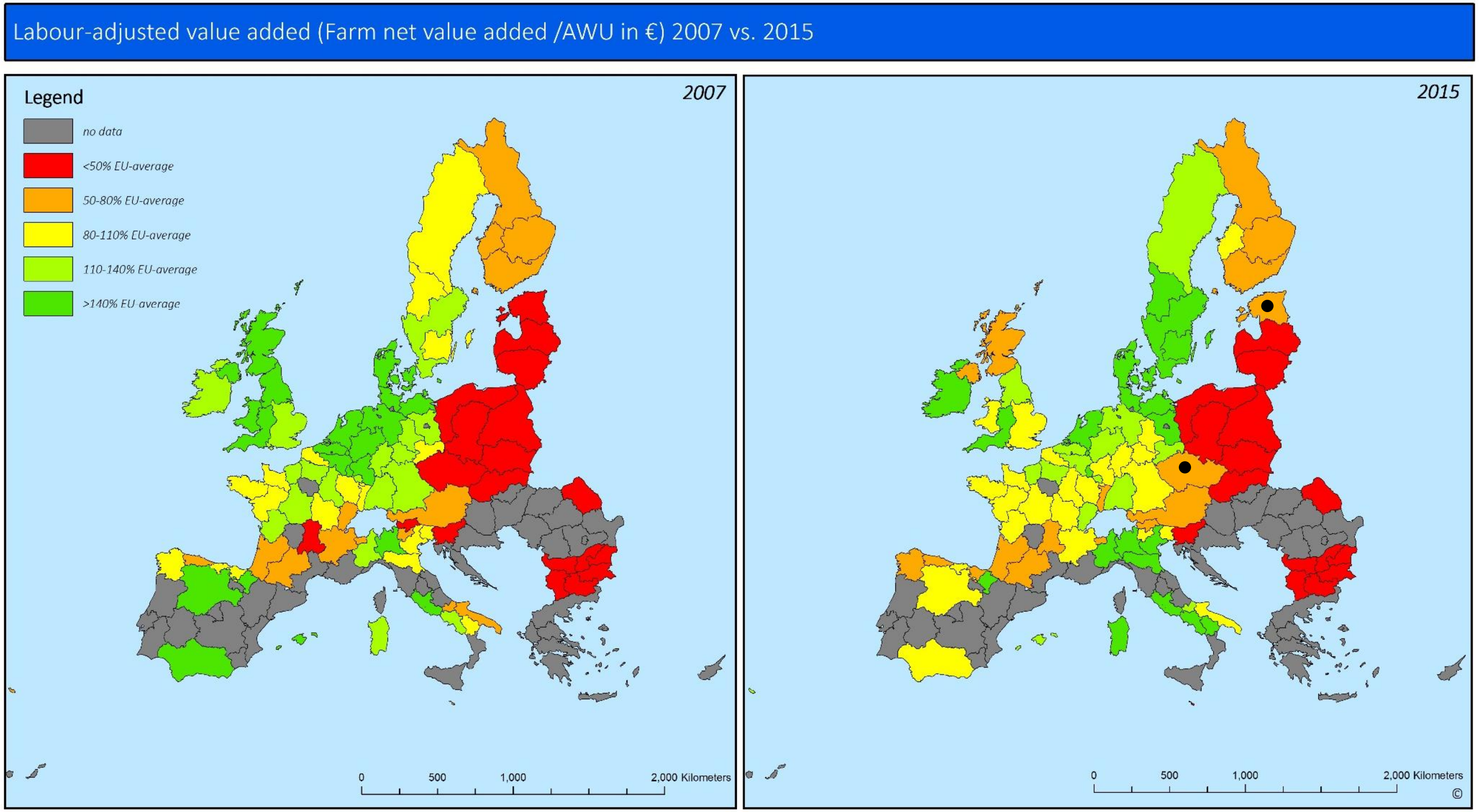


Figure 6.3. Class transition per region for labour adjusted value added (farm net value added/AWU) for 91 FADN regions from 2007 vs. 2015
Note: EU-sample average=100. The black dot represents NMS regions that have caught up.
(Source: author, based on FADN data)

6.5.2. Labour income

In Figure 6.4, the Kernel density plot of labour income is shown. There is a large difference between the density plots of the years 2004-2006 and 2017-2015, because the distribution is less dense at the middle. Instead, from 2007 onwards the spread of labour income is wider. In the years 2011-2015, the first small peak at the lefts seems to have disappeared, which could hint towards convergence. Like in the plot of labour-adjusted value added, there is skewed to the right, since the tail on the right is longer. This can be caused by lower bound around zero income. The spread seems to be equal for most years, except for the years 2009 and 2014.

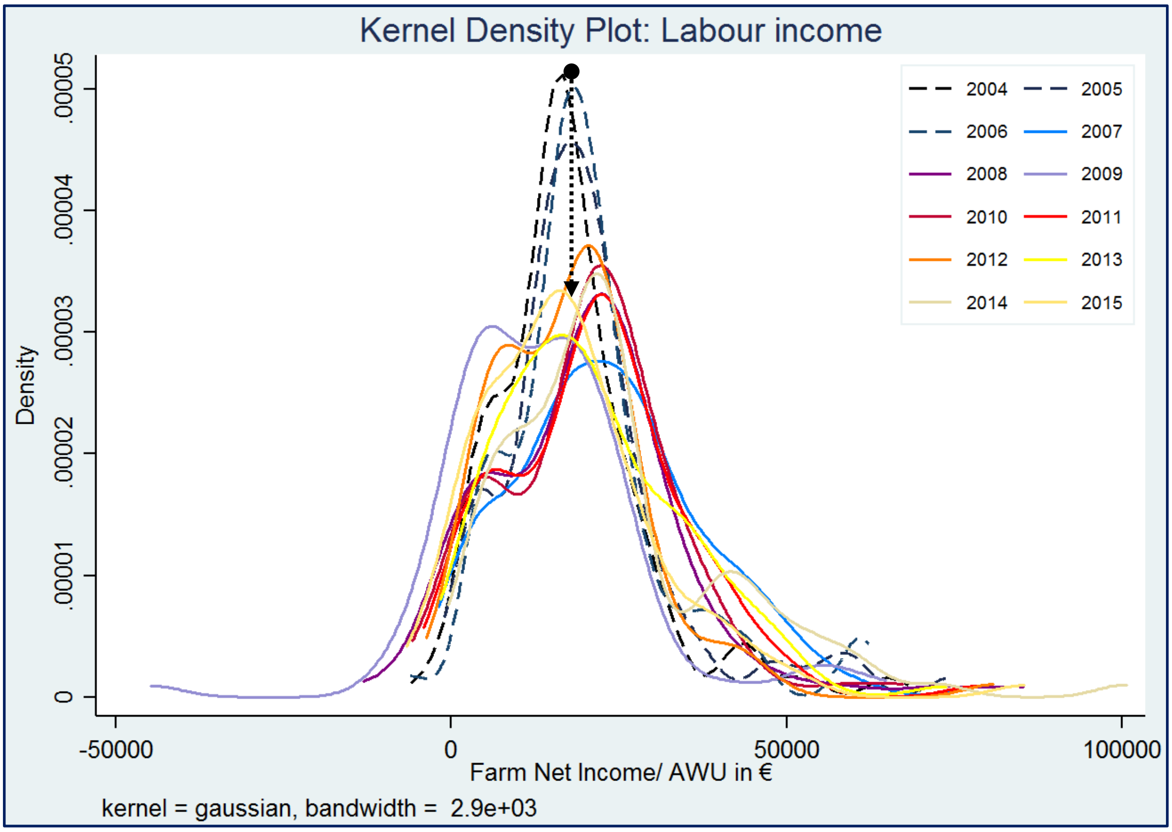


Figure 6.4. Kernel density labour income (farm net income/AWU), for 91 FADN regions from 2004-2015
Note: the dashed arrow shows the decrease in density at the main peak.
(Source: author, based on FADN data)

In Table 6.4, the transition matrix of labour income is shown. Again, the probabilities $\hat{p}_{11}$ and $\hat{p}_{55}$ show are higher than the other probabilities along the diagonal. Hence at both ends there is a high probability to remain in the same category. This also results in high shares at the tails of the stationary distribution. The mobility index is considerable higher if we compare it to mobility indices the productivity measures. The half-life for labour income is considerably lower than the half-life of labour-adjusted value added. This means that it takes a shorter time to close the gap towards the stationary distribution.

Table 6.4. Transition probability matrix of labour income (farm net income/AWU) for 84 FADN regions from 2004-2015

Transition probability matrix						
Initial distribution (2004)	*Class*	*<50*	*50-80*	*80-110*	*110-140*	*>140*
17	*<50*	0.816	0.116	0.058	0.005	0.005
11	*50-80*	0.141	0.456	0.336	0.060	0.007
28	*80-110*	0.042	0.187	0.515	0.183	0.073
12	*110-140*	0.037	0.059	0.324	0.338	0.243
16	*>140*	0.005	0.016	0.086	0.160	0.733
Summary statistics						
Class		*<50*	*50-80*	*80-110*	*110-140*	*>140*
Stationary distribution		0.219	0.161	0.269	0.142	0.210
Half-life		3.516				
Mobility index		0.535				
Upward mobility		0.507				
Downward mobility		0.493				

Note: EU-sample average=100, estimates are rounded to three decimals. Standard errors can be found in Appendix IV.
(Source: author, based on FADN data)

As shown in Table 6.5, the transition probabilities for the period 2007-2015 show again that the probabilities to stay at both ends of the distribution is higher compared to the 2004-2015 period. It appears that in the stationary distribution around 30% of the regions would be in the lowest income class.

Table 6.5. Transition probability matrix of labour income (farm net income/AWU) for 91 FADN regions from 2007-2015

Transition probability matrix						
Initial distribution (2007)	*Class*	*<50*	*50-80*	*80-110*	*110-140*	*>140*
19	*<50*	0.876	0.062	0.041	0.021	0.000
19	*50-80*	0.163	0.337	0.337	0.152	0.011
17	*80-110*	0.060	0.180	0.413	0.240	0.107
15	*110-140*	0.040	0.089	0.290	0.331	0.250
21	*>140*	0.006	0.024	0.101	0.154	0.716
Summary statistics						
Class		*<50*	*50-80*	*80-110*	*110-140*	*>140*
Stationary distribution		0.307	0.113	0.202	0.158	0.220
Half-life		4.257				
Mobility index		0.582				
Upward mobility		0.525				
Downward mobility		0.475				

Note: EU-sample average=100, estimates are rounded to three decimals. Standard errors can be found in Appendix IV.
(Source: author, based on FADN data)

If we look at Figure 6.5, there are no signs of regions in the NMS that catch-up. Mainly, the regions in southern Sweden have done considerably well compared to the EU-average. In the north-western regions of the EU, a fall in the labour incomes can be observed. The Italian regions on the contrary have shown an increase in their labour income. From the NMS regions, only Pomorze i Mazury (Poland) has shown an increase.

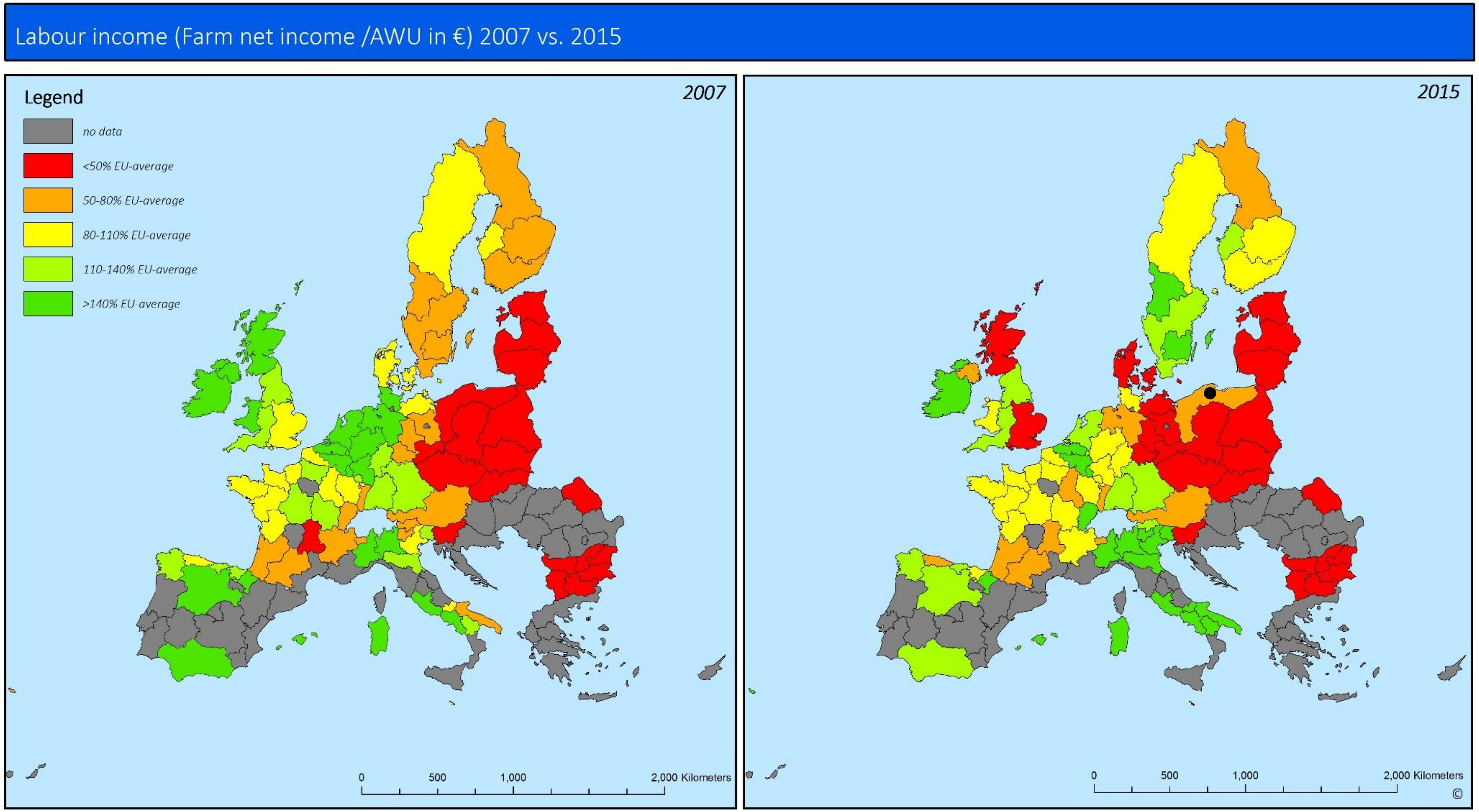


Figure 6.5. Class transition per region for labour income (farm net income/AWU) for 91 FADN regions from 2007 vs. 2015
Note: EU-sample average=100. The black dot represents NMS regions that have caught up.
(Source: author, based on FADN data)

6.5.3. Family farm income

The Kernel density plot of family farm income is shown in Figure 6.6. The distribution of family farm income is skewed to the right. The data appears to suggest that there is not a stable distribution over time. Moreover, bimodality cannot be seen in this distribution. The main peak has moved slightly to the right, which means that for a large group the family farm income has grown. The figure also shows that negative family incomes are present in the sample. Except for the year 2009, the short left tail suggests that there are only a few regions that face a small negative family farm income. As in the labour income, the spread is more or less equal over time, except for some larger spread in 2009 and 2014.

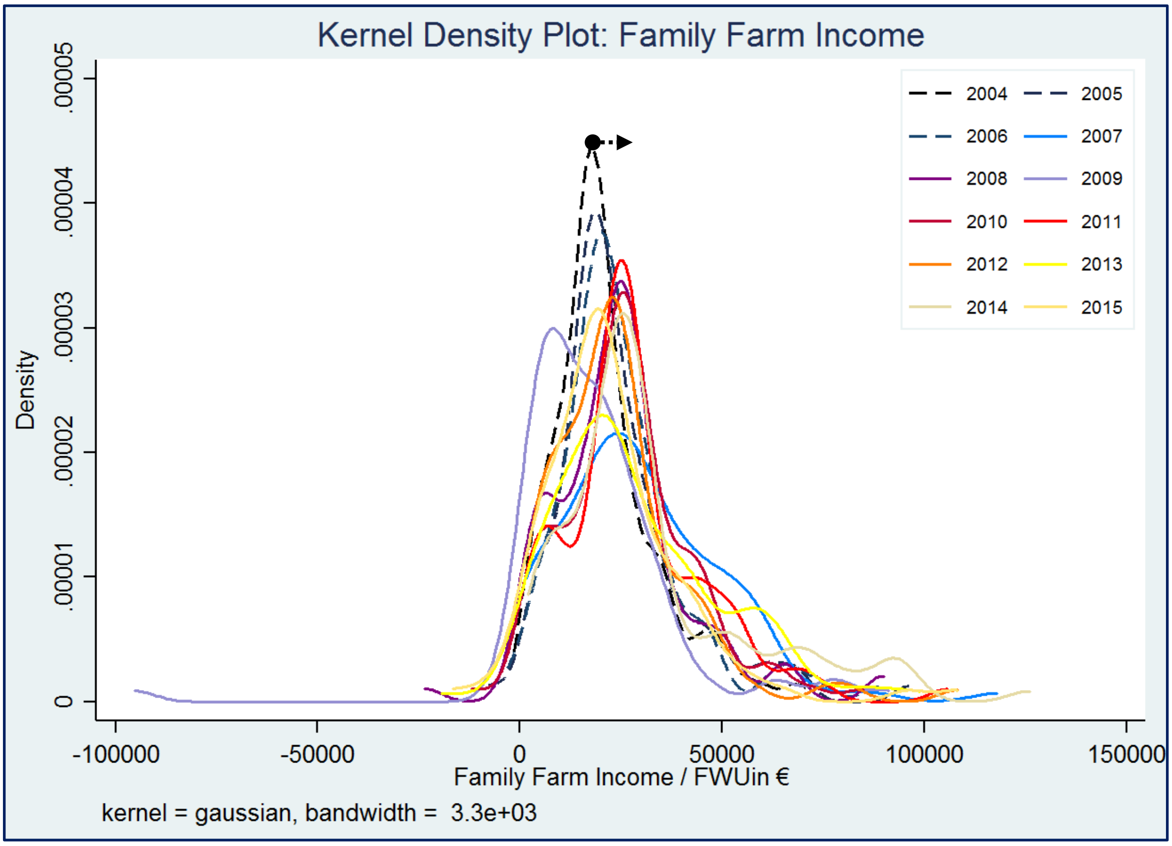


Figure 6.6. Kernel density family farm income (family farm income income/FWU), for 91 FADN regions from 2004-2015
Note: the dashed arrow shows the slight movement of the peak to the right.
(Source: author, based on FADN data)

In Table 6.6 the transition probabilities for family farm income since 2004 can be seen. Of the three variables for income, family farm income has the lowest probability ($\hat{p}_{11}$) to stay in the lowest class. There is a probability of 0.189 to move from the lowest class to the second class ($\hat{p}_{12}$). The same holds for regions in the second class (50-80) which have a probability of 0.324 to move to a higher class. So there seems to be considerable mobility in family farm income if we compare it to the other two income variables.

Table 6.6. Transition probability matrix of farmily farm income (family farm income/FWU) for 84 FADN regions from 2004-2015

Transition probability matrix						
Initial distribution (2004)	*Class*	*<50*	*50-80*	*80-110*	*110-140*	*>140*
16	*<50*	0.711	0.189	0.082	0.013	0.006
19	*50-80*	0.104	0.576	0.281	0.026	0.013
22	*80-110*	0.060	0.239	0.504	0.132	0.064
11	*110-140*	0.060	0.078	0.172	0.397	0.293
16	*>140*	0.005	0.016	0.087	0.168	0.723
Summary statistics						
Class		*<50*	*50-80*	*80-110*	*110-140*	*>140*
Stationary distribution		0.171	0.247	0.249	0.127	0.207
Half-life		2.927				
Mobility index		0.522				
Upward mobility		0.526				
Downward mobility		0.474				

Note: EU-sample average=100, estimates are rounded to three decimals. Standard errors can be found in Appendix IV.
(Source: author, based on FADN data)

In Table 6.7 the transition matrix for 2007-2015 is shown. As for every variable, the mobility in the lower class is less for the period 2007-2015. For the lowest class, the probability to stay in the lowest class is now around 0.8 ($\hat{p}_{11}$). The stationary distribution shows that also the share of the lowest class has increased compared to the previous matrix. The half-life is somewhat higher than in the previous matrix. Still the half-life is quite low. Compared to the previous matrix, we now see that the upward mobility is lower than the downward mobility, which means that downward mobility dominates for this variable.

Table 6.7. Transition probability matrix of family farm income (family farm income/FWU) for 91 FADN regions from 2007-2015

Transition probability matrix						
Initial distribution (2007)	*Class*	*<50*	*50-80*	*80-110*	*110-140*	*>140*
18	*<50*	0.812	0.079	0.085	0.012	0.012
21	*50-80*	0.094	0.493	0.341	0.043	0.029
20	*80-110*	0.071	0.217	0.489	0.158	0.065
10	*110-140*	0.091	0.143	0.312	0.234	0.221
22	*>140*	0.006	0.024	0.067	0.134	0.768
Summary statistics						
Class		*<50*	*50-80*	*80-110*	*110-140*	*>140*
Stationary distribution		0.247	0.187	0.256	0.103	0.207
Half-life		3.073				
Mobility index		0.551				
Upward mobility		0.474				
Downward mobility		0.526				

Note: EU-sample average=100, estimates are rounded to three decimals. Standard errors can be found in Appendix IV.
(Source: author, based on FADN data)

In Figure 6.7, the class transitions are portrayed on the map. In north-western Europe there is a major decline compared to the average. Regions in Germany, Netherlands, Belgium, Denmark and the UK are in a worse position than they used to be. On the other hand, improvements can be found in southern Sweden, north- and southern regions in Italy. Amongst the NMS, Slovakia, the Czech Republic and Severozapaden (Bulgaria) have improved their position.

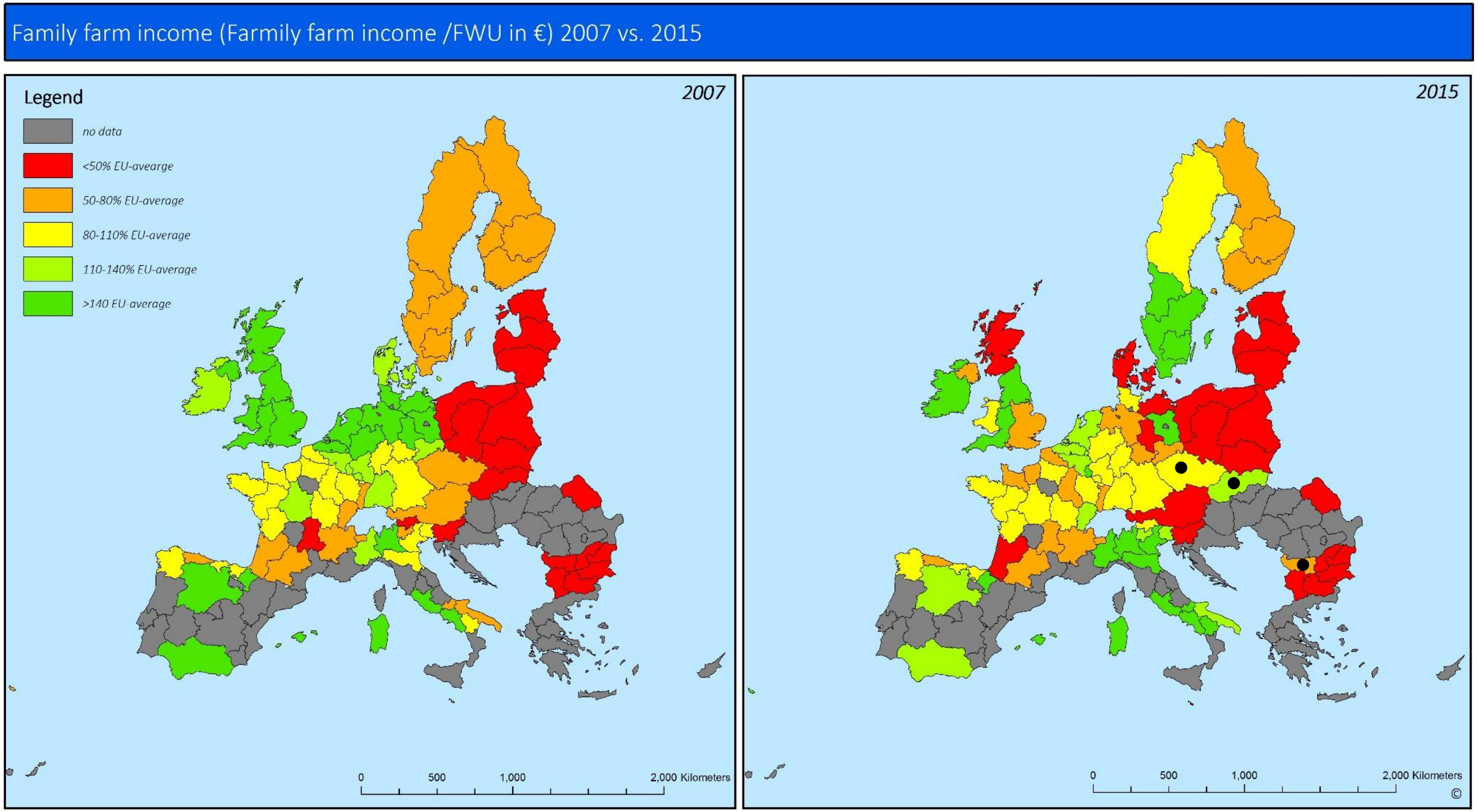


Figure 6.7. Class transition per region for family farm income (family farm income/FWU) for 91 FADN regions from 2007 vs. 2015
Note: EU-sample average=100. The black dot represents NMS regions that have caught up.
(Source: author, based on FADN data)

6.6 DISCUSSION

The available evidence seems to indicate that there has been only very limited income convergence since 2004. Particularly for labour-adjusted value added the transition probabilities along the diagonal of the matrix are very high, which points at a stable process. For labour income and family farm income there is more mobility in the lower classes, but still a strong catch-up process cannot be observed. If we look at the share of upward or downward mobility in the overall mobility, we find that the differences between upward and downward mobility is small. In the productivity analysis, we observe that upward mobility seems to be dominant for biological and mechanical productivity, but for income convergence we could not find such a pattern. Still the process towards convergence seems to be faster compared to productivity convergence, reflected in the lower half-lives. In general, the mobility index is also higher for income than for productivity, which means that there is less mobility in productivity than income. This can be explained by the fact that income fluctuates more than productivity does due to the price variation over the years. Also, non-zero probabilities for p_{15} are found, meaning that there is a chance to move from the lowest to the highest category in one step. It is not very realistic to go from the lowest to the highest income class in one step, but these transitions could be caused by region-specific extraordinary events. Remarkable relative income growth is observed in the southern Swedish regions (Slattbygdslan and Skogs-och mellanbygdslan), which moved to a higher class for all three income variables. Referring back to Table 2.3, Sweden had one of the highest scaling-up processes in the EU, since the average number of dairy cows per dairy farm almost doubled between 2004 and 2015. For the NMS regions, only the Czech Republic, Estonia and Slovakia, Pomorze i Mazury (Poland) and Severozapaden (Bulgaria) have shown an upward pattern. For the other NMS regions there has been a status-quo in their income position. The stationary distributions have also shown that in most cases, around 20% stays in the lowest income class. These distributions also show that the lowest, middle and highest class have the highest shares. In the light of a convergence process, one would like to see that the share of the lowest class declines while the second income class would increase.

Like in the analysis of productivity, it could be noticed that the probability of remaining at the tails (p_{11} and p_{55}) was structurally higher for the 2007-2015 period compared to the 2004-2015 period. It means that the six Bulgarian and one Romanian region do not contribute to a process of convergence, but rather the opposite.

The fact that we find that the convergence process is the strongest for family farm income is in line with the findings of Brasili et al. (2006), who also found that convergence was stronger for family farm income than for net value added per hectare. Perhaps this could be explained by the fact that immobile (hired) labour could hamper the process of income convergence. The probabilities of the transition matrices do not show odd values if we compare it to previous studies. Quah (1996b) has somewhat higher values along the diagonal, but this seems reasonable since he analysed cross-country convergence. Monfort (2008) has similar transition probabilities for GDP/capita convergence across EU regions along the diagonal. In his study also more zero probabilities can be found, which could be caused by the fact that GDP/capita is less fluctuating than farm income. The half-lives that are reported by Monfort show somewhat higher values than this research. In this research, the half-lives are in the range 2.927-19.866 for one-year periods, while Monfort (2008) reports half-lives between 2.6-8.4 for five-year periods. This difference could be caused by the choice of t, the transition period. In our study this is one year, while Monfort uses ten- and five-year intervals. In this study it means that half of the gap towards the stationary distribution is already closed after three years.

7 CONCLUSION

As stated in the introduction, this research was conducted in order to examine to what extent convergence took place between the dairy sectors of the EU-27 Member States since 2004. This chapter answers the sub-questions in order to define to the main conclusion of this thesis. First the sub-questions stated in the beginning are answered, followed by the final conclusion on the main research question. In the discussion we reflect on the main limitations and weaknesses of this thesis. Lastly, some suggestions for further research are given.

7.1 CONCLUSION

i) What disparities existed in the EU dairy sector and how did they change over time?

In chapter 2 it was identified that there exists substantial disparity between the regional dairy sectors in the EU. Crucially, the number of dairy cows per farm in the NMS is still very low if we compare it to the OMS. Although, there has been some scaling-up process in the NMS between 2004-2015, this went at a slower rate than the scaling-up process in the OMS. As stated in the introduction, Van Berkum (2009) pointed out that in Bulgaria, Poland and Lithuania still so many small farms existed, that these regions probably still would have to face a structural scaling-up process. However, we find that in these regions this structural scaling-up process went even slower at a slower rate than the OMS. With respect to the milk prices it is found that between 1996-2003, the milk prices of the OMS fluctuated steadily within a small bandwidth. From 2004-2017, with the introduction of the NMS this bandwidth increased. Moreover, since 2007 there has been a lot of variation in the national milk prices due to the severe price volatility in the dairy market. Given that, structural disparity existed between the NMS and the OMS, and it remained largely in place over time.

ii) How can convergence be modelled and measured?

The identification of the models and measures of convergence in economics has shown that the theory on convergence is based on different paradigms. This research has used the σ-convergence definition, i.e. whether the degree of dispersion across income levels or growth has decreased over time. The definition of β-convergence, whether poorer regions grow faster than richer regions, is not suitable for the dairy sector, because prices and output can vary heavily by year, which would not give a statistically consistent outcome. As a result, the empirical measures have followed the σ-convergence definition. The empirical approach towards convergence does also depend on the variable under enquiry. For prices it was straightforward to measure σ-convergence with the coefficient of variation, but it is hard to find the 'true' convergence process due to the volatility.

For productivity and income, we have chosen to follow the distributional approach by using Kernel density estimation and Markov chains. This distributional approach has the ability to study the shape and spread of the cross-section distribution of income (or productivity) over time. At the same time, this approach can identify individual movements along the cross-section distribution and it can give statistical interference about the process of convergence.

iii) What is the empirical evidence on convergence in prices, productivity and farm income of the dairy sector across the MS and regions of the EU?

In chapter 4, we find that in the period 2004-2007, under stable market conditions, the degree of dispersion between OMS and NMS(2004) has decreased over time. Interestingly, we find that for the 12 largest milk producing countries, the dispersion measures are much lower over time. The regression results showed that there is a significant negative linear time trend for the coefficient of variation. Hence, price dispersion decreased so there is σ-convergence. However, closer inspection of the residual

plots show that the assumptions of the linear regression are violated, thus the regression results should be treated with the utmost caution.
For productivity convergence analysis, we find that there is persistent bimodality in the distribution of biological and mechanical productivity across regions. This implicates that there is an enduring gap between a large group of productive regions and a small group of less productive regions. Moreover, Markov chains show that particularly for mechanical productivity the probability of a region that is in the least productive class (<50%-average) to transition to a higher productivity class is less than 10 percent. The overall likelihood to stay in the same class is very large for productivity. For both productivity measures it is found that the evidence for convergence decreased when we added the NMS(2007) regions to the model.
In chapter 6, the convergence of labour-adjusted value added, labour income, family farm income was analysed. Kernel density plots of all income measures show a right-skewed distribution, meaning that there is more variation in the higher income regions than in the lower income regions. From the Markov chain analysis, it is found that the probability of lowest income regions to catch up with higher income regions is around 20%, which is higher than productivity. Furthermore, we find that convergence is greater for family farm income and labour income than for labour-adjusted net value added.
Considering the three variables for which we analysed convergence, the evidence for convergence between the is scarce. Only for a small group of NMS regions there seems to be some catch-up. Perhaps, it is more likely that there is only convergence within some clubs of regions. For example, in chapter 4 we have seen that the disparity in prices is much lower for the largest producing countries. In chapter 5 and 6, we have seen bimodality in the Kernel density plots which points at the existence of two considerable groups with a clear difference in productivity and income. Taken together, the empirical evidence indicates no considerable structural convergence process across the MS and regions of the EU.

Final conclusion

As stated in the introduction, the aim of the research was to find to what extent convergence took place between the dairy sectors of the EU-27 Member States since 2004. Before interpreting our results, we remind the reader of Figure 1.1 that illustrated disparity, convergence and cohesion. Returning to this figure, we can conclude that there has been a status-quo between the dairy sectors of the EU-27. When the NMS entered there was structural disparity, which remained largely in place over the period 2004-2017. Following the definition of σ-convergence, this study has shown that the dispersion did not decrease over time. The empirical findings show that with respect to price, productivity and income convergence there is hardly evidence for a structural convergence of the regional dairy sectors. Exceptions are the NMS regions: Estonia, Slovakia, the Czech Republic; who have shown a process of catch-up in terms of income and productivity between 2004-2015. For Lithuania, Latvia, Slovenia, Poland, Romania and Bulgaria there is no evidence of a catch-up process. Especially for Poland, Lithuania, Romania and Bulgaria, this is a worrying result given the size of the dairy sector in these countries as we have seen in section 2.2.3. Our empirical results seem to correspond with Cechura et al. (2017) who also finds no signs for convergence in terms of productivity for the period 2004-2011. Likewise, Jansik and Irz (2014) find catch-up for Estonian farms, but also mention that Polish and Lithuanian farms keep lagging behind. Thus, this EU-wide study for an extended time period and adding regions lends support to the previous findings in the literature that there is only very limited evidence on convergence across the regional EU dairy sectors.

With regard to the CAP 2020 reform this research has given several policy implications. As was mentioned in the introduction, the Bulgarian minister said that a key priority for his country was to move towards greater convergence in support “to ensure a level playing field on the single market”. From chapter 2, we can conclude that at least the players in this level playing field are still structurally different. In chapter 3, it is also clear that convergence is an important goal in the European policy, but that it is hard to judge if this goal is achieved, particularly within the framework of the CAP. There is a lack of well-defined measurable targets that aim to achieve structural upward convergence of regional dairy sectors that are lagging behind. So in the new CAP, a clear structural policy goal on structural convergence accompanied with measurable targets should be designed. This should include a methodological toolbox with sufficient data sources in order to empirically evaluate the impact of structural policy. Secondly, there should be debated whether the main instrument in the CAP, the direct payments per hectare are currently distributed in the best way. A report from the European Commission (2017d) shows that in 2015 on average in the EU still about 80% of the direct payments goes to 20% of the biggest beneficiaries. In many of the NMS this percentage is even higher. It is doubtful whether such a distribution is beneficial for the long-run structural convergence process.

Considering these points, the effectiveness and accountability of policy on structural convergence in the CAP could still be improved.

Above all, this thesis shows that there is not much evidence for convergence, while most of the NMS are already an EU MS for more than 10 years now. For the regions that were in the sample it could be seen that Bulgaria, Romania, Poland, Lithuania and Latvia have not caught up with the rest of the EU. Given the size of the dairy sectors and the size of farms in these countries this is a worrying result. Convergence should therefore not be regarded as an automatic process.

7.2 Discussion

A key problem of the convergence literature, is first that there is no single theoretical definition of convergence. Hence, multiple definition leads to multiple methods that measure convergence. Moreover, convergence studies have examined numerous variables which makes it even more complex. The lack of a clear theoretical concept can lead to a weakened link between the theoretical propositions and the empirical applications. The diversity in definitions and measures can therefore lead to confusion in the academic debate on economic convergence.

From an empirical aspect this study has a number of weaknesses. First, we have chosen in this thesis to focus on specialist dairy farms. Nevertheless, this approach ignores the mixed farm types that still are much present in especially the NMS. However, to measure the income of dairy farms it is better to only use farms that gain most of their income by milk collection, since otherwise the results would be troubled by developments in the crop markets. Still, we acknowledge that by studying specialist dairy farms, we do not account for all dairy farming in the EU. Moreover, there are some weaknesses in the FADN data. The first is that certain regions do not have data, which can lead to a selection bias. For example, we have only one Romanian region in the sample, which makes it hard to draw a conclusion for that country. Out of 148 FADN, only 91 regions are in the sample. The second is that some regions only have very small sample sizes and that they therefore do not give a good representation of the population. Moreover, since the data is collected at a national level it can be that they measure certain variables differently. Also the fact that we measure convergence across territorial units (FADN) can be criticized on the fact that not every region has an equally large dairy sector. A counterargument for this is that the dairy sector is an economic sector that is in particular bound to land. Next to that, the policy and policy debate is often linked to territorial units.

Secondly, there might exists several drawbacks in the empirical models. The empirical models that were used have not been used a lot in the literature. For that reason, they might have weaknesses that we have overlooked. As a counter-argument, these models have shown to fit best to this specific research question. We could have used a neoclassical growth model with initial income vs. income growth regressions (the common workhorse), but it would not fit the specific circumstances that were present in the dairy sector. A drawback of the CV method is that we are not sure about the functional form of the model. The market volatility makes it difficult to empirically find evidence for convergence, since the volatility troubles the true convergence process. The structural break test indicated a break in the series, we tried to resolve this problem by adding dummies. However, the residual plots show that the model is far from perfect. The Kernel density plot was able to show some interesting dynamics in the distribution. Still, outliers in the distribution can change the shape of the distribution substantially (e.g. the crisis year 2009). A particular weakness of the Markov chain is that we arbitrarily chose the several classes. Although the classes are chosen by inspection of the quantiles and shares between the classes are equal. It is clear that changing the class size also influences the results. Secondly, convergence is in the real world a continuous process but in the Markov chain case this process is captured in a discrete model. Given the data this compromise was unavoidable. Since the data is aggregated over a year it is probably not a major issue since it is therefore not a quick snapshot. Thirdly, we assume stationarity in the Markov chain, however it can be questioned whether transition probabilities are identical over a turbulent period of twelve years.

7.3. RECOMMENDATIONS FOR FURTHER RESEARCH

Since this study has only studied a delineated area of convergence across dairy sectors, some suggestions for follow-up studies could be specified. First, this study has focussed on agricultural sectors in a specific territorial unit, but this spatial component has not been studied. In highly innovative regions, there are spill-overs that flow to neighbouring regions. As could be seen throughout this study is that some larger regions like north-western Europe seem to represent a cluster of very competitive regions. It would therefore be interesting to see if such clusters exist and whether they influence the convergence process. This research would fit more into the framework of the 'new economic geography' as was discussed in section 1.3.

Secondly, this study has largely neglected the national- and regional policy differentiation in the CAP. The implementation of the CAP has been different in the different MS. It would therefore be interesting to find what the implication is of the regional differentiation in implementation is on the economic performance of a region. In the European Commission (2017a, p. 10) draft on the new CAP, it is stated that: *"In line with the logic of the Commission's "budget focused on results" approach, a future delivery system should thus be more result-driven, boost subsidiarity by giving Member States a much greater role in rolling out CAP schemes,…".* This means that national or regional implementation in the future CAP, might even become more important. It is therefore most interesting to study how the regional differentiation in implementation has influenced the regional economic performance vis-à-vis other regions.

More broadly, in this thesis we did not try to identify or explain the underlying reasons of convergence. An importance issue is still how structural convergence between agricultural sectors could be achieved. This study has descriptively studied convergence, but for a better understanding of this process, it would be needed to study convergence in an explanatory way. An example could be the approach of Zimmerman and Heckelei (2012) who added a second stage in Markov chain analysis, by regressing the transition probabilities on a set of explanatory variables.

LITERATURE

- Agrafocus. (2017). Bulgarian Presidency takes over at crucial time. November 2017 – No. 260.
- Alexiadis, S. (2010). Convergence in agriculture: Evidence from the European Regions. *Agricultural Economics Review*, 11(2), 84-96.
- Anderson, T. W., and Goodman, L. A. (1957). Statistical inference about Markov chains. *The Annals of Mathematical Statistics*, 89-110.
- Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*: *Journal of the Econometric Society*, 821-856.
- Bach, C. F., Frandsen, S. E., & Jensen, H. G. (2000). Agricultural and Economy-Wide Effects of European Enlargement: Modelling the Common Agricultural Policy. *Journal of Agricultural Economics*, 51(2), 162-180.
- Baráth, L., and Ferto, I. (2014). Agricultural Productivity in the EU: A TFP Comparison between the Old (EU-15) and New (EU-10) EU Member States. In: *Agricultural Economics Society: 88th Annual Conference*.
- Baráth, L., and Fertő, I. (2017). Productivity and Convergence in European Agriculture. *Journal of Agricultural Economics*, 68(1), 228-248.
- Barro, R. J. (1991). Economic growth in a cross section of countries. *The quarterly journal of economics*, 106(2), 407-443.
- Barro, R. J., and Sala-i-Martin, X. (1992). Convergence. *Journal of political Economy,* 100(2), 223-251.
- Ben Arfa, N., Daniel, K., Jacquet, F., and Karantininis, K. (2015). Agricultural policies and structural change in French dairy farms: a nonstationary Markov model. *Canadian Journal of Agricultural Economics*, 63(1), 19-42.
- Bergmann, D., O'Connor, D., and Thümmel, A. (2015). Seasonal and cyclical behaviour of farm gate milk prices. *British Food Journal,* 117(12), 2899-2913.
- Bernard, A. B., and Durlauf, S. N. (1996). Interpreting tests of the convergence hypothesis. *Journal of Econometrics*, 71(1), 161-173.
- Bijman, J., Iliopoulos, C., Poppe, K. J., Gijselinckx, C., Hagedorn, K., Hanish, M., Hendrikse, G.W.J., Kühl, R.,Ollila, P., Pyykkönen, P. and van der Sangen, G. (2012). Support for Farmers' Co-operatives. Final report.
- Bivand, R. S., and Brunstad, R. J. (2003). Regional growth in Western Europe: an empirical exploration of interactions with agriculture and agricultural policy. In: Fingleton, B. (ed.). *European Regional Growth*. Berlin, Heidelberg: Springer, pp.351-373.

- Bonnet, C., Corre, T., & Réquillart, V. (2015). "Price Transmission in Food Chains: The Case of the Dairy Industry. S McCorriston. *Food Price Dynamics and Price Adjustment in the EU*, 65-101.
- Brasili, C., Fanfani, R., & Gutierrez, L. (2006). Convergence in the Agricultural Incomes: A Comparison between the US and EU. In: Contributed Paper to the International Association of *Agricultural Economists Conference Gold Coast, Australia,* 12(18.08), 2006.
- Brümmer, B., Glauben, T., & 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(3), 628-644.
- Caves, D. W., Christensen, L. R., & Diewert, W. E. (1982). Multilateral comparisons of output, input, and productivity using superlative index numbers. *The Economic Journal*, 92(365), 73-86.
- Cechura, L., Grau, A., Hockmann, H., Levkovych, I., & Kroupova, Z. (2017). Catching Up or Falling Behind in European Agriculture: The Case of Milk Production. *Journal of Agricultural Economics*, 68(1), 206-227.
- Chavas, J. P., and Kim, K. (2001). Hedonic Pricing of Components and Cointegration Relationships: A Dynamic Analysis of Dairy Product Prices (No. 201572).
- CLAL. (2017). EU-28: SMP - Stocks, market prices and intervention prices. Available from: https://www.clal.it/en/?section=magazzino_smp_confronto# [Last update: 14-09-17]
- Cochrane, W. W. (1958). *Farm prices: myth and reality*. U of Minnesota Press.
- De Bont, C.J.A.M., van Everdingen, W.H., de Jager, J.H., Sengers, H.H.W.J.M. and de Vlieger, J.J. (2003). De melkprijs in beweging: Gevolgen van Europese zuivelhervorming voor de melkveehouderij. *LEI Wageningen UR* (Rapport 6.03.14).
- Deutsche Milk Kontor. (2017). DMK GROUP reacts to crisis year with far-reaching measures. Press release 29 June 2017. Available from: https://www.dmk.de/en/press/press-release/dmk-group-reagiert-mit-tiefgreifenden-massnahmen-auf-krisenjahr 1/?no_cache=1&cHash=975af43d1456859e013eda5f6b48dc50 [Accessed: 29-12-2017]
- Dreger, C., Kholodilin, K., Lommatzsch, K., Slačálek, J., and Wozniak, P. (2008). Price convergence in an enlarged internal market. *Eastern European Economics*, 46(5), 57-68.
- Dunford, M., and Smith, A. (2000). Catching up or falling behind? Economic performance and regional trajectories in the "New Europe". *Economic Geography*, 76(2), 169-195.
- Ernst and Young. (2013). Analysis on future developments in the milk sector. Prepared for the European Commission - DG Agriculture and Rural Development, AGRI-2012-C4-04. Available from: https://ec.europa.eu/agriculture/sites/agriculture/files/events/2013/milk-conference/ernst-and-young-report_en.pdf [Accessed: 02-11-2017]
- European Commission. (2002). Enlargement and agriculture: Summit adopts fair and tailor-made package which benefits countries in accession countries. Brussel/Kopenhagen, IP/02/1882

- European Commission. (2009). Communication from the Commission to the Council - Dairy market situation 2009 {SEC(2009) 1050}. Brussels, 22.7.2009, COM(2009) 385 final
- European Commission. (2010a). Commission proposes new measures to improve future stability in the dairy sector. IP/10/1691
- European Commission. (2010b). Report from the European Commission to the European Parliament and the Council: Evolution of the market situation and the consequent conditions for smoothly phasing out the milk quota system. Available from: https://ec.europa.eu/agriculture/milk/quota-report/com-2010-727_en.pdf
- European Commission. (2012). Agriculture in the European Union Statistical and Economic Information. Available from: https://ec.europa.eu/agriculture/sites/agriculture/files/statistics/agricultural/2012/pdf/full-report_en.pdf [Accessed: 02-07-2017]
- European Commission. (2014). European Structural and Investment Funds. Available from: http://ec.europa.eu/regional_policy/en/funding/ [Last update: 15-08-14]
- European Commission. (2015). European Commission presents a €500 million comprehensive package of measures to support European farmers. Press release IP/15/5599.
- European Commission. (2016a). European Commission activates exceptional measures to further support European farmers in crisis. Press release IP/16/806.
- European Commission. (2016b). EU AGRICULTURAL OUTLOOK: Prospect for the EU agricultural markets and income 2016-2026. Medium-term Outlook report. Available from: https://ec.europa.eu/agriculture/markets-and-prices/medium-term-outlook_en
- European Commission. (2017a). Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions The Future Of Food And Farming. Brussels, 29-11-2017 COM(2017) 713 final. Available from: https://ec.europa.eu/agriculture/sites/agriculture/files/future-of-cap/future_of_food_and_farming_communication_en.pdf [Accessed: 29-11-2017]
- European Commission. (2017b). Commodity Price Dashboard No 57 – February 2017 edition. Available from: https://ec.europa.eu/agriculture/markets-and-prices/price-monitoring_en
- European Commission. (2017c). My Region, My Europe Our Future: Seventh report on economic, social and territorial cohesion. Available from: http://ec.europa.eu/regional_policy/sources/docoffic/official/reports/cohesion7/7cr.pdf
- European Commission. (2017d). Report on the distribution of direct payments. Available from: https://ec.europa.eu/agriculture/sites/agriculture/files/cap-funding/beneficiaries/direct-aid/pdf/annex2-2016_en.pdf [Accessed: 18-02-2018]

- European Union. (2012). Consolidated version of the Treaty on the Functioning of the European Union: Article 39. OJ C 326, 26.10.2012: 47–390.
- European Union. (1992). Treaty on European Union (Consolidated Version), Treaty of Maastricht. *Official Journal of the European Communities C 325/5; 24 December 2002*. Available from: http://www.refworld.org/docid/3ae6b39218.html [Accessed: 12-02-2018]
- Eurostat (2015). Figure 5: Food and non-alcoholic beverages, Member States, 2015, % of total expenditure. Available from: http://ec.europa.eu/eurostat/statistics-explained/index.php/File:Figure_5_Food_MSs.png [Accessed: 04-12-2017]
- Eurostat. (2017). Collection of cow's milk in 1000 tonnes. Available from: http://ec.europa.eu/eurostat/tgm/download.do?tab=table&plugin=1&language=en&pcode=tag00037
- Fackler, P. L., and Goodwin, B. K. (2001). Spatial price analysis. In: Gardner, B. and Rausser, G. (eds.). *Handbook of agricultural economics* 1. North Holland, pp. 971-1024.
- FADN. (2010). FADN Standard Results : income indicators. Available from: http://ec.europa.eu/agriculture/rica/annex003_en.cfm#ii [Last update 25-01-2010]
- FADN. (2014). Diffusion. Available from: http://ec.europa.eu/agriculture/rica/diffusion_en.cfm [Accessed: 08-06-2017].
- FADN. (2017a). Farm Net Income (SE420)FADN Public Database: Year(2004), Country(All), Regions(All), Type of farming(5: Milk). Available from: http://ec.europa.eu/agriculture/rica/database/report_en.cfm?dwh=SO [Accessed: 02-07-2017]
- FADN. (2017b). Dairy cows (SE085). FADN Public Database: Year(2004-2015), Country(All), Type of farming(45: Specialist dairying). Available from: http://ec.europa.eu/agriculture/rica/database/report_en.cfm?dwh=SO [Accessed: 02-10-2017]
- FADN. (2018a). Standard Results Database: YEAR.COUNTRY.REGION.TF14.zip. Available from: http://ec.europa.eu/agriculture/rica/database/consult_std_reports_en.cfm [Downloaded: 11-01-2018]
- FADN. (2018b). Definition of variables. Available from: http://ec.europa.eu/agriculture/rica/definitions_en.cfm [Accessed: 9-1-2018]
- Fingleton, B., and López-Bazo, E. (2003). Explaining the distribution of manufacturing productivity in the EU regions. In: Fingleton, B. (ed.). *European Regional Growth*. Berlin, Heidelberg: Springer, pp. 375-409.
- Fischer, C. (2012). Price convergence in the EMU? Evidence from micro data. *European Economic Review,* 56(4), 757-776.
- Friedman, M. (1992). Do Old Fallacies Ever Die? *Journal of Economic Literature*, 30(4), 2129-2132.

- Fujita, M., and Krugman, P. (2004). The new economic geography: Past, present and the future. *Papers in regional science*, 83(1), 139-164.
- Fujita, M., Krugman, P. R., Venables, A. J., and Fujita, M. (1999). *The spatial economy: cities, regions and international trade* (Vol. 213). Cambridge, MA: MIT press.
- Geppert, K., and Stephan, A. (2008). Regional disparities in the European Union: Convergence and agglomeration. *Papers in Regional Science,* 87(2), 193-217.
- Gil-Pareja, S., and Sosvilla-Rivero, S. (2008). Price convergence in the European car market. *Applied Economics,* 40(2), 241-250.
- Gocht, A., Röder, N., Neuenfeldt, S., Storm, H., and Heckelei, T. (2012). Modelling farm structural change: A feasibility study for ex-post modelling utilizing FADN and FSS data in Germany and developing an ex-ante forecast module for the CAPRI farm type layer baseline. *Institute for Prospective and Technological Studies, Joint Research Centre,* (No. JRC75524).
- Goldberg, P. K., and Verboven, F. (2005). Market integration and convergence to the Law of One Price: evidence from the European car market. *Journal of international Economics*, 65(1), 49-73.
- Gujarati, D. (1992). *Essentials of Econometrics.* Singapore: McGraw-Hill.
- Guolo, A., & Varin, C. (2014). Beta regression for time series analysis of bounded data, with application to Canada Google® Flu Trends. *The Annals of Applied Statistics*, 74-88.
- Hansen, H., and Teuber, R. (2011). Assessing the impacts of EU's common agricultural policy on regional convergence: sub-national evidence from Germany. *Applied Economics*, 43(26), 3755-3765.
- Hayami, Y., and Ruttan, V. W. (1971*). Agricultural development: an international perspective*. Baltimore, Md/London: The Johns Hopkins Press.
- Heijman, W. J., and Schipper, R. A. (2010). *Space and Economics: An introduction to regional economics* (Vol. 7). Wageningen Academic Publishers.
- Huettel, S., and Jongeneel, R. (2011). How has the EU milk quota affected patterns of herd-size change? *European Review of Agricultural Economics*, 38(4), 497-527.
- IEG Policy. 2015. Five predictions for the EU dairy sector post-milk quotas. Available from: https://iegpolicy.agribusinessintelligence.informa.com/PL025393/Five-predictions-for-the-EU-dairy-sector-postmilk-quotas [Accessed: 25-01-2018]
- Ihle, R., Dries, L., Jongeneel, R., Venus, T. and Wesseler, J. (2017). The EU Cattle Sector: Challenges and Opportunities - Milk and Meat. Research for Agri Committee, IP/B/AGRI/IC/2016-014. Available from: http://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL_STU(2017)585911
- Islam, N. (2003). What have we learnt from the convergence debate? *Journal of economic surveys*, 17(3), 309-362.

- Jambor, A., Somai, M., and Kovács, S. (2016). 10 Years of EU Membership: Diverging Performances in New Member States Agriculture. *Ekonomický časopis*, 64(5), 458-476.
- Jansik, C., and Irz, X. (2015a). Competitiveness Makes a Difference in the European Dairy Sector. *EuroChoices*, 14(3), 12-19.
- Jansik, C., and Irz, X. (2015b). Competitiveness of dairy farms in northern Europe: A cross-country analysis. *Agricultural and Food Science*, 24(3), 206-218.
- Jansik, C., Irz, X., and Kuosmanen, N. (2014). Competitiveness of Northern European dairy chains. *MTT Agrifood Research.*
- Jongeneel, R. (2011). European Union dairy policy reform: impact and challenges. In: Moreddu, C. (ed.), *Disaggregated Impacts of CAP Reforms: Proceedings of an OECD Workshop*. OECD publishing, pp. 93-110.
- Jongeneel, R., Van Berkum, S. (2015). What will happen after the EU milk quota system expires in 2015?: an assessment of the Dutch dairy sector. *LEI Wageningen UR*, (No. 2015-041).
- Kinnucan, H. W., and Forker, O. D. (1987). Asymmetry in farm-retail price transmission for major dairy products. *American Journal of Agricultural Economics*, 69(2), 285-292.
- Kumbhakar, S. C., Ghosh, S., and McGuckin, J. T. (1991). A generalized production frontier approach for estimating determinants of inefficiency in US dairy farms. *Journal of Business & Economic Statistics*, 9(3), 279-286.
- Kuokkanen, K., and Vihinen, H. (2006). Contribution of the CAP to the general objectives of the EU. *MTT Economic Research.*
- Lafougère, C. (2012). World and EU dairy through 2016, Presentation on the joint meeting of the Management Committee and the Advisory Group on Milk on 12/12/2012, European Commission, Agriculture and Rural Development, http://ec.europa.eu/agriculture/milk/background/ [Accessed: 13-12-2017]
- Lawler, G. F. (2006). *Introduction to stochastic processes.* CRC Press.
- Le Gallo, J. (2004). Space-time analysis of GDP disparities among European regions: A Markov chains approach. *International Regional Science Review*, 27(2), 138-163.
- Liu, Y., Richard Shumway, C., Rosenman, R., and Ball, V. E. (2011). Productivity growth and convergence in US agriculture: new cointegration panel data results. *Applied economics,* 43(1), 91-102.
- Mankiw, N. G., Romer, D., and Weil, D. N. (1992). A contribution to the empirics of economic growth. *The quarterly journal of economics*, 107(2), 407-437.
- Matthews, A. (2014). What is Happening to EU Agricultural Productivity Growth? Available from: http://capreform.eu/what-is-happening-to-eu-agricultural-productivity-growth/ [Last update: 04-05-2014]

- Meester, G., Berkhout, P., and Dries, L.K.E. (Eds.). (2013). *EU-beleid voor landbouw, voedsel en groen: van politiek naar praktijk*. 4th edition. Wageningen Academic Publishers.
- Merriam-Webster. (2018). Convergence. Available from: https://www.merriam-webster.com/dictionary/convergence [Accessed: 09-01-2018]
- Milk Market Observatory. (2017a). Historical EU Price Serie Of Cow's Raw Milk In Euro/100 Kg. Regulation (EU) No 2017/1185 Article 12(a) - Annex II.4(a)
- Milk Market Observatory. (2017b). EU historical intervention / stocks. Available from: https://ec.europa.eu/agriculture/sites/agriculture/files/market-observatory/milk/pdf/eu-historical-stocks-series_en.pdf [Last update: 23-02-2017]
- Miller, S. M., and Upadhyay, M. P. (2002). Total factor productivity and the convergence hypothesis. *Journal of Macroeconomics*, 24(2), 267-286.
- Monfort, P. (2008). Convergence of EU regions: Measures and evolution. Brussels: European Commission, Regional Policy.
- O'Connor, D. and Keane, M. (2011). Empirical Issues Relating to Dairy Commodity Price Volatility. In: Piot-Lepetit, I and M'Barek, R. (eds.). *Methods to Analyse Agricultural Commodity Price Volatility*. New York: Springer, pp. 63-84.
- O'Connor, D., Bergmann, D., and Keane, M. (2015). The challenges posed by price volatility in the EU dairy sector. Conference Paper: *Agrarian Perspectives XXIV and 25th Annual Conference of the Austrian Society of Agricultural Economics, at Prague*.
- Officer, L. H. (1986). The law of one price cannot be rejected: two tests based on the tradable/nontradable price ratio. *Journal of Macroeconomics*, 8(2), 159-182.
- Parsley, D. C., and Wei, S. J. (1996). Convergence to the law of one price without trade barriers or currency fluctuations. *The Quarterly Journal of Economics*, 111(4), 1211-1236.
- Pellegrini, G. (2002). Proximity, polarization, and local labor market performances. *Networks and Spatial Economics,* 2(2), 151-173.
- Phillips, P. C., and Sul, D. (2007). Some empirics on economic growth under heterogeneous technology. *Journal of Macroeconomics,* 29(3), 455-469.
- Polet, Y. (2015). A Perfect Storm for EU Dairy Price. Brussels, 14 August 2015. USDA Gain Report No. E15033. Available from: https://gain.fas.usda.gov/Recent%20GAIN%20Publications/A%20Perfect%20Storm%20for%20EU%20Dairy%20Prices_Brussels%20USEU_EU-28_8-19-2015.pdf [Accessed: 04-12-2017]
- Polet, Y. and Kuypers, K. (2017). EU Dairy Markets Still Looking for Equilibrium after Dairy Crisis. Brussels, 3 March 2017. USDA Gain Report No. E 17022. Available from: https://gain.fas.usda.gov/Recent%20GAIN%20Publications/EU%20Dairy%20Markets%20Still%20L

ooking%20for%20Equilibrium%20after%20Dairy%20Crisis_Brussels%20USEU_Belgium%20EU-28_3-10-2017.pdf [Accessed: 04-12-2017]

- Quah, D. (1993). Galton's fallacy and tests of the convergence hypothesis. *The Scandinavian Journal of Economics*, 427-443.
- Quah, D. T. (1996a). Regional convergence clusters across Europe. *European Economic review*, 40(3), 951-958.
- Quah, D. T. (1996b). Empirics for economic growth and convergence. *European Economic review,* 40(6), 1353-1375.
- Quah, D. T. (1997). Empirics for growth and distribution: stratification, polarization, and convergence clubs. *Journal of Economic Growth*, 2(1), 27-59.
- Ragonnaud, G. (2014). Trends In EU-Third Countries Trade Of Milk And Dairy Products. 09 September 2014, IP/B/AGRI/NT/2014-09. Available from: http://www.europarl.europa.eu/RegData/etudes/STUD/2015/529076/IPOL_STU(2014)529076_EN.pdf
- Ray, S. C. (2004). *Data envelopment analysis: theory and techniques for economics and operations research*. Cambridge University Press.
- Rey, S. J. (2001). Spatial empirics for economic growth and convergence. *Geographical Analysis,* 33(3), 195-214.
- Rezitis, A. N. (2010). Agricultural productivity and convergence: Europe and the United States. *Applied Economics,* 42(8), 1029-1044.
- Richardson, J. D. (1978). Some empirical evidence on commodity arbitrage and the law of one price. *Journal of International Economics,* 8(2), 341-351.
- Rogers, J. H. (2007). Monetary union, price level convergence, and inflation: How close is Europe to the USA? *Journal of Monetary economics,* 54(3), 785-796.
- Romer, P. M. (1990). Endogenous technological change. *Journal of political Economy*, 98(5, Part 2), S71-S102.
- Sala-i-Martin, X. X. (1996). The classical approach to convergence analysis. *The Economic Journal*, 1019-1036.
- Salou, T., van der Werf, H. M., Levert, F., Forslund, A., Hercule, J., and Le Mouël, C. (2017). Could EU dairy quota removal favour some dairy production systems over others? The case of French dairy production systems. *Agricultural Systems*, 153, 1-10.
- Schmutzler, A. (1999). The new economic geography. *Journal of Economic Surveys*, 13(4), 355-379.
- Shorrocks, A. F. (1978). The measurement of mobility. *Econometrica: Journal of the Econometric Society,* 1013-1024.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis (Vol. 26).* CRC press.

- Smit, R. (2015). FrieslandCampina kan toevoer melk niet meer aan. *Financieel Dagblad*, 28 December 2015. Available from: https://fd.nl/ondernemen/1133272/frieslandcampina-kan-de-toevoer-van-melk-niet-meer-aan
- Soares, F. B., and Ronco, R. (2000). Agricultural Income and Productivity in the European Union: Convergence or Divergence among Members? (No. 0-20). *International Centre for Economic Research.*
- Solow, R. M. (1956). A contribution to the theory of economic growth. *The quarterly journal of economics,* 70(1), 65-94.
- Sosvilla-Rivero, S., and Gil-Pareja, S. (2004). Price convergence in the European Union. *Applied Economics Letters*, 11(1), 39-47.
  to agricultural producers (financial year2016). Available from: https://ec.europa.eu/agriculture/sites/agriculture/files/cap-funding/beneficiaries/direct-aid/pdf/annex2-2016_en.pdf [Accessed: 16-02-2018]
- Tonini, A., and Jongeneel, R. (2009). The distribution of dairy farm size in Poland: a Markov approach based on information theory. *Applied Economics,* 41(1), 55-69.
- Van Ammelrooy, P. (2017). Melkveehouders stoppen massaal: 'stoppot' van ministerie binnen 24 uur leeg. *Volkskrant,* 21 February 2017. Available from: https://www.volkskrant.nl/economie/melkveehouders-stoppen-massaal-stoppot-van-ministerie-binnen-24-uur-leeg~a4465413/
- Van Berkum, S. (2009). Dairy chain competitiveness in EU's new Member States, candidate and potential candidate countries. *Agrarwirtschaft,* 58 (7).
- Van Berkum, S., and Helming, J. (2006). European dairy policy in the years to come: impact of quota abolition on the dairy sector. *LEI Wageningen UR*.
- Van Dam, M.H.P. (2017). Regeling van de Staatssecretaris van Economische Zaken van 16 februari 2017, nr. WJZ/17023701, houdende invoering van de verplichting tot betaling van een geldsom (Regeling fosfaatreductieplan 2017). *Staatscourant*, 17 February 2017.
- Versteijlen, H. (2013). Market situation and perspectives. Presentation held at the conference "The EU dairy sector: developing beyond 2015 Brussels, 24.09.13 Available from: https://ec.europa.eu/agriculture/sites/agriculture/files/events/2013/milk-conference/versteijlen_en.pdf [Accessed 22-10-2017]
- Williamson, J. G. (1965). Regional inequality and the process of national development: a description of the patterns. *Economic development and cultural change*, 13(4, Part 2), 1-84.
- Zimmermann, A., and Heckelei, T. (2012). Structural change of European dairy farms–A cross-regional analysis. *Journal of Agricultural Economics*, 63(3), 576-603.

- Zimmermann, A., Heckelei, T., & Domínguez, I. P. (2009). Modelling farm structural change for integrated ex-ante assessment: review of methods and determinants. *Environmental Science & Policy*, 12(5), 601-618.

APPENDICES

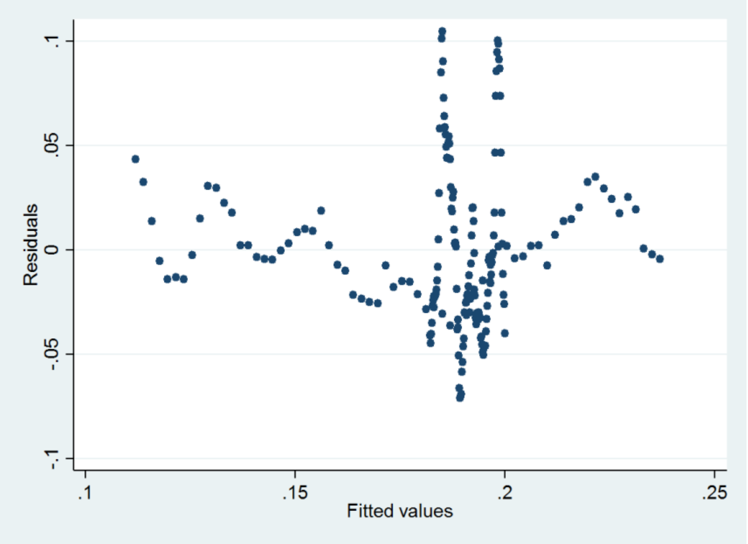
APPENDIX I: FADN REGIONS INCLUDED IN THE SAMPLE

Region name	*Code*	*Region name*	*Code*	*Region name*	*Code*	*Region name*	*Code*
Schleswig-Holstein	10	Bretagne	163	Denmark	370	Lan i norra	730
Niedersachsen	30	Poitou-Charentes	164	Ireland	380	Czech Republic	745
Nordrhein-Westfalen	50	Aquitaine	182	England-North	411	Estonia	755
Hessen	60	Midi-Pyrénées	183	England-East	412	Latvia	770
Rheinland-Pfalz	70	Rhônes-Alpes	192	England-West	413	Lithuania	775
Baden-Württemberg	80	Auvergne	193	Wales	421	Malta	780
Bayern	90	Valle d'Aoste	221	Scotland	431	Pomorze-Muzurie	785
Saarland	100	Piemonte	222	Northern Ireland	441	Wielkpolska-Slask	790
Brandenburg	112	Lombardia	230	Galicia	500	Mazowsze-Podlasie	795
Mecklenburg-Vorpommern	113	Trentino	241	Asturias	505	Malopolska-Pogórze	800
Sachsen	114	Alto-Adige	242	Cantabria	510	Slovakia	810
Sachsen-Anhalt	115	Veneto	243	Pais Vasco	515	Slovenia	820
Thueringen	116	Friuli-Venezia	244	Navarra	520	*Severozapaden*	*831*
Champagne-Ardenne	131	Emilia-Romagna	260	Baleares	540	*Severen tsentralen*	*832*
Picardie	132	Lazio	291	Castilla-León	545	*Severoiztochten*	*833*
Haute-Normandie	133	Molise	301	Andalucia	575	*Yugozapaden*	*834*
Centre	134	Campania	302	Açores e da Madeira	650	*Yuzhen tsentralen*	*835*
Basse-Normandie	135	Puglia	311	Austria	660	*Yugoiztochen*	*836*
Bourgogne	136	Basilicata	312	Etela-Suomi	670	*Nord-Est*	*840*
Nord-Pas-de-Calais	141	Sardegna	330	Sisa-Suomi	680		
Lorraine	151	Vlaanderen	341	Pohjanmaa	690		
Alsace	152	Wallonie	343	Pohjois-Suomi	700		
Franche-Comté	153	Luxembourg	350	Slattbygdslan	710		
Pays de la Loire	162	The Netherlands	360	Skogs-och mellanbygdslan	720		

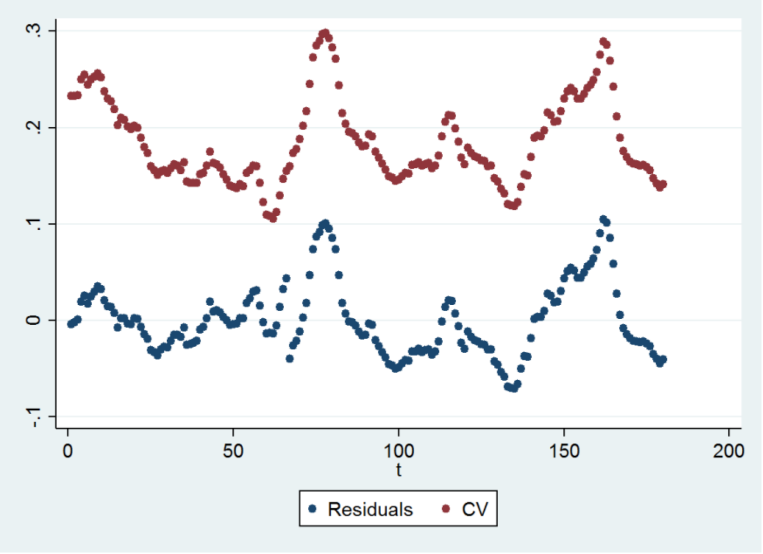
APPENDIX II CHAPTER 4 STATA-OUTPUT

```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   14.1   Copyright 1985-2015 StataCorp LP
  Statistics/Data Analysis            StataCorp
                                      4905 Lakeway Drive
                                      College Station, Texas 77845 USA
                                      800-STATA-PC        http://www.stata.com
                                      979-696-4600        stata@stata.com
                                      979-696-4601 (fax)

10-user Stata network perpetual license:
       Serial number:  301406234271
         Licensed to:  WUR
                       Wageningen UR

Notes:
      1.  Unicode is supported; see help unicode_advice.

. do "C:\Users\jong285\AppData\Local\Temp\STD01000000.tmp"

. *import file OMSNMS2004*
. import excel "\\WURNET.NL\Homes\jong285\My Documents\MScThesis\Text thesis\Data files\Chapter 4\Files for STATA\RegOMSNMS2004.xlsx",
sheet("Sheet1") firstrow

. *assign as time series*
. tsset t
        time variable:  t, 1 to 180
                delta:  1 unit

. *structural break test*
. reg CV t

      Source |       SS           df       MS      Number of obs   =       180
-------------+----------------------------------   F(1, 178)       =      0.01
       Model |  .000018424         1  .000018424   Prob > F        =    0.9235
    Residual |  .354393097       178  .001990972   R-squared       =    0.0001
-------------+----------------------------------   Adj R-squared   =   -0.0056
       Total |  .354411521       179  .001979953   Root MSE        =    .04462

------------------------------------------------------------------------------
          CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |  -6.16e-06     .000064    -0.10   0.923    -.0001325    .0001202
       _cons |   .1854866   .0066794    27.77   0.000     .1723055    .1986676
------------------------------------------------------------------------------

. *unknown break test*
. estat sbsingle
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100
..........................

Test for a structural break: Unknown break date

                             Number of obs =        180

Full sample:                          1 -        180
Trimmed sample:                      28 -        154
Estimated break date:                66
Ho: No structural break

     Test            Statistic          p-value
-----------------------------------------------
     swald           72.4089            0.0000
-----------------------------------------------
Exogenous variables:           t
Coefficients included in test: t _cons

. *test known break dates from intervention in SMP*
. estat sbknown, break(75 82) breakvars(t)

Wald test for a structural break: Known break date

                                  Number of obs  =        180
Sample:               1 -        180
Break date:  75         82
Ho: No structural break

             chi2(2)      =   76.8021
             Prob > chi2  =    0.0000
```

```
Exogenous variables:           t
Coefficients included in test: t

. estat sbknown, break(151 168) breakvars(t)

Wald test for a structural break: Known break date

                                  Number of obs  =        180
Sample:               1 -        180
Break date:  151        168
Ho: No structural break

             chi2(2)      =   55.2430
             Prob > chi2  =    0.0000

Exogenous variables:           t
Coefficients included in test: t

. *create dummy variable based on break test*
. gen D= 0

. replace D= 1 if t>66
(114 real changes made)

. *run regression for linear model with dummy based on structural break test OMS+NMS2004*
. gen Dt=D*t

. reg CV t D Dt

      Source |       SS           df       MS      Number of obs   =       180
-------------+----------------------------------   F(3, 176)       =     24.13
       Model |  .103284775         3  .034428258   Prob > F        =    0.0000
    Residual |  .251126746       176  .001426857   R-squared       =    0.2914
-------------+----------------------------------   Adj R-squared   =    0.2793
       Total |  .354411521       179  .001979953   Root MSE        =    .03777

------------------------------------------------------------------------------
          CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |   -.001924   .0002441    -7.88   0.000    -.0024057   -.0014423
           D |  -.0283658   .0166515    -1.70   0.090    -.0612281    .0044965
          Dt |   .0017664   .0002667     6.62   0.000     .0012401    .0022927
       _cons |    .238853   .0094059    25.39   0.000       .22029    .2574159
------------------------------------------------------------------------------

. eststo linearOMSNMS2004_1

. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of CV

         chi2(1)      =     3.43
         Prob > chi2  =   0.0640

. estat dwatson

Durbin-Watson d-statistic(  4,   180) =  .0988882

. rvfplot

. graph export resvsfitted1.png
(file resvsfitted1.png written in PNG format)
```



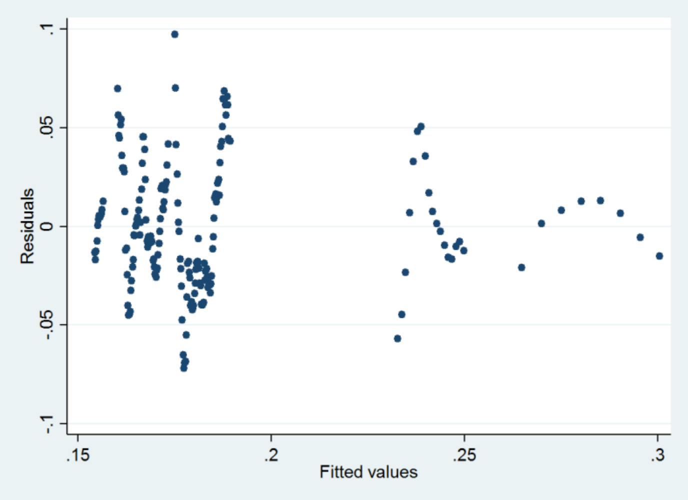

```
. predict res1, residuals

. twoway (scatter res1 t) (scatter CV t)

. graph export resvsCV1.png
(file resvsCV1.png written in PNG format)
```



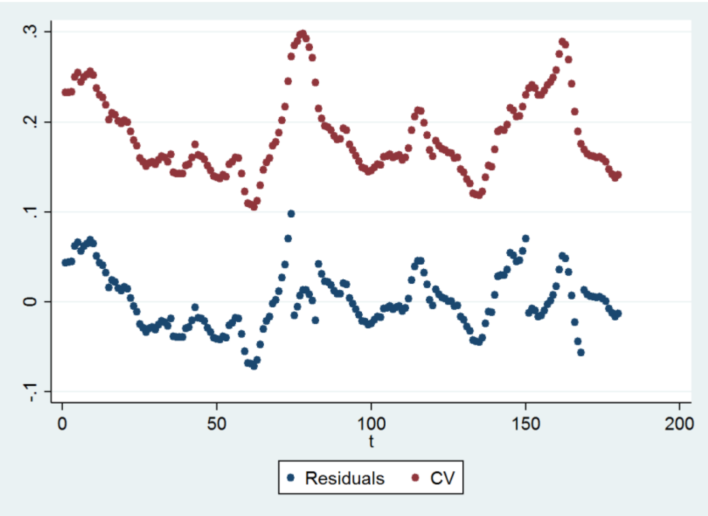

```
. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
linearOMSN~1 |        180  305.3139    336.319       4   -664.6379  -651.8661
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

. test _b[t]=0

 ( 1)  t = 0

       F(  1,   176) =   62.14
            Prob > F =    0.0000

. local sign_t = sign(_b[t])

. display "Ho: coef <= 0  p-value = "ttail(r(df_r),`sign_t'*sqrt(r(F)))
Ho: coef <= 0  p-value = 1

. display "Ho: coef >= 0  p-value = " 1-ttail(r(df_r),`sign_t'*sqrt(r(F)))
Ho: coef >= 0  p-value = 1.609e-13

.
. *run regresssion for linear-log model with dummy based on structural break test OMS+NMS2004*
. gen ln_t=ln(t)
```

```
. gen Dln_t=D*ln_t

. reg CV ln_t Dln_t D

      Source |       SS           df       MS      Number of obs   =       180
-------------+----------------------------------   F(3, 176)       =     23.96
       Model |  .102783174         3  .034261058   Prob > F        =    0.0000
    Residual |  .251628347       176  .001429707   R-squared       =    0.2900
-------------+----------------------------------   Adj R-squared   =    0.2779
       Total |  .354411521       179  .001979953   Root MSE        =    .03781

------------------------------------------------------------------------------
          CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ln_t |  -.0400263   .0051739    -7.74   0.000    -.0502371   -.0298155
       Dln_t |   .0149684   .0136134     1.10   0.273    -.0118982     .041835
           D |   .0068593   .0627253     0.11   0.913    -.1169313    .1306498
       _cons |   .3038982   .0173742    17.49   0.000     .2696096    .3381868
------------------------------------------------------------------------------

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        180  305.3139   336.1394       4   -664.2787  -651.5069
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

. *run regression for log-linear model with dummy based on structural break test OMS+NMS2004*
. gen ln_CV=ln(CV)

. reg ln_CV t D Dt

      Source |       SS           df       MS      Number of obs   =       180
-------------+----------------------------------   F(3, 176)       =     29.33
       Model |  3.22811974         3  1.07603991   Prob > F        =    0.0000
    Residual |  6.45599604       176  .036681796   R-squared       =    0.3333
-------------+----------------------------------   Adj R-squared   =    0.3220
       Total |  9.68411579       179  .054101206   Root MSE        =    .19152

------------------------------------------------------------------------------
       ln_CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |  -.0107996   .0012375    -8.73   0.000    -.0132418   -.0083573
           D |   -.169494   .0844284    -2.01   0.046    -.3361164   -.0028716
          Dt |   .0099866   .0013522     7.39   0.000     .0073179    .0126553
       _cons |  -1.411578   .0476911   -29.60   0.000    -1.505698   -1.317458
------------------------------------------------------------------------------

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        180  7.613351   44.10634       4   -80.21268  -67.44086
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

. *run regression for linear model with dummy's based on intervention data OMS+NMS2004*
. gen D_1t=D_1*t

. gen D_2t=D_2*t

. reg CV t D_1 D_2 D_1t D_2t

      Source |       SS           df       MS      Number of obs   =       180
-------------+----------------------------------   F(5, 174)       =     32.48
       Model |  .171102458         5  .034220492   Prob > F        =    0.0000
    Residual |  .183309063       174    .0010535   R-squared       =    0.4828
-------------+----------------------------------   Adj R-squared   =    0.4679
       Total |  .354411521       179  .001979953   Root MSE        =    .03246

------------------------------------------------------------------------------
          CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |  -.0001955    .000052    -3.76   0.000    -.0002981    -.000093
         D_1 |   .4937023   .3933535     1.26   0.211    -.2826561    1.270061
         D_2 |    .211524   .2353752     0.90   0.370    -.2530339     .676082
        D_1t |  -.0049065   .0050086    -0.98   0.329    -.0147919     .004979
        D_2t |   -.000806   .0014755    -0.55   0.586    -.0037181    .0021062
```

```
       _cons |   .1895039    .005047    37.55   0.000     .1795426    .1994651
------------------------------------------------------------------------------

. eststo linearOMSNMS2004_2

. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of CV

         chi2(1)      =     0.83
         Prob > chi2  =   0.3626

. estat dwatson

Durbin-Watson d-statistic(  6,   180) =  .2441764

. rvfplot

. graph export resvsfitted2.png
(file resvsfitted2.png written in PNG format)
```




```
. predict res2, residuals

. twoway (scatter res2 t) (scatter CV t)

. graph export resvsCV2.png
(file resvsCV2.png written in PNG format)
```

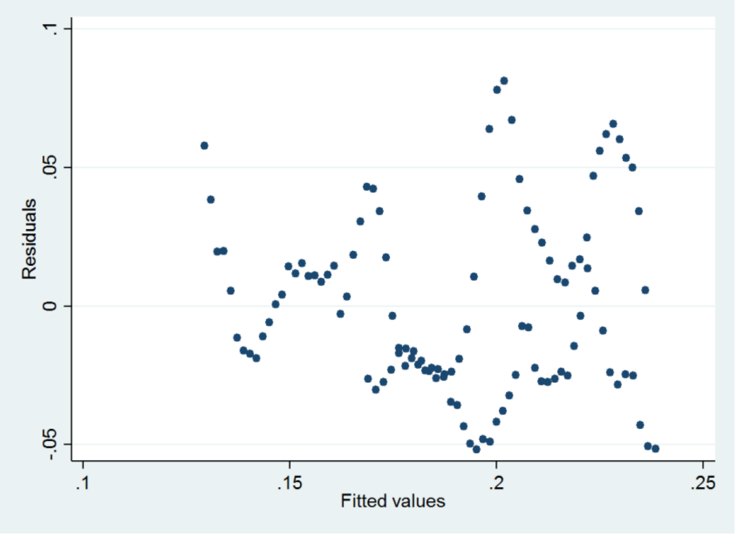



```
. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
linearOMSN~2 |        180  305.3139   364.6495       6   -717.2991  -698.1413
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.
```

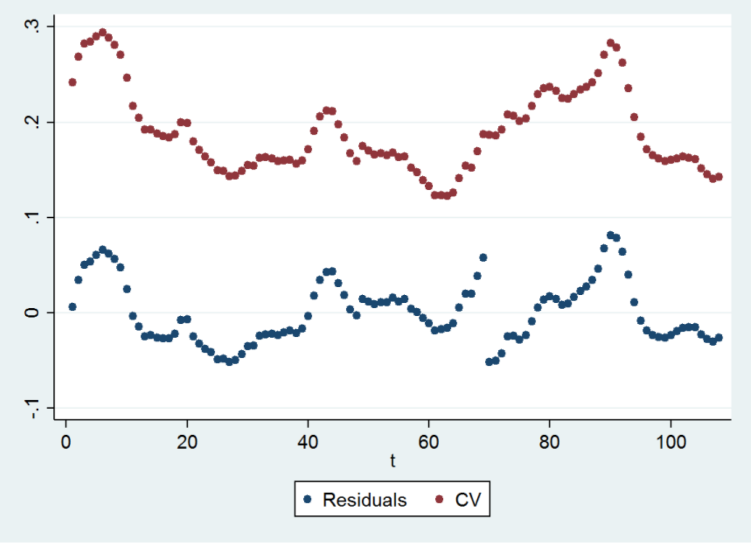
```
. test _b[t]=0

 ( 1)  t = 0

       F(  1,   174) =   14.16
            Prob > F =    0.0002

. local sign_t2 = sign(_b[t])

. display "Ho: coef <= 0  p-value = "ttail(r(df_r),`sign_t2'*sqrt(r(F)))
Ho: coef <= 0  p-value = .9998852

. display "Ho: coef >= 0  p-value = " 1-ttail(r(df_r),`sign_t2'*sqrt(r(F)))
Ho: coef >= 0  p-value = .0001148

. *run regresssion for linear-log model with dummy based on intervention data OMS+NMS2004*
. gen D_1ln_t=D_1*ln_t

. gen D_2ln_t=D_2*ln_t

. reg CV ln_t D_1 D_2 D_1ln_t D_2ln_t

      Source |       SS           df       MS      Number of obs   =       180
-------------+----------------------------------   F(5, 174)       =     43.90
       Model |  .197689218         5  .039537844   Prob > F        =    0.0000
    Residual |  .156722303       174  .000900703   R-squared       =    0.5578
-------------+----------------------------------   Adj R-squared   =    0.5451
       Total |  .354411521       179  .001979953   Root MSE        =    .03001

------------------------------------------------------------------------------
          CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ln_t |  -.0168392   .0024784    -6.79   0.000    -.0217308   -.0119476
         D_1 |   1.764211   1.585093     1.11   0.267    -1.364274    4.892696
         D_2 |   .7648318   1.102174     0.69   0.489     -1.41052    2.940184
     D_1ln_t |  -.3783054   .3633232    -1.04   0.299    -1.095393    .3387826
     D_2ln_t |  -.1341844   .2173264    -0.62   0.538    -.5631196    .2947508
       _cons |   .2423756   .0104555    23.18   0.000     .2217396    .2630116
------------------------------------------------------------------------------

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        180  305.3139   378.7524       6   -745.5047   -726.347
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

. *run regression for log-linear model with dummy's based on intervention data OMS+NMS2004*
. reg ln_CV t D_1 D_2 D_1t D_2t

      Source |       SS           df       MS      Number of obs   =       180
-------------+----------------------------------   F(5, 174)       =     24.51
       Model |  4.00234793         5  .800469587   Prob > F        =    0.0000
    Residual |  5.68176785       174  .032653838   R-squared       =    0.4133
-------------+----------------------------------   Adj R-squared   =    0.3964
       Total |  9.68411579       179  .054101206   Root MSE        =     .1807

------------------------------------------------------------------------------
       ln_CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |  -.0009916   .0002893    -3.43   0.001    -.0015626   -.0004205
         D_1 |   1.906633   2.189943     0.87   0.385    -2.415638    6.228905
         D_2 |   1.175552    1.31042     0.90   0.371    -1.410812    3.761916
        D_1t |  -.0178906   .0278847    -0.64   0.522    -.0729264    .0371451
        D_2t |  -.0047463   .0082147    -0.58   0.564    -.0209595    .0114669
       _cons |  -1.689603   .0280986   -60.13   0.000    -1.745061   -1.634145
------------------------------------------------------------------------------

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        180  7.613351   55.60356       6   -99.20712  -80.04938
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

. clear

.
```

```
. *import file OMSNMS20042007*
.    import     excel     "\\WURNET.NL\Homes\jong285\My     Documents\MScThesis\Text     thesis\Data     files\Chapter     4\Files     for
STATA\RegOMSNMS2004NMS2007.xlsx", sheet("Sheet1") firstrow

. *assign as time series*
. tsset t
        time variable:  t, 1 to 108
                delta:  1 unit

. *structural break test*
. reg CV t

      Source |       SS           df       MS      Number of obs   =       108
-------------+----------------------------------   F(1, 106)       =      2.81
       Model |  .005315795         1  .005315795   Prob > F        =    0.0968
    Residual |  .200763043       106  .001893991   R-squared       =    0.0258
-------------+----------------------------------   Adj R-squared   =    0.0166
       Total |  .206078837       107   .00192597   Root MSE        =    .04352

------------------------------------------------------------------------------
          CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |   -.000225   .0001343    -1.68   0.097    -.0004914    .0000413
       _cons |   .2025054   .0084339    24.01   0.000     .1857843    .2192265
------------------------------------------------------------------------------

. *unknown break test*
. estat sbsingle
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
........................

Test for a structural break: Unknown break date

                             Number of obs =        108

Full sample:                         1 -        108
Trimmed sample:                     18 -         92
Estimated break date:               69
Ho: No structural break

     Test            Statistic          p-value
-----------------------------------------------
     swald             85.4337           0.0000
-----------------------------------------------
Exogenous variables:           t
Coefficients included in test: t _cons

. *test known break dates from intervention in SMP*
. estat sbknown, break(79 96) breakvars(t)

Wald test for a structural break: Known break date

                                  Number of obs  =        108
Sample:               1 -        108
Break date:  79         96
Ho: No structural break

             chi2(2)      =   76.7230
             Prob > chi2  =    0.0000

Exogenous variables:           t
Coefficients included in test: t

. *no sbknown test possible for first period, not enough obs. on the left*
. *create dummy variable based on break test*
. gen D= 0

. replace D= 1 if t>69
(39 real changes made)

. *run regression for linear model with dummy based on structural break test OMS+NMS20042007*
. gen Dt=D*t

. reg CV t D Dt

      Source |       SS           df       MS      Number of obs   =       108
-------------+----------------------------------   F(3, 104)       =     29.57
       Model |  .094861951         3   .03162065   Prob > F        =    0.0000
    Residual |  .111216886       104  .001069393   R-squared       =    0.4603
-------------+----------------------------------   Adj R-squared   =    0.4448
       Total |  .206078837       107   .00192597   Root MSE        =     .0327

------------------------------------------------------------------------------
          CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
```

```
           t |  -.0015681    .0001977    -7.93   0.000    -.0019601   -.0011762
           D |   .1288839    .0424911     3.03   0.003     .0446226    .2131453
          Dt |  -.0002601    .0005055    -0.51   0.608    -.0012626    .0007424
       _cons |   .2375222      .00796    29.84   0.000     .2217373    .2533072
------------------------------------------------------------------------------

. eststo linearOMSNMS20042007_1

. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of CV

         chi2(1)      =     7.65
         Prob > chi2  =   0.0057

. estat dwatson

Durbin-Watson d-statistic(  4,    108) =  .2099776

. rvfplot

. graph export resvsfitted3.png
(file resvsfitted3.png written in PNG format)
```



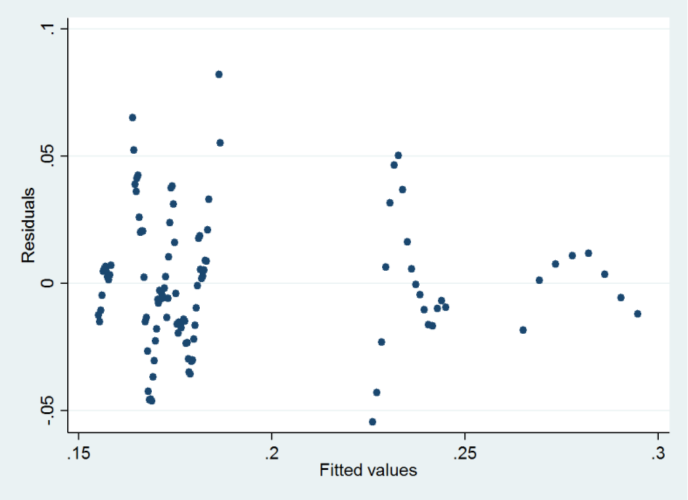

```
. predict res3, residuals

. twoway (scatter res3 t) (scatter CV t)

. graph export resvsCV3.png
(file resvsCV3.png written in PNG format)
```



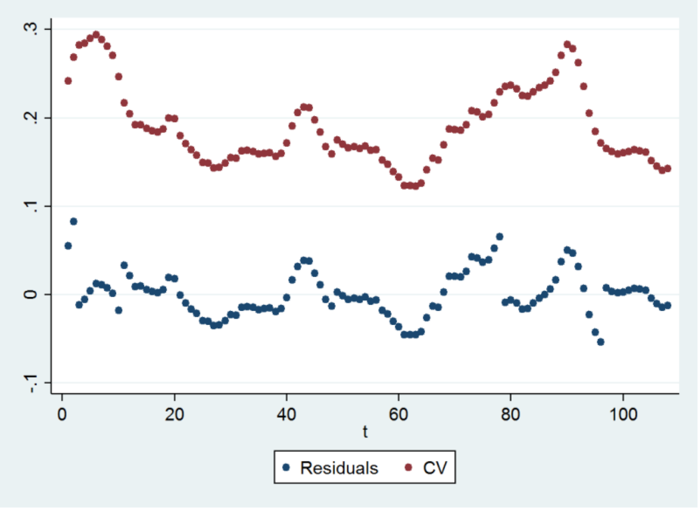

```
. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
```

```
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
linearOMSN~1 |        108  184.8825   218.1885       4   -428.3769  -417.6484
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

. test _b[t]=0

 ( 1)  t = 0

       F(  1,   104) =   62.94
            Prob > F =    0.0000

. local sign_t = sign(_b[t])

. display "Ho: coef <= 0  p-value = "ttail(r(df_r),`sign_t'*sqrt(r(F)))
Ho: coef <= 0  p-value = 1

. display "Ho: coef >= 0  p-value = " 1-ttail(r(df_r),`sign_t'*sqrt(r(F)))
Ho: coef >= 0  p-value = 1.287e-12

. *run regresssion for linear-log model with dummy based on structural break teest OMS+NMS20042007*
. gen ln_t=ln(t)

. gen Dln_t=D*ln_t

. reg CV ln_t Dln_t D

      Source |       SS           df       MS      Number of obs   =       108
-------------+----------------------------------   F(3, 104)       =     41.21
       Model |  .111921426         3  .037307142   Prob > F        =    0.0000
    Residual |  .094157411       104   .00090536   R-squared       =    0.5431
-------------+----------------------------------   Adj R-squared   =    0.5299
       Total |  .206078837       107   .00192597   Root MSE        =    .03009

------------------------------------------------------------------------------
          CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ln_t |  -.0393039   .0040141    -9.79   0.000    -.0472639   -.0313438
       Dln_t |  -.1092259   .0378832    -2.88   0.005    -.1843497    -.034102
           D |   .5577031   .1694006     3.29   0.001     .2217754    .8936308
       _cons |   .3114805   .0136481    22.82   0.000     .2844158    .3385452
------------------------------------------------------------------------------

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        108  184.8825   227.1802       4   -446.3605   -435.632
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

. *run regression for log-linear model with dummy based on structural break test OMS+NMS20042007*
. gen ln_CV=ln(CV)

. reg ln_CV t D Dt

      Source |       SS           df       MS      Number of obs   =       108
-------------+----------------------------------   F(3, 104)       =     33.50
       Model |  2.56237659         3  .854125529   Prob > F        =    0.0000
    Residual |  2.65200252       104  .025500024   R-squared       =    0.4914
-------------+----------------------------------   Adj R-squared   =    0.4767
       Total |   5.2143791       107  .048732515   Root MSE        =    .15969

------------------------------------------------------------------------------
       ln_CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |  -.0079431   .0009652    -8.23   0.000    -.0098572    -.006029
           D |   .7365573    .207491     3.55   0.001     .3250949     1.14802
          Dt |  -.0021611   .0024685    -0.88   0.383    -.0070563     .002734
       _cons |  -1.447869   .0388699   -37.25   0.000    -1.524949   -1.370788
------------------------------------------------------------------------------

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        108  10.41304   46.92271       4   -85.84543   -75.1169
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.
```

```
. *run regression for linear model with dummy's based on intervention data OMS+NMS20042007*
. gen D_1t=D_1*t

. gen D_2t=D_2*t

. reg CV t D_1 D_2 D_1t D_2t

      Source |       SS           df       MS      Number of obs   =       108
-------------+----------------------------------   F(5, 102)       =     40.13
       Model |  .136628809         5  .027325762   Prob > F        =    0.0000
    Residual |  .069450028       102  .000680883   R-squared       =    0.6630
-------------+----------------------------------   Adj R-squared   =    0.6465
       Total |  .206078837       107   .00192597   Root MSE        =    .02609

------------------------------------------------------------------------------
          CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |  -.0002935   .0001017    -2.89   0.005    -.0004952   -.0000919
         D_1 |   .1204527   .0283943     4.24   0.000     .0641327    .1767726
         D_2 |   .1457953   .1040845     1.40   0.164    -.0606557    .3522464
        D_1t |  -.0039362   .0040276    -0.98   0.331     -.011925    .0040526
        D_2t |  -.0008162   .0011898    -0.69   0.494    -.0031762    .0015438
       _cons |   .1868154   .0060157    31.05   0.000     .1748832    .1987476
------------------------------------------------------------------------------

. eststo linearOMSNMS20042007_2

. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of CV

         chi2(1)      =     0.84
         Prob > chi2  =   0.3608

. estat dwatson

Durbin-Watson d-statistic(  6,    108) =  .4435891

. rvfplot

. graph export resvsfitted4.png
(file resvsfitted4.png written in PNG format)
```



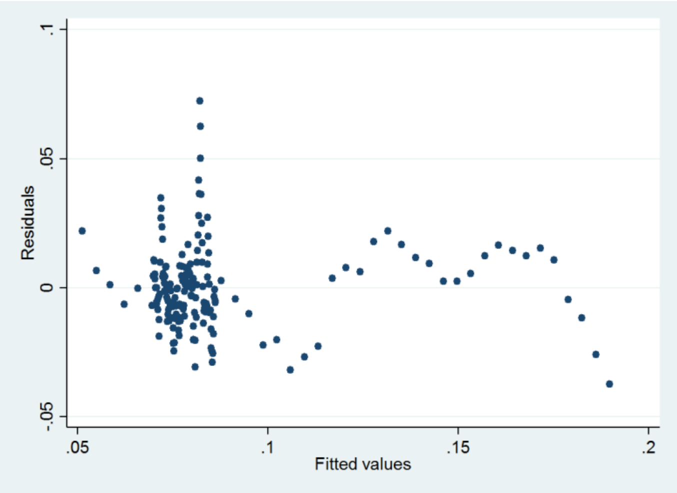

```
. predict res4, residuals

. twoway (scatter res4 t) (scatter CV t)

. graph export resvsCV4.png
(file resvsCV4.png written in PNG format)
```



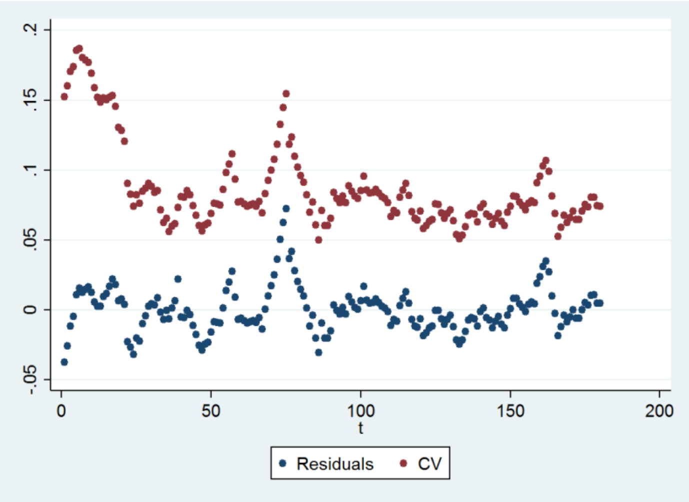

```
. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
linearOMSN~2 |        108  184.8825   243.6157       6   -475.2314  -459.1386
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

. test _b[t]=0

 ( 1)  t = 0

       F(  1,   102) =    8.34
            Prob > F =    0.0047

. local sign_t2 = sign(_b[t])

. display "Ho: coef <= 0  p-value = "ttail(r(df_r),`sign_t2'*sqrt(r(F)))
Ho: coef <= 0  p-value = .99762735

. display "Ho: coef >= 0  p-value = " 1-ttail(r(df_r),`sign_t2'*sqrt(r(F)))
Ho: coef >= 0  p-value = .00237265

. *run regresssion for linear-log model with dummy based on intervention data OMS+NMS20042007*
. gen D_1ln_t=D_1*ln_t

. gen D_2ln_t=D_2*ln_t

. reg CV ln_t D_1 D_2 D_1ln_t D_2ln_t

      Source |       SS           df       MS      Number of obs   =       108
-------------+----------------------------------   F(5, 102)       =     48.23
       Model |  .144826947         5  .028965389   Prob > F        =    0.0000
    Residual |   .06125189       102  .000600509   R-squared       =    0.7028
-------------+----------------------------------   Adj R-squared   =    0.6882
       Total |  .206078837       107   .00192597   Root MSE        =    .02451

------------------------------------------------------------------------------
          CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ln_t |  -.0165304   .0033982    -4.86   0.000    -.0232708   -.0097901
         D_1 |   .0835009   .0430673     1.94   0.055     -.001923    .1689247
         D_2 |   .3995687   .4345905     0.92   0.360    -.4624396    1.261577
     D_1ln_t |  -.0040058   .0225423    -0.18   0.859    -.0487183    .0407068
     D_2ln_t |  -.0723597    .097234    -0.74   0.458     -.265223    .1205035
       _cons |    .233268   .0129692    17.99   0.000     .2075436    .2589923
------------------------------------------------------------------------------

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        108  184.8825   250.3988       6   -488.7976  -472.7048
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

. *run regression for log-linear model with dummy's based on intervention data OMS+NMS20042007*
```

```
. reg ln_CV t D_1 D_2 D_1t D_2t

      Source |       SS           df       MS      Number of obs   =       108
-------------+----------------------------------   F(5, 102)       =     31.74
       Model |  3.17430439         5  .634860878   Prob > F        =    0.0000
    Residual |  2.04007471       102  .020000732   R-squared       =    0.6088
-------------+----------------------------------   Adj R-squared   =    0.5896
       Total |   5.2143791       107  .048732515   Root MSE        =    .14142

------------------------------------------------------------------------------
       ln_CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |  -.0016308    .000551    -2.96   0.004    -.0027237   -.0005378
         D_1 |   .5173074   .1538926     3.36   0.001     .2120622    .8225525
         D_2 |   .7889433   .5641217     1.40   0.165    -.3299893    1.907876
        D_1t |  -.0141115   .0218291    -0.65   0.519    -.0574095    .0291865
        D_2t |  -.0046824   .0064486    -0.73   0.469    -.0174732    .0081085
       _cons |  -1.690099   .0326044   -51.84   0.000    -1.754769   -1.625428
------------------------------------------------------------------------------

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        108  10.41304   61.08846       6   -110.1769  -94.08413
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

. clear

.
. *import file 12 Largest*
. import excel "\\WURNET.NL\Homes\jong285\My Documents\MScThesis\Text thesis\Data files\Chapter 4\Files for STATA\Reg12Largest.xlsx",
sheet("Sheet1") firstrow

. *assign as time series*
. tsset t
        time variable:  t, 1 to 180
                delta:  1 unit

. *structural break test*
. reg CV t

      Source |       SS           df       MS      Number of obs   =       180
-------------+----------------------------------   F(1, 178)       =     82.90
       Model |  .054135597         1  .054135597   Prob > F        =    0.0000
    Residual |  .116234722       178  .000653004   R-squared       =    0.3178
-------------+----------------------------------   Adj R-squared   =    0.3139
       Total |  .170370319       179  .000951789   Root MSE        =    .02555

------------------------------------------------------------------------------
          CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |  -.0003338   .0000367    -9.11   0.000    -.0004061   -.0002614
       _cons |   .1173912   .0038253    30.69   0.000     .1098424    .1249399
------------------------------------------------------------------------------

. *unknown break test*
. estat sbsingle
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100
..........................

Test for a structural break: Unknown break date

                             Number of obs =        180

Full sample:                         1 -        180
Trimmed sample:                     28 -        154
Estimated break date:               39
Ho: No structural break

     Test           Statistic          p-value
----------------------------------------------
     swald          268.0704           0.0000
----------------------------------------------
Exogenous variables:            t
Coefficients included in test: t _cons

. *test known break dates from intervention in SMP*
. estat sbknown, break(75 82) breakvars(t)
```

```
Wald test for a structural break: Known break date

                                  Number of obs  =        180
Sample:               1 -        180
Break date:  75         82
Ho: No structural break

             chi2(2)      =   27.7126
             Prob > chi2  =    0.0000

Exogenous variables:           t
Coefficients included in test: t

. estat sbknown, break(151 168) breakvars(t)

Wald test for a structural break: Known break date

                                  Number of obs  =        180
Sample:               1 -        180
Break date:  151        168
Ho: No structural break

             chi2(2)      =   22.1336
             Prob > chi2  =    0.0000

Exogenous variables:           t
Coefficients included in test: t

. *create dummy variable based on break test*
. gen D= 0

. replace D= 1 if t>39
(141 real changes made)

. *run regression for linear model with dummy based on structural break test 12 largest*
. gen Dt=D*t

. reg CV t D Dt

      Source |       SS           df       MS      Number of obs   =       180
-------------+----------------------------------   F(3, 176)       =    156.56
       Model |  .123931389         3  .041310463   Prob > F        =    0.0000
    Residual |  .046438929       176  .000263858   R-squared       =    0.7274
-------------+----------------------------------   Adj R-squared   =    0.7228
       Total |  .170370319       179  .000951789   Root MSE        =    .01624

------------------------------------------------------------------------------
          CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |  -.0036425   .0002311   -15.76   0.000    -.0040986   -.0031864
           D |  -.1024747   .0066083   -15.51   0.000    -.1155164   -.0894329
          Dt |   .0035254   .0002335    15.10   0.000     .0030645    .0039863
       _cons |   .1933296   .0053038    36.45   0.000     .1828623    .2037969
------------------------------------------------------------------------------

. eststo linear12largest_1

. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of CV

         chi2(1)      =     1.44
         Prob > chi2  =   0.2303

. estat dwatson

Durbin-Watson d-statistic(  4,   180) =  .2732659

. rvfplot

. graph export resvsfitted5.png
(file resvsfitted5.png written in PNG format)
```



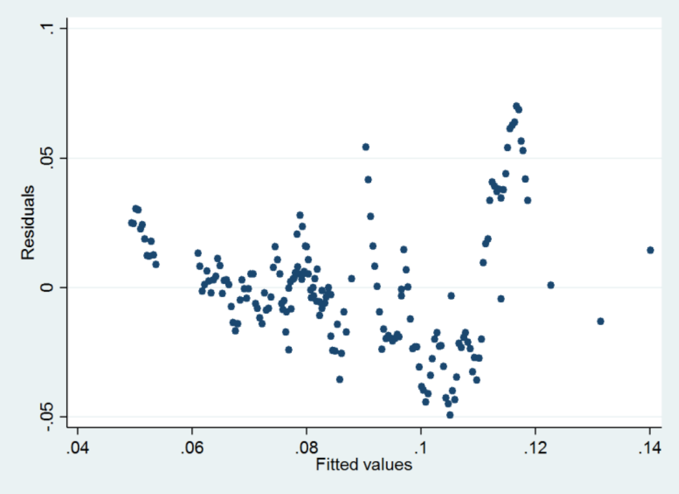

```
. predict res5, residuals

. twoway (scatter res5 t) (scatter CV t)

. graph export resvsCV5.png
(file resvsCV5.png written in PNG format)
```



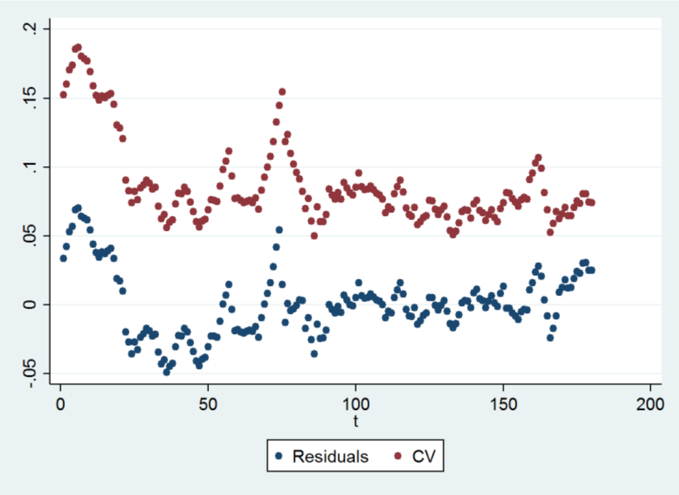

```
. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
linear12la~1 |        180  371.2375   488.2227       4   -968.4455  -955.6736
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

. test _b[t]=0

 ( 1)  t = 0

       F(  1,   176) =  248.40
            Prob > F =    0.0000

. local sign_t = sign(_b[t])

. display "Ho: coef <= 0  p-value = "ttail(r(df_r),`sign_t'*sqrt(r(F)))
Ho: coef <= 0  p-value = 1

. display "Ho: coef >= 0  p-value = " 1-ttail(r(df_r),`sign_t'*sqrt(r(F)))
Ho: coef >= 0  p-value = 0

. *run regresssion for linear-log model with dummy based on structural break test 12 largest*
. gen ln_t=ln(t)

. gen Dln_t=D*ln_t

. reg CV ln_t Dln_t D

      Source |       SS           df       MS      Number of obs   =       180
```

```
-------------+----------------------------------   F(3, 176)       =     94.24
       Model |  .105003917         3  .035001306   Prob > F        =    0.0000
    Residual |  .065366402       176   .0003714    R-squared       =    0.6163
-------------+----------------------------------   Adj R-squared   =    0.6098
       Total |  .170370319       179  .000951789   Root MSE        =    .01927

------------------------------------------------------------------------------
          CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ln_t |  -.0404048   .0035873   -11.26   0.000    -.0474845   -.0333252
       Dln_t |   .0298562   .0053084     5.62   0.000       .01938    .0403324
           D |  -.1042259   .0208648    -5.00   0.000    -.1454032   -.0630486
       _cons |   .2309526   .0102822    22.46   0.000     .2106604    .2512448
------------------------------------------------------------------------------

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        180  371.2375   457.4544       4   -906.9088   -894.137
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

. *run regression for log-linear model with dummy based on structural break test 12 largest*
. gen ln_CV=ln(CV)

. reg ln_CV t D Dt

      Source |       SS           df       MS      Number of obs   =       180
-------------+----------------------------------   F(3, 176)       =    101.77
       Model |   10.116203         3  3.37206767   Prob > F        =    0.0000
    Residual |  5.83179848       176  .033135219   R-squared       =    0.6343
-------------+----------------------------------   Adj R-squared   =    0.6281
       Total |  15.9480015       179   .08909498   Root MSE        =    .18203

------------------------------------------------------------------------------
       ln_CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |  -.0324465   .0025899   -12.53   0.000    -.0375578   -.0273353
           D |  -.8835308   .0740543   -11.93   0.000     -1.02968    -.737382
          Dt |   .0310946   .0026171    11.88   0.000     .0259296    .0362597
       _cons |  -1.539808    .059436   -25.91   0.000    -1.657107   -1.422509
------------------------------------------------------------------------------

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        180 -37.28284   53.25789       4   -98.51578  -85.74395
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

. *run regression for linear model with dummy's based on intervention data 12 largest*
. gen D_1t=D_1*t

. gen D_2t=D_2*t

. reg CV t D_1 D_2 D_1t D_2t

      Source |       SS           df       MS      Number of obs   =       180
-------------+----------------------------------   F(5, 174)       =     22.36
       Model |  .066649972         5  .013329994   Prob > F        =    0.0000
    Residual |  .103720347       174  .000596094   R-squared       =    0.3912
-------------+----------------------------------   Adj R-squared   =    0.3737
       Total |  .170370319       179  .000951789   Root MSE        =    .02442

------------------------------------------------------------------------------
          CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |  -.0003869   .0000391    -9.90   0.000     -.000464   -.0003097
         D_1 |   .6744546   .2958851     2.28   0.024     .0904687     1.25844
         D_2 |   .0390208   .1770519     0.22   0.826     -.310425    .3884667
        D_1t |  -.0083243   .0037675    -2.21   0.028    -.0157603   -.0008884
        D_2t |  -.0001022   .0011099    -0.09   0.927    -.0022928    .0020884
       _cons |   .1189914   .0037964    31.34   0.000     .1114984    .1264843
------------------------------------------------------------------------------

. eststo linear12largest_2

. estat hettest
```

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of CV

         chi2(1)      =    60.11
         Prob > chi2  =   0.0000

. estat dwatson

Durbin-Watson d-statistic(  6,   180) =  .1311977

. rvfplot

. graph export resvsfitted6.png
(file resvsfitted6.png written in PNG format)
```




```
. predict res6, residuals

. twoway (scatter res6 t) (scatter CV t)

. graph export resvsCV6.png
(file resvsCV6.png written in PNG format)
```

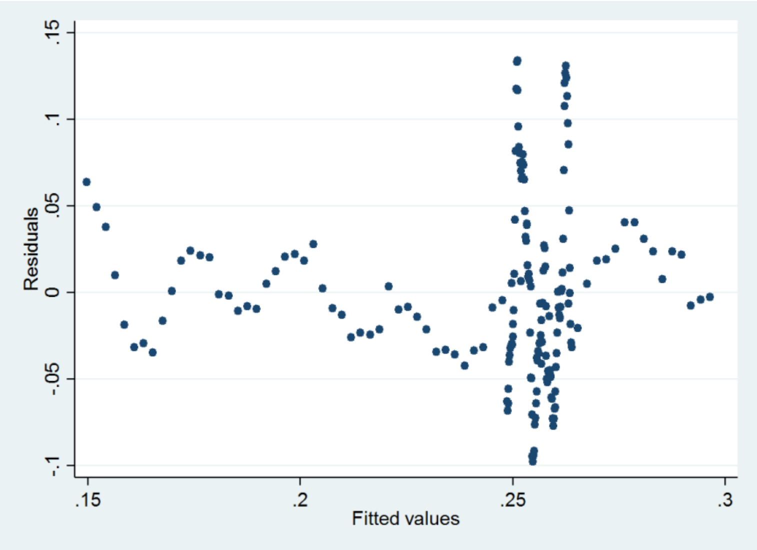



```
. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
linear12la~2 |        180  371.2375   415.9023       6   -819.8046  -800.6469
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

. test _b[t]=0

 ( 1)  t = 0

       F(  1,   174) =   97.93
            Prob > F =    0.0000
```

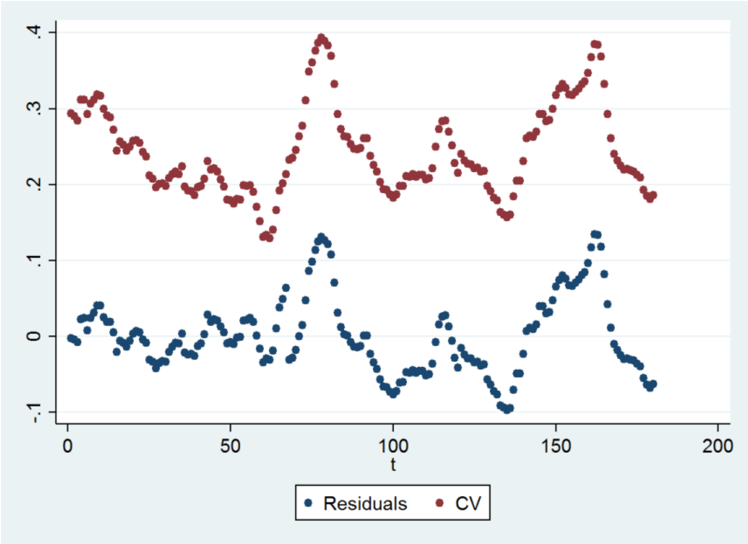
```
. local sign_t2 = sign(_b[t])

. display "Ho: coef <= 0  p-value = "ttail(r(df_r),`sign_t2'*sqrt(r(F)))
Ho: coef <= 0  p-value = 1

. display "Ho: coef >= 0  p-value = " 1-ttail(r(df_r),`sign_t2'*sqrt(r(F)))
Ho: coef >= 0  p-value = 0

. *run regresssion for linear-log model with dummy based on intervention data 12 largest*
. gen D_1ln_t=D_1*ln_t

. gen D_2ln_t=D_2*ln_t

. reg CV ln_t D_1 D_2 D_1ln_t D_2ln_t

      Source |       SS           df       MS      Number of obs   =       180
-------------+----------------------------------   F(5, 174)       =     57.48
       Model |  .106122878         5  .021224576   Prob > F        =    0.0000
    Residual |  .064247441       174  .000369238   R-squared       =    0.6229
-------------+----------------------------------   Adj R-squared   =    0.6121
       Total |  .170370319       179  .000951789   Root MSE        =    .01922

------------------------------------------------------------------------------
          CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ln_t |  -.0258313   .0015868   -16.28   0.000    -.0289632   -.0226994
         D_1 |   2.906052   1.014886     2.86   0.005     .9029796    4.909124
         D_2 |   .2644079   .7056882     0.37   0.708    -1.128403    1.657219
     D_1ln_t |  -.6593711   .2326246    -2.83   0.005      -1.1185   -.2002418
     D_2ln_t |  -.0485602   .1391474    -0.35   0.728    -.3231942    .2260737
       _cons |   .1928789   .0066944    28.81   0.000     .1796663    .2060915
------------------------------------------------------------------------------

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        180  371.2375   459.0084       6   -906.0168   -886.859
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

. *run regression for log-linear model with dummy's based on intervention data 12 largest*
. reg ln_CV t D_1 D_2 D_1t D_2t

      Source |       SS           df       MS      Number of obs   =       180
-------------+----------------------------------   F(5, 174)       =     22.22
       Model |  6.21411372         5  1.24282274   Prob > F        =    0.0000
    Residual |  9.73388776       174  .055941884   R-squared       =    0.3896
-------------+----------------------------------   Adj R-squared   =    0.3721
       Total |  15.9480015       179   .08909498   Root MSE        =    .23652

------------------------------------------------------------------------------
       ln_CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |  -.0036938   .0003787    -9.75   0.000    -.0044412   -.0029463
         D_1 |   6.078956   2.866383     2.12   0.035     .4215998    11.73631
         D_2 |   1.015746   1.715188     0.59   0.554    -2.369507    4.400998
        D_1t |  -.0742397   .0364978    -2.03   0.043    -.1462752   -.0022043
        D_2t |  -.0048805    .010752    -0.45   0.650    -.0261017    .0163407
       _cons |  -2.189603   .0367778   -59.54   0.000    -2.262191   -2.117014
------------------------------------------------------------------------------

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        180 -37.28284   7.151976       6   -2.303953   16.85379
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

. clear

.
. *import file 12Smallest*
. import excel "\\WURNET.NL\Homes\jong285\My Documents\MScThesis\Text thesis\Data files\Chapter 4\Files for STATA\Reg12Smallest.xlsx",
sheet("Sheet1") firstrow

. *assign as time series*
. tsset t
        time variable:  t, 1 to 180
```

```
              delta:  1 unit

. *structural break test*
. reg CV t

      Source |       SS           df       MS      Number of obs   =       180
-------------+----------------------------------   F(1, 178)       =      1.69
       Model |  .005889801         1  .005889801   Prob > F        =    0.1957
    Residual |  .621628253       178  .003492294   R-squared       =    0.0094
-------------+----------------------------------   Adj R-squared   =    0.0038
       Total |  .627518054       179  .003505687   Root MSE        =     .0591

------------------------------------------------------------------------------
          CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |   .0001101   .0000848     1.30   0.196    -.0000572    .0002774
       _cons |   .2339046   .0088463    26.44   0.000     .2164475    .2513617
------------------------------------------------------------------------------

. *unknown break test*
. estat sbsingle
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100
..........................

Test for a structural break: Unknown break date

                             Number of obs =       180

Full sample:                          1 -       180
Trimmed sample:                      28 -       154
Estimated break date:                67
Ho: No structural break

     Test           Statistic          p-value
-----------------------------------------------
     swald           65.2199           0.0000
-----------------------------------------------
Exogenous variables:           t
Coefficients included in test: t _cons

. *test known break dates from intervention in SMP*
. estat sbknown, break(75 82) breakvars(t)

Wald test for a structural break: Known break date

                                  Number of obs  =       180
Sample:               1 -       180
Break date:  75         82
Ho: No structural break

             chi2(2)      =   75.9540
             Prob > chi2  =    0.0000

Exogenous variables:           t
Coefficients included in test: t

. estat sbknown, break(151 168) breakvars(t)

Wald test for a structural break: Known break date

                                  Number of obs  =       180
Sample:               1 -       180
Break date:  151        168
Ho: No structural break

             chi2(2)      =   60.1704
             Prob > chi2  =    0.0000

Exogenous variables:           t
Coefficients included in test: t

. *create dummy variable based on break test*
. gen D= 0

. replace D= 1 if t>67
(113 real changes made)

. *run regression for linear model with dummy based on structural break test 12 smallest*
. gen Dt=D*t

. reg CV t D Dt

      Source |       SS           df       MS      Number of obs   =       180
-------------+----------------------------------   F(3, 176)       =     22.17
```

```
       Model |  .172103497         3  .057367832   Prob > F        =    0.0000
    Residual |  .455414557       176  .002587583   R-squared       =    0.2743
-------------+----------------------------------   Adj R-squared   =    0.2619
       Total |  .627518054       179  .003505687   Root MSE        =    .05087

------------------------------------------------------------------------------
          CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |  -.0022209   .0003213    -6.91   0.000    -.0028551   -.0015867
           D |  -.0254217   .0226232    -1.12   0.263    -.0700693    .0192259
          Dt |   .0020844   .0003532     5.90   0.000     .0013873    .0027816
       _cons |   .2985551   .0125695    23.75   0.000     .2737486    .3233615
------------------------------------------------------------------------------

. eststo linear12smallest_1

. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of CV

         chi2(1)      =     7.35
         Prob > chi2  =   0.0067

. estat dwatson

Durbin-Watson d-statistic(  4,    180) =  .0939949

. rvfplot

. graph export resvsfitted7.png
(file resvsfitted7.png written in PNG format)
```

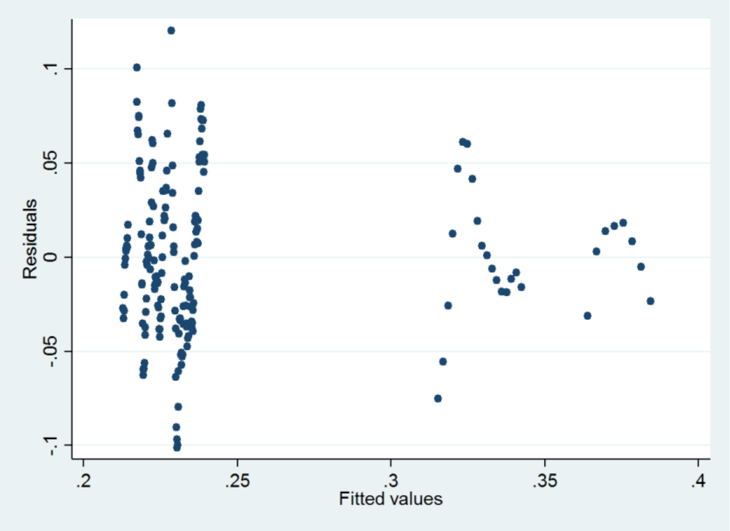



```
. predict res7, residuals

. twoway (scatter res7 t) (scatter CV t)

. graph export resvsCV7.png
(file resvsCV7.png written in PNG format)
```

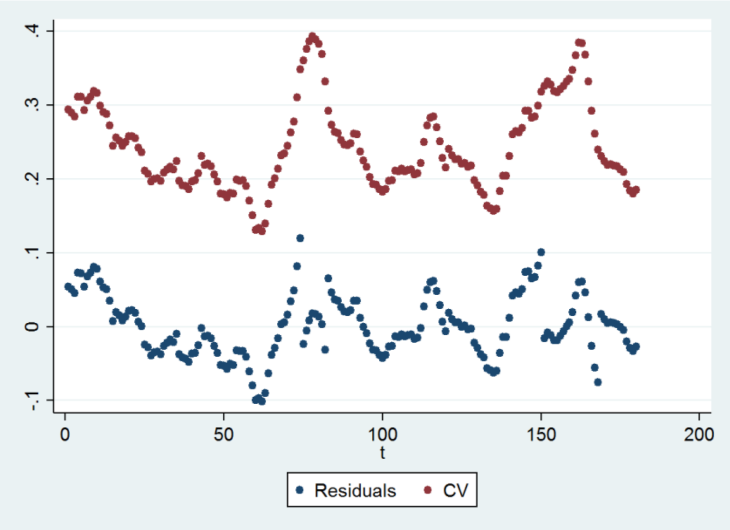



```
. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
linear12sm~1 |        180  253.8956   282.7464       4   -557.4929   -544.721
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

. test _b[t]=0

 ( 1)  t = 0

       F(  1,   176) =   47.76
            Prob > F =    0.0000

. local sign_t = sign(_b[t])

. display "Ho: coef <= 0  p-value = "ttail(r(df_r),`sign_t'*sqrt(r(F)))
Ho: coef <= 0  p-value = 1

. display "Ho: coef >= 0  p-value = " 1-ttail(r(df_r),`sign_t'*sqrt(r(F)))
Ho: coef >= 0  p-value = 4.245e-11

. *run regresssion for linear-log model with dummy based on structural break teest 12 smallest*
. gen ln_t=ln(t)

. gen Dln_t=D*ln_t

. reg CV ln_t Dln_t D

      Source |       SS           df       MS      Number of obs   =       180
-------------+----------------------------------   F(3, 176)       =     21.26
       Model |  .166905263         3  .055635088   Prob > F        =    0.0000
    Residual |  .460612791       176  .002617118   R-squared       =    0.2660
-------------+----------------------------------   Adj R-squared   =    0.2535
       Total |  .627518054       179  .003505687   Root MSE        =    .05116

------------------------------------------------------------------------------
          CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ln_t |  -.0460021   .0069402    -6.63   0.000    -.0596989   -.0323054
       Dln_t |   .0204969   .0186949     1.10   0.274    -.0163981    .0573919
           D |    .005669   .0864022     0.07   0.948    -.1648488    .1761867
       _cons |   .3725434   .0234042    15.92   0.000     .3263543    .4187324
------------------------------------------------------------------------------

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        180  253.8956    281.725       4   -555.4499  -542.6781
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

. *run regression for log-linear model with dummy based on structural break test 12 smallest*
```

```
. gen ln_CV=ln(CV)

. reg ln_CV t D Dt

      Source |       SS           df       MS      Number of obs   =       180
-------------+----------------------------------   F(3, 176)       =     28.33
       Model |  3.29672524         3  1.09890841   Prob > F        =    0.0000
    Residual |  6.82771764       176   .03879385   R-squared       =    0.3256
-------------+----------------------------------   Adj R-squared   =    0.3141
       Total |  10.1244429       179  .056561133   Root MSE        =    .19696

------------------------------------------------------------------------------
       ln_CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |  -.0099656   .0012443    -8.01   0.000    -.0124211     -.00751
           D |  -.1341472   .0875967    -1.53   0.127    -.3070224     .038728
          Dt |   .0094065   .0013678     6.88   0.000     .0067071    .0121058
       _cons |  -1.185608   .0486692   -24.36   0.000    -1.281658   -1.089557
------------------------------------------------------------------------------

. *run regression for linear model with dummy's based on intervention data 12 smallest*
. gen D_1t=D_1*t

. gen D_2t=D_2*t

. reg CV t D_1 D_2 D_1t D_2t

      Source |       SS           df       MS      Number of obs   =       180
-------------+----------------------------------   F(5, 174)       =     35.84
       Model |  .318360913         5  .063672183   Prob > F        =    0.0000
    Residual |  .309157141       174  .001776765   R-squared       =    0.5073
-------------+----------------------------------   Adj R-squared   =    0.4932
       Total |  .627518054       179  .003505687   Root MSE        =    .04215

------------------------------------------------------------------------------
          CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |  -.0001479   .0000675    -2.19   0.030    -.0002811   -.0000147
         D_1 |   .3641776   .5108351     0.71   0.477    -.6440532    1.372408
         D_2 |   .3429618   .3056739     1.12   0.263    -.2603441    .9462677
        D_1t |   -.002777   .0065045    -0.43   0.670    -.0156149    .0100609
        D_2t |  -.0014423   .0019162    -0.75   0.453    -.0052243    .0023396
       _cons |   .2394661   .0065544    36.54   0.000     .2265298    .2524025
------------------------------------------------------------------------------

. eststo linear12smallest_2

. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of CV

         chi2(1)      =     2.90
         Prob > chi2  =   0.0887

. estat dwatson

Durbin-Watson d-statistic(  6,   180) =  .2718313

. rvfplot

. graph export resvsfitted8.png
(file resvsfitted8.png written in PNG format)
```




```
. predict res8, residuals

. twoway (scatter res8 t) (scatter CV t)

. graph export resvsCV8.png
(file resvsCV8.png written in PNG format)
```




```
. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
linear12sm~2 |        180  253.8956   317.6087       6   -623.2174  -604.0596
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

. test _b[t]=0

 ( 1)  t = 0

       F(  1,   174) =    4.80
            Prob > F =    0.0297

. local sign_t2 = sign(_b[t])

. display "Ho: coef <= 0  p-value = "ttail(r(df_r),`sign_t2'*sqrt(r(F)))
Ho: coef <= 0  p-value = .98513825

. display "Ho: coef >= 0  p-value = " 1-ttail(r(df_r),`sign_t2'*sqrt(r(F)))
Ho: coef >= 0  p-value = .01486175

. *run regresssion for linear-log model with dummy based on intervention data 12 smallest*
. gen D_1ln_t=D_1*ln_t

. gen D_2ln_t=D_2*ln_t
```

```
. reg CV ln_t D_1 D_2 D_1ln_t D_2ln_t

      Source |       SS           df       MS      Number of obs   =       180
-------------+----------------------------------   F(5, 174)       =     42.91
       Model |   .34650613         5  .069301226   Prob > F        =    0.0000
    Residual |  .281011925       174  .001615011   R-squared       =    0.5522
-------------+----------------------------------   Adj R-squared   =    0.5393
       Total |  .627518054       179  .003505687   Root MSE        =    .04019

------------------------------------------------------------------------------
          CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ln_t |  -.0158424   .0033187    -4.77   0.000    -.0223924   -.0092923
         D_1 |   1.050496    2.12252     0.49   0.621    -3.138704    5.239696
         D_2 |   1.268647   1.475867     0.86   0.391     -1.64426    4.181553
     D_1ln_t |  -.2061911   .4865082    -0.42   0.672    -1.166408     .754026
     D_2ln_t |  -.2270954    .291011    -0.78   0.436    -.8014613    .3472705
       _cons |   .2922015   .0140005    20.87   0.000     .2645688    .3198341
------------------------------------------------------------------------------

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        180  253.8956   326.1994       6   -640.3988  -621.2411
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

. *run regression for log-linear model with dummy's based on intervention data 12 smallest*
. reg ln_CV t D_1 D_2 D_1t D_2t

      Source |       SS           df       MS      Number of obs   =       180
-------------+----------------------------------   F(5, 174)       =     24.99
       Model |  4.23140669         5  .846281339   Prob > F        =    0.0000
    Residual |  5.89303619       174  .033868024   R-squared       =    0.4179
-------------+----------------------------------   Adj R-squared   =    0.4012
       Total |  10.1244429       179  .056561133   Root MSE        =    .18403

------------------------------------------------------------------------------
       ln_CV |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |  -.0005784   .0002947    -1.96   0.051    -.0011599    3.19e-06
         D_1 |   1.120151   2.230286     0.50   0.616    -3.281746    5.522048
         D_2 |   1.341957    1.33456     1.01   0.316    -1.292053    3.975967
        D_1t |  -.0077315   .0283984    -0.27   0.786    -.0637812    .0483181
        D_2t |  -.0057437    .008366    -0.69   0.493    -.0222556    .0107682
       _cons |  -1.452517   .0286162   -50.76   0.000    -1.508997   -1.396038
------------------------------------------------------------------------------

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        180  3.611448   52.31776       6   -92.63552  -73.47778
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

.
. *regression tables for thesis document*
. estout linearOMSNMS2004_1 linearOMSNMS20042007_1 linear12largest_1 linear12smallest_1  , cells(b(star fmt(5) label(Coefficient)) se(par
fmt(2) label(std.errors))) starl( * 0.10
>  ** 0.05 *** 0.010) label stats(N r2 aic F, labels ("No. of Obs." "R-Squared" "AIC" "F" fmt(8 2 2 2)))

----------------------------------------------------------------------------------
                    linearOMSN~1    linearOMSN~1    linear12la~1    linear12sm~1
                    Coeff..err~s    Coeff..err~s    Coeff..err~s    Coeff..err~s
----------------------------------------------------------------------------------
t                       -0.00192***     -0.00157***     -0.00364***     -0.00222***
                          (0.00)          (0.00)          (0.00)          (0.00)
D                       -0.02837*        0.12888***     -0.10247***     -0.02542
                          (0.02)          (0.04)          (0.01)          (0.02)
Dt                       0.00177***     -0.00026         0.00353***      0.00208***
                          (0.00)          (0.00)          (0.00)          (0.00)
_cons                    0.23885***      0.23752***      0.19333***      0.29856***
                          (0.01)          (0.01)          (0.01)          (0.01)
----------------------------------------------------------------------------------
No. of Obs.            180.00000       108.00000       180.00000       180.00000
R-Squared                0.29143         0.46032         0.72742         0.27426
AIC                   -664.63791      -428.37694      -968.44545      -557.49285
F                       24.12875        29.56878       156.56350        22.17043
```

```
-----------------------------------------------------------------------------------

. estout linearOMSNMS2004_2 linearOMSNMS20042007_2 linear12largest_2 linear12smallest_2  , cells(b(star fmt(5) label(Coefficient)) se(par
fmt(2) label(std.errors))) starl( * 0.10
>  ** 0.05 *** 0.010) label stats(N r2 aic F, labels ("No. of Obs." "R-Squared" "AIC" "F" fmt(0 2 2 2)))

-----------------------------------------------------------------------------------
                    linearOMSN~2    linearOMSN~2    linear12la~2    linear12sm~2
                    Coeff..err~s    Coeff..err~s    Coeff..err~s    Coeff..err~s
-----------------------------------------------------------------------------------
t                    -0.00020***     -0.00029***     -0.00039***     -0.00015**
                       (0.00)          (0.00)          (0.00)          (0.00)
D_1                   0.49370         0.12045***      0.67445**       0.36418
                       (0.39)          (0.03)          (0.30)          (0.51)
D_2                   0.21152         0.14580         0.03902         0.34296
                       (0.24)          (0.10)          (0.18)          (0.31)
D_1t                 -0.00491        -0.00394        -0.00832**      -0.00278
                       (0.01)          (0.00)          (0.00)          (0.01)
D_2t                 -0.00081        -0.00082        -0.00010        -0.00144
                       (0.00)          (0.00)          (0.00)          (0.00)
_cons                 0.18950***      0.18682***      0.11899***      0.23947***
                       (0.01)          (0.01)          (0.00)          (0.01)
-----------------------------------------------------------------------------------
No. of Obs.         180.00000       108.00000       180.00000       180.00000
R-Squared             0.48278         0.66299         0.39121         0.50733
AIC                -717.29906      -475.23141      -819.80462      -623.21737
F                    32.48266        40.13285        22.36224        35.83601
-----------------------------------------------------------------------------------

.
end of do-file

.
```

APPENDIX III CHAPTER 5 R-OUTPUT

```
> #loadpackage
> library(markovchain)
Package:  markovchain
Version:  0.6.9.8-1
Date:     2017-08-15
BugReport: http://github.com/spedygiorgio/markovchain/issues

Warning message:
package ‘markovchain’ was built under R version 3.3.3
> library(matrixcalc)
Warning message:
package ‘matrixcalc’ was built under R version 3.3.2
> library(rJava)
Warning message:
package ‘rJava’ was built under R version 3.3.3
> library(xlsxjars)
Warning message:
package ‘xlsxjars’ was built under R version 3.3.3
> library(xlsx)
Warning message:
package ‘xlsx’ was built under R version 3.3.3
> library(plotly)
Loading required package: ggplot2

Attaching package: ‘plotly’

The following object is masked from ‘package:ggplot2’:

    last_plot

The following object is masked from ‘package:stats’:

    filter

The following object is masked from ‘package:graphics’:

    layout

Warning messages:
1: package ‘plotly’ was built under R version 3.3.3
2: package ‘ggplot2’ was built under R version 3.3.3
>
> #load data from computer Biological productivity 2004-2015
>          a          <-read.table("//WURNET.NL/Homes/jong285/My
Documents/MScThesis/Text  thesis/Data  files/Chapter  5/File  for
R/Biological_productivity_2004-2015.txt", header=T)
>
> #define as matrix
> datamatrix <-as.matrix(a)
> datamatrix
      X2004 X2005 X2006 X2007 X2008 X2009 X2010 X2011 X2012 X2013
X2014 X2015
 [1,] "d"   "d"   "d"   "d"   "d"   "e"   "d"   "d"   "d"   "d"
"d"   "d"
 [2,] "d"   "d"   "d"   "d"   "d"   "e"   "d"   "d"   "d"   "e"
"d"   "d"
 [3,] "d"   "d"   "d"   "d"   "d"   "e"   "d"   "d"   "d"   "d"
"d"   "d"
 [4,] "c"   "c"   "c"   "c"   "c"   "d"   "d"   "d"   "d"   "d"
"c"   "d"
 [5,] "c"   "c"   "d"   "c"   "d"   "d"   "d"   "d"   "d"   "d"
"d"   "d"
 [6,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
"c"   "c"
 [7,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
"c"   "c"
 [8,] "d"   "d"   "d"   "c"   "d"   "d"   "d"   "d"   "c"   "d"
"d"   "d"
 [9,] "d"   "d"   "d"   "e"   "d"   "e"   "e"   "e"   "d"   "e"
"d"   "e"
[10,] "d"   "e"   "e"   "e"   "d"   "e"   "e"   "e"   "e"   "e"
"d"   "d"
[11,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
"e"   "e"
[12,] "d"   "d"   "e"   "d"   "e"   "e"   "e"   "d"   "e"   "e"
"e"   "e"
[13,] "e"   "d"   "d"   "d"   "d"   "d"   "e"   "e"   "e"   "e"
"e"   "e"
[14,] "b"   "b"   "b"   "b"   "c"   "c"   "c"   "c"   "c"   "c"
"c"   "c"
[15,] "d"   "d"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
"c"   "c"
[16,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
"c"   "c"
[17,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
"d"   "d"
[18,] "c"   "b"   "b"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
"c"   "c"
[19,] "c"   "d"   "d"   "d"   "c"   "d"   "c"   "d"   "c"   "c"
"c"   "c"
[20,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "d"   "d"   "c"
"d"   "d"
[21,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
"c"   "c"
[22,] "c"   "c"   "c"   "c"   "c"   "c"   "d"   "c"   "c"   "c"
"c"   "c"
[23,] "c"   "c"   "c"   "c"   "b"   "c"   "c"   "c"   "c"   "c"
"c"   "c"
[24,] "c"   "d"   "c"   "c"   "d"   "c"   "c"   "c"   "c"   "c"
"c"   "c"
[25,] "d"   "d"   "c"   "d"   "c"   "c"   "d"   "d"   "c"   "c"
"d"   "d"
[26,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "c"   "d"   "d"
"d"   "d"
[27,] "c"   "d"   "c"   "c"   "c"   "c"   "c"   "c"   "d"   "d"
"d"   "d"
[28,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
"c"   "c"
[29,] "c"   "c"   "c"   "c"   "b"   "c"   "c"   "c"   "c"   "c"
"c"   "c"
[30,] "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"
"b"   "b"
[31,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
"a"   "a"
[32,] "d"   "c"   "c"   "c"   "b"   "b"   "b"   "b"   "a"   "b"
"c"   "a"
[33,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
"d"   "e"
[34,] "c"   "c"   "c"   "c"   "c"   "c"   "b"   "c"   "c"   "c"
"c"   "c"
[35,] "b"   "b"   "c"   "c"   "a"   "b"   "c"   "c"   "c"   "c"
"c"   "c"
[36,] "d"   "c"   "c"   "d"   "c"   "d"   "d"   "c"   "c"   "c"
"c"   "c"
[37,] "c"   "c"   "c"   "c"   "b"   "c"   "c"   "b"   "c"   "c"
"b"   "b"
[38,] "d"   "d"   "d"   "d"   "c"   "c"   "c"   "c"   "d"   "c"
"d"   "d"
[39,] "b"   "b"   "b"   "b"   "a"   "a"   "a"   "a"   "a"   "a"
"a"   "a"
[40,] "a"   "e"   "a"   "a"   "b"   "a"   "a"   "a"   "b"   "a"
"a"   "a"
[41,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
"a"   "a"
[42,] "c"   "b"   "b"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
"c"   "c"
[43,] "a"   "b"   "b"   "a"   "a"   "a"   "b"   "a"   "a"   "a"
"a"   "a"
[44,] "e"   "e"   "d"   "d"   "d"   "d"   "d"   "e"   "d"   "c"
"d"   "d"
[45,] "c"   "c"   "c"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
"d"   "d"
[46,] "b"   "b"   "b"   "b"   "b"   "c"   "c"   "c"   "b"   "c"
"c"   "c"
[47,] "d"   "d"   "c"   "c"   "c"   "d"   "d"   "d"   "d"   "d"
"c"   "d"
[48,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
"d"   "d"
[49,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
"e"   "e"
[50,] "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"
"b"   "b"
[51,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
"d"   "d"
[52,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
"d"   "d"
[53,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "c"   "d"
"c"   "c"
[54,] "c"   "c"   "c"   "c"   "c"   "c"   "d"   "d"   "c"   "d"
"c"   "c"
[55,] "d"   "c"   "d"   "d"   "d"   "d"   "d"   "c"   "c"   "c"
"c"   "d"
[56,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
"c"   "c"
```

```
[57,] "b"   "b"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "d"
"d"   "d"
[58,] "c"   "c"   "c"   "c"   "c"   "b"   "b"   "b"   "b"   "a"
"a"   "a"
[59,] "b"   "c"   "c"   "c"   "c"   "b"   "b"   "b"   "b"   "b"
"b"   "b"
[60,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "d"   "e"
"e"   "e"
[61,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
"e"   "e"
[62,] "c"   "b"   "b"   "a"   "a"   "b"   "b"   "b"   "b"   "a"
"b"   "b"
[63,] "d"   "c"   "c"   "e"   "e"   "d"   "c"   "e"   "c"   "c"
"b"   "c"
[64,] "d"   "c"   "c"   "c"   "c"   "c"   "b"   "c"   "c"   "c"
"c"   "c"
[65,] "b"   "b"   "b"   "b"   "c"   "c"   "c"   "c"   "c"   "b"
"c"   "c"
[66,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
"c"   "c"
[67,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
"e"   "e"
[68,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
"e"   "e"
[69,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
"e"   "e"
[70,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
"e"   "e"
[71,] "d"   "d"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
"e"   "e"
[72,] "e"   "d"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
"e"   "e"
[73,] "d"   "d"   "e"   "e"   "d"   "d"   "d"   "d"   "d"   "d"
"d"   "d"
[74,] "c"   "b"   "b"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
"c"   "c"
[75,] "b"   "c"   "c"   "c"   "d"   "d"   "d"   "d"   "d"   "d"
"d"   "e"
[76,] "a"   "a"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"
"b"   "b"
[77,] "a"   "a"   "a"   "b"   "a"   "b"   "b"   "b"   "b"   "b"
"b"   "b"
[78,] "b"   "b"   "b"   "b"   "b"   "b"   "c"   "c"   "c"   "c"
"c"   "c"
[79,] "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"
"b"   "b"
[80,] "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"
"b"   "b"
[81,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "b"   "b"
"b"   "b"
[82,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
"a"   "a"
[83,] "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"
"b"   "b"
[84,] "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"
"a"   "a"
>
> #estimate transition matrix with MLE
>   result   <-markovchainFit(data=datamatrix,   method="mle",
name="datamatrix")
> result$estimate
datamatrix
 A  5 - dimensional discrete Markov Chain defined by the following
states:
 a, b, c, d, e
 The transition matrix  (by rows)  is defined as follows:
            a          b           c         d           e
a 0.833333333 0.15384615 0.000000000 0.0000000 0.012820513
b 0.070063694 0.78980892 0.140127389 0.0000000 0.000000000
c 0.006430868 0.05466238 0.819935691 0.1125402 0.006430868
d 0.000000000 0.00000000 0.150214592 0.7596567 0.090128755
e 0.006896552 0.00000000 0.006896552 0.1241379 0.862068966

> result$standardError
            a          b           c          d          e
a 0.103362279 0.04441156 0.000000000 0.00000000 0.01282051
b 0.021124999 0.07092693 0.029875260 0.00000000 0.00000000
c 0.004547310 0.01325757 0.051346365 0.01902276 0.00454731
d 0.000000000 0.00000000 0.025390900 0.05709929 0.01966771
e 0.006896552 0.00000000 0.006896552 0.02925959 0.07710579
> transmatrix <-result$estimate[1:5,1:5]
> upward <-result$estimate[1:5,1:5]
> downward <-result$estimate[1:5,1:5]
>   write.xlsx(transmatrix,   "M:/My   Documents/MScThesis/Text
thesis/Data           files/Chapter           5/File           for
R/Matrix_Biol_prod_2004_2015.xlsx.xlsx")
>
> #define transition matrix as markov object
> showClass("markovchain")
Class "markovchain" [package "markovchain"]

Slots:

Name:                 states               byrow transitionMatrix
name
Class:          character            logical               matrix
character
> simpleMc<-new("markovchain", states=c("a","b","c","d","e"),
+ transitionMatrix=transmatrix,
+ name="simpleMc")
>
>
> #stationary distribution
> steadyStates(result$estimate)
              a        b         c         d         e
[1,] 0.08205205 0.146261 0.3314779 0.2522797 0.1879294
>
> #half life
> eigenvalues <-eigen(transmatrix)
> eigenvaluevector <-eigenvalues$values
> secondeigenvalue <-eigenvaluevector[[2]]
> secondeigenvalue
[1] 0.9157691
> halflife <- (-log(2))/log((abs(secondeigenvalue)))
> halflife
[1] 7.877477
>
> #mobilityindex
> trace <-matrix.trace(transmatrix)
> mobility <- (5-trace)*(5-1)^-1
> mobility
[1] 0.2337991
>
> #upward mobility
> upward[lower.tri(upward)] <- 0
> upward[lower.tri(upward,diag=TRUE)] <- 0
> sum(upward)*(5-trace)^-1
[1] 0.5516423
>
> #downward mobility
> downward[upper.tri(downward)] <- 0
> downward[upper.tri(downward,diag=TRUE)] <- 0
> sum(downward)*(5-trace)^-1
[1] 0.4483577
>
> #load data from computer Biological productivty 2007-2015
>          b          <-read.table("//WURNET.NL/Homes/jong285/My
Documents/MScThesis/Text  thesis/Data  files/Chapter  5/File  for
R/Biological_productivity_2007-2015.txt", header=T)
>
> #define as matrix
> datamatrix2 <-as.matrix(b)
> datamatrix2
      X2007 X2008 X2009 X2010 X2011 X2012 X2013 X2014 X2015
 [1,] "e"   "e"   "e"   "e"   "e"   "d"   "e"   "d"   "d"
 [2,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "d"   "e"
 [3,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "d"   "e"
 [4,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
 [5,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
 [6,] "c"   "c"   "c"   "c"   "c"   "c"   "d"   "c"   "c"
 [7,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
 [8,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
 [9,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[10,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[11,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[12,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[13,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[14,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
[15,] "c"   "c"   "c"   "d"   "d"   "c"   "c"   "c"   "d"
[16,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
[17,] "d"   "e"   "d"   "d"   "d"   "e"   "d"   "e"   "d"
[18,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
[19,] "d"   "d"   "d"   "d"   "d"   "d"   "c"   "d"   "c"
[20,] "c"   "c"   "c"   "c"   "d"   "d"   "d"   "d"   "d"
[21,] "c"   "c"   "c"   "c"   "d"   "c"   "c"   "d"   "d"
[22,] "c"   "d"   "d"   "d"   "d"   "d"   "c"   "d"   "d"
[23,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
[24,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
[25,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
[26,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
[27,] "c"   "d"   "d"   "c"   "d"   "d"   "d"   "d"   "d"
[28,] "c"   "c"   "c"   "c"   "c"   "d"   "d"   "c"   "c"
```

```
[29,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
[30,] "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"
[31,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[32,] "d"   "b"   "c"   "b"   "b"   "b"   "b"   "c"   "b"
[33,] "d"   "d"   "d"   "d"   "e"   "d"   "d"   "d"   "e"
[34,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
[35,] "c"   "b"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
[36,] "d"   "c"   "d"   "d"   "d"   "d"   "c"   "d"   "d"
[37,] "c"   "b"   "c"   "c"   "c"   "c"   "c"   "c"   "b"
[38,] "d"   "d"   "d"   "c"   "c"   "d"   "c"   "d"   "d"
[39,] "b"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "b"
[40,] "a"   "b"   "a"   "a"   "b"   "b"   "a"   "b"   "b"
[41,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[42,] "c"   "c"   "c"   "d"   "d"   "d"   "c"   "c"   "d"
[43,] "a"   "a"   "b"   "b"   "a"   "a"   "a"   "a"   "a"
[44,] "d"   "d"   "d"   "e"   "e"   "d"   "c"   "d"   "d"
[45,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
[46,] "b"   "b"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
[47,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
[48,] "e"   "e"   "e"   "e"   "d"   "d"   "e"   "d"   "e"
[49,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[50,] "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"
[51,] "d"   "d"   "d"   "e"   "d"   "d"   "d"   "d"   "d"
[52,] "d"   "d"   "d"   "e"   "d"   "d"   "e"   "d"   "d"
[53,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
[54,] "d"   "d"   "c"   "d"   "d"   "d"   "d"   "d"   "c"
[55,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
[56,] "c"   "c"   "c"   "d"   "c"   "d"   "d"   "d"   "c"
[57,] "c"   "c"   "c"   "c"   "d"   "d"   "d"   "d"   "d"
[58,] "c"   "c"   "c"   "c"   "c"   "b"   "b"   "b"   "a"
[59,] "c"   "c"   "c"   "b"   "c"   "c"   "c"   "c"   "c"
[60,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[61,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[62,] "a"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"
[63,] "e"   "e"   "d"   "c"   "e"   "c"   "c"   "b"   "d"
[64,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "d"   "c"
[65,] "b"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
[66,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
[67,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[68,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[69,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[70,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[71,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[72,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[73,] "e"   "e"   "d"   "e"   "e"   "e"   "d"   "e"   "e"
[74,] "c"   "c"   "c"   "c"   "c"   "d"   "c"   "d"   "d"
[75,] "d"   "d"   "d"   "d"   "d"   "d"   "e"   "e"   "e"
[76,] "b"   "b"   "b"   "b"   "b"   "b"   "b"   "c"   "b"
[77,] "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"
[78,] "c"   "c"   "b"   "c"   "c"   "c"   "c"   "c"   "c"
[79,] "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"
[80,] "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"
[81,] "b"   "b"   "b"   "b"   "a"   "b"   "b"   "b"   "b"
[82,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[83,] "b"   "b"   "b"   "b"   "b"   "c"   "c"   "c"   "c"
[84,] "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"
[85,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[86,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[87,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[88,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[89,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[90,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[91,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
>
> #estimate transition matrix with MLE
>  result2  <-markovchainFit(data=datamatrix2,  method="mle",
name="datamatrix2")
> result2$estimate
datamatrix2
 A  5 - dimensional discrete Markov Chain defined by the following
states:
 a, b, c, d, e
 The transition matrix  (by rows)  is defined as follows:
           a           b           c          d           e
a 0.92929293 0.070707071 0.000000000 0.00000000 0.000000000
b 0.06185567 0.824742268 0.103092784 0.01030928 0.000000000
c 0.00000000 0.053763441 0.801075269 0.13978495 0.005376344
d 0.00000000 0.005291005 0.111111111 0.79365079 0.089947090
e 0.00000000 0.000000000 0.006369427 0.10828025 0.885350318

> result2$standardError
           a           b           c          d           e
a 0.09688549 0.026724761 0.000000000 0.00000000 0.000000000
b 0.02525247 0.092208989 0.032600801 0.01030928 0.000000000
c 0.00000000 0.017001493 0.065626643 0.02741408 0.005376344
d 0.00000000 0.005291005 0.024246432 0.06480132 0.021815374
e 0.00000000 0.000000000 0.006369427 0.02626182 0.075094434
> transmatrix2<- result2$estimate[1:5,1:5]
> transmatrix2
           a           b           c          d           e
a 0.92929293 0.070707071 0.000000000 0.00000000 0.000000000
b 0.06185567 0.824742268 0.103092784 0.01030928 0.000000000
c 0.00000000 0.053763441 0.801075269 0.13978495 0.005376344
d 0.00000000 0.005291005 0.111111111 0.79365079 0.089947090
e 0.00000000 0.000000000 0.006369427 0.10828025 0.885350318
> upward2 <-result2$estimate[1:5,1:5]
> downward2 <-result2$estimate[1:5,1:5]
>  write.xlsx(transmatrix2,  "M:/My  Documents/MScThesis/Text
thesis/Data          files/Chapter          5/File          for
R/Matrix_Biol_prod_2007_2015.xlsx")
>
> #define transition matrix as markov object
> showClass("markovchain")
Class "markovchain" [package "markovchain"]

Slots:

Name:               states               byrow transitionMatrix
name
Class:          character             logical              matrix
character
> simpleMc2<-new("markovchain", states=c("a","b","c","d","e"),
+ transitionMatrix=transmatrix2,
+ name="simpleMc2")
>
> #stationary distribution
> steadyStates(simpleMc2)
             a         b         c         d         e
[1,] 0.1093753 0.1250267 0.2350665 0.2911163 0.2394151
>
> #half life
> eigenvalues2 <-eigen(transmatrix2)
> eigenvaluevector2 <-eigenvalues2$values
> secondeigenvalue2 <-eigenvaluevector2[[2]]
> secondeigenvalue2
[1] 0.9572976
> halflife2 <- (-log(2))/log((abs(secondeigenvalue2)))
> halflife2
[1] 15.88294
>
> #mobilityindex
> trace2 <-matrix.trace(transmatrix2)
> mobility2 <- (5-trace2)*(5-1)^-1
> mobility2
[1] 0.1914721
>
> #upward mobility
> upward2[lower.tri(upward2)] <- 0
> upward2[lower.tri(upward2,diag=TRUE)] <- 0
> sum(upward2)*(5-trace2)^-1
[1] 0.5473611
>
> #downward mobility
> downward2[upper.tri(downward2)] <- 0
> downward2[upper.tri(downward2,diag=TRUE)] <- 0
> sum(downward2)*(5-trace2)^-1
[1] 0.4526389
>
> #load data from computer Mechanical productivity 2004-2015
>          c         <-read.table("//WURNET.NL/Homes/jong285/My
Documents/MScThesis/Text  thesis/Data  files/Chapter  5/File  for
R/Mechanical_productivity_2004-2015.txt", header=T)
>
> #define as matrix
> datamatrix3 <-as.matrix(c)
> datamatrix3
      X2004 X2005 X2006 X2007 X2008 X2009 X2010 X2011 X2012 X2013
X2014 X2015
 [1,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
"e"   "e"
 [2,] "e"   "d"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
"e"   "e"
 [3,] "d"   "d"   "d"   "e"   "d"   "e"   "d"   "d"   "e"   "e"
"e"   "e"
 [4,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "d"   "d"
"c"   "c"
 [5,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
"d"   "d"
 [6,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "d"   "d"
"c"   "c"
 [7,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
"c"   "c"
```

```
 [8,] "d"  "d"  "d"  "d"  "d"  "d"  "d"  "d"  "d"  "d"
"d"  "d"
 [9,] "c"  "c"  "c"  "c"  "c"  "b"  "b"  "b"  "c"  "b"
"b"  "c"
[10,] "d"  "d"  "d"  "d"  "d"  "c"  "c"  "d"  "d"  "c"
"d"  "d"
[11,] "b"  "b"  "b"  "b"  "b"  "b"  "b"  "b"  "b"  "b"
"b"  "b"
[12,] "c"  "c"  "c"  "c"  "c"  "c"  "c"  "c"  "c"  "c"
"c"  "c"
[13,] "b"  "b"  "b"  "c"  "b"  "b"  "b"  "b"  "b"  "b"
"b"  "b"
[14,] "d"  "d"  "d"  "d"  "d"  "d"  "d"  "d"  "c"  "c"
"c"  "c"
[15,] "d"  "d"  "d"  "d"  "d"  "d"  "d"  "d"  "d"  "d"
"d"  "d"
[16,] "d"  "d"  "d"  "d"  "d"  "d"  "d"  "d"  "d"  "d"
"d"  "d"
[17,] "d"  "d"  "d"  "d"  "d"  "d"  "c"  "d"  "d"  "d"
"c"  "d"
[18,] "d"  "d"  "d"  "d"  "d"  "d"  "d"  "d"  "d"  "d"
"d"  "d"
[19,] "c"  "c"  "c"  "c"  "c"  "c"  "c"  "c"  "c"  "c"
"c"  "c"
[20,] "c"  "c"  "c"  "c"  "d"  "c"  "d"  "d"  "d"  "d"
"d"  "d"
[21,] "c"  "c"  "c"  "c"  "c"  "c"  "c"  "c"  "c"  "c"
"c"  "c"
[22,] "c"  "c"  "c"  "c"  "c"  "d"  "d"  "d"  "d"  "c"
"c"  "c"
[23,] "c"  "c"  "c"  "c"  "c"  "c"  "c"  "c"  "c"  "c"
"c"  "c"
[24,] "c"  "c"  "c"  "c"  "c"  "c"  "c"  "c"  "c"  "c"
"c"  "c"
[25,] "c"  "c"  "c"  "c"  "c"  "c"  "d"  "d"  "c"  "d"
"d"  "d"
[26,] "c"  "c"  "c"  "c"  "c"  "c"  "d"  "d"  "d"  "d"
"d"  "d"
[27,] "c"  "d"  "d"  "d"  "c"  "c"  "c"  "c"  "c"  "c"
"c"  "d"
[28,] "c"  "c"  "c"  "c"  "c"  "c"  "d"  "d"  "c"  "d"
"d"  "c"
[29,] "c"  "c"  "c"  "c"  "c"  "c"  "c"  "c"  "c"  "c"
"c"  "c"
[30,] "d"  "c"  "c"  "c"  "c"  "c"  "c"  "d"  "c"  "c"
"c"  "c"
[31,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "b"  "a"  "a"
"b"  "b"
[32,] "d"  "d"  "d"  "d"  "c"  "c"  "c"  "d"  "d"  "c"
"c"  "c"
[33,] "d"  "d"  "d"  "d"  "e"  "e"  "d"  "d"  "d"  "d"
"d"  "d"
[34,] "b"  "b"  "b"  "b"  "b"  "b"  "a"  "b"  "a"  "b"
"b"  "b"
[35,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[36,] "b"  "b"  "c"  "b"  "b"  "c"  "c"  "c"  "c"  "c"
"c"  "c"
[37,] "b"  "b"  "b"  "b"  "c"  "c"  "b"  "b"  "b"  "c"
"c"  "b"
[38,] "c"  "c"  "c"  "c"  "d"  "d"  "c"  "c"  "c"  "c"
"c"  "c"
[39,] "c"  "c"  "c"  "c"  "c"  "c"  "c"  "c"  "c"  "c"
"d"  "d"
[40,] "a"  "a"  "b"  "a"  "a"  "b"  "a"  "a"  "b"  "b"
"b"  "b"
[41,] "c"  "c"  "c"  "c"  "c"  "d"  "d"  "d"  "d"  "d"
"e"  "e"
[42,] "b"  "c"  "b"  "b"  "b"  "b"  "b"  "b"  "b"  "b"
"b"  "b"
[43,] "d"  "d"  "d"  "b"  "b"  "b"  "b"  "b"  "b"  "c"
"e"  "d"
[44,] "d"  "d"  "d"  "d"  "c"  "b"  "b"  "c"  "c"  "d"
"c"  "c"
[45,] "d"  "d"  "d"  "d"  "d"  "d"  "d"  "d"  "d"  "d"
"d"  "d"
[46,] "d"  "d"  "d"  "d"  "e"  "e"  "d"  "d"  "d"  "d"
"e"  "d"
[47,] "c"  "c"  "c"  "c"  "d"  "d"  "d"  "d"  "d"  "d"
"d"  "d"
[48,] "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"
"e"  "e"
[49,] "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"
"e"  "e"
[50,] "d"  "d"  "d"  "e"  "e"  "e"  "e"  "e"  "e"  "e"
"e"  "e"
[51,] "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"
"e"  "e"
[52,] "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "d"  "d"
"d"  "d"
[53,] "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"
"e"  "e"
[54,] "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"
"e"  "e"
[55,] "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"
"e"  "e"
[56,] "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"
"e"  "e"
[57,] "b"  "b"  "b"  "b"  "b"  "c"  "b"  "b"  "c"  "c"
"c"  "c"
[58,] "b"  "b"  "b"  "b"  "c"  "c"  "c"  "c"  "c"  "c"
"c"  "c"
[59,] "c"  "c"  "c"  "c"  "c"  "d"  "d"  "c"  "c"  "c"
"c"  "c"
[60,] "c"  "c"  "c"  "c"  "d"  "c"  "c"  "d"  "b"  "c"
"c"  "c"
[61,] "e"  "e"  "e"  "e"  "e"  "e"  "e"  "d"  "d"  "e"
"d"  "d"
[62,] "e"  "e"  "e"  "e"  "e"  "d"  "d"  "d"  "d"  "d"
"d"  "d"
[63,] "d"  "d"  "d"  "d"  "d"  "d"  "c"  "c"  "c"  "c"
"c"  "c"
[64,] "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"
"e"  "e"
[65,] "b"  "c"  "b"  "c"  "b"  "b"  "b"  "b"  "b"  "b"
"b"  "c"
[66,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[67,] "a"  "a"  "a"  "a"  "a"  "b"  "b"  "b"  "b"  "b"
"b"  "b"
[68,] "a"  "a"  "a"  "a"  "a"  "b"  "b"  "b"  "b"  "b"
"b"  "b"
[69,] "a"  "a"  "a"  "b"  "b"  "b"  "b"  "b"  "b"  "b"
"b"  "b"
[70,] "a"  "a"  "a"  "a"  "b"  "b"  "b"  "b"  "b"  "b"
"b"  "b"
[71,] "c"  "c"  "c"  "c"  "c"  "d"  "c"  "c"  "d"  "d"
"c"  "d"
[72,] "c"  "c"  "c"  "c"  "c"  "d"  "c"  "d"  "d"  "d"
"d"  "d"
[73,] "b"  "b"  "b"  "c"  "c"  "c"  "c"  "c"  "c"  "c"
"c"  "c"
[74,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[75,] "a"  "a"  "a"  "a"  "a"  "a"  "b"  "b"  "b"  "b"
"b"  "b"
[76,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[77,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[78,] "d"  "d"  "d"  "c"  "c"  "c"  "c"  "c"  "c"  "c"
"c"  "a"
[79,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[80,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[81,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[82,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[83,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[84,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
>
> #estimate transition matrix with MLE
>  result3  <-markovchainFit(data=datamatrix3,  method="mle",
name="datamatrix3")
> result3$estimate
datamatrix3
 A  5 - dimensional discrete Markov Chain defined by the following
states:
 a, b, c, d, e
 The transition matrix  (by rows)  is defined as follows:
            a           b         c          d           e
a 0.925465839 0.074534161 0.0000000 0.00000000 0.000000000
b 0.039370079 0.818897638 0.1417323 0.00000000 0.000000000
c 0.003649635 0.040145985 0.8394161 0.11313869 0.003649635
d 0.000000000 0.009433962 0.1273585 0.81603774 0.047169811
e 0.000000000 0.000000000 0.0000000 0.07333333 0.926666667

> result3$standardError
```

```
            a           b          c          d           e
a 0.075817116 0.021516159 0.00000000 0.00000000 0.000000000
b 0.017606834 0.080299520 0.03340662 0.00000000 0.000000000
c 0.003649635 0.012104470 0.05534946 0.02032031 0.003649635
d 0.000000000 0.006670819 0.02451015 0.06204220 0.014916404
e 0.000000000 0.000000000 0.00000000 0.02211083 0.078598841
> transmatrix3<- result3$estimate[1:5,1:5]
> transmatrix3
            a           b         c          d           e
a 0.925465839 0.074534161 0.0000000 0.00000000 0.000000000
b 0.039370079 0.818897638 0.1417323 0.00000000 0.000000000
c 0.003649635 0.040145985 0.8394161 0.11313869 0.003649635
d 0.000000000 0.009433962 0.1273585 0.81603774 0.047169811
e 0.000000000 0.000000000 0.0000000 0.07333333 0.926666667
> upward3 <-result3$estimate[1:5,1:5]
> downward3 <-result3$estimate[1:5,1:5]
>   write.xlsx(transmatrix3,   "M:/My   Documents/MScThesis/Text
thesis/Data           files/Chapter          5/File          for
R/Matrix_Mech_prod_2004_2015.xlsx")
>
> #define transition matrix as markov object
> showClass("markovchain")
Class "markovchain" [package "markovchain"]

Slots:

Name:               states               byrow transitionMatrix
name
Class:          character             logical             matrix
character
> simpleMc3<-new("markovchain", states=c("a","b","c","d","e"),
+ transitionMatrix=transmatrix3,
+ name="simpleMc3")
>
>
> #stationary distribution
> steadyStates(simpleMc3)
              a         b         c         d         e
[1,] 0.07913259 0.1195241 0.3267199 0.2789413 0.1956821
>
> #half life
> eigenvalues3 <-eigen(transmatrix3)
> eigenvaluevector3 <-eigenvalues3$values
> secondeigenvalue3 <-eigenvaluevector3[[2]]
> secondeigenvalue3
[1] 0.9484024
> halflife3 <- (-log(2))/log((abs(secondeigenvalue3)))
> halflife3
[1] 13.08409
>
> #mobilityindex
> trace3 <-matrix.trace(transmatrix3)
> mobility3 <- (5-trace3)*(5-1)^-1
> mobility3
[1] 0.168379
>
> #upward mobility
> upward3[lower.tri(upward3)] <- 0
> upward3[lower.tri(upward3,diag=TRUE)] <- 0
> sum(upward3)*(5-trace3)^-1
[1] 0.5645368
>
> #downward mobility
> downward3[upper.tri(downward3)] <- 0
> downward3[upper.tri(downward3,diag=TRUE)] <- 0
> sum(downward3)*(5-trace3)^-1
[1] 0.4354632
>
> #load data from computer Mechanical productivity 2007-2015
>          d          <-read.table("//WURNET.NL/Homes/jong285/My
Documents/MScThesis/Text  thesis/Data  files/Chapter  5/File  for
R/Mechanical_productivity_2007-2015.txt", header=T)
>
> #define as matrix
> datamatrix4 <-as.matrix(d)
> datamatrix4
      X2007 X2008 X2009 X2010 X2011 X2012 X2013 X2014 X2015
 [1,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
 [2,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
 [3,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
 [4,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
 [5,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
 [6,] "c"   "c"   "c"   "c"   "c"   "d"   "d"   "d"   "d"
 [7,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
 [8,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
 [9,] "c"   "c"   "b"   "c"   "c"   "c"   "b"   "b"   "c"
[10,] "d"   "d"   "d"   "c"   "d"   "d"   "d"   "d"   "d"
[11,] "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"
[12,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
[13,] "c"   "c"   "c"   "c"   "b"   "b"   "b"   "b"   "b"
[14,] "d"   "d"   "d"   "d"   "d"   "c"   "c"   "c"   "d"
[15,] "d"   "d"   "d"   "d"   "e"   "e"   "d"   "d"   "d"
[16,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
[17,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
[18,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
[19,] "d"   "d"   "c"   "c"   "c"   "c"   "c"   "c"   "d"
[20,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
[21,] "c"   "d"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
[22,] "c"   "d"   "d"   "d"   "d"   "d"   "d"   "c"   "c"
[23,] "c"   "d"   "d"   "d"   "d"   "c"   "d"   "c"   "c"
[24,] "d"   "d"   "c"   "d"   "d"   "d"   "d"   "d"   "c"
[25,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
[26,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
[27,] "d"   "d"   "c"   "c"   "c"   "c"   "c"   "c"   "d"
[28,] "d"   "d"   "d"   "d"   "d"   "c"   "d"   "d"   "d"
[29,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
[30,] "c"   "d"   "c"   "d"   "d"   "d"   "c"   "c"   "c"
[31,] "a"   "a"   "a"   "a"   "b"   "a"   "a"   "b"   "b"
[32,] "d"   "c"   "c"   "d"   "d"   "d"   "c"   "d"   "d"
[33,] "d"   "e"   "e"   "e"   "d"   "e"   "e"   "e"   "e"
[34,] "b"   "b"   "b"   "a"   "b"   "b"   "b"   "b"   "b"
[35,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[36,] "b"   "b"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
[37,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "b"
[38,] "c"   "e"   "e"   "c"   "c"   "c"   "c"   "c"   "c"
[39,] "c"   "c"   "c"   "c"   "c"   "c"   "d"   "d"   "d"
[40,] "a"   "a"   "b"   "a"   "a"   "b"   "b"   "b"   "b"
[41,] "c"   "d"   "d"   "d"   "d"   "e"   "e"   "e"   "e"
[42,] "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"
[43,] "b"   "b"   "b"   "b"   "b"   "b"   "c"   "e"   "d"
[44,] "d"   "c"   "c"   "b"   "c"   "c"   "d"   "d"   "c"
[45,] "e"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
[46,] "e"   "e"   "e"   "d"   "e"   "e"   "e"   "e"   "e"
[47,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
[48,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[49,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[50,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[51,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[52,] "e"   "e"   "e"   "e"   "e"   "d"   "d"   "e"   "e"
[53,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[54,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[55,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[56,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[57,] "b"   "b"   "c"   "b"   "b"   "c"   "c"   "c"   "c"
[58,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
[59,] "c"   "c"   "d"   "d"   "d"   "d"   "d"   "c"   "c"
[60,] "c"   "d"   "d"   "d"   "d"   "b"   "c"   "c"   "c"
[61,] "e"   "e"   "e"   "e"   "e"   "d"   "e"   "e"   "d"
[62,] "e"   "e"   "e"   "e"   "e"   "d"   "e"   "d"   "d"
[63,] "d"   "d"   "d"   "d"   "c"   "c"   "c"   "c"   "c"
[64,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[65,] "c"   "c"   "b"   "c"   "c"   "c"   "c"   "c"   "c"
[66,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[67,] "a"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"
[68,] "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"
[69,] "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"
[70,] "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"
[71,] "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
[72,] "c"   "d"   "d"   "d"   "d"   "d"   "d"   "d"   "d"
[73,] "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
[74,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[75,] "a"   "a"   "a"   "b"   "b"   "b"   "b"   "b"   "b"
[76,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[77,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[78,] "d"   "d"   "c"   "c"   "c"   "c"   "c"   "c"   "a"
[79,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[80,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[81,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[82,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[83,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[84,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[85,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[86,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[87,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[88,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[89,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[90,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[91,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
>
> #estimate transition matrix with MLE
>   result4   <-markovchainFit(data=datamatrix4,   method="mle",
name="datamatrix4")
```

```
> result4$estimate
datamatrix4
 A  5 - dimensional discrete Markov Chain defined by the following
states:
 a, b, c, d, e
 The transition matrix  (by rows)  is defined as follows:
            a           b          c         d          e
a 0.955974843 0.044025157 0.00000000 0.0000000 0.00000000
b 0.034482759 0.862068966 0.10344828 0.0000000 0.00000000
c 0.006622517 0.046357616 0.79470199 0.1390728 0.01324503
d 0.000000000 0.005154639 0.10309278 0.8505155 0.04123711
e 0.000000000 0.000000000 0.00729927 0.0729927 0.91970803

> result4$standardError
            a           b          c          d           e
a 0.077539799 0.016639945 0.00000000 0.00000000 0.000000000
b 0.019908630 0.099543150 0.03448276 0.00000000 0.000000000
c 0.006622517 0.017521532 0.07254603 0.03034818 0.009365653
d 0.000000000 0.005154639 0.02305225 0.06621254 0.014579521
e 0.000000000 0.000000000 0.00729927 0.02308232 0.081934103
> transmatrix4<- result4$estimate[1:5,1:5]
> transmatrix4
            a           b          c         d          e
a 0.955974843 0.044025157 0.00000000 0.0000000 0.00000000
b 0.034482759 0.862068966 0.10344828 0.0000000 0.00000000
c 0.006622517 0.046357616 0.79470199 0.1390728 0.01324503
d 0.000000000 0.005154639 0.10309278 0.8505155 0.04123711
e 0.000000000 0.000000000 0.00729927 0.0729927 0.91970803
> upward4 <-result4$estimate[1:5,1:5]
> downward4 <-result4$estimate[1:5,1:5]
>   write.xlsx(transmatrix4,   "M:/My   Documents/MScThesis/Text
thesis/Data           files/Chapter           5/File          for
R/Matrix_Mech_prod_2007_2015.xlsx")
>
> #define transition matrix as markov object
> showClass("markovchain")
Class "markovchain" [package "markovchain"]

Slots:

Name:               states               byrow transitionMatrix
name
Class:          character             logical             matrix
character
> simpleMc4<-new("markovchain", states=c("a","b","c","d","e"),
+ transitionMatrix=transmatrix4,
+ name="simpleMc4")
>
> #stationary distribution
> steadyStates(simpleMc4)
             a         b         c         d         e
[1,] 0.1376105 0.1319515 0.2277487 0.3072958 0.1953935
>
> #half life
> eigenvalues4 <-eigen(transmatrix4)
> eigenvaluevector4 <-eigenvalues4$values
> secondeigenvalue4 <-eigenvaluevector4[[2]]
> secondeigenvalue4
[1] 0.9657109
> halflife4 <- (-log(2))/log((abs(secondeigenvalue4)))
> halflife4
[1] 19.86623
>
> #mobilityindex
> trace4 <-matrix.trace(transmatrix4)
> mobility4 <- (5-trace4)*(5-1)^-1
> mobility4
[1] 0.1542577
>
> #upward mobility
> upward4[lower.tri(upward4)] <- 0
> upward4[lower.tri(upward4,diag=TRUE)] <- 0
> sum(upward4)*(5-trace4)^-1
[1] 0.5526928
>
> #downward mobility
> downward4[upper.tri(downward4)] <- 0
> downward4[upper.tri(downward4,diag=TRUE)] <- 0
> sum(downward4)*(5-trace4)^-1
[1] 0.4473072
```

APPENDIX IV CHAPTER 6 R-OUTPUT

```
> #loadpackage
> library(markovchain)
Package:  markovchain
Version:  0.6.9.8-1
Date:     2017-08-15
BugReport: http://github.com/spedygiorgio/markovchain/issues

Warning message:
package ‘markovchain’ was built under R version 3.3.3
> library(matrixcalc)
Warning message:
package ‘matrixcalc’ was built under R version 3.3.2
> library(rJava)
Warning message:
package ‘rJava’ was built under R version 3.3.3
> library(xlsxjars)
Warning message:
package ‘xlsxjars’ was built under R version 3.3.3
> library(xlsx)
Warning message:
package ‘xlsx’ was built under R version 3.3.3
> library(plotly)
Loading required package: ggplot2

Attaching package: ‘plotly’

The following object is masked from ‘package:ggplot2’:

    last_plot

The following object is masked from ‘package:stats’:

    filter

The following object is masked from ‘package:graphics’:

    layout

Warning messages:
1: package ‘plotly’ was built under R version 3.3.3
2: package ‘ggplot2’ was built under R version 3.3.3
>
> #load data from computer Labour-adjusted value added 2004-2015
> a <-read.table("//WURNET.NL/Homes/jong285/My Documents/MScThesis/Text thesis/Data files/Chapter 6/File for R/Labour_adjusted_value_added_2004-2015.txt", header=T)
>
> #define as matrix
> datamatrix <-as.matrix(a)
> datamatrix
      X2004 X2005 X2006 X2007 X2008 X2009 X2010 X2011 X2012 X2013
X2014 X2015
 [1,] "e"    "e"    "e"    "e"    "d"    "e"    "e"    "e"    "e"    "e"
"d"    "d"
 [2,] "e"    "e"    "e"    "e"    "d"    "e"    "e"    "e"    "e"    "e"
"d"    "d"
 [3,] "e"    "e"    "e"    "e"    "d"    "e"    "e"    "d"    "e"    "e"
"d"    "d"
 [4,] "c"    "c"    "c"    "c"    "b"    "c"    "c"    "c"    "c"    "d"
"c"    "c"
 [5,] "d"    "d"    "d"    "e"    "c"    "d"    "d"    "c"    "d"    "d"
"c"    "c"
 [6,] "c"    "c"    "c"    "d"    "b"    "c"    "c"    "c"    "c"    "d"
"c"    "d"
 [7,] "c"    "c"    "c"    "c"    "b"    "c"    "c"    "c"    "c"    "c"
"c"    "c"
 [8,] "d"    "d"    "d"    "d"    "c"    "c"    "d"    "c"    "d"    "d"
"c"    "c"
 [9,] "d"    "d"    "d"    "d"    "c"    "c"    "c"    "c"    "d"    "d"
"d"    "d"
[10,] "e"    "e"    "e"    "e"    "d"    "d"    "d"    "d"    "d"    "d"
"d"    "d"
[11,] "c"    "c"    "c"    "c"    "c"    "c"    "c"    "c"    "d"    "c"
"c"    "c"
[12,] "e"    "d"    "d"    "d"    "c"    "d"    "d"    "d"    "d"    "d"
"c"    "c"
[13,] "c"    "c"    "c"    "d"    "b"    "b"    "c"    "c"    "d"    "c"
"c"    "c"
[14,] "d"    "c"    "c"    "c"    "c"    "b"    "d"    "d"    "c"    "c"
"c"    "c"
[15,] "c"    "c"    "c"    "d"    "d"    "b"    "d"    "d"    "d"    "c"
"c"    "d"
[16,] "d"    "c"    "d"    "c"    "c"    "c"    "d"    "d"    "d"    "d"
"c"    "c"
[17,] "c"    "d"    "c"    "d"    "c"    "b"    "c"    "c"    "c"    "b"
"c"    "c"
[18,] "c"    "c"    "c"    "c"    "c"    "b"    "d"    "d"    "c"    "c"
"c"    "c"
[19,] "c"    "c"    "c"    "c"    "c"    "b"    "c"    "c"    "c"    "b"
"b"    "c"
[20,] "c"    "b"    "c"    "c"    "d"    "b"    "d"    "d"    "c"    "c"
"c"    "c"
[21,] "c"    "c"    "c"    "c"    "c"    "b"    "c"    "c"    "c"    "b"
"c"    "c"
[22,] "c"    "c"    "c"    "c"    "c"    "b"    "d"    "c"    "c"    "b"
"b"    "b"
[23,] "c"    "c"    "c"    "b"    "b"    "c"    "d"    "c"    "d"    "c"
"c"    "d"
[24,] "c"    "c"    "c"    "c"    "c"    "c"    "c"    "d"    "c"    "c"
"c"    "c"
[25,] "c"    "c"    "c"    "c"    "c"    "b"    "c"    "c"    "c"    "c"
"c"    "c"
[26,] "c"    "c"    "c"    "c"    "c"    "b"    "c"    "c"    "c"    "c"
"c"    "c"
[27,] "b"    "b"    "b"    "b"    "a"    "a"    "b"    "a"    "b"    "a"
"b"    "b"
[28,] "b"    "b"    "b"    "b"    "b"    "a"    "b"    "b"    "b"    "b"
"b"    "b"
[29,] "b"    "b"    "b"    "b"    "b"    "b"    "c"    "c"    "b"    "b"
"b"    "c"
[30,] "b"    "b"    "a"    "a"    "b"    "b"    "c"    "b"    "b"    "b"
"b"    "b"
[31,] "b"    "b"    "b"    "b"    "b"    "b"    "b"    "b"    "b"    "b"
"b"    "b"
[32,] "d"    "d"    "d"    "d"    "d"    "d"    "e"    "d"    "d"    "d"
"d"    "d"
[33,] "e"    "e"    "e"    "e"    "e"    "e"    "e"    "e"    "e"    "e"
"e"    "e"
[34,] "a"    "b"    "b"    "b"    "a"    "b"    "b"    "b"    "b"    "a"
"b"    "c"
[35,] "a"    "a"    "a"    "a"    "a"    "a"    "a"    "a"    "a"    "a"
"a"    "b"
[36,] "c"    "c"    "c"    "c"    "c"    "d"    "d"    "d"    "d"    "d"
"e"    "d"
[37,] "b"    "c"    "c"    "c"    "c"    "d"    "c"    "c"    "b"    "b"
"c"    "c"
[38,] "d"    "c"    "c"    "c"    "e"    "e"    "e"    "e"    "e"    "e"
"e"    "e"
[39,] "d"    "e"    "e"    "d"    "c"    "d"    "d"    "d"    "c"    "c"
"e"    "e"
[40,] "b"    "c"    "c"    "b"    "b"    "c"    "b"    "b"    "d"    "d"
"d"    "d"
[41,] "d"    "d"    "e"    "d"    "e"    "e"    "e"    "d"    "d"    "c"
"e"    "e"
[42,] "b"    "b"    "c"    "b"    "b"    "b"    "b"    "b"    "b"    "b"
"b"    "c"
[43,] "d"    "e"    "e"    "c"    "c"    "c"    "c"    "b"    "b"    "b"
"e"    "e"
[44,] "e"    "e"    "e"    "d"    "d"    "e"    "d"    "e"    "e"    "e"
"e"    "e"
[45,] "d"    "e"    "e"    "e"    "d"    "d"    "e"    "d"    "d"    "e"
"e"    "d"
[46,] "e"    "d"    "d"    "e"    "e"    "d"    "d"    "d"    "d"    "d"
"e"    "d"
[47,] "d"    "d"    "d"    "d"    "e"    "c"    "c"    "d"    "d"    "d"
"d"    "e"
[48,] "e"    "e"    "e"    "e"    "e"    "e"    "e"    "e"    "e"    "e"
"e"    "e"
[49,] "e"    "e"    "e"    "e"    "e"    "e"    "e"    "e"    "e"    "e"
"e"    "e"
[50,] "d"    "d"    "c"    "d"    "d"    "c"    "d"    "e"    "d"    "d"
"e"    "e"
[51,] "e"    "d"    "d"    "d"    "d"    "d"    "d"    "e"    "d"    "e"
"e"    "d"
[52,] "d"    "e"    "d"    "d"    "d"    "e"    "d"    "d"    "d"    "d"
"e"    "c"
[53,] "e"    "e"    "d"    "e"    "e"    "e"    "d"    "e"    "e"    "e"
"e"    "e"
[54,] "e"    "e"    "d"    "d"    "e"    "e"    "d"    "e"    "d"    "e"
"e"    "c"
[55,] "d"    "d"    "d"    "e"    "e"    "e"    "d"    "d"    "d"    "d"
"c"    "a"
[56,] "d"    "c"    "c"    "e"    "c"    "b"    "d"    "d"    "c"    "d"
"c"    "b"
[57,] "b"    "c"    "c"    "c"    "c"    "c"    "b"    "b"    "b"    "b"
"b"    "b"
[58,] "b"    "b"    "b"    "b"    "b"    "b"    "b"    "a"    "b"    "a"
"b"    "b"
```

```
[59,] "c"   "c"   "c"   "c"   "c"   "c"   "b"   "b"   "b"   "b"
"b"   "c"
[60,] "b"   "b"   "c"   "c"   "c"   "c"   "b"   "b"   "a"   "b"
"c"   "b"
[61,] "d"   "d"   "d"   "e"   "e"   "c"   "c"   "c"   "c"   "d"
"e"   "d"
[62,] "e"   "e"   "e"   "e"   "d"   "d"   "c"   "b"   "b"   "c"
"c"   "d"
[63,] "e"   "e"   "e"   "e"   "e"   "e"   "c"   "e"   "c"   "c"
"b"   "c"
[64,] "e"   "e"   "e"   "e"   "e"   "e"   "d"   "e"   "e"   "e"
"d"   "c"
[65,] "b"   "b"   "b"   "b"   "b"   "c"   "b"   "b"   "b"   "b"
"b"   "c"
[66,] "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"
"b"   "a"
[67,] "b"   "b"   "b"   "b"   "c"   "c"   "c"   "b"   "c"   "b"
"c"   "b"
[68,] "b"   "b"   "b"   "b"   "b"   "c"   "c"   "b"   "c"   "b"
"b"   "b"
[69,] "b"   "b"   "b"   "b"   "c"   "d"   "c"   "c"   "d"   "c"
"c"   "c"
[70,] "b"   "b"   "b"   "b"   "c"   "c"   "c"   "b"   "c"   "b"
"b"   "b"
[71,] "c"   "c"   "c"   "c"   "d"   "c"   "d"   "d"   "d"   "d"
"c"   "e"
[72,] "b"   "c"   "c"   "c"   "d"   "c"   "d"   "d"   "d"   "d"
"d"   "e"
[73,] "b"   "b"   "c"   "c"   "e"   "c"   "c"   "c"   "c"   "c"
"c"   "d"
[74,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "b"   "a"
"b"   "b"
[75,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "b"   "b"   "b"
"a"   "a"
[76,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
"a"   "a"
[77,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
"a"   "a"
[78,] "b"   "b"   "b"   "b"   "c"   "a"   "a"   "a"   "a"   "a"
"a"   "a"
[79,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
"a"   "a"
[80,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
"a"   "a"
[81,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
"a"   "a"
[82,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
"a"   "a"
[83,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
"a"   "a"
[84,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
"a"   "a"
>
> #estimate transition matrix with MLE
> result <-markovchainFit(data=datamatrix, method="mle", name="datamatrix")
> result$estimate
datamatrix
 A 5 - dimensional discrete Markov Chain defined by the following states:
 a, b, c, d, e
 The transition matrix  (by rows)  is defined as follows:
            a         b          c          d           e
a 0.888888889 0.1111111 0.00000000 0.00000000 0.000000000
b 0.070270270 0.6540541 0.23243243 0.03783784 0.005405405
c 0.007722008 0.1505792 0.66023166 0.15444015 0.027027027
d 0.000000000 0.0212766 0.23404255 0.54787234 0.196808511
e 0.000000000 0.0000000 0.06369427 0.24203822 0.694267516

> result$standardError
            a          b          c          d           e
a 0.081144083 0.02868877 0.00000000 0.00000000 0.000000000
b 0.019489466 0.05945946 0.03544561 0.01430136 0.005405405
c 0.005460284 0.02411196 0.05048918 0.02441913 0.010215256
d 0.000000000 0.01063830 0.03528324 0.05398347 0.032355120
e 0.000000000 0.00000000 0.02014190 0.03926378 0.066498768
> transmatrix<- result$estimate[1:5,1:5]
> transmatrix
            a         b          c          d           e
a 0.888888889 0.1111111 0.00000000 0.00000000 0.000000000
b 0.070270270 0.6540541 0.23243243 0.03783784 0.005405405
c 0.007722008 0.1505792 0.66023166 0.15444015 0.027027027
d 0.000000000 0.0212766 0.23404255 0.54787234 0.196808511
e 0.000000000 0.0000000 0.06369427 0.24203822 0.694267516
> upward <-result$estimate[1:5,1:5]
> downward <-result$estimate[1:5,1:5]
> write.xlsx(transmatrix, "M:/My Documents/MScThesis/Text thesis/Data files/Chapter 6/File for R/Matrix_Labour_adjusted_value_added_2004_2015.xlsx.xlsx")
>
> #define transition matrix as matrix object
> showClass("markovchain")
Class "markovchain" [package "markovchain"]

Slots:

Name:          states          byrow transitionMatrix name
Class:      character        logical           matrix character
> simpleMc<-new("markovchain", states=c("a","b","c","d","e"),
+ transitionMatrix=transmatrix,
+ name="simpleMc")
>
> #stationary distribution
> steadyStates(simpleMc)
             a         b         c         d         e
[1,] 0.1405985 0.1891566 0.3017308 0.2059322 0.1625819
>
> #half life
> eigenvalues <-eigen(transmatrix)
> eigenvaluevector <-eigenvalues$values
> secondeigenvalue <-eigenvaluevector[[2]]
> secondeigenvalue
[1] 0.9010496
> halflife <- (-log(2))/log((abs(secondeigenvalue)))
> halflife
[1] 6.652406
>
> #mobilityindex
> trace <-matrix.trace(transmatrix)
> mobility <- (5-trace)*(5-1)^-1
> mobility
[1] 0.3886714
>
> #upward mobility
> upward[lower.tri(upward)] <- 0
> upward[lower.tri(upward,diag=TRUE)] <- 0
> sum(upward)*(5-trace)^-1
[1] 0.4921011
>
> #downward mobility
> downward[upper.tri(downward)] <- 0
> downward[upper.tri(downward,diag=TRUE)] <- 0
> sum(downward)*(5-trace)^-1
[1] 0.5078989
>
>
> #load data from computer Labour-adjusted value added 2007-2015
> b <-read.table("//WURNET.NL/Homes/jong285/My Documents/MScThesis/Text thesis/Data files/Chapter 6/File for R/Labour_adjusted_value_added_2007-2015.txt", header=T)
>
> #define as matrix
> datamatrix2 <-as.matrix(b)
> datamatrix2
      X2007 X2008 X2009 X2010 X2011 X2012 X2013 X2014 X2015
 [1,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "d"   "e"
 [2,] "e"   "d"   "e"   "e"   "e"   "e"   "e"   "d"   "d"
 [3,] "e"   "d"   "e"   "e"   "e"   "e"   "e"   "d"   "d"
 [4,] "d"   "c"   "c"   "d"   "c"   "c"   "d"   "c"   "c"
 [5,] "e"   "d"   "d"   "d"   "c"   "d"   "e"   "c"   "c"
 [6,] "d"   "b"   "c"   "c"   "d"   "d"   "d"   "c"   "d"
 [7,] "d"   "b"   "c"   "c"   "c"   "c"   "d"   "c"   "c"
 [8,] "d"   "c"   "d"   "d"   "c"   "d"   "e"   "c"   "d"
 [9,] "d"   "c"   "c"   "d"   "d"   "d"   "d"   "d"   "e"
[10,] "e"   "d"   "d"   "d"   "d"   "e"   "d"   "d"   "e"
[11,] "c"   "c"   "c"   "c"   "c"   "d"   "d"   "c"   "d"
[12,] "d"   "c"   "d"   "d"   "d"   "d"   "e"   "c"   "c"
[13,] "d"   "c"   "c"   "c"   "c"   "d"   "d"   "c"   "c"
[14,] "d"   "d"   "c"   "d"   "d"   "c"   "c"   "c"   "c"
[15,] "d"   "d"   "c"   "d"   "d"   "d"   "d"   "c"   "d"
[16,] "d"   "d"   "c"   "d"   "d"   "d"   "d"   "d"   "d"
[17,] "d"   "c"   "b"   "c"   "c"   "c"   "b"   "c"   "c"
[18,] "c"   "d"   "b"   "d"   "d"   "d"   "d"   "c"   "c"
[19,] "c"   "c"   "b"   "c"   "d"   "c"   "c"   "b"   "c"
[20,] "c"   "d"   "c"   "d"   "d"   "d"   "c"   "c"   "c"
[21,] "c"   "d"   "c"   "d"   "d"   "c"   "b"   "c"   "c"
[22,] "c"   "c"   "c"   "d"   "d"   "c"   "b"   "b"   "b"
[23,] "b"   "c"   "c"   "d"   "d"   "d"   "c"   "c"   "d"
[24,] "c"   "d"   "c"   "d"   "d"   "d"   "c"   "c"   "c"
[25,] "c"   "c"   "b"   "c"   "d"   "c"   "c"   "c"   "c"
```

```
[26,] "d"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
[27,] "b"   "a"   "a"   "b"   "a"   "b"   "b"   "b"   "b"
[28,] "b"   "b"   "b"   "c"   "c"   "b"   "b"   "b"   "b"
[29,] "b"   "b"   "b"   "c"   "c"   "b"   "b"   "b"   "c"
[30,] "a"   "b"   "b"   "c"   "b"   "b"   "b"   "b"   "b"
[31,] "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"
[32,] "d"   "d"   "d"   "e"   "d"   "e"   "d"   "e"   "e"
[33,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[34,] "b"   "a"   "b"   "b"   "b"   "c"   "a"   "b"   "c"
[35,] "a"   "b"   "b"   "a"   "a"   "a"   "a"   "a"   "b"
[36,] "c"   "c"   "e"   "d"   "d"   "d"   "d"   "e"   "e"
[37,] "c"   "c"   "e"   "c"   "c"   "b"   "b"   "d"   "c"
[38,] "c"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[39,] "e"   "c"   "d"   "d"   "d"   "d"   "c"   "e"   "e"
[40,] "b"   "b"   "c"   "b"   "c"   "e"   "d"   "d"   "d"
[41,] "d"   "e"   "e"   "e"   "e"   "d"   "d"   "e"   "e"
[42,] "b"   "b"   "c"   "b"   "b"   "c"   "b"   "b"   "c"
[43,] "c"   "c"   "d"   "c"   "c"   "b"   "b"   "e"   "e"
[44,] "d"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[45,] "e"   "d"   "e"   "e"   "e"   "e"   "e"   "e"   "d"
[46,] "e"   "e"   "e"   "d"   "d"   "d"   "e"   "e"   "d"
[47,] "e"   "e"   "c"   "c"   "e"   "d"   "e"   "d"   "e"
[48,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[49,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[50,] "d"   "d"   "c"   "d"   "e"   "d"   "e"   "e"   "e"
[51,] "e"   "e"   "d"   "d"   "e"   "e"   "e"   "e"   "d"
[52,] "d"   "e"   "e"   "d"   "d"   "d"   "e"   "e"   "c"
[53,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[54,] "e"   "e"   "e"   "e"   "e"   "d"   "e"   "e"   "c"
[55,] "e"   "e"   "e"   "d"   "d"   "d"   "d"   "d"   "b"
[56,] "e"   "c"   "b"   "d"   "d"   "c"   "d"   "c"   "b"
[57,] "c"   "c"   "d"   "b"   "b"   "b"   "b"   "b"   "b"
[58,] "b"   "c"   "b"   "b"   "b"   "b"   "b"   "b"   "b"
[59,] "c"   "c"   "c"   "b"   "b"   "b"   "c"   "b"   "c"
[60,] "c"   "c"   "c"   "b"   "b"   "a"   "b"   "c"   "b"
[61,] "e"   "e"   "c"   "c"   "c"   "c"   "d"   "e"   "e"
[62,] "e"   "e"   "d"   "c"   "c"   "b"   "c"   "c"   "d"
[63,] "e"   "e"   "e"   "d"   "e"   "c"   "c"   "b"   "c"
[64,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "d"   "c"
[65,] "b"   "c"   "c"   "b"   "b"   "c"   "b"   "c"   "d"
[66,] "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"   "b"
[67,] "b"   "c"   "c"   "c"   "c"   "c"   "c"   "c"   "b"
[68,] "b"   "c"   "c"   "c"   "c"   "c"   "b"   "c"   "b"
[69,] "b"   "c"   "d"   "c"   "c"   "d"   "c"   "c"   "c"
[70,] "b"   "c"   "c"   "c"   "b"   "c"   "b"   "b"   "b"
[71,] "d"   "e"   "c"   "d"   "d"   "d"   "e"   "d"   "e"
[72,] "c"   "d"   "c"   "d"   "d"   "d"   "e"   "d"   "e"
[73,] "c"   "e"   "d"   "d"   "c"   "c"   "c"   "c"   "d"
[74,] "a"   "a"   "a"   "a"   "b"   "b"   "b"   "b"   "b"
[75,] "a"   "a"   "a"   "b"   "b"   "b"   "b"   "a"   "b"
[76,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[77,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[78,] "b"   "c"   "a"   "a"   "a"   "b"   "a"   "a"   "a"
[79,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[80,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[81,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[82,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[83,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[84,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[85,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[86,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[87,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[88,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[89,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[90,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[91,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
>
> #estimate transition matrix with MLE
>  result2  <-markovchainFit(data=datamatrix2,  method="mle",
name="datamatrix2")
> result2$estimate
datamatrix2
 A  5 - dimensional discrete Markov Chain defined by the following
states:
 a, b, c, d, e
 The transition matrix  (by rows)  is defined as follows:
           a          b          c          d           e
a 0.91780822 0.08219178 0.00000000 0.00000000 0.000000000
b 0.06086957 0.59130435 0.31304348 0.02608696 0.008695652
c 0.01176471 0.18235294 0.50588235 0.25882353 0.041176471
d 0.00000000 0.03311258 0.28476821 0.46357616 0.218543046
e 0.00000000 0.00000000 0.08219178 0.21232877 0.705479452

> result2$standardError
            a          b          c          d           e
a 0.079286554 0.02372672 0.00000000 0.00000000 0.000000000
b 0.023006533 0.07170618 0.05217391 0.01506131 0.008695652
c 0.008318903 0.03275156 0.05455070 0.03901912 0.015563243
d 0.000000000 0.01480840 0.04342675 0.05540795 0.038043461
e 0.000000000 0.00000000 0.02372672 0.03813537 0.069512956
> transmatrix2<- result2$estimate[1:5,1:5]
> transmatrix2
           a          b          c          d           e
a 0.91780822 0.08219178 0.00000000 0.00000000 0.000000000
b 0.06086957 0.59130435 0.31304348 0.02608696 0.008695652
c 0.01176471 0.18235294 0.50588235 0.25882353 0.041176471
d 0.00000000 0.03311258 0.28476821 0.46357616 0.218543046
e 0.00000000 0.00000000 0.08219178 0.21232877 0.705479452
> upward2 <-result2$estimate[1:5,1:5]
> downward2 <-result2$estimate[1:5,1:5]
>  write.xlsx(transmatrix2,  "M:/My  Documents/MScThesis/Text
thesis/Data         files/Chapter        6/File          for
R/Matrix_Labour_adjusted_value_added_2007_2015.xlsx")
>
> #define transition matrix as matrix object
> showClass("markovchain")
Class "markovchain" [package "markovchain"]

Slots:

Name:              states              byrow transitionMatrix
name
Class:          character            logical            matrix
character
> simpleMc2<-new("markovchain", states=c("a","b","c","d","e"),
+ transitionMatrix=transmatrix2,
+ name="simpleMc2")
>
> #stationary distribution
> steadyStates(simpleMc2)
             a         b         c         d         e
[1,] 0.1603272 0.1660172 0.2611357 0.2130327 0.1994872
>
> #half life
> eigenvalues2 <-eigen(transmatrix2)
> eigenvaluevector2 <-eigenvalues2$values
> secondeigenvalue2 <-eigenvaluevector2[[2]]
> secondeigenvalue2
[1] 0.9178993
> halflife2 <- (-log(2))/log((abs(secondeigenvalue2)))
> halflife2
[1] 8.091121
>
> #mobilityindex
> trace2 <-matrix.trace(transmatrix2)
> mobility2 <- (5-trace2)*(5-1)^-1
> mobility2
[1] 0.4539874
>
> #upward mobility
> upward2[lower.tri(upward2)] <- 0
> upward2[lower.tri(upward2,diag=TRUE)] <- 0
> sum(upward2)*(5-trace2)^-1
[1] 0.5223498
>
> #downward mobility
> downward2[upper.tri(downward2)] <- 0
> downward2[upper.tri(downward2,diag=TRUE)] <- 0
> sum(downward2)*(5-trace2)^-1
[1] 0.4776502
>
>
> #load data from computer Family farm income 2004-2015
>          c          <-read.table("//WURNET.NL/Homes/jong285/My
Documents/MScThesis/Text  thesis/Data  files/Chapter  6/File  for
R/Family_farm_income_2004-2015.txt", header=T)
>
> #define as matrix
> datamatrix3 <-as.matrix(c)
> datamatrix3
      X2004 X2005 X2006 X2007 X2008 X2009 X2010 X2011 X2012 X2013
X2014 X2015
 [1,] "e"   "d"   "e"   "e"   "c"   "e"   "e"   "d"   "e"   "e"
"b"   "c"
 [2,] "d"   "d"   "d"   "e"   "c"   "e"   "e"   "d"   "e"   "e"
"c"   "b"
 [3,] "d"   "d"   "d"   "e"   "c"   "d"   "d"   "d"   "d"   "e"
"c"   "c"
 [4,] "b"   "b"   "b"   "c"   "b"   "c"   "c"   "c"   "c"   "d"
"b"   "c"
 [5,] "c"   "c"   "d"   "d"   "c"   "c"   "d"   "b"   "c"   "d"
"b"   "c"
```

```
 [6,] "c"  "b"  "c"  "c"  "b"  "b"  "c"  "c"  "c"  "d"
"c"  "c"
 [7,] "b"  "b"  "c"  "c"  "b"  "c"  "c"  "c"  "c"  "c"
"b"  "c"
 [8,] "c"  "c"  "c"  "d"  "b"  "c"  "d"  "c"  "c"  "d"
"b"  "c"
 [9,] "d"  "e"  "e"  "e"  "b"  "b"  "e"  "e"  "e"  "e"
"e"  "e"
[10,] "e"  "e"  "e"  "e"  "d"  "e"  "e"  "d"  "e"  "e"
"c"  "a"
[11,] "b"  "b"  "c"  "d"  "a"  "c"  "c"  "c"  "d"  "c"
"b"  "b"
[12,] "e"  "d"  "d"  "e"  "c"  "e"  "d"  "b"  "c"  "b"
"c"  "a"
[13,] "c"  "b"  "b"  "c"  "a"  "b"  "b"  "b"  "b"  "b"
"b"  "b"
[14,] "d"  "c"  "b"  "c"  "c"  "a"  "c"  "c"  "b"  "b"
"b"  "b"
[15,] "b"  "b"  "b"  "c"  "d"  "a"  "c"  "c"  "b"  "b"
"b"  "b"
[16,] "c"  "c"  "c"  "c"  "c"  "a"  "c"  "c"  "c"  "c"
"b"  "b"
[17,] "c"  "d"  "c"  "d"  "c"  "a"  "c"  "c"  "c"  "b"
"c"  "c"
[18,] "c"  "b"  "b"  "b"  "c"  "a"  "c"  "c"  "c"  "b"
"b"  "b"
[19,] "b"  "c"  "b"  "c"  "c"  "a"  "b"  "c"  "c"  "b"
"a"  "c"
[20,] "c"  "b"  "b"  "b"  "c"  "a"  "c"  "c"  "c"  "b"
"b"  "b"
[21,] "c"  "c"  "c"  "c"  "c"  "b"  "c"  "c"  "c"  "b"
"b"  "c"
[22,] "b"  "b"  "c"  "b"  "c"  "b"  "c"  "c"  "c"  "a"
"b"  "a"
[23,] "c"  "b"  "b"  "b"  "b"  "c"  "c"  "c"  "c"  "b"
"b"  "d"
[24,] "b"  "c"  "b"  "b"  "c"  "c"  "c"  "c"  "c"  "b"
"b"  "c"
[25,] "c"  "b"  "b"  "b"  "c"  "a"  "c"  "c"  "c"  "c"
"b"  "c"
[26,] "c"  "c"  "c"  "c"  "b"  "a"  "b"  "b"  "b"  "b"
"b"  "b"
[27,] "a"  "b"  "a"  "b"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[28,] "b"  "b"  "b"  "b"  "b"  "a"  "b"  "b"  "b"  "a"
"b"  "b"
[29,] "b"  "b"  "b"  "a"  "b"  "b"  "b"  "b"  "b"  "b"
"b"  "b"
[30,] "a"  "b"  "a"  "a"  "b"  "a"  "b"  "b"  "a"  "b"
"b"  "b"
[31,] "b"  "b"  "b"  "b"  "c"  "c"  "c"  "c"  "c"  "b"
"b"  "b"
[32,] "e"  "d"  "e"  "d"  "d"  "e"  "e"  "e"  "e"  "d"
"e"  "e"
[33,] "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"
"e"  "e"
[34,] "b"  "b"  "c"  "b"  "a"  "b"  "c"  "b"  "c"  "a"
"b"  "d"
[35,] "a"  "a"  "b"  "a"  "b"  "b"  "a"  "a"  "a"  "a"
"a"  "c"
[36,] "c"  "c"  "d"  "c"  "c"  "e"  "d"  "e"  "e"  "d"
"e"  "e"
[37,] "c"  "c"  "c"  "c"  "c"  "e"  "c"  "c"  "b"  "b"
"c"  "d"
[38,] "d"  "c"  "c"  "c"  "e"  "e"  "e"  "e"  "e"  "e"
"e"  "e"
[39,] "e"  "e"  "e"  "e"  "d"  "e"  "e"  "e"  "e"  "d"
"e"  "e"
[40,] "c"  "c"  "c"  "b"  "c"  "d"  "b"  "c"  "e"  "d"
"d"  "e"
[41,] "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"
"e"  "e"
[42,] "c"  "c"  "c"  "b"  "b"  "c"  "b"  "c"  "c"  "b"
"b"  "d"
[43,] "e"  "e"  "e"  "c"  "c"  "e"  "c"  "c"  "c"  "c"
"e"  "e"
[44,] "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"
"e"  "e"
[45,] "d"  "d"  "d"  "e"  "d"  "d"  "d"  "d"  "d"  "e"
"d"  "d"
[46,] "d"  "d"  "d"  "d"  "d"  "d"  "d"  "d"  "c"  "d"
"e"  "d"
[47,] "d"  "d"  "d"  "d"  "e"  "c"  "c"  "d"  "c"  "d"
"e"  "e"
[48,] "d"  "d"  "d"  "d"  "d"  "a"  "d"  "e"  "d"  "e"
"d"  "d"
[49,] "b"  "c"  "d"  "d"  "a"  "a"  "a"  "b"  "a"  "e"
"e"  "a"
[50,] "d"  "d"  "c"  "d"  "d"  "d"  "d"  "e"  "d"  "d"
"e"  "e"
[51,] "e"  "e"  "c"  "e"  "e"  "e"  "e"  "e"  "e"  "e"
"e"  "e"
[52,] "c"  "e"  "d"  "d"  "e"  "e"  "e"  "e"  "d"  "e"
"e"  "b"
[53,] "e"  "e"  "d"  "e"  "e"  "e"  "e"  "e"  "e"  "e"
"e"  "e"
[54,] "e"  "e"  "d"  "d"  "e"  "e"  "e"  "e"  "d"  "e"
"e"  "c"
[55,] "d"  "d"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "d"
"d"  "a"
[56,] "c"  "c"  "c"  "e"  "c"  "c"  "d"  "d"  "b"  "d"
"c"  "b"
[57,] "b"  "c"  "c"  "c"  "c"  "d"  "b"  "b"  "b"  "b"
"b"  "c"
[58,] "c"  "c"  "b"  "b"  "c"  "c"  "b"  "a"  "b"  "b"
"b"  "b"
[59,] "c"  "c"  "c"  "c"  "c"  "c"  "b"  "b"  "b"  "c"
"b"  "c"
[60,] "b"  "c"  "c"  "c"  "d"  "c"  "b"  "b"  "a"  "c"
"c"  "c"
[61,] "e"  "e"  "e"  "e"  "e"  "e"  "e"  "d"  "c"  "d"
"e"  "e"
[62,] "e"  "e"  "e"  "d"  "d"  "d"  "c"  "b"  "b"  "c"
"c"  "e"
[63,] "e"  "e"  "e"  "e"  "e"  "e"  "c"  "e"  "c"  "c"
"c"  "d"
[64,] "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"
"e"  "d"
[65,] "b"  "b"  "b"  "b"  "c"  "d"  "c"  "c"  "c"  "c"
"c"  "d"
[66,] "b"  "b"  "b"  "b"  "b"  "b"  "b"  "b"  "b"  "b"
"a"  "a"
[67,] "b"  "b"  "a"  "b"  "c"  "c"  "c"  "b"  "c"  "b"
"b"  "b"
[68,] "b"  "b"  "b"  "b"  "c"  "c"  "c"  "b"  "c"  "b"
"b"  "b"
[69,] "c"  "b"  "b"  "b"  "c"  "d"  "c"  "c"  "d"  "c"
"c"  "c"
[70,] "c"  "b"  "b"  "b"  "c"  "c"  "c"  "b"  "c"  "b"
"b"  "b"
[71,] "a"  "b"  "a"  "b"  "d"  "a"  "c"  "c"  "a"  "b"
"b"  "e"
[72,] "a"  "b"  "a"  "b"  "c"  "a"  "c"  "b"  "b"  "c"
"c"  "e"
[73,] "a"  "b"  "b"  "b"  "e"  "d"  "c"  "b"  "a"  "b"
"b"  "c"
[74,] "a"  "a"  "b"  "b"  "b"  "b"  "b"  "b"  "b"  "b"
"b"  "c"
[75,] "a"  "a"  "a"  "a"  "b"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[76,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[77,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[78,] "b"  "b"  "b"  "b"  "d"  "a"  "b"  "b"  "b"  "b"
"b"  "a"
[79,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[80,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[81,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[82,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[83,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "d"
[84,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
>
> #estimate transition matrix with MLE
>   result3   <-markovchainFit(data=datamatrix3,   method="mle",
name="datamatrix3")
> result3$estimate
datamatrix3
 A  5 - dimensional discrete Markov Chain defined by the following
states:
 a, b, c, d, e
 The transition matrix  (by rows)  is defined as follows:
            a           b          c          d           e
a 0.710691824 0.18867925 0.08176101 0.01257862 0.006289308
b 0.103896104 0.57575758 0.28138528 0.02597403 0.012987013
c 0.059829060 0.23931624 0.50427350 0.13247863 0.064102564
```

```
d 0.060344828 0.07758621 0.17241379 0.39655172 0.293103448
e 0.005434783 0.01630435 0.08695652 0.16847826 0.722826087

> result3$standardError
            a          b          c           d           e
a 0.066856263 0.03444796 0.02267642 0.008894425 0.006289308
b 0.021207703 0.04992451 0.03490155 0.010603852 0.007498055
c 0.015989989 0.03197998 0.04642214 0.023793865 0.016551211
d 0.022808201 0.02586207 0.03855290 0.058468362 0.050266827
e 0.005434783 0.00941332 0.02173913 0.030259589 0.062676971
> transmatrix3<- result3$estimate[1:5,1:5]
> transmatrix3
            a          b          c           d           e
a 0.710691824 0.18867925 0.08176101 0.01257862 0.006289308
b 0.103896104 0.57575758 0.28138528 0.02597403 0.012987013
c 0.059829060 0.23931624 0.50427350 0.13247863 0.064102564
d 0.060344828 0.07758621 0.17241379 0.39655172 0.293103448
e 0.005434783 0.01630435 0.08695652 0.16847826 0.722826087
> upward3 <-result3$estimate[1:5,1:5]
> downward3 <-result3$estimate[1:5,1:5]
>  write.xlsx(transmatrix3,  "M:/My  Documents/MScThesis/Text
thesis/Data         files/Chapter        6/File         for
R/Matrix_Family_farm_income_2004_2015.xlsx")
>
> #define transition matrix as matrix object
> showClass("markovchain")
Class "markovchain" [package "markovchain"]

Slots:

Name:               states               byrow transitionMatrix
name
Class:          character              logical             matrix
character
> simpleMc3<-new("markovchain", states=c("a","b","c","d","e"),
+ transitionMatrix=transmatrix3,
+ name="simpleMc3")
>
> #stationary distribution
> steadyStates(simpleMc3)
             a         b         c         d         e
[1,] 0.1705376 0.2472825 0.2487887 0.1265618 0.2068294
>
> #half life
> eigenvalues3 <-eigen(transmatrix3)
> eigenvaluevector3 <-eigenvalues3$values
> secondeigenvalue3 <-eigenvaluevector3[[2]]
> secondeigenvalue3
[1] 0.7891327
> halflife3 <- (-log(2))/log((abs(secondeigenvalue3)))
> halflife3
[1] 2.926885
>
> #mobilityindex
> trace3 <-matrix.trace(transmatrix3)
> mobility3 <- (5-trace3)*(5-1)^-1
> mobility3
[1] 0.5224748
>
> #upward mobility
> upward3[lower.tri(upward3)] <- 0
> upward3[lower.tri(upward3,diag=TRUE)] <- 0
> sum(upward3)*(5-trace3)^-1
[1] 0.5260249
>
> #downward mobility
> downward3[upper.tri(downward3)] <- 0
> downward3[upper.tri(downward3,diag=TRUE)] <- 0
> sum(downward3)*(5-trace3)^-1
[1] 0.4739751
>
> #load data from computer Family farm income 2007-2015
>         d         <-read.table("//WURNET.NL/Homes/jong285/My
Documents/MScThesis/Text  thesis/Data  files/Chapter  6/File  for
R/Family_farm_income_2007-2015.txt", header=T)
>
> #define as matrix
> datamatrix4 <-as.matrix(d)
> datamatrix4
      X2007 X2008 X2009 X2010 X2011 X2012 X2013 X2014 X2015
 [1,] "e"   "d"   "e"   "e"   "e"   "e"   "e"   "b"   "c"
 [2,] "e"   "c"   "e"   "e"   "e"   "e"   "e"   "c"   "b"
 [3,] "e"   "c"   "e"   "e"   "d"   "d"   "e"   "c"   "c"
 [4,] "c"   "b"   "c"   "c"   "c"   "c"   "d"   "b"   "c"
 [5,] "d"   "c"   "c"   "d"   "b"   "c"   "e"   "b"   "c"
 [6,] "d"   "b"   "b"   "c"   "c"   "c"   "d"   "c"   "c"
 [7,] "c"   "b"   "c"   "c"   "c"   "c"   "c"   "c"   "c"
 [8,] "d"   "b"   "c"   "d"   "c"   "c"   "d"   "b"   "c"
 [9,] "e"   "b"   "b"   "e"   "e"   "e"   "e"   "e"   "e"
[10,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "c"   "a"
[11,] "d"   "a"   "c"   "c"   "c"   "d"   "d"   "c"   "b"
[12,] "e"   "c"   "e"   "d"   "c"   "c"   "c"   "c"   "a"
[13,] "c"   "a"   "b"   "b"   "b"   "b"   "b"   "b"   "b"
[14,] "c"   "c"   "a"   "d"   "c"   "b"   "b"   "c"   "b"
[15,] "c"   "d"   "a"   "c"   "c"   "b"   "b"   "b"   "c"
[16,] "c"   "c"   "a"   "c"   "c"   "c"   "c"   "b"   "b"
[17,] "d"   "c"   "a"   "c"   "c"   "c"   "b"   "c"   "c"
[18,] "c"   "c"   "a"   "c"   "c"   "c"   "b"   "b"   "b"
[19,] "c"   "c"   "a"   "c"   "c"   "c"   "b"   "a"   "c"
[20,] "c"   "d"   "b"   "d"   "c"   "c"   "b"   "c"   "b"
[21,] "c"   "c"   "b"   "c"   "c"   "c"   "b"   "b"   "c"
[22,] "b"   "c"   "b"   "c"   "c"   "c"   "a"   "b"   "b"
[23,] "b"   "b"   "c"   "d"   "c"   "c"   "b"   "c"   "d"
[24,] "c"   "d"   "c"   "c"   "c"   "c"   "b"   "b"   "c"
[25,] "c"   "c"   "b"   "c"   "c"   "c"   "c"   "c"   "c"
[26,] "c"   "c"   "b"   "b"   "b"   "b"   "b"   "c"   "c"
[27,] "b"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[28,] "b"   "b"   "a"   "b"   "b"   "b"   "a"   "b"   "b"
[29,] "b"   "b"   "b"   "c"   "c"   "b"   "b"   "b"   "b"
[30,] "a"   "b"   "a"   "c"   "b"   "b"   "b"   "b"   "b"
[31,] "b"   "c"   "c"   "c"   "c"   "c"   "b"   "b"   "b"
[32,] "d"   "e"   "e"   "e"   "e"   "e"   "d"   "e"   "e"
[33,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[34,] "b"   "a"   "c"   "c"   "c"   "c"   "a"   "b"   "d"
[35,] "a"   "b"   "b"   "a"   "a"   "a"   "a"   "a"   "c"
[36,] "c"   "c"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[37,] "c"   "c"   "e"   "d"   "c"   "b"   "b"   "d"   "d"
[38,] "c"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[39,] "e"   "d"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[40,] "b"   "c"   "d"   "b"   "c"   "e"   "d"   "e"   "e"
[41,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[42,] "b"   "b"   "d"   "c"   "c"   "c"   "b"   "c"   "d"
[43,] "c"   "d"   "e"   "c"   "c"   "c"   "c"   "e"   "e"
[44,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[45,] "e"   "d"   "d"   "d"   "d"   "d"   "e"   "e"   "d"
[46,] "d"   "d"   "d"   "d"   "d"   "c"   "d"   "e"   "d"
[47,] "d"   "e"   "c"   "c"   "d"   "c"   "d"   "e"   "e"
[48,] "e"   "d"   "a"   "e"   "e"   "e"   "e"   "d"   "d"
[49,] "d"   "a"   "a"   "a"   "b"   "a"   "e"   "e"   "a"
[50,] "d"   "d"   "e"   "d"   "e"   "d"   "e"   "e"   "e"
[51,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[52,] "e"   "e"   "e"   "e"   "e"   "d"   "e"   "e"   "b"
[53,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[54,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "c"
[55,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "d"   "a"
[56,] "e"   "c"   "c"   "d"   "d"   "b"   "e"   "c"   "b"
[57,] "c"   "d"   "d"   "b"   "c"   "c"   "c"   "b"   "c"
[58,] "b"   "c"   "c"   "b"   "b"   "b"   "b"   "b"   "b"
[59,] "c"   "c"   "d"   "b"   "b"   "b"   "c"   "b"   "c"
[60,] "c"   "d"   "c"   "b"   "c"   "a"   "c"   "c"   "c"
[61,] "e"   "e"   "e"   "e"   "d"   "d"   "d"   "e"   "e"
[62,] "e"   "e"   "d"   "c"   "c"   "b"   "c"   "c"   "e"
[63,] "e"   "e"   "e"   "d"   "e"   "d"   "d"   "c"   "d"
[64,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "d"
[65,] "b"   "c"   "d"   "c"   "c"   "d"   "c"   "c"   "e"
[66,] "b"   "c"   "b"   "b"   "b"   "b"   "b"   "a"   "a"
[67,] "b"   "c"   "c"   "c"   "c"   "d"   "b"   "c"   "b"
[68,] "b"   "c"   "d"   "c"   "c"   "c"   "b"   "b"   "b"
[69,] "b"   "c"   "d"   "c"   "c"   "d"   "c"   "c"   "c"
[70,] "b"   "c"   "d"   "c"   "b"   "c"   "b"   "b"   "b"
[71,] "b"   "d"   "a"   "c"   "c"   "a"   "c"   "b"   "e"
[72,] "b"   "c"   "a"   "c"   "b"   "b"   "c"   "c"   "e"
[73,] "b"   "e"   "d"   "c"   "b"   "b"   "b"   "b"   "c"
[74,] "b"   "b"   "b"   "b"   "c"   "b"   "b"   "b"   "c"
[75,] "a"   "b"   "a"   "a"   "b"   "a"   "a"   "a"   "a"
[76,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[77,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[78,] "b"   "d"   "a"   "b"   "b"   "b"   "b"   "b"   "a"
[79,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[80,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[81,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[82,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[83,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "d"
[84,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[85,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "b"
[86,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[87,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[88,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "b"   "a"
[89,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[90,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[91,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
>
```

```
> #estimate transition matrix with MLE
>   result4   <-markovchainFit(data=datamatrix4,   method="mle",
name="datamatrix4")
> result4$estimate
datamatrix4
 A  5 - dimensional discrete Markov Chain defined by the following
states:
 a, b, c, d, e
 The transition matrix  (by rows)  is defined as follows:
            a          b          c          d          e
a 0.812121212 0.07878788 0.08484848 0.01212121 0.01212121
b 0.094202899 0.49275362 0.34057971 0.04347826 0.02898551
c 0.070652174 0.21739130 0.48913043 0.15760870 0.06521739
d 0.090909091 0.14285714 0.31168831 0.23376623 0.22077922
e 0.006097561 0.02439024 0.06707317 0.13414634 0.76829268

> result4$standardError
            a          b          c           d           e
a 0.070156587 0.02185183 0.02267671 0.008570991 0.008570991
b 0.026127183 0.05975515 0.04967866 0.017749926 0.014492754
c 0.019595387 0.03437258 0.05155887 0.029267200 0.018826639
d 0.034360407 0.04307305 0.06362311 0.055099230 0.053546826
e 0.006097561 0.01219512 0.02022332 0.028600096 0.068444952
> transmatrix4<- result4$estimate[1:5,1:5]
> transmatrix4
            a          b          c          d          e
a 0.812121212 0.07878788 0.08484848 0.01212121 0.01212121
b 0.094202899 0.49275362 0.34057971 0.04347826 0.02898551
c 0.070652174 0.21739130 0.48913043 0.15760870 0.06521739
d 0.090909091 0.14285714 0.31168831 0.23376623 0.22077922
e 0.006097561 0.02439024 0.06707317 0.13414634 0.76829268
> upward4 <-result4$estimate[1:5,1:5]
> downward4 <-result4$estimate[1:5,1:5]
>   write.xlsx(transmatrix4,   "M:/My   Documents/MScThesis/Text
thesis/Data          files/Chapter          6/File          for
R/Matrix_Family_farm_income_2007_2015.xlsx")
>
> #define transition matrix as matrix object
> showClass("markovchain")
Class "markovchain" [package "markovchain"]

Slots:

Name:                states               byrow transitionMatrix
name
Class:          character             logical              matrix
character
> simpleMc4<-new("markovchain", states=c("a","b","c","d","e"),
+ transitionMatrix=transmatrix4,
+ name="simpleMc4")
>
> #stationary distribution
> steadyStates(simpleMc4)
             a         b         c         d        e
[1,] 0.2467755 0.1870741 0.2559297 0.1033748 0.206846
>
> #half life
> eigenvalues4 <-eigen(transmatrix4)
> eigenvaluevector4 <-eigenvalues4$values
> secondeigenvalue4 <-eigenvaluevector4[[2]]
> secondeigenvalue4
[1] 0.7980802
> halflife4 <- (-log(2))/log((abs(secondeigenvalue4)))
> halflife4
[1] 3.073194
>
> #mobilityindex
> trace4 <-matrix.trace(transmatrix4)
> mobility4 <- (5-trace4)*(5-1)^-1
> mobility4
[1] 0.550984
>
> #upward mobility
> upward4[lower.tri(upward4)] <- 0
> upward4[lower.tri(upward4,diag=TRUE)] <- 0
> sum(upward4)*(5-trace4)^-1
[1] 0.4739374
>
> #downward mobility
> downward4[upper.tri(downward4)] <- 0
> downward4[upper.tri(downward4,diag=TRUE)] <- 0
> sum(downward4)*(5-trace4)^-1
[1] 0.5260626
>
> #load data from computer Labour income 2004-2015
```

```
>          e          <-read.table("//WURNET.NL/Homes/jong285/My
Documents/MScThesis/Text  thesis/Data  files/Chapter  6/File  for
R/Labour_income_2004-2015.txt", header=T)
>
> #define as matrix
> datamatrix5 <-as.matrix(e)
> datamatrix5
      X2004 X2005 X2006 X2007 X2008 X2009 X2010 X2011 X2012 X2013
X2014 X2015
 [1,] "e"   "d"   "e"   "e"   "c"   "e"   "e"   "d"   "d"   "e"
"b"   "c"
 [2,] "d"   "d"   "d"   "e"   "c"   "e"   "e"   "e"   "e"   "e"
"c"   "b"
 [3,] "d"   "d"   "d"   "e"   "c"   "e"   "e"   "d"   "d"   "e"
"c"   "c"
 [4,] "c"   "b"   "c"   "c"   "b"   "c"   "c"   "c"   "c"   "e"
"b"   "c"
 [5,] "d"   "c"   "d"   "e"   "c"   "d"   "d"   "c"   "d"   "e"
"c"   "c"
 [6,] "c"   "c"   "c"   "d"   "b"   "c"   "c"   "c"   "d"   "d"
"c"   "d"
 [7,] "c"   "c"   "c"   "d"   "b"   "c"   "c"   "c"   "d"   "d"
"c"   "d"
 [8,] "c"   "c"   "c"   "d"   "b"   "c"   "d"   "c"   "c"   "e"
"b"   "c"
 [9,] "a"   "b"   "a"   "b"   "a"   "a"   "a"   "a"   "a"   "a"
"a"   "a"
[10,] "c"   "b"   "c"   "c"   "a"   "a"   "a"   "a"   "a"   "b"
"a"   "a"
[11,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
"a"   "a"
[12,] "b"   "b"   "b"   "b"   "a"   "a"   "a"   "a"   "a"   "a"
"a"   "a"
[13,] "a"   "a"   "a"   "b"   "a"   "a"   "a"   "a"   "a"   "a"
"a"   "a"
[14,] "d"   "c"   "c"   "c"   "c"   "a"   "d"   "c"   "b"   "b"
"c"   "b"
[15,] "c"   "c"   "b"   "d"   "d"   "a"   "c"   "d"   "c"   "b"
"b"   "c"
[16,] "d"   "c"   "c"   "c"   "c"   "a"   "c"   "c"   "d"   "d"
"c"   "c"
[17,] "d"   "d"   "c"   "d"   "c"   "b"   "c"   "c"   "c"   "b"
"c"   "c"
[18,] "c"   "c"   "c"   "c"   "c"   "a"   "c"   "c"   "c"   "c"
"b"   "c"
[19,] "b"   "c"   "c"   "c"   "c"   "b"   "c"   "c"   "c"   "b"
"a"   "c"
[20,] "c"   "b"   "b"   "c"   "d"   "b"   "d"   "d"   "c"   "b"
"c"   "c"
[21,] "c"   "c"   "c"   "c"   "d"   "b"   "d"   "d"   "c"   "b"
"c"   "c"
[22,] "b"   "b"   "c"   "b"   "c"   "b"   "d"   "c"   "c"   "a"
"b"   "a"
[23,] "c"   "c"   "c"   "b"   "b"   "c"   "d"   "d"   "d"   "c"
"c"   "e"
[24,] "c"   "c"   "c"   "c"   "d"   "c"   "d"   "d"   "d"   "c"
"c"   "c"
[25,] "c"   "c"   "c"   "c"   "c"   "b"   "c"   "d"   "c"   "c"
"c"   "c"
[26,] "c"   "c"   "c"   "c"   "c"   "b"   "b"   "b"   "b"   "b"
"c"   "c"
[27,] "b"   "b"   "b"   "b"   "a"   "a"   "b"   "a"   "a"   "a"
"a"   "b"
[28,] "b"   "b"   "b"   "b"   "b"   "a"   "b"   "b"   "b"   "b"
"b"   "b"
[29,] "c"   "b"   "b"   "b"   "b"   "b"   "c"   "c"   "c"   "b"
"b"   "c"
[30,] "b"   "b"   "a"   "a"   "b"   "b"   "c"   "c"   "b"   "b"
"b"   "b"
[31,] "b"   "b"   "c"   "b"   "c"   "c"   "c"   "c"   "c"   "b"
"b"   "b"
[32,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
"e"   "e"
[33,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
"e"   "e"
[34,] "b"   "b"   "c"   "b"   "a"   "c"   "c"   "c"   "d"   "a"
"c"   "e"
[35,] "b"   "b"   "b"   "a"   "b"   "c"   "b"   "a"   "b"   "a"
"b"   "c"
[36,] "d"   "c"   "d"   "c"   "d"   "e"   "e"   "e"   "e"   "e"
"e"   "e"
[37,] "c"   "d"   "c"   "c"   "d"   "e"   "d"   "c"   "c"   "c"
"e"   "e"
[38,] "d"   "c"   "c"   "c"   "e"   "e"   "e"   "e"   "e"   "e"
"e"   "e"
```

```
[39,] "e"  "e"  "e"  "e"  "c"  "e"  "d"  "d"  "d"  "d"
"e"  "e"
[40,] "c"  "d"  "c"  "c"  "c"  "d"  "b"  "d"  "e"  "e"
"e"  "e"
[41,] "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "d"
"e"  "e"
[42,] "c"  "c"  "c"  "b"  "b"  "d"  "c"  "c"  "c"  "c"
"c"  "e"
[43,] "e"  "e"  "e"  "d"  "c"  "e"  "c"  "c"  "c"  "c"
"e"  "e"
[44,] "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"
"e"  "e"
[45,] "e"  "e"  "e"  "e"  "d"  "e"  "e"  "e"  "e"  "e"
"e"  "e"
[46,] "e"  "d"  "e"  "e"  "e"  "e"  "e"  "e"  "d"  "e"
"e"  "e"
[47,] "d"  "d"  "d"  "e"  "e"  "d"  "c"  "e"  "d"  "e"
"e"  "e"
[48,] "d"  "e"  "e"  "e"  "d"  "a"  "e"  "e"  "e"  "e"
"e"  "d"
[49,] "a"  "b"  "c"  "c"  "a"  "a"  "a"  "a"  "a"  "c"
"d"  "a"
[50,] "e"  "d"  "d"  "d"  "d"  "e"  "d"  "e"  "e"  "e"
"e"  "e"
[51,] "d"  "d"  "c"  "d"  "d"  "d"  "d"  "e"  "d"  "e"
"e"  "c"
[52,] "b"  "c"  "c"  "c"  "d"  "c"  "c"  "c"  "b"  "c"
"e"  "a"
[53,] "e"  "d"  "c"  "d"  "d"  "e"  "d"  "e"  "d"  "e"
"e"  "d"
[54,] "e"  "d"  "d"  "d"  "e"  "e"  "e"  "e"  "d"  "e"
"e"  "c"
[55,] "c"  "c"  "d"  "e"  "e"  "e"  "e"  "d"  "d"  "d"
"c"  "a"
[56,] "d"  "d"  "d"  "e"  "c"  "d"  "d"  "d"  "c"  "e"
"c"  "b"
[57,] "c"  "d"  "d"  "d"  "d"  "e"  "c"  "c"  "c"  "c"
"c"  "d"
[58,] "c"  "c"  "c"  "c"  "c"  "c"  "b"  "b"  "b"  "b"
"c"  "b"
[59,] "c"  "c"  "c"  "c"  "d"  "d"  "b"  "b"  "c"  "c"
"c"  "d"
[60,] "c"  "c"  "d"  "d"  "d"  "d"  "b"  "c"  "a"  "c"
"c"  "c"
[61,] "e"  "e"  "e"  "e"  "e"  "e"  "e"  "d"  "c"  "d"
"e"  "e"
[62,] "e"  "e"  "e"  "e"  "e"  "d"  "c"  "b"  "b"  "c"
"c"  "e"
[63,] "e"  "e"  "e"  "e"  "e"  "e"  "d"  "e"  "c"  "c"
"b"  "d"
[64,] "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"  "e"
"e"  "d"
[65,] "c"  "b"  "b"  "b"  "c"  "e"  "c"  "c"  "d"  "c"
"c"  "e"
[66,] "b"  "b"  "b"  "b"  "c"  "c"  "b"  "c"  "c"  "b"
"b"  "b"
[67,] "c"  "b"  "b"  "b"  "c"  "d"  "c"  "c"  "d"  "c"
"c"  "b"
[68,] "c"  "b"  "b"  "b"  "c"  "d"  "c"  "c"  "d"  "c"
"c"  "b"
[69,] "c"  "c"  "b"  "b"  "d"  "e"  "d"  "c"  "d"  "c"
"c"  "c"
[70,] "c"  "c"  "b"  "b"  "c"  "d"  "c"  "b"  "c"  "c"
"c"  "b"
[71,] "a"  "b"  "a"  "b"  "c"  "a"  "c"  "c"  "a"  "b"
"b"  "d"
[72,] "a"  "b"  "a"  "b"  "c"  "a"  "c"  "b"  "a"  "c"
"c"  "d"
[73,] "a"  "b"  "b"  "b"  "e"  "d"  "c"  "b"  "a"  "b"
"b"  "c"
[74,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[75,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[76,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[77,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[78,] "c"  "b"  "c"  "c"  "d"  "a"  "b"  "b"  "b"  "b"
"b"  "a"
[79,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "b"  "b"  "a"
"a"  "a"
[80,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[81,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[82,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[83,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
[84,] "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"  "a"
"a"  "a"
>
> #estimate transition matrix with MLE
>  result5  <-markovchainFit(data=datamatrix5,  method="mle",
name="datamatrix5")
> result5$estimate
datamatrix5
 A 5 - dimensional discrete Markov Chain defined by the following
states:
 a, b, c, d, e
 The transition matrix  (by rows)  is defined as follows:
            a          b          c           d           e
a 0.815789474 0.11578947 0.05789474 0.005263158 0.005263158
b 0.140939597 0.45637584 0.33557047 0.060402685 0.006711409
c 0.041984733 0.18702290 0.51526718 0.183206107 0.072519084
d 0.036764706 0.05882353 0.32352941 0.338235294 0.242647059
e 0.005347594 0.01604278 0.08556150 0.160427807 0.732620321

> result5$standardError
            a           b          c           d           e
a 0.065525787 0.024686399 0.01745592 0.005263158 0.005263158
b 0.030755542 0.055343700 0.04745683 0.020134228 0.006711409
c 0.012658873 0.026717557 0.04434714 0.026443524 0.016637019
d 0.016441676 0.020797258 0.04877389 0.049870073 0.042239431
e 0.005347594 0.009262304 0.02139037 0.029289976 0.062591978
> transmatrix5<- result5$estimate[1:5,1:5]
> transmatrix5
            a          b          c           d           e
a 0.815789474 0.11578947 0.05789474 0.005263158 0.005263158
b 0.140939597 0.45637584 0.33557047 0.060402685 0.006711409
c 0.041984733 0.18702290 0.51526718 0.183206107 0.072519084
d 0.036764706 0.05882353 0.32352941 0.338235294 0.242647059
e 0.005347594 0.01604278 0.08556150 0.160427807 0.732620321
> upward5 <-result5$estimate[1:5,1:5]
> downward5 <-result5$estimate[1:5,1:5]
>   write.xlsx(transmatrix5,   "M:/My   Documents/MScThesis/Text
thesis/Data          files/Chapter          6/File          for
R/Matrix_Labour_income_2004_2015.xlsx.xlsx")
>
> #define transition matrix as matrix object
> showClass("markovchain")
Class "markovchain" [package "markovchain"]

Slots:

Name:               states               byrow transitionMatrix
name
Class:          character             logical            matrix
character
> simpleMc5<-new("markovchain", states=c("a","b","c","d","e"),
+ transitionMatrix=transmatrix5,
+ name="simpleMc5")
>
> #stationary distribution
> steadyStates(simpleMc5)
             a         b         c        d         e
[1,] 0.2186134 0.1606475 0.2690087 0.141774 0.2099564
>
> #half life
> eigenvalues5 <-eigen(transmatrix5)
> eigenvaluevector5 <-eigenvalues5$values
> secondeigenvalue5 <-eigenvaluevector5[[2]]
> secondeigenvalue5
[1] 0.8210813
> halflife5 <- (-log(2))/log((abs(secondeigenvalue5)))
> halflife5
[1] 3.516137
>
> #mobilityindex
> trace5 <-matrix.trace(transmatrix5)
> mobility5 <- (5-trace5)*(5-1)^-1
> mobility5
[1] 0.535428
>
> #upward mobility
> upward5[lower.tri(upward5)] <- 0
> upward5[lower.tri(upward5,diag=TRUE)] <- 0
> sum(upward5)*(5-trace5)^-1
[1] 0.5067289
>
> #downward mobility
```

```
> downward5[upper.tri(downward5)] <- 0
> downward5[upper.tri(downward5,diag=TRUE)] <- 0
> sum(downward5)*(5-trace5)^-1
[1] 0.4932711
>
> #load data from computer Labour income 2007-2015
>          f          <-read.table("//WURNET.NL/Homes/jong285/My
Documents/MScThesis/Text  thesis/Data  files/Chapter  6/File  for
R/Labour_income_2007-2015.txt", header=T)
>
> #define as matrix
> datamatrix6 <-as.matrix(f)
> datamatrix6
      X2007 X2008 X2009 X2010 X2011 X2012 X2013 X2014 X2015
 [1,] "e"   "c"   "e"   "e"   "e"   "e"   "e"   "b"   "c"
 [2,] "e"   "c"   "e"   "e"   "e"   "e"   "e"   "c"   "b"
 [3,] "e"   "d"   "e"   "e"   "d"   "d"   "e"   "c"   "c"
 [4,] "d"   "b"   "c"   "d"   "c"   "c"   "e"   "c"   "c"
 [5,] "e"   "c"   "d"   "e"   "c"   "d"   "e"   "c"   "c"
 [6,] "d"   "b"   "c"   "d"   "d"   "d"   "e"   "c"   "d"
 [7,] "d"   "b"   "d"   "d"   "c"   "d"   "d"   "c"   "d"
 [8,] "d"   "b"   "d"   "d"   "c"   "d"   "e"   "b"   "c"
 [9,] "b"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[10,] "c"   "a"   "a"   "a"   "a"   "a"   "b"   "a"   "a"
[11,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[12,] "b"   "a"   "a"   "a"   "a"   "a"   "b"   "a"   "a"
[13,] "b"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[14,] "c"   "c"   "a"   "d"   "d"   "b"   "b"   "c"   "b"
[15,] "d"   "d"   "a"   "d"   "d"   "c"   "b"   "b"   "c"
[16,] "c"   "c"   "a"   "d"   "c"   "d"   "d"   "c"   "c"
[17,] "d"   "c"   "b"   "c"   "d"   "d"   "b"   "c"   "c"
[18,] "c"   "c"   "a"   "c"   "c"   "c"   "c"   "b"   "c"
[19,] "d"   "c"   "b"   "c"   "d"   "c"   "b"   "a"   "c"
[20,] "c"   "d"   "b"   "d"   "d"   "d"   "c"   "c"   "c"
[21,] "c"   "d"   "b"   "d"   "d"   "d"   "b"   "c"   "c"
[22,] "b"   "c"   "b"   "d"   "d"   "d"   "a"   "b"   "b"
[23,] "b"   "c"   "d"   "e"   "d"   "d"   "c"   "c"   "e"
[24,] "c"   "d"   "d"   "d"   "d"   "d"   "c"   "c"   "c"
[25,] "c"   "c"   "b"   "d"   "d"   "d"   "d"   "c"   "c"
[26,] "c"   "c"   "b"   "c"   "c"   "c"   "b"   "c"   "c"
[27,] "b"   "a"   "a"   "b"   "a"   "a"   "a"   "a"   "b"
[28,] "b"   "b"   "a"   "c"   "c"   "b"   "b"   "b"   "b"
[29,] "b"   "b"   "b"   "c"   "c"   "c"   "b"   "c"   "c"
[30,] "a"   "b"   "b"   "c"   "c"   "b"   "b"   "b"   "b"
[31,] "b"   "d"   "d"   "c"   "c"   "c"   "b"   "b"   "b"
[32,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[33,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[34,] "b"   "a"   "c"   "d"   "c"   "d"   "a"   "c"   "e"
[35,] "b"   "b"   "c"   "b"   "a"   "b"   "a"   "b"   "d"
[36,] "c"   "d"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[37,] "d"   "d"   "e"   "d"   "d"   "c"   "c"   "e"   "e"
[38,] "d"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[39,] "e"   "d"   "e"   "e"   "e"   "e"   "d"   "e"   "e"
[40,] "c"   "c"   "e"   "b"   "d"   "e"   "e"   "e"   "e"
[41,] "e"   "e"   "e"   "e"   "e"   "e"   "d"   "e"   "e"
[42,] "b"   "b"   "d"   "c"   "c"   "d"   "c"   "c"   "e"
[43,] "d"   "d"   "e"   "d"   "c"   "c"   "c"   "e"   "e"
[44,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[45,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[46,] "e"   "e"   "e"   "e"   "e"   "d"   "e"   "e"   "e"
[47,] "e"   "e"   "d"   "d"   "e"   "d"   "e"   "e"   "e"
[48,] "e"   "d"   "b"   "e"   "e"   "e"   "e"   "e"   "d"
[49,] "c"   "a"   "a"   "a"   "a"   "a"   "d"   "d"   "a"
[50,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"
[51,] "d"   "d"   "d"   "d"   "e"   "e"   "e"   "e"   "d"
[52,] "c"   "d"   "d"   "c"   "c"   "b"   "d"   "e"   "a"
[53,] "d"   "e"   "e"   "d"   "e"   "d"   "e"   "e"   "d"
[54,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "c"
[55,] "e"   "e"   "e"   "e"   "d"   "e"   "d"   "c"   "a"
[56,] "e"   "d"   "d"   "e"   "e"   "c"   "e"   "c"   "b"
[57,] "d"   "e"   "e"   "c"   "c"   "c"   "d"   "c"   "d"
[58,] "c"   "c"   "d"   "b"   "b"   "b"   "b"   "c"   "b"
[59,] "d"   "d"   "e"   "c"   "b"   "c"   "d"   "c"   "d"
[60,] "d"   "d"   "d"   "c"   "c"   "a"   "c"   "d"   "c"
[61,] "e"   "e"   "e"   "e"   "d"   "c"   "d"   "e"   "e"
[62,] "e"   "e"   "d"   "c"   "c"   "b"   "c"   "c"   "e"
[63,] "e"   "e"   "e"   "d"   "e"   "c"   "c"   "b"   "d"
[64,] "e"   "e"   "e"   "e"   "e"   "e"   "e"   "e"   "d"
[65,] "b"   "c"   "e"   "c"   "c"   "d"   "d"   "d"   "e"
[66,] "b"   "c"   "c"   "c"   "c"   "c"   "c"   "b"   "b"
[67,] "b"   "c"   "d"   "c"   "c"   "d"   "c"   "c"   "c"
[68,] "b"   "c"   "d"   "c"   "c"   "d"   "c"   "c"   "c"
[69,] "c"   "d"   "e"   "d"   "c"   "e"   "c"   "c"   "d"
[70,] "b"   "c"   "d"   "c"   "c"   "d"   "c"   "c"   "b"
[71,] "b"   "d"   "a"   "c"   "c"   "a"   "b"   "b"   "d"
[72,] "b"   "c"   "a"   "c"   "b"   "b"   "c"   "c"   "e"
[73,] "c"   "e"   "d"   "c"   "b"   "b"   "b"   "b"   "c"
[74,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[75,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[76,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[77,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[78,] "c"   "e"   "b"   "b"   "b"   "c"   "b"   "b"   "a"
[79,] "a"   "a"   "a"   "a"   "b"   "b"   "b"   "a"   "b"
[80,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[81,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[82,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[83,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[84,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[85,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[86,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[87,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[88,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "b"   "a"
[89,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[90,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
[91,] "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"   "a"
>
> #estimate transition matrix with MLE
>   result6   <-markovchainFit(data=datamatrix6,   method="mle",
name="datamatrix6")
> result6$estimate
datamatrix6
 A  5 - dimensional discrete Markov Chain defined by the following
states:
 a, b, c, d, e
 The transition matrix  (by rows)  is defined as follows:
           a          b          c          d          e
a 0.87564767 0.06217617 0.04145078 0.02072539 0.00000000
b 0.16304348 0.33695652 0.33695652 0.15217391 0.01086957
c 0.06000000 0.18000000 0.41333333 0.24000000 0.10666667
d 0.04032258 0.08870968 0.29032258 0.33064516 0.25000000
e 0.00591716 0.02366864 0.10059172 0.15384615 0.71597633

> result6$standardError
           a          b          c          d          e
a 0.06735751 0.01794871 0.01465506 0.01036269 0.00000000
b 0.04209765 0.06051918 0.06051918 0.04067019 0.01086957
c 0.02000000 0.03464102 0.05249339 0.04000000 0.02666667
d 0.01803281 0.02674697 0.04838710 0.05163810 0.04490133
e 0.00591716 0.01183432 0.02439707 0.03017171 0.06508876
> transmatrix6<- result6$estimate[1:5,1:5]
> transmatrix6
           a          b          c          d          e
a 0.87564767 0.06217617 0.04145078 0.02072539 0.00000000
b 0.16304348 0.33695652 0.33695652 0.15217391 0.01086957
c 0.06000000 0.18000000 0.41333333 0.24000000 0.10666667
d 0.04032258 0.08870968 0.29032258 0.33064516 0.25000000
e 0.00591716 0.02366864 0.10059172 0.15384615 0.71597633
> upward6 <-result6$estimate[1:5,1:5]
> downward6 <-result6$estimate[1:5,1:5]
>   write.xlsx(transmatrix6,   "M:/My   Documents/MScThesis/Text
thesis/Data           files/Chapter           6/File           for
R/Matrix_Labour_income_2007_2015.xlsx")
>
> #define transition matrix as matrix object
> showClass("markovchain")
Class "markovchain" [package "markovchain"]

Slots:

Name:                states               byrow transitionMatrix
name
Class:           character             logical             matrix
character
> simpleMc6<-new("markovchain", states=c("a","b","c","d","e"),
+ transitionMatrix=transmatrix6,
+ name="simpleMc6")
>
> #stationary distribution
> steadyStates(simpleMc6)
             a         b         c         d         e
[1,] 0.3071963 0.1127426 0.2023697 0.1581611 0.2195304
>
> #half life
> eigenvalues6 <-eigen(transmatrix6)
> eigenvaluevector6 <-eigenvalues6$values
> secondeigenvalue6 <-eigenvaluevector6[[2]]
> secondeigenvalue6
[1] 0.8497384
> halflife6 <- (-log(2))/log((abs(secondeigenvalue6)))
> halflife6
[1] 4.256963
>
```

```
> #mobilityindex
> trace6 <-matrix.trace(transmatrix6)
> mobility6 <- (5-trace6)*(5-1)^-1
> mobility6
[1] 0.5818602
>
> #upward mobility
> upward6[lower.tri(upward6)] <- 0
> upward6[lower.tri(upward6,diag=TRUE)] <- 0
> sum(upward6)*(5-trace6)^-1
[1] 0.5246187
>
> #downward mobility
> downward6[upper.tri(downward6)] <- 0
> downward6[upper.tri(downward6,diag=TRUE)] <- 0
> sum(downward6)*(5-trace6)^-1
[1] 0.4753813
>
```