Modelling Twentieth Century Economic Growth in Industrialized Nations:

Exploiting Cross-Country Similarities

Rengert Segers

Master's Thesis
Econometrics & Management Science
Erasmus University Rotterdam

Supervisors
Herman K. van Dijk and Richard Paap
Econometric Institute
Erasmus University Rotterdam

May 13th, 2004

Contents

Chapter 1

Introduction

"Doing econometrics is like trying to learn the laws of electricity by playing the radio."
– Guy Henderson Orcutt

1.1 Motivation

In the 20th century, the world economy has grown faster than at any time in the past. Nowadays we do not only hope for economic progress, but it is considered a must to sustain a high quality of life and a stable social environment[1]. For the economic policy-maker to develop efficient policies that facilitate growth it is of key importance to have good insights into the determinants of growth. Secondly, he or she should be provided with reliable estimates of long-run economic growth in order to identify structural changes in economic performance. In the 20th century, many national as well as international economic policy research institutes were founded, with the objective of providing this information. For a formal analysis of economic growth, models needed to be developed and tested empirically. Not surprisingly, economic growth has therefore been one of the major subdisciplines in 20th century macroeconomic research.

To understand economic growth, it has proved to be very useful to analyze macroeconomic time series using dynamic models. Among the first to develop such models was Jan Tinbergen (1939). Using a model, initially to describe the economy of the United States over the period 1919-1932, he aimed at distinguishing long-run growth from short-run fluctuations. In Tinbergen's model, long-run growth is assumed to be constant, whereas short-run fluctuations in growth, which are cyclical in nature, are analyzed in terms of quantitative relations between production, wages, prices and investment, among other variables. It was mainly for this work that he was the first to receive the Nobel Prize in Economic Science, together with Ragnar Frisch, in 1969.

[1] To illustrate this, in many industrialized countries, enduring slowdowns in economic growth during 1973 and 1995 have shown to shake political foundations (Reich, 1992, and DeLong, 2000).

Early ideas for further improvement of Tinbergen's model were put forward by Guy Orcutt (1948), another pioneer in time series econometrics. It is interesting to discuss these ideas here, as they will prove to be of particular relevance for this research. Though he was one of the greatest econometricians of his time, Orcutt used to be quite pessimistic about the role of econometrics in economic policy-making (Orcutt, 1952). As the above quote illustrates, he wished to emphasize that econometricians, macroeconometricians in particular, should not claim a high degree of precision of their estimates. This is due to the fact that macroeconometricians aim at describing highly aggregated quantities. It is practically impossible to measure these quantities frequently and accurately[2]. Orcutt argued that one could reduce the effect of this difficulty by treating a large number of time series in one model, and thus using a whole number to obtain more accurate estimates. He realized, however, that the validity of this approach depends on the level of similarity in growth across the series under consideration. Though Orcutt could not measure the level of similarity across the time series used in Tinbergen's study, he did show that these 52 series appear to have the same underlying autoregressive structure. By pooling the time series data over all economic quantities, he improved the efficiency of Tinbergen's estimates. Based on these estimates, Orcutt concluded that the series do not have a tendency to revert to a constant long-run growth rate, as suggested by Tinbergen.

Since the seminal work of Tinbergen and Orcutt, a vast body of literature has been devoted to the latter issue of persistence in economic growth, see Stock (1994) for an excellent survey. In line with Orcutt's findings, one may advocate the Schumpeterian view that growth is fully driven by persistent, stochastic shocks (Schumpeter, 1939). However, this view does not coincide with the slowly evolving trends that macroeconomic time series usually exhibit. Therefore, nowadays, a widely expressed vision is that growth tends to revert to a persistent long-run growth path. This path is typically nonlinear, see, for instance, Granger *et al.* (1997), and Van Dijk (2004). However, it has proved to be difficult to model these paths correctly. On the one hand, the specification of the growth path should be flexible enough to describe the time variation in long-run growth, whereas, on the other hand, it should evolve slowly over time. The key problem is to find a balance between these two properties.

In contrast to the issue of persistence, little attention has been paid to Orcutt's idea to improve estimation efficiency by exploiting possible similarities in economic growth across series in a legitimate manner. The aim of this thesis is to model 20th century economic growth in industrialized countries by exactly doing so. In particular, we will model the time series of annual real per capita gross domestic product (GDP) levels of 17 industrialized countries by exploiting possible similarities in long-run growth across the

[2]For example, reliable estimates of annual economic growth, in terms of gross domestic product (GDP), are only available for a small group of industrialized countries from the end of the 19th century onwards. Quarterly estimates are available from 1949.

series[3]. Similarities in long-run growth across countries have been studied extensively. For example, Baumol (1986, 1988) found that in the 19th century groups of countries have emerged, catching up with one another and eventually having common long-run growth paths. One of these so called *convergence clubs* is the group of industrialized countries studied here. Note, however, that within this group, countries did not catch up at the same speed. Some countries seem to have joined the club later, during the 20th century (Abramovitz, 1986, Quah, 1997). Studying the 20th century in particular, obviously, there are many other causes of disparities in growth, such as differences in involvement in the Second World War (Cook, 2002). This illustrates that, even though there is evidence for similarities in long-run growth across the economies under consideration, it is unrealistic to assume that these economies all have similar growth paths during the 20th century. Instead, it is likely that within the group of industrialized countries, only smaller subgroups of countries grow similarly (Hobijn and Franses, 2000). Ideally, the data should tell us which countries belong to which groups and how many groups there are, if any. For this purpose, we will use the methodology for endogenous clustering of time series, as proposed recently in Paap *et al.* (2004). Using a latent class panel data model, countries are classified to clusters, based on similarities in their growth characteristics[4]. In particular, each country has some probability of getting assigned to a latent class, such that within each class countries have the same long-run growth characteristics, while these characteristics are different across groups. Note that if we find the number groups to be smaller than the number of countries, the model parameters that govern the characteristics of long-run growth within each group are estimated over a larger number of observations, compared to the univariate case. This improves the efficiency of the estimates, as desired. Apart from the gain in efficiency, the fact that we pool groups of two or more time series might also enable us to estimate more complex models. Using these models we may in turn be able to obtain more realistic estimates of long-run economic growth, compared to the estimates obtained using univariate models.

Additionally, if we find evidence for the existence of groups of countries that show similar growth, this modelling framework allows us to address questions such as:

- Is the composition of the groups equal to a grouping based on partnerships due to common history, culture, and/or geographical proximity?
- Does the number and/or the composition of groups change over time, or can we consider these to be constant over the entire 20th century?

[3]Note that we will not try to establish causality. For recent empirical work on the determinants of economic growth in industrialized countries, we refer to Bassanini and Scarpetta (2001), OECD (2003), and the references cited there.

[4]In the economic growth literature, other applications of clustering techniques include Durlauf and Johnson (1995), and Desdoigts (1999). Note, however, that in these studies countries are clustered based on particular economic variables and not on growth characteristics themselves.

- If the composition of groups is likely to have changed during the 20th century, which economies tend to have switched from a lower growth group to a higher growth group, and vice versa?

- During which periods of the century did the disparity in growth across industrialized countries change?

- Do the long-run growth paths of established economies typically evolve more slowly than the growth paths of emerging economies?

Prior to our model based analysis, we perform an exploratory data analysis. This analysis will provide us some first guidance regarding the answers to the above questions and might help us to refine these. However, let us first briefly outline the thesis.

1.2 Outline

Strictly speaking, long-run economic growth cannot be observed. Therefore, to be able to distinguish long-run economic growth from short-run fluctuations, we need to define what we consider to be the characteristics of both processes. The next chapter covers empirical observations, theoretical, and economic historian's viewpoints on this subject. Firstly, the data set is introduced that is used throughout the thesis. Using graphical and descriptive statistical techniques, we establish two sets of stylized facts of economic growth. The first set concerns observations about the process of economic growth over time, whereas the second set is related to cross-country similarities in growth.

Chapter 3 is devoted to univariate time series models for economic growth. We review several models that have been proposed in the literature, discussing their properties and economic implications. The conclusions of this review as well as the first set of stylized facts established in Chapter 2 will guide the formulation of a novel nonlinear time series model. This model may strike a better balance between flexibility and slowness of evolvement of the long-run growth process. The model parameters are estimated for each of the 17 series of real per capita GDP levels, and we discuss the results in detail.

In Chapter 4, the focus is shifted towards panel data models. Panel data sets are defined and the main advantages of analyzing panel data instead of aggregated data are discussed. We proceed with discussing the issue of model specification. Apart from the functional form, an important step in the specification of a panel data model is to impose a structure on the parameters. We provide a modelling framework that may facilitate this step. Finally, as we will use latent class panel data models in the remainder of the thesis, this class of models is discussed more in depth.

Chapter 5 covers the formulation, estimation, model selection and evaluation of two very basic static latent class models for the panel of economic growth rates of our series. The first model is introduced to describe clusters of countries with common average growth rates and growth rate volatility. We estimate the model separately for different subperiods of the 20th century. As a consequence, the clustering obtained for this model may be different for each subperiod. In contrast to this, the second model can be used to obtain a constant clustering over time, by incorporating shifts in the average growth rate and growth rate volatility. The results of the models are compared to the results of the exploratory data analysis of Chapter 2.

In Chapter 6 we hypothesize that the similarities in the long-run growth paths of economies, established in Chapter 2, are due to cross-country similarities in the dynamics of long-run growth. To evaluate this hypothesis, the univariate nonlinear time series model of Chapter 3 is extended to a latent class panel data model. Essentially, in this model, the building blocks of Chapters 3-5 are brought together. This results in a relatively complex model, which incorporates dynamics, nonlinear growth over time and similarities in the dynamics of long-run growth across countries. Special attention will be paid to the economic interpretation of the estimation results of the model.

Chapter 7 concludes the thesis, summarizing the main findings of our research. Finally, some directions for further research are suggested.

Chapter 2

Stylized Facts of Economic Growth

2.1 Introduction

Before diving into an analysis of economic growth using time series models, let us first take a detailed look at what it is we are trying to understand. In the next section of this chapter, we will discuss the measurement of economic growth briefly, and introduce the data set that is used throughout the thesis. This data set consists of levels of average output, produced per member of the population of a country, which are measured annually over the 20th century. We will study the time series of these levels and the corresponding growth rates for 17 industrialized countries.

To establish stylized facts of economic growth, we continue with a graphical presentation of the series and compute several summary statistics over the entire century. Average growth and growth volatility are also computed for subperiods of the century, in order to identify possible time variation. Traditionally, economic growth is believed to be stable in the long run, that is, over decades, whereas it is subject to shocks in the short run. Very large shocks, however, which can be caused by political or economic events, may also affect growth in the long run. In that case, long-run economic growth varies over time. Our graphical analysis may shed some light on this issue.

As pointed out earlier, a central question we will be researching is whether there is reason to believe that economic growth is similar across groups of countries. In addressing this question, we distinguish between cross-country similarities in long-run economic growth paths, or trends, on the one hand, and similarities in short-run fluctuations on the other hand. Firstly, similarities in long-run growth are studied, comparing the levels of our series across countries. To identify the long-run growth paths of the series, a very intuitive visual estimation technique is used. Similarities in short-run fluctuations are studied in Section 2.4. This is a more formal analysis of estimated correlations of growth rates across countries. These correlations are visualized using multidimensional scaling (MDS) techniques. Finally, we conclude the chapter by summing up the stylized facts established. These stylized facts will guide the choice of our time series models in subsequent chapters.

2.2 The Data Set and Notation

As an indicator of economic development we will look at the Gross Domestic Product (GDP) of a country. Obviously this is not a perfect measure, as, for example, it does not convey information on income distribution with a country, nor does it reflect qualitative dimensions of economic development, such as health or literacy. Nevertheless, it is widely accepted as the best indicator available. In order to compare GDP levels over time and across countries, we account for differences in purchasing power parity and convert the levels to a single currency. Furthermore, as we are not interested in cross-country differences in scale, we divide the levels by the total population of each country to obtain a measure for the average output that is produced per member of the population. This indicates how many goods and services people are likely to have, and consequently, what their standard of living is. The resultant real per capita GDP levels were measured annually over the 20th century by the Organization for Economic Cooperation and Development (OECD) for 17 industrialized countries. These countries include Australia, Canada, New Zealand, Japan, the USA and 12 Western European countries (See Table 2.1).

The real per capita GDP levels for 1900-1949 were obtained from Maddison (1995) whereas those for 1950-1998 were obtained from Maddison (2001). For 1999 and 2000, the data were obtained from the GGDC Total Economy Database[1]. All the levels are measured in 1990 US dollars converted at Geary-Khamis purchasing power parities (See Maddison, 1995, for a full description). Consequently, our data set consists of 17 series measured annually over the period 1900-2000. Since all GDP series show an exponential trend across time, we apply a log transformation. The resultant data set will be denoted by $y_{i,t}$:

$$y_{i,t} = \ln(GDP_{i,t}), \qquad i = 1, ..., 17, \quad t = 1, ..., 101, \tag{2.1}$$

where i is the cross-sectional index, t the time index and $GDP_{i,t}$ the real per capita GDP level of country i in year t. Needless to say, the annual growth rates of the real per capita GDP levels are a measure of economic growth. Measured in percentage terms, these growth rates, denoted by $\Delta y_{i,t}$, are computed as:

$$\Delta y_{i,t} = 100\big(\ln(GDP_{i,t}) - \ln(GDP_{i,t-1})\big), \qquad i = 1, ..., 17, \quad t = 1, ..., 100. \tag{2.2}$$

For brevity, we will refer to the series of the logarithmic levels, $y_{i,t}$, and the series of the growth rates in percentage terms, $\Delta y_{i,t}$, as the levels and the growth rates, respectively, in the remainder of the thesis.

[1]Groningen Growth and Development Centre and The Conference Board, Total Economy Database, July 2003, `http://www.ggdc.net`.

Table 2.1: 17 industrialized countries

Country	Abbr.	Country	Abbr.
Western Europe			
Austria	AT	Italy	IT
Belgium	BE	Netherlands	NL
Denmark	DM	Norway	NO
Finland	FI	Sweden	SE
France	FR	Switzerland	SW
Germany	GE	United Kingdom	UK
Western Offshoots			
Australia	AU	New Zealand	NZ
Canada	CA	USA	US
Japan	JP		

2.3 Graphical and Descriptive Analysis of the Data

The levels and the growth rates of the 17 countries are plotted in Appendix A.1. Furthermore, for the growth rates, we have computed several summary statistics, which are given in Table 2.2. As we also want to gain some insights in the changes of average growth and growth rate volatility over the 20th century, these two statistics are computed for subperiods as well (See Table 2.3). The subperiods, discussed later on, correspond to five distinct phases of economic development that are often considered in the literature. Let us first discuss the tables, which provide the headline numbers of 20th century economic growth. We will then discuss some stylized facts of economic growth that can be deduced from the graphs. In particular, we distinguish between stylized facts of short-run and long-run economic growth.

The distributions of economic growth rates

Strikingly, the average growth rates of the 17 economies over the 20th century are very similar. They are close to 2%, which seems to have become a magic number in economic growth research. For most countries, the median of the series is higher, which indicates that the means are strongly influenced by some very low, or even negative observations. Canada, Finland, Germany, Italy, Japan and Norway have the highest median growth rate, whereas Denmark, New Zealand, Switzerland and the UK fall behind. For all the countries, maximum and minimum growth rates were observed around the two world wars, which is not surprising. In contrast to the means, there is much cross-country variation in the standard deviations of the series, though a quick look at Table 2.3 reveals that this is due to some extreme observations in the Second World War period. Although some countries, such as the USA and Canada, benefited from the war because of their weapons industry, growth was at a nadir in general. Due to this period, the skewness

Table 2.2: Summary statistics for the percentage growth rate series.

	Mean $\Delta\bar{y}_i$	Median $\Delta\tilde{y}_i$	Maximum Δy_i^M	Minimum Δy_i^m	S.D. s_i	Skew. S_i	Kurt. K_i
Australia	1.61	2.37	10.05 ('41)	-10.30 ('13)	3.54	-0.62	3.79
Austria	1.94	2.60	24.37 ('47)	-87.87 ('44)	10.65	-6.04	52.19
Belgium	1.74	2.26	16.94 ('18)	-20.67 ('17)	4.73	-1.33	9.89
Canada	2.09	2.81	15.05 ('41)	-18.37 ('30)	5.36	-0.93	5.18
Denmark	2.04	2.01	13.11 ('45)	-15.83 ('39)	4.22	-1.02	6.82
Finland	2.52	2.80	19.11 ('18)	-18.01 ('16)	4.84	-0.93	7.31
France	1.99	2.27	40.43 ('45)	-21.80 ('17)	7.39	0.48	11.11
Germany	1.78	2.68	16.47 ('26)	-54.72 ('45)	8.95	-3.52	20.70
Italy	2.37	2.80	26.40 ('45)	-24.79 ('44)	6.17	-1.17	10.14
Japan	2.92	3.19	13.96 ('38)	-68.07 ('44)	8.49	-5.87	50.04
Netherlands	1.81	2.18	50.62 ('45)	-40.71 ('43)	7.83	0.82	25.05
New Zealand	1.32	1.49	16.35 ('35)	-12.43 ('47)	5.01	-0.17	3.66
Norway	2.64	3.07	14.69 ('18)	-10.68 ('16)	4.05	-0.70	6.16
Sweden	2.07	2.15	9.17 ('45)	-13.86 ('16)	3.12	-1.36	8.51
Switzerland	1.83	1.58	24.16 ('44)	-11.71 ('16)	3.87	1.47	13.88
UK	1.47	2.10	9.06 ('39)	-11.43 ('18)	3.29	-1.13	5.81
USA	1.93	2.35	17.16 ('41)	-24.15 ('45)	5.83	-0.85	7.16
Average	2.00	2.39	19.83	-27.38	5.73	1.67	14.55

of the growth rates is mostly negative. Finally, the kurtosis of the series exceeds three, which indicates that the tails of the distributions of the growth rates are fatter than the tails of the normal distribution.

Short-run economic growth and the impact of shocks

Let us take a detailed look at the graphs of our growth rates in Appendix A.1 and make some general observations. Firstly, we observe that in the short run, economic growth fluctuates in patterns. To illustrate this, we compare the series of growth rates to the average rates during subperiods, which are indicated by dashed horizontal lines. Typically, the series evolve either above or below this average for two up to eight years at a time. This cyclical process is commonly referred to as the economic cycle[2]. Secondly, during some periods, economic growth tends to fluctuate more than in others. This suggests that growth volatility is not constant over time. In particular, we observe that volatility is relatively high before the Second World War, whereas is it low during the postwar period. Finally, we observe that dramatic negative spikes are followed by catch-up periods. Clearly, there is variation in the length of these periods across countries and over time. We will now analyze growth during subperiods of the century in more detail, and touch upon some important historical events, which are generally thought to be the

[2]Note that we do not refer to this process as the *business* cycle. Research on business cycles usually deals with shorter run fluctuations in growth, such as quarterly fluctuations.

Table 2.3: Average growth rates and sample standard deviations over different sub-periods of the 20th century.

	1901-1918		1919-1939		1940-1945		1946-1973		1974-2000	
	$\Delta\bar{y}_i$	s_i	$\Delta\bar{y}_i$	s_i	$\Delta\bar{y}_i$	s_i	$\Delta\bar{y}_i$	s_i	$\Delta\bar{y}_i$	s_i
Australia	0.89	4.75	0.53	3.91	2.80	6.94	2.32	2.52	1.95	1.76
Austria	-0.67	5.39	2.25	7.19	-14.42	36.24	6.67	5.11	2.17	1.56
Belgium	-1.47	6.48	2.80	5.53	-2.88	7.09	3.76	1.95	1.99	1.76
Canada	2.35	5.88	0.34	8.43	6.71	7.41	2.56	2.93	1.75	2.20
Denmark	0.76	3.80	2.62	3.96	-2.80	10.72	3.75	3.20	1.75	1.90
Finland	-1.10	6.38	4.34	5.37	0.20	6.26	4.27	2.64	2.21	3.05
France	-1.02	7.74	3.30	6.97	-10.37	9.85	5.85	7.07	1.71	1.34
Germany	0.03	4.95	2.69	9.80	-4.15	16.11	3.63	11.89	1.63	1.86
Italy	3.57	4.37	0.18	5.98	-10.09	10.59	6.19	4.80	2.10	1.78
Japan	1.92	5.03	2.49	5.52	-12.30	27.45	7.78	2.45	2.26	1.85
Netherlands	-0.43	3.76	2.40	5.61	-12.08	14.98	5.74	9.28	1.86	1.50
New Zealand	0.74	4.98	1.30	6.84	1.17	4.68	2.09	5.50	0.93	2.55
Norway	0.92	3.59	3.24	5.39	-1.90	7.18	4.00	3.00	2.91	1.78
Sweden	-0.06	4.40	3.27	3.51	1.70	2.93	3.16	1.93	1.52	1.90
Switzerland	0.33	3.13	2.46	3.84	3.30	10.36	3.10	2.86	0.71	2.20
UK	1.08	2.43	0.33	5.06	1.99	6.07	2.07	2.22	1.87	1.91
USA	1.80	6.27	0.70	6.86	9.65	8.74	1.26	5.68	1.94	2.03
Average	0.57	4.90	2.07	5.87	-2.56	11.39	4.01	4.41	1.84	1.94

reason for these spikes. A chronological list of key political and economic events in the 20th century can be found in Appendix A.2. The events that are listed are believed to have affected the economies of at least two of the countries considered in our study. Note that, obviously, the list is not exhaustive.

The first period of the century, 1901-1918, is usually referred to as the Liberal Order. Growth was particularly low, and very volatile. Clearly, its most important event is the First World War, which affected growth negatively in most of the countries. Many economies, however, recovered very rapidly. Only for Austria, Canada, Germany and the UK, recovery took longer than three years[3]. The second period, the Roaring Twenties and Dirty Thirties, 1919-1939, showed very volatile economic growth as well. In the late twenties, the USA and the Canadian economies were booming, but this came to a swift end with the stock market crash in 1929. It is believed that this event plunged many countries in the decade-long Great Depression. Prices deflated and business activity fell sharply. This caused very high unemployment in the Western Offshoots. It took these economies, which are typically stable, about seven years to recover. Needless to say, the third period, which is the Second World War, was the most serious event of the century. Average growth fell below -10% in Austria, France, Italy, Japan and the Netherlands.

[3]At this stage, let us define the recovery period of an economy loosely as the period needed after a trough to return to the real per capita GDP level before the depression.

In the period thereafter, the Golden Age, 1946-1973, some countries, like Belgium, the Netherlands and France caught up with their original output levels very quickly, whereas for Austria, Germany, Italy and Japan recovery took much longer. Interestingly, most economies managed to keep up the high growth rates for a very long time. Additionally, growth rate volatility decreased significantly after the war. This may be the success of postwar counter-cyclical economic policy and intensified economic relationships. Sadly, the Golden Age did not last forever. In the final period of the century, the Neoliberal Order, 1974-2000, average growth rates decreased gradually. For some countries, such as Belgium, Denmark, France and Japan, this decrease seems to have occurred instantaneously in 1973. Therefore one may be inclined to conclude that the 1973 OPEC oil crisis caused growth rates to change structurally. However, many economists agree that it is more plausible to consider this event the straw that broke the camel's back (See, for instance, Baumol *et al.*, 1989, and Maddison, 1991).

Long-run economic growth

To conclude this section, we will now take a bird's eye view. If we would have been asked to draw the long-run economic growth paths of the 17 countries in the graphs of the levels as continuous lines, would these lines be linear[4]? Stated alternatively, is long-run growth constant over time? To be able to answer this question, let us first try to estimate these long-run growth paths visually. The following procedure is applied. Initially, the paths are drawn as linear lines. In case the original series deviates from this line for a period longer than a decade, the slope of the line is adjusted as to shorten this deviation, resulting in a piecewise linear path. This procedure is continued until deviations longer than a decade are no longer observed. Finally, during periods of 30 years or longer, where the slope of the growth path had to be changed every decade, the line is curved, such as to represent a continuous change[5]. The resultant long-run growth paths are represented by the dotted lines in the left panels of the graphs in Appendix A.1. To be able to compare these paths across countries, the paths are sorted by the number and magnitude of adjustments needed, and shown together in four graphs in Appendix A.3.

Considering the growth paths obtained, we conclude that, in contrast to the traditional view, long-run economic growth clearly varies over time. However, it tends to change only very slowly. This observation is consistent with economic theory. After all, changes in long-run output, which is a highly aggregated quantity, are due to changes in behaviour of a very large number of economic agents. These agents may have different objectives, and they may operate in markets with a different degree of efficiency. Moreover, changes in their behaviour may require a different degree of institutional change. Therefore we do

[4]Note that if the logarithmic levels of the series, briefly referred to as the levels here, would be linear, the untransformed series would show an exponential trend.

[5]Obviously, the outcome of this exercise would have been different if deviations from the long-run growth path exceeding a decade had been tolerated. For a thorough analysis, long-run growth has to be defined formally. This topic will be discussed in the next chapter.

not expect them to react simultaneously to an economic stimulus. Instead, it takes a long period of time for all agents to adjust. This consideration suggests that long-run output, and, consequently, long-run output growth changes slowly, rather than abruptly (See, for instance, Teräsvirta *et al.*, 1994, and Greenaway *et al.*, 2000, for a similar discussion).

Let us now compare long-run growth paths across countries. For the first group shown in Appendix A.3, long-run growth is most stable. More precisely, growth tends to be constant over periods of 50 years or longer. Remarkably, during the final 60 years of the century, the growth paths of Australia, the UK and the US are equally sloped. Recently, the two Scandinavian countries seem to have caught up with this group. In the next two graphs, paths with minor changes in long-run growth are shown. Though nonlinear, these growth paths have a typical curvature as well. In particular, for Belgium, France, Germany and the Netherlands, the paths tend to overlap. The latter economies are characterized by very fast growth after the Second World War, and a gradual decline in long-run growth during the final three decades. The fourth graph shows growth paths with relatively large changes in the slope. Interestingly, these changes seem to have occurred at different points in time. Also for this group, the so called catch-up effect appears to be strong, as the output levels of these countries are relatively low around 1900, whereas they are relatively high around 2000. Finally, closely related to this, we observe that, although there was relatively much cross-country disparity in growth at the beginning of the century, output levels are very similar near the end. This observation is frequently used as an indication of economic convergence across industrialized countries.

Though beyond the scope of our research, it is very interesting to analyze what caused strong similarities within, and differences between, groups of countries to emerge. Let us mention a few economic historians' visions on this issue here. Firstly, similarities in long-run growth are believed to be the result of strong historical (trading) partnerships. Many of these partnerships are due to common history, culture, and/or geographical proximity. Examples are the Scandinavian countries, and the first member states of the EU. Differences arise as some economies are more mature than others. As mentioned earlier, during the 20th century, economies that were relatively immature have caught up with others (Abramovitz, 1986). Examples are Austria, Italy and Japan. Secondly, shocks, such as the Second World War, seem to have had a more persistent effect on some economies than on others. This may be simply because some countries were more involved in the war, or because some economies managed to recover more quickly than others (Cook, 2002). Some examples of the latter kind are Austria and the Netherlands. Finally, it could be that some groups of countries were more successful in developing efficient policies to facilitate growth, because of better institutions (Acemoglu *et al.*, 2001).

2.4 Time-varying Correlations Across Growth Rates

The graphical analysis of the real per capita GDP *levels* in the previous section suggests that there are strong similarities in the long-run growth paths of some groups of countries. In this section we will study cross-country similarities in the *growth rates*. As mentioned earlier, growth rates represent annual percentage growth fluctuations. Therefore one may consider this a study of short-run comovements in growth. To measure how strongly growth rates are related across countries, we will analyze estimated correlations between pairs of series[6]. It is not only interesting to look at correlations measured over the entire sample period, but also to look at correlations in smaller subperiods. Especially, to look how these correlations vary over time. An elegant way to study time-varying correlations, is to visualize them using multidimensional scaling (MDS). Before we apply this technique to our data set, let us first discuss the technique briefly. For details we refer to Kruskal (1964), Groenen and Franses (2000), and the references cited there.

Multidimensional scaling

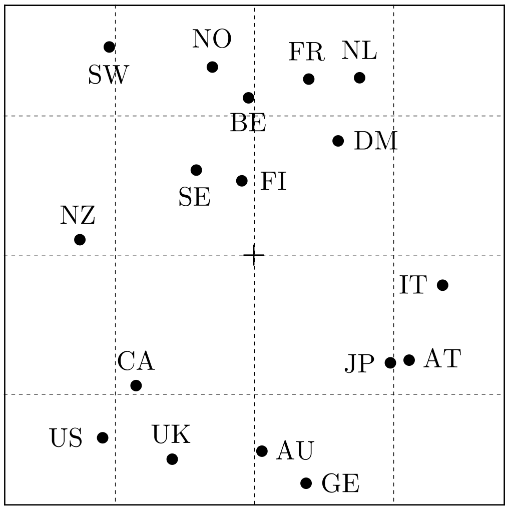
MDS can be used to visualize distances between entities, in our case correlations between growth rates of countries, in a one- or multidimensional plot. In this plot each country is represented by a point. Points at close distance correlate highly while points at large distance correlate less or even negatively. Ideally, for each pair of countries, the Euclidean distance between the two corresponding points in the plot is equal to one minus the correlation. However, in most applications it is not possible to find such perfect coordinates, and one has to resolve to an approximate solution of the problem. One way to find such a solution is to minimize a criterion function that measures the disparity between the desired and true distances between the points for a given matrix of coordinates. This $N \times P$ matrix of coordinates will be denoted by X, where N is the number of countries and P the dimensionality. A natural choice for the criterion function, called the Stress function, $S(X)$, is:

$$S(X) = \frac{\sum_{j=2}^{N} \sum_{i=1}^{j-1} \left((1-\rho_{i,j}) - d_{i,j}(X)\right)^2}{\sum_{j=2}^{N} \sum_{i=1}^{j-1} (1-\rho_{i,j})^2}, \tag{2.3}$$

where $\rho_{i,j}$ denotes the correlation between countries i and j, and $d_{i,j}(X)$ the Euclidean distance in the P-dimensional space. In the numerator, the sum of the squared differences between one minus the correlations and the Euclidean distances is computed[7]. The denominator of the function is a normalization factor, which ensures that the value of Stress

[6]Note that correlation is a linear measure of similarity. Therefore, estimated correlations should be interpreted cautiously in case the underlying process is nonlinear.

[7]As both the matrix of ones minus the correlations and the matrix of Euclidean distances are symmetric, that is $(1-\rho_{i,j}) = (1-\rho_{j,i})$ and $d_{i,j}(X) = d_{j,i}(X)$ for all i and j, and furthermore the diagonal elements of these matrices are zero, that is $(1-\rho_{ii}) = 0$ and $d_{ii}(X) = 0$ for all i, we only sum over the upper triangular elements.



$S(X) = 0.0243$

Figure 2.1: Two-dimensional scaling solution for the matrix of estimated correlations, measured over the 20th century. The dashed grid lines represent half unit distances, where the origin is indicated by a cross.

lies in the unit-interval. This allows us to compare the values of the Stress function for different MDS models. The value of $(1 - \text{Stress})$ is a measure for the goodness-of-fit of the model, similar to the R^2 measure in a linear regression model. The solution of the problem is not unique, as one can freely rotate the coordinates in the plane without affecting the value of the Stress function. To overcome this, it is imposed that the coordinates sum to zero per dimension. As it is not possible to solve the problem analytically, one has to use a numerical algorithm[8] to obtain this solution. MDS is an exploratory technique. Therefore the choice of the dimensionality P may be based on the specific nature of the problem. Alternatively, one can base the choice on the incremental contribution of each dimension to the goodness-of-fit of the model.

Results of the MDS analysis

Let us first analyze a two-dimensional[9] scaling solution for the matrix of estimated correlations measured over the entire century, which is shown in Figure 2.1. As desired, there seem to emerge groups, or clusters, of countries, which show similar growth rates during the 20th century. The number of clusters and the classification of countries can be determined by visual inspection. Note, however, that this is arbitrary. One may

[8]Following Groenen and Franses (2000) we use the SMACOF algorithm, which appears in the PROX-SCAL program in SPSS 10+. We use the classical scaling solution as a starting configuration and stop when the decrease in Stress between two iterations is less than 10^{-8}.

[9]Based on the incremental contributions criterium, we established that the appropriate dimensionality of our problem is two.

opt for a division in three clusters. The first cluster covers Scandinavia, the majority of the Western European countries and New Zealand. The second cluster covers North America, the UK, Australia and Germany, and the third covers Austria, Italy and Japan. Looking in more detail, one may consider Australia and Germany to be a separate cluster. Furthermore, within the large first cluster, Switzerland and New Zealand tend to behave quite differently compared to the other countries.

Although this graph provides us with an elegant overall picture, it does not tell us anything about the stability of the correlations over time. Therefore, we compute the two-dimensional scaling solutions for matrices of estimated correlations measured over subperiods of 20 years[10]. This results in a sequence of 80 matrices of coordinates[11], which correspond to the periods 1901-1920 through 1980-1999. As mentioned earlier, one can freely rotate the coordinates without affecting the distances between the countries. This freedom is used to match the solutions in subsequent subperiods[12]. Finally, eight graphs were constructed, each showing the coordinates of ten subsequent subperiods. The graphs are shown in Appendix A.4. For each country the coordinates are connected by straight lines. At the top of each graph, the midpoint of the window is given. The first position is indicated by a white circle. From that point, a country moves along the line towards the final position of the window, which is indicated by a black circle.

Considering the large changes in the coordinates of countries, the correlations between their growth rates varied strongly over time. The first graph shows the solutions centered around the First World War. During this period, correlations were relatively stable, and strong clustering behaviour can be observed. A classification based on this graph would be fairly equal to the classification obtained above for the entire century. Perhaps the most prominent difference is that Italy is not highly correlated to Austria and Japan during this period. The subsequent 20 solutions show great instability in the correlations. However, the instability disappears all of a sudden in the period centered around the Second World War. Also some clusters tend to emerge again. Compared to the classification based on the First World War, Australia and Germany now joined the cluster of Australia, the UK and the USA. Italy tends to move towards the European countries. The next three graphs, which roughly cover the Golden Age, show a very unstable correlational structure. These graphs are shown merely for the sake of completeness. A thorough analysis of the individual solutions is needed to identify comovements of groups of countries during this period. Only towards the end of the century, stability tends to re-emerge. However, in contrast to earlier periods, this tendency does not coincide with the formation of distinct

[10]If the subperiods are long, the estimates of the correlations are less influenced by aberrant observations. However, if there is a structural change in the true correlation, its exact timing cannot be observed. The choice of the length of our subperiods is a compromise between these two considerations.

[11]The values of the stress obtained for these 80 solutions are only slightly higher than the value of the stress obtained for the solution of the entire century: $\bar{S} = 0.0263$ and $s_S = 0.0052$.

[12]To match the solutions, Procrustes rotation is applied (Cliff, 1966).

clusters. This could be interpreted as another sign of convergence across the countries under consideration.

Obviously, interesting conclusions can also be drawn from the graphs regarding the behaviour of individual countries, or small groups of countries. However, these are less relevant for the purpose of this chapter. We will come back to the results in more detail in Section 5.3, where we compare the results of a model based analysis to this analysis.

2.5 Conclusions

Throughout this thesis, we analyze the real per capita GDP levels and the corresponding growth rates of 17 industrialized countries, measured annually over the 20th century. A detailed analysis of the graphs and summary statistics of these series revealed some important stylized facts of economic growth. These stylized facts will guide the choice of our time series models in subsequent chapters.

Our first set of stylized facts concerns observations about the process of economic growth over time. Firstly, we observed that short-run economic growth fluctuates in cyclical patterns around a long-run level. The volatility of these fluctuations is not constant over time. As regards the 20th century, summary statistics computed over subperiods of the century indicated that growth volatility was very high before the Second World War, whereas it was low after this period. This observation suggests the presence of a structural break in volatility located around 1945-1950. Secondly, we established that long-run economic growth tends to vary over time. We observed, however, that it changes only slowly. Finally, economic growth is subject to shocks. These shocks are possibly due to key political or economic events. For most shocks that occurred during the 20th century, it took countries up to at most seven years to recover. Therefore we consider the effects of these shocks to be temporary. Some shocks, however, also had an impact on long-run economic growth. Typically, these shocks were due to events that concerned many countries. Examples are the Great Depression and the Second World War.

The second set of stylized facts is related to cross-country similarities in economic growth. Firstly, we established that, in general, short-run movements in economic growth are very different across countries. Only for some pairs of countries, correlations are stable over prolonged periods of time. However, stable correlations and strong clustering behaviour occurs during periods surrounding dramatic events. Interestingly, this is also one of the conclusions of Groenen and Franses (2000), who studied time-varying correlations across 13 stock markets. In contrast to the short run, long-run economic growth is similar across countries for very long periods of time. For the 20th century, three groups of countries seem to emerge. Firstly, there is a group for which the long-run growth paths appear to be close to linear. In particular, for these countries, long-run growth was constant over periods of 50 years or longer. For the second group of countries, only minor changes in long-run growth can be observed. These economies are characterized by faster growth after the Second World War, and a gradual decline in long-run growth during the final three decades. For the remainder of the countries, changes in long-run growth are larger. During the 20th century, these countries seem to have caught up with others, as average output was relatively low around 1900, whereas it is high around 2000.

Chapter 3

Univariate Analysis

3.1 Introduction

In this chapter we discuss univariate time series models to describe economic growth. A basic premise of these models is that each of our series of real per capita GDP levels, denoted by y_t, can be decomposed as:

$$y_t = n_t + z_t, \tag{3.1}$$

where n_t represents the trend of the series and z_t the deviations from this trend. The z_t component is usually labelled as the cycle. For the trend to represent the long-run growth path of a series, we established earlier that it should satisfy two conditions. Firstly, it should evolve slowly over time. Secondly, it should be flexible enough to describe the time variation in long-run growth. Recall that we satisfied the latter condition in our visual estimation exercise by adjusting the trend in case the series deviates for a period longer than a decade. This intuitive approach guides our thinking about trend modelling in two ways. Firstly, to anticipate to structural changes in long-run growth, we might allow the direction of our trend to adjust during periods of large deviations. Secondly, to verify whether a trend is flexible enough, we could analyze whether the deviations from the trend are temporary. Stated alternatively, we could analyze whether the series reverts to the trend in the long run. For the trend to be flexible enough to consider its deviations temporary, one may be inclined to specify a trend function that allows for abrupt adjustments to unanticipated shocks. However, it is very unlikely that this trend will evolve slowly over time. Such a trend function is not only inconsistent with economic theory, it is also undesirable from a forecasting point of view. After all, the more flexibility one introduces in the trend, the more forecast uncertainty is implied. The key problem of trend modelling is therefore to find a balance between flexibility and slowness of evolvement.

The chapter is organized as follows. At first, attention will be paid to models for the deviations from the trend, or the economic cycle. This enables us to discuss two methods to check whether deviations from a trend are temporary. Subsequently, in Section 3.3, we provide a concise overview of various trend specifications suggested in the literature, and discuss whether these meet the two conditions stated above. Both this overview and the first set of stylized facts established in Chapter 2 guide the formulation of a new time series model for macroeconomic growth which is proposed in Section 3.4. Basically, our model corresponds to the vision that one should allow the direction of the trend to change stochastically over time. However, the extend to which long-run growth can change over a year is limited and should be modelled accordingly. Finally, in Section3.5 we estimate the model parameters for our 17 time series, and discuss the results in detail. Section 3.6 concludes the chapter.

3.2 Modelling Economic Cycles

In the previous chapter we observed that in the short run, economic growth fluctuates in cyclical patterns around a long-run level. This level is represented by the trend in our models. An implication of this observation is that the deviations from the trend may be autocorrelated. To describe the autocorrelation in the trend deviations z_t, one usually considers autoregressive (AR) models:

$$\phi(L)z_t = \varepsilon_t, \tag{3.2}$$

where ε_t is a stochastic error term, $\phi(L)$ is a lag polynomial, defined as $\phi(L) = 1 - \phi_1 L - ... - \phi_p L^p$, L is the lag operator, defined by $L^k y_t = y_{t-k}$, and p is the number of lags. Alternatively, one can use an autoregressive moving average (ARMA) model. This model is of particular interest when the number of lags p is large, in case of an AR representation. It can be formulated as:

$$\phi(L)z_t = \theta(L)\varepsilon_t, \tag{3.3}$$

where $\theta(L) = 1 - \theta_1 L - ... - \theta_q L^q$. The model approximates an AR polynomial with a high lag order by a ratio of two polynomials of order p and q, which involves less parameters.

To analyze whether the deviations from a particular trend are temporary, one could simply check whether these deviations always revert to zero within a tolerable period of time. However, often, a more formal treatment is desired. Therefore, one usually tests whether the cycle is stationary. Note, however, that this is a somewhat stronger condition than mean reversion[1]. Within the framework of an ARMA model, stationarity of the cycle requires that all the p roots of the polynomial $\phi(x)$ are greater than one in absolute value. Moreover, the process is invertible if the q roots of $\theta(x)$ are outside the unit circle. This implies that y_t is a stationary and invertible ARMA process around the trend n_t. In case one of the roots is equal to one, the series is said to have a unit root.

[1]Loosely speaking, stationarity requires that the mean, variance and autocovariance of a time series are constant over time, see, for instance, Franses and van Dijk (2000) for a formal definition.

Completing the specification of the cycle, it is usually assumed that the errors ε_t are independently and identically normally distributed with zero mean and variance σ^2:

$$\varepsilon_t \sim NID(0, \sigma^2). \tag{3.4}$$

Obviously, one could also impose other distributions than the normal. For example, Geweke (1993) and Hoek *et al.* (1995) suggest to use the Student-t distribution, as this distribution is more robust to outlying observations. Considering the high kurtosis of our growth rates (See Table 2.2), the assumption of normality might indeed be questioned in our case as well. A second issue is related to the variance of the cycle. In the specification above, the cycle is assumed to be homoscedastic. However, as we observed structural changes in growth volatility over time, a relaxation of this assumption is desired.

3.3 Modelling Economic Trends

Let us first discuss two extreme choices for the trend. This will illustrate the key problem of trend modelling, which we discussed above. Traditionally, a very popular choice for the trend component has been the linear deterministic trend, or time trend,

$$n_t = n_1 + \alpha(t-1), \tag{3.5}$$

written recursively as,

$$n_t = n_{t-1} + \alpha, \tag{3.6}$$

where both the slope α and the initial value n_1 are unknown parameters. It corresponds to the vision that in the long run, an economy grows at a constant rate of α per year, which is also referred to as the long-run sustainable rate of growth. Clearly, this trend satisfies the condition of slowness of evolvement, as its direction does not change over time. Another advantage is that it has no forecast uncertainty, since it is fully deterministic. However, Nelson and Plosser (1982) concluded that for most macroeconomic time series, the deviations from this trend function are nonstationary[2]. Indeed, applying their procedure to our series, using one lagged differenced term, we find that only for the series of the USA, the deviations from the trend can be considered stationary. This suggests that the linear deterministic trend is not flexible enough to represent the long-run growth path of the 16 other series.

Alternatively, one could introduce a random variable in the trend, to allow for unanticipated changes in its direction:

$$n_t = n_{t-1} + \alpha + \beta u_t, \qquad u_t \sim NID(0,1), \tag{3.7}$$

[2]The conclusion is based on the results of the Dickey-Fuller (1979) test for a unit root, carried out on 14 long annual macroeconomic time series for the USA. For 13 of these series the null hypothesis of a unit root could not be rejected at the 5% significance level.

where α is a drift term, and u_t denotes an unanticipated shock, which is multiplied by a factor β. This trend corresponds to the Schumpeterian view that growth is fully driven by persistent, stochastic shocks, or innovations, and is referred to as the random walk with drift. Essentially, the direction of the trend is adjusted to every change in the level of the series. Although this trend is very flexible, naturally, it does not evolve slowly over time. Moreover, it leads to forecast uncertainty, which increases linearly over time (See Paap, 1997, among others). For this reason, the random walk with drift is not particularly useful to describe and forecast long-run economic growth.

The conclusions of Nelson and Plosser have spawned a vast literature on long-run inference and the specification of variable macroeconomic trends (See Stock and Watson, 1988, and Stock, 1994, Section 5, for a survey). A first attempt to model the variability of macroeconomic trends deterministically was made independently by Perron (1989) and Rappoport and Reichlin (1989). They suggested to allow for structural breaks and/or changes of the slope of the linear deterministic trend. These structural changes in the series might be due to important political or economic events such as the ones discussed in the previous chapter. Perron illustrated that if one models a structural break in 1929 (the Great Crash) and a change of the slope in 1973 (the OPEC oil crisis), the resultant cycle is stationary for most of the series used in the work of Nelson and Plosser. This result, however, is very sensitive to the choice of the breakpoints. Therefore Zivot and Andrews (1992) suggested to treat these breakpoints as unknown parameters and adjust the test for the null hypothesis of a unit root accordingly. Unfortunately, this approach leads to an overturn of the verdict again.

The ease with which the null hypothesis of a unit root is not rejected, lead to the suspicion that unit root tests have low power in small samples. In the late 1980s, a large body of Monte Carlo evidence has been found in favour of this (See Stock, 1994, Section 3.2.4, for a survey). Perron (1991) refined this statement by showing that the power of unit root tests is more influenced by the span of the data than the number of observations. Stated alternatively, the power of a unit root test will not increase by using higher frequency data if this is at the cost of a smaller time span. Regrettably, this situation applies to most quantities studied in macroeconomics. Quantities for which data are available over a relatively large time span, are only measured annually, whereas for quantities measured quarterly, data are only available from 1951 onwards. As a consequence, in this field of research, attention was largely directed away from unit root tests. Apart from this, for unit root testing in nonlinear models a formal distribution theory has not yet been developed, as this is a very difficult problem. This is the reason why we will not be able to perform unit root tests to the cycles estimated using our models.

To conclude this section, we discuss two more recent approaches to trend modelling. The first trend specification, which is an alternative to the random walk with drift, is proposed by Hamilton (1989). Hamilton suggested to exchange the continuous random variable u_t in (3.7) for an unobserved first-order Markov process s_t:

$$n_t = n_{t-1} + \alpha + \beta s_t, \qquad s_t = 0, 1. \tag{3.8}$$

The process s_t takes either the value zero or one, depending on the previous state of the process, s_{t-1}, where the transition probabilities are given by:

$$\begin{aligned} \Pr[s_t = 0 | s_{t-1} = 0] &= p, \qquad \Pr[s_t = 1 | s_{t-1} = 0] = 1 - p \qquad \text{with } 0 < p < 1, \\ \Pr[s_t = 1 | s_{t-1} = 1] &= q, \qquad \Pr[s_t = 0 | s_{t-1} = 1] = 1 - q \qquad \text{with } 0 < q < 1. \end{aligned} \tag{3.9}$$

Consequently, the slope of this trend takes only two values, rather than a continuous spectrum. If s_t equals zero, the slope of the trend is α, whereas, if s_t equals one, it is $\alpha + \beta$. Without going into further details, the reader will probably notice that this trend, which is known in the literature as the Markov trend, is less flexible than the random walk with drift but more flexible than the linear deterministic trend (Paap, 1997). Therefore it may strike a better balance between flexibility and slowness of evolvement. Markov trends were found to be particularly useful to describe alternating periods of expansion and recession. For example, such periods can be detected from macroeconomic quantities measured quarterly. However, to describe the long-run growth paths of our time series, two different values for the slope of the trend might not be enough to describe the variability over time.

The final trend specification we discuss is an alternative to the linear deterministic trend given by (3.6). Greenaway *et al.* (2000) suggested to introduce a flexible nonlinear deterministic function in the trend:

$$n_t = n_{t-1} + \alpha + \beta G(t|\gamma, \tau), \tag{3.10}$$

where $G(\cdot)$ denotes the logistic transition function, as considered in the class of logistic smooth transition autoregressive (LSTAR) models, popularized by Teräsvirta (1994):

$$G(t|\gamma, \tau) = \frac{1}{1 + \exp\big(-\gamma(t - \tau)\big)}. \tag{3.11}$$

This function changes monotonically from 0 to 1 as t varies from $-\infty$ to $+\infty$, while $G(\tau|\gamma, \tau) = \frac{1}{2}$. The parameter τ determines the timing of the transition midpoint. The smoothness of the change in the value of the transition function is determined by the parameter γ, which is restricted to be positive, where a low value of γ corresponds to a smooth change and a high value to an abrupt change[3]. As a consequence, this specification allows the slope of the trend to vary smoothly from α to $\alpha + \beta$. This creates a rich class of trends, which evolve slowly and smoothly over time by design, provided that the estimate of γ is low. An obvious drawback of this trend is that it is fully deterministic. Although the trend might be flexible enough to describe long-run growth within the time span of the data under consideration, it does anticipate to any structural changes in future growth.

[3]A slightly different logistic transition function will be discussed in more detail in the next section.

3.4 Introducing the STAAR Model

Our modelling approach aims at combining the beneficial properties of the Markov trend and the nonlinear deterministic trend, as proposed by Greenaway *et al.* (2000). Similar to the Markov trend specification, we allow the direction of the trend to change stochastically over time. However, the direction does not depend on the previous state of a Markov process. Instead it depends on the standardized difference between the level of the series and the trend in the previous year, $(y_{t-1} - n_{t-1})/\sigma$. We will refer to this value as the standardized trend deviation. Furthermore, in contrast to the Markov trend, where the slope can take only two different values, we allow the slope to take all values on a fixed continuous interval. Note that this requires us to specify a function that maps standardized trend deviations on this interval. As we do not want to fully impose the shape of this function, a flexible transition function is used, similar to the function used in Greenaway *et al.* (2000). The trend is formulated as follows:

$$n_t = n_{t-1} + \alpha + \beta F\Big((y_{t-1} - n_{t-1})/\sigma \Big| \gamma\Big), \tag{3.12}$$

where $F(\cdot)$ is a logistic transition function[4], defined as:

$$F(x_t|\gamma) = \frac{1 - \exp(-\gamma x_t)}{1 + \exp(-\gamma x_t)}. \tag{3.13}$$

Different to the transition function discussed earlier, this function changes monotonically from -1 to 1 as x_t varies from $-\infty$ to $+\infty$, while $F(0) = 0$. In Figure 3.1, the function is plotted for different values of the smoothness parameter γ. As γ becomes very large, the change of the function from -1 to 1 becomes almost instantaneous at $x_t = 0$, while, as γ approaches zero, the function becomes approximately zero for all values of x_t in a finite region around zero. If, in the previous year, the actual output level was higher than the trend, the output of the transition function is a positive value between 0 and 1. Logically, in case of a negative difference, the output of the transition function varies between -1 and 0. The parameter β determines the maximum positive or negative adjustment of the trend allowed. Consequently, the slope of the trend may take all values on the closed interval $[\alpha - \beta, \alpha + \beta]$. It is important to make clear at this point that one may not expect the estimate of the constant α to be the same for this trend as for the linear deterministic trend, since the estimate of β will be nonzero in general. Nevertheless, if β equals zero, the trend function reduces to a linear deterministic trend. In contrast, if β is nonzero and γ is large, the trend allows for abrupt adjustments to the long-run growth rate, similar to the random walk with drift. Our trend therefore embeds these two extremes as limiting cases.

[4]The function $F(\cdot)$ is a transformed version of the function $G(\cdot)$ in (3.11). In particular, $F(\cdot) = 2G(\cdot) - 1$, where the threshold τ is set equal to zero.

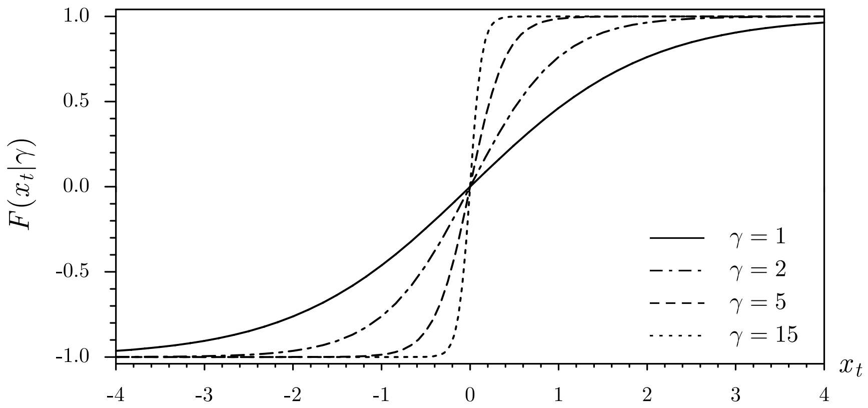


Figure 3.1: The logistic transition function $F(x_t|\gamma)$ as given in (3.13) for different values of the smoothness parameter γ.

Finally, let us discuss two potential weaknesses of this trend specification. Firstly, the expectation of the argument of the transition function is zero[5], while, in the immediate neighbourhood of zero, the transition function changes most rapidly from -1 to 1. This implies that trend adjustments due to small deviations from the trend are relatively large, which is not desired in most cases. Ideally, the transition function should be stable around zero, as an adjustment to the trend might not be strictly necessary for small deviations. Note, however, that this problem is less severe if γ is small. Secondly, using the transition function above, trend adjustments are assumed to be symmetric. That is, an adjustment due to a positive difference between the level of the series and the trend has the same magnitude as an adjustment due to an equal negative difference. Obviously, one could also consider specifications which do not impose these restrictions. However, these specifications are often heavily parameterized[6]. In the interest of parsimony and ease of interpretation, we believe the trend specification given by (3.12) is a good starting point.

[5]Assuming that z_t is a stationary ARMA process with $\varepsilon_t \sim N(0, \sigma^2)$, we obtain: $E\big((y_t - n_t)/\sigma\big) = E\big(z_t/\sigma\big) = \sigma^{-1}E\big(z_t\big) = 0$, for $t = 1, ..., T$.

[6]To illustrate this, a possible way to incorporate both stability around zero, and asymmetry in downwards and upwards adjustments to the trend, is to use an LSTAR specification with three regimes (See Franses and van Dijk, 2000, among others):

$$n_t = n_{t-1} + (\alpha - \beta_1) + \beta_1 G_1(x_{t-1}|\gamma_1, \tau_1) + \beta_2 G_2(x_{t-1}|\gamma_2, \tau_2),$$

where $x_t = (y_t - n_t)/\sigma$, $G_j(\cdot)$ is specified as in (3.11) , and β_j and γ_j are positive, for $j = 1, 2$. The parameters β_1 and β_2 respectively denote the magnitude of the maximum downwards and upwards adjustments allowed, whereas τ_1 and τ_2 denote the thresholds between the middle and low regime, and middle and high regime, where $\tau_1 < \tau_2$. Although this specification is interesting from a theoretical point of view, it is too heavily parameterized for most empirical work in macroeconomics.

To complete our time series model for macroeconomic growth, we assume that the cycle can be described as an ARMA(p, q) process, where the error term ε_t is independently and identically normally distributed with zero mean and variance σ^2. In summary, the model can now be written as:

$$\begin{aligned} y_t &= n_t + z_t \\ n_t &= n_{t-1} + \alpha + \beta F\Big((y_{t-1} - n_{t-1})/\sigma \Big| \gamma\Big) \\ \phi(L) z_t &= \theta(L)\varepsilon_t \\ \varepsilon_t &\sim NID(0, \sigma^2) \end{aligned} \tag{3.14}$$

where n_1 is treated as an unknown parameter and $F(\cdot)$ is specified in (3.13). We refer to this model as the smooth trend adjustments autoregressive (STAAR) model.

3.5 Results

Due to the small sample size of the individual time series, one additional restriction has to be imposed to estimate the parameters of the STAAR model. It is assumed that the cycle z_t is an AR process with a lag order of one[7]. Consequently, the specification for the cycle reduces to:

$$z_t = \rho z_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim NID(0, \sigma^2) \tag{3.15}$$

The parameter estimates of the model, shown in Table 3.1, are obtained by Maximum Likelihood (ML), using the Newton-Raphson method to maximize the log-likelihood function. These estimates are asymptotically normally distributed with mean equal to the true parameter vector and covariance matrix equal to the inverse of the information matrix. The information matrix is estimated as minus the Hessian matrix evaluated at the estimated vector of parameters, where the Hessian matrix is defined as the matrix of second partial derivatives of the log-likelihood function. We set the estimate of the initial value of a series, $\hat{y}_1$, equal to the estimate of the initial value of the trend, $\hat{n}_1$. To analyze the performance of the model in more detail, the estimated trend components are presented graphically in Appendix A.5. Also, the original series and the output series of the transition function are shown.

A first look at Table 3.1 suggests that the univariate STAAR model has considerable potential. For 13 of the 17 series, the estimated value of β is highly significant, which indicates an important role for the nonlinear adjustment mechanism of the model. Moreover, the estimates of the autoregressive parameter ρ are relatively small, or even insignificant. This indicates a low degree of persistence of the cycle. Therefore, the trend appears to be flexible enough to describe long-run growth, as desired.

[7] Note that the AR(1) process does not generate cyclical behaviour. If the purpose of ones analysis is to study economic cycles, other specifications should be considered.

Initial optimism fades, as we take a closer look at the graphs of the trends. For many countries, the estimated trends tend to overlap the original series. Clearly, in these cases, the trend does not evolve slowly over time and the STAAR model suffers from overfitting the trend component. Typically, for these countries, the estimates of both β and γ are high. Note that this is the limiting case where the model allows for large, abrupt adjustments to the long-run growth rate, similar to the random walk with drift. This is also confirmed by the output series of the transition function, which resemble the series of growth rates of the particular countries. Scaled with an appropriate factor β, the trend adjustment nearly equals the growth rate in the previous period. In these cases, it is likely that the estimates of β and γ are highly influenced by shocks, due to, for example, the Second World War and relatively high volatility in the first half of the century.

Finally, let us mention some more promising cases as well. For Australia and New Zealand, the trend does evolve slower than the series, as deviations from the trend over medium-run periods up to eight years are frequently observed. These deviation are even longer for Denmark, France and Germany. For the latter group of countries, long-run growth is close to zero or decreases slowly during the periods of the two world wars, but, as expected, it does not fall sharply. During the Golden Age, trend growth is lower than actual growth, indicating that the 'golden growth rates' would not last in the long run. A possible drawback of these relatively stable trends is that the resultant cycles are more persistent. However, the estimates of ρ are still smaller than 0.9, which is generally considered to be not unacceptably close to one, given the small sample size. Ironically, for Denmark, France and Germany, the estimates of β appear to be statistically insignificant[8]. It is conjectured that more observations are needed in order to accurately estimate the parameters of the adjustment mechanism in its desired role.

3.6 Conclusions

A basic premise of the models discussed in this thesis is that each of our time series of real per capita GDP levels can be decomposed as the sum of a trend and a cycle component. For the trend component to represent the long-run growth path of the series, it should evolve slowly over time. Moreover, it should be flexible enough to describe the time variation in long-run growth. The key problem of trend modelling is to specify a trend that strikes a balance between these two possibly conflicting conditions. To illustrate this, we have provided a concise overview of trend specifications that have been proposed in the literature.

Traditionally, deterministic trends have been used very frequently, as these trends can be made arbitrarily smooth. A popular choice in this class is the linear deterministic trend,

[8] Note that we cannot apply the standard t-test here, as the asymptotic distribution is nonstandard. This is due to the fact that γ is not identified under the null hypothesis $H_0 : \beta = 0$. The problem is known as the Davies (1977) problem. See Hansen (1996), among others, for a solution.

Table 3.1: Parameter estimates of the univariate STAAR model. Estimated standard errors are in parenthesis.

Country	$\hat{n}_1$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\rho}$	$\hat{\sigma}$
Australia	8.382***	0.007*	0.031***	1.802**	0.743***	0.034***
	(0.024)	(0.004)	(0.007)	(0.801)	(0.093)	(0.002)
Austria	7.984***	0.010	0.742***	0.485***	−0.050	0.099***
	(0.061)	(0.017)	(0.098)	(0.088)	(0.138)	(0.007)
Belgium	8.209***	0.014**	0.153***	0.570***	0.375**	0.044***
	(0.042)	(0.007)	(0.042)	(0.166)	(0.154)	(0.003)
Canada	7.910***	0.018***	0.115***	1.581***	0.304**	0.048***
	(0.026)	(0.005)	(0.011)	(0.287)	(0.127)	(0.003)
Denmark	7.973***	0.020***	0.019	1.092	0.811***	0.041***
	(0.042)	(0.004)	(0.012)	(0.883)	(0.133)	(0.003)
Finland	7.406***	0.020**	0.099***	1.012**	0.437***	0.045***
	(0.038)	(0.008)	(0.035)	(0.441)	(0.133)	(0.003)
France	7.963***	0.017	0.047	1.093***	0.819***	0.072***
	(0.090)	(0.090)	(0.045)	(0.099)	(0.090)	(0.006)
Germany	8.052***	0.018***	0.012	2.011***	0.899***	0.087***
	(0.104)	(0.004)	(0.011)	(0.256)	(0.059)	(0.006)
Italy	7.449***	0.019*	0.182***	0.650***	0.432***	0.057***
	(0.051)	(0.010)	(0.042)	(0.189)	(0.140)	(0.004)
Japan	7.015***	0.011	0.077***	3.576*	0.715***	0.080***
	(0.039)	(0.009)	(0.010)	(1.905)	(0.079)	(0.006)
Netherlands	8.188***	0.017	0.077***	1.523***	0.605***	0.072***
	(0.062)	(0.011)	(0.023)	(0.535)	(0.109)	(0.005)
New Zealand	8.374***	0.013***	0.048***	0.651	0.582**	0.049***
	(0.049)	(0.003)	(0.005)	(0.522)	(0.267)	(0.003)
Norway	7.476***	0.021***	0.055***	1.625	0.404	0.039***
	(0.047)	(0.003)	(0.017)	(1.839)	(0.325)	(0.003)
Sweden	7.855***	0.011**	0.067***	1.766***	0.151	0.029***
	(0.014)	(0.005)	(0.011)	(0.497)	(0.139)	(0.002)
Switzerland	8.169***	0.021***	0.215***	0.524***	0.087	0.036***
	(0.028)	(0.006)	(0.051)	(0.155)	(0.151)	(0.003)
UK	8.440***	0.014***	0.417***	0.147***	0.369**	0.030***
	(0.028)	(0.005)	(0.042)	(0.024)	(0.182)	(0.002)
USA	8.291***	0.020**	0.640	0.213	0.126	0.055***
	(0.051)	(0.007)	(0.887)	(0.312)	(0.257)	(0.004)

*** Significant at the 1% level, ** at the 5% level, * at the 10% level.

which corresponds to the vision that in the long run an economy grows at a constant rate. However, in line with our findings in Chapter 2, this trend is found to be too restrictive to describe long-run growth. As a result, more flexible nonlinear deterministic trends have been proposed. For example, Greenaway *et al.* (2000) considered deterministic trends with an LSTAR component, as considered by Teräsvirta (1994). Although (nonlinear) deterministic trends might be flexible enough to describe long-run growth within the time span of the data under consideration, an obvious drawback of these trends is that they do not anticipate to any structural changes in future growth. Therefore it may be necessary to allow the slope of the trend to change stochastically over time. We concluded that the random walk with drift is inappropriate for this purpose, as it tends to overreact to short-run fluctuations. A way to overcome this problem is to restrict the number and/or magnitude of changes in the direction of the trend allowed. For example, the Markov trend, proposed by Hamilton (1989), allows for only two possible directions of the trend. This trend is found to be particularly useful to describe alternating periods of expansion and recession. However, to describe long-run economic growth a trend that incorporates more than two possible directions is needed.

In Section 3.4 we proposed a trend function that allows the slope of the trend to vary within a certain closed interval. The slope of the trend depends on the standardized difference between the level of the series and the trend in the previous year, referred to as the standardized trend deviation. In case the trend was higher than the level of the series, we expect the slope of the trend to be adjusted downwards, and vice versa. As the standardized trend deviations may take any value on the open interval $(-\infty, \infty)$, a logistic transition function is used to map these values to the closed interval $[-1, 1]$. An important property of our trend specification is that it embeds the linear deterministic trend and the random walk with drift as limiting cases. Completing the specification of our model, which is referred to as the STAAR model, we assume that the cycle can be described as an ARMA process, where the error term is independently and identically normally distributed with zero mean and constant variance. The model parameters were estimated using Maximum Likelihood. Regrettably, for 12 of our 17 series, the estimated trends act like a random walk, as the model suffers from overfitting the trend component. In these cases, it is likely that the parameter estimates are highly influenced by shocks, due to, for example, the Second World War and relatively high volatility in the first half of the century. The latter indicates that a careful treatment of changes in volatility is needed. For the remainder of the series, the estimated trends appear to be substantially better. However, for some of these, the parameter that governs the nonlinear adjustment mechanism appears to be insignificant. This suggests that, in order to accurately estimate the parameters of the adjustment mechanism in its desired role, more observations are needed.

Chapter 4

Latent Class Panel Data Models

4.1 Introduction

In the previous chapter we concluded that it is very difficult to detect a trend from a single time series that satisfies the desired characteristics of a long-run economic growth path. Perhaps the most important reason for this difficulty is that the sample size of our individual time series is relatively small. To overcome this problem, we will try to exploit the similarities in long-run economic growth across countries, which we established in Chapter 2. These similarities suggest that some groups of our time series belong to a single population. By treating several time series in one model, we obtain more accurate estimates of the parameters which describe trends of time series that belong to this population. These trends may better represent the long-run economic growth path of the time series under consideration than the trends obtained using univariate models. Obviously, the quality of the results obtained depends critically on the level of similarity across these time series. This issue immediately raises another question, which is how to classify the time series in groups that supposedly belong to the same population. Ideally, the data should tell us which time series belong to which groups, and, moreover, how many groups there are. Multivariate models well suited for this type of problem are called latent class panel data models. This chapter covers the theoretical material needed to understand and estimate these models.

In order to make this thesis self-explanatory, we provide a brief introduction to panel data modelling in the next two sections of the chapter. Firstly, panel data sets are defined and the main advantages of analyzing panel data sets instead of aggregated data are discussed. Apart from the functional form, an important step in the specification of a panel data model is to impose a structure on the parameters. In section 4.3, a modelling framework is provided that may facilitate this step. As we focus on latent class panel data models, we discuss the estimation of these models in more detail in Section 4.4. Finally, some practical considerations in latent class modelling are discussed, such as convergence and model selection. Section 4.6 concludes the chapter.

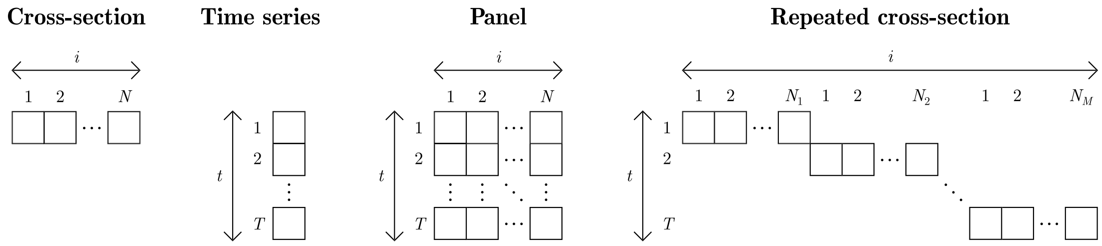


Figure 4.1: Four different types of data sets

4.2 The Analysis of Panel Data Sets

The data sets of the levels and growth rates of real per capita GDP, studied in this thesis, are two typical examples of panel data sets. Panel data sets contain a unit, or cross-sectional dimension, and a time dimension. Needless to say, in our case, these units are countries. Formally, a panel data set, also called a longitudinal data set, is defined as a data set that contains observations of a fixed set of units over a certain period of time. In most panels that are used in microeconomics and in the field of marketing, the units are individuals or households whereas in finance usually firms, stocks or stock indices are considered. In macroeconomics, one focuses on countries or industry sectors. Sometimes these units are *a priori* one of a kind, such as in our case of countries. In other situations, the set of units observed can be considered as a random sample from an underlying population. In that case, one is not interested in the characteristics of a specific unit, but rather in the common characteristics of arbitrary units. Note that a panel data set contains repeated observations on the same units. This is in contrast to a repeated cross-section, where, at each point in time, a new random sample from the population is drawn (Figure 4.1).

The advantage of analyzing panel data sets is that the researcher is able to study the characteristics of units over time, incorporating the possible presence of individual characteristics and/or common characteristics within subgroups of the units. If one would aggregate the data over the cross-sectional dimension, it is very often uncertain to which extent the common characteristics are preserved. Furthermore, the individual characteristics might bias the common characteristics one is mainly interested in.

In the 1980s and 1990s, most panel data models were developed for the analysis of data sets where either the cross-sectional population size, or the number of time periods is small. For the latter class of data sets, especially vector autoregressions were popular. Due to limited computer capacity, larger data sets could not be analyzed before aggregation over the cross-sectional dimension, or at least over subgroups of the units. However, as information technology has developed very quickly in recent decades, there is no need to aggregate over units anymore. Moreover, very large panel data sets have become available, especially in finance, marketing and labor economics. To analyze panels where both the population size and the number of time periods are large, new models need to be developed (Granger, 1998, Franses, 2002).

4.3 A Panel Data Modelling Framework

Consider a panel data set $y_{i,t}$, where $i = 1, ..., N$ is the cross-sectional dimension and $t = 1, ..., T$ the time dimension. Suppose that $y_{i,t}$ can be described by $k = 1, ..., K$ explanatory variables $x^k_{i,t}$. These explanatory variables can be constant over time, that is $x^k_{i,t} = x^k_i$ for all t, or, for example, include lagged values of the dependent variable $y_{i,t}$, in case of a dynamic model. In general terms, a descriptive model for $y_{i,t}$ can be formulated as:

$$y_{i,t} = h(x^k_{i,t}, \theta^k_{i,t}, \varepsilon_{i,t}), \tag{4.1}$$

where $h(\cdot)$ is a possibly nonlinear function, $\theta^k_{i,t}$ are the model parameters and $\varepsilon_{i,t}$ is an error term. Usually, this term is assumed to be independent and identically distributed over the cross-sectional and time dimension, with zero mean and variance σ^2_i. A simple example of such a model is the linear panel data model with a single explanatory variable and an additive error term, written as:

$$y_{i,t} = x_{i,t}\beta_{i,t} + \varepsilon_{i,t}. \tag{4.2}$$

The researcher has to make two important steps in the specification of a panel data model. Firstly, he or she should specify an appropriate functional form for $h(\cdot)$, that incorporates the features of the time series under consideration, such as trends, seasonality, outliers and nonlinearity. Secondly, he or she has to impose a structure on each of the k groups of parameters $\theta^k_{i,t}$. This structure may be different for each group, as the nature of the corresponding explanatory variables $x^k_{i,t}$ usually differs. Different parametric structures require different techniques to estimate these parameters. As a result, models are usually classified by their parametric structure.

In the remainder of this section we review several structures that are widely used in practice. Let us denote the parameters of a particular group k by $\theta_{i,t}$, for ease of notation. In most applications of panel data models, the parameters are assumed to be constant over time, that is $\theta_{i,t} = \theta_i$ for $i = 1, ..., N$. For ease of exposition, we restrict our attention to this class of models. Firstly, one could make one of two extreme assumptions:

- **Pooled parameters**

  The parameters θ_i are assumed to be constant over the cross-sectional dimension:

  $$\theta_i = \theta \qquad \text{for } i = 1, ..., N. \tag{4.3}$$

  In case this assumption is made for all K explanatory variables, another way to estimate the parameters consistently is to append the columns of the $y_{i,t}$ and $x^k_{i,t}$

matrices to obtain the vectors y_l and x_l^k, $l = 1, ..., NT$. Accounting for possible heteroscedasticity, one can now estimate the single equation:

$$y_l = h(x_l^k, \theta^k, \varepsilon_l). \tag{4.4}$$

For most choices of the function $h(\cdot)$, this equation can be estimated using conventional estimation techniques, such as Maximum Likelihood.

- **Fixed effects**

  The parameters θ_i are assumed to be different across the N time series. In case one assumes fixed effects for all K explanatory variables, and for the variance, one could have also estimated the N separate equations:

  $$\boldsymbol{y_i} = h(\boldsymbol{x_i^k}, \theta_i^k, \boldsymbol{\varepsilon_i}) \qquad \text{for } i = 1, ..., N, \tag{4.5}$$

  where $\boldsymbol{y_i}$, $\boldsymbol{x_i^k}$ and $\boldsymbol{\varepsilon_i}$ are $T \times 1$ vectors.

For some variables, these specifications are very useful. However, one can imagine that for others it is more convenient to choose a parametric structure that is in between these two extremes. In broad outlines, there are three possibilities:

- **Discrete random effects**

  Assume that the θ_i parameters can take J different values with probability π_j:

  $$\theta_i \in \{\alpha_1, ..., \alpha_J\}, \qquad \Pr[\theta_i = \alpha_j] = \pi_j \qquad \text{for } j = 1, ..., J, \tag{4.6}$$

  where $J < N$ and $\sum_{j=1}^{J} \pi_j = 1$. As the true classification of series is unobserved, models incorporating discrete random effects are also called latent class models. Another frequently used name is mixture models. The parameters can be estimated by ML, using the Expectation-Maximization (EM) algorithm by Dempster *et al.* (1977), or using the simulated EM (SEM) algorithm, see Nielsen (2000).

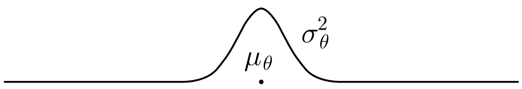
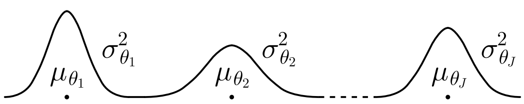
- **Continuous random effects**

  The parameters θ_i are assumed to be drawings from a continuous population distribution:

  $$\theta_i \sim IID(\mu_\theta, \sigma_\theta^2), \tag{4.7}$$

  where, in practice, the normal distribution is often used. This approach is also referred to as the Hierarchical Bayes approach. The parameters can be estimated by Simulated Maximum Likelihood (SML). If the function $h(\cdot)$ is linear, it is also convenient to use Bayesian techniques, such as the Gibbs sampler, see Gelfand *et al.* (1990).

Table 4.1: Overview of parametric structures, used in panel data modelling.

Type	Graphical representation	Number of parameters	Estimation method
Pooled parameters	θ	1	ML
Fixed effects	$\theta_1 \quad \theta_2 \quad \theta_3 \quad \cdots \quad \theta_N$	N	ML
Discrete random effects	$\theta_1 \quad \theta_2 \quad \cdots \quad \theta_J$	J $(J < N)$	(S)EM
Continuous random effects		2	(S)ML or Bayesian techniques
Discrete continuous random effects		$2J$	SEM or Bayesian techniques

- **Discrete continuous random effects**

  Basically, the discrete continuous random effects specification is a combination of the latter two specifications. Similar to the discrete random effects specification, it is assumed that the θ_i parameters can take J different values with probability π_j as given in (4.6). These parameters θ_j are in turn drawings from a continuous population distribution:

  $$\theta_j \sim IID(\mu_{\theta_j}, \sigma^2_{\theta_j}) \tag{4.8}$$

  Parameter estimates can be obtained by the simulated EM algorithm or Bayesian techniques.

The five parametric structures are summarized in Table 4.1. Recently, a number of extensions have been proposed, some of which are discussed in Franses (2002). In general, a thorough knowledge of the data underlying a specific problem is crucial for specifying a proper parametric structure. Given the strong similarities in long-run growth across groups of countries, it is evident that latent class panel data models might be appropriate for our problem. Therefore, the estimation of these models is discussed in more detail in the next section. Extensive discussions on panel data models can be found in recent textbooks by Baltagi (2001), Arellano (2002) and Hsiao (2003), among others.

4.4 Estimation of Latent Class Panel Data Models

Before we discuss the basic techniques that are used to estimate the parameters of the latent class panel data model, we first have to introduce some notation for finite mixtures and mixture distributions.

Mixtures and mixture distributions

For ease of notation, let us denote a panel data set $y_{i,t}$ by $\boldsymbol{y}$, where $\boldsymbol{y} = \{\{y_{i,t}\}_{i=1}^{N}\}_{t=1}^{T}$. The data vector of a particular unit i will be denoted by $\boldsymbol{y_i}$, where $\{y_{i,t}\}_{t=1}^{T}$. Suppose that the N units of this data set arise from a population that is a mixture of J clusters, $J < N$, where, for the moment, we assume J is known *a priori*. Each unit is classified to one of these clusters. However, it is unknown to which cluster a unit is classified. The unobserved classification of unit i is represented by s_i, $s_i \in \{1, ..., J\}$. Furthermore, the *ex ante* probability that a unit belongs to cluster j is specified as:

$$\Pr[s_i = j] = \pi_j, \qquad \sum_{j=1}^{J} \pi_j = 1, \qquad \pi_j \geq 0, \qquad i = 1, ..., N. \tag{4.9}$$

Given that unit i belongs to cluster j, the conditional distribution function of $\boldsymbol{y_i}$ is defined as $f_j(\boldsymbol{y_i}|\boldsymbol{\vartheta_j})$. The vector $\boldsymbol{\vartheta_j}$ contains all the unknown parameters associated with the distribution function $f_j(\cdot)$. The unconditional distribution of $\boldsymbol{y_i}$ is now obtained as:

$$f(\boldsymbol{y_i}|\boldsymbol{\theta}) = \sum_{j=1}^{J} \pi_j f_j(\boldsymbol{y_i}|\boldsymbol{\vartheta_j}), \tag{4.10}$$

with $\boldsymbol{\theta} = (\boldsymbol{\pi}, \boldsymbol{\vartheta})'$, the vector of all unknown parameters, $\boldsymbol{\pi} = \{\pi_j\}_{j=1}^{J}$ and $\boldsymbol{\vartheta} = \{\boldsymbol{\vartheta_j}\}_{j=1}^{J}$. It lies upon the surface to estimate the parameters $\boldsymbol{\theta}$ using Maximum Likelihood. In this context, the likelihood function of $\boldsymbol{\theta}$ is defined as the product over the densities of the N vectors $\boldsymbol{y}_i$, which are assumed to be independent. It is formulated as:

$$L(\boldsymbol{y}|\boldsymbol{\theta}) = \prod_{i=1}^{N} f(\boldsymbol{y_i}|\boldsymbol{\theta}). \tag{4.11}$$

The maximum likelihood estimator of $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}}_{\boldsymbol{ML}}$, is obtained by maximizing the likelihood function with respect to the restrictions in (4.9). Once the estimator $\hat{\boldsymbol{\theta}}_{\boldsymbol{ML}}$ is obtained, for each vector $\boldsymbol{y_i}$ the *ex post* probability that unit i belongs to cluster j after observing the data, denoted $\hat{p}_{i,j}$, can be calculated using Bayes' theorem:

$$\hat{p}_{i,j} = \frac{\hat{\pi}_j f_j(\boldsymbol{y_i}|\hat{\boldsymbol{\vartheta}}_{\boldsymbol{j}})}{\sum_{j=1}^{J} \hat{\pi}_j f_j(\boldsymbol{y_i}|\hat{\boldsymbol{\vartheta}}_{\boldsymbol{j}})}. \tag{4.12}$$

Subsequently, an estimate of the classification of unit i, denoted $\hat{s}_i$, can be computed as:

$$\hat{s}_i = \arg\max_j \hat{p}_{i,j}. \tag{4.13}$$

Ideally, for each unit, the highest *ex post* probability, $\hat{p}_{i,\hat{s}_i}$, is close to one. In that case, unit i almost certainly belongs to cluster $\hat{s}_i$.

Maximization of the likelihood function using the EM algorithm

For computational purposes, one usually maximizes the logarithm of the likelihood function, instead of the likelihood function itself. As this is a monotone transformation, the maxima are obtained for the same parameter values. The log-likelihood function of a latent class model can be maximized directly, using numerical methods, such as the Newton-Raphson method. However, in practice, these methods are often computationally unattractive, as the log-likelihood function is relatively complicated. In that case, one usually resolves to the Expectation-Maximization (EM) algorithm. The EM algorithm was initially formulated by Dempster, Laird and Rubin (1977). It was presented as a technique to compute maximum likelihood estimators from incomplete data problems. 'Incomplete' means that the data is partially unobserved. In the context of latent class models, the observed part of the data is $\boldsymbol{y}$, whereas the unobserved part is the vector $\boldsymbol{s}$, where $\boldsymbol{s} = \{\boldsymbol{s}_i\}_{i=1}^N$ (Laird, 1978).

Assume that the vector $\boldsymbol{s}$, which represents the classification of the units, is known. One could now facilitate this information by maximizing the complete data likelihood function, defined as:

$$L_c(\boldsymbol{y}, \boldsymbol{s}|\boldsymbol{\theta}) = \prod_{i=1}^{N}\prod_{j=1}^{J}\left(\pi_j f_j(\boldsymbol{y_i}|\boldsymbol{\vartheta_j})\right)^{I[s_i=j]}, \tag{4.14}$$

instead of the normal likelihood function, where $I[\cdot]$ denotes the indicator function, whose value is one if the argument in brackets is true and zero otherwise. The complete data log-likelihood function is written as:

$$\ln L_c(\boldsymbol{y}, \boldsymbol{s}|\boldsymbol{\theta}) = \sum_{i=1}^{N}\sum_{j=1}^{J}\left(I[s_i = j]\ln \pi_j + I[s_i = j]\ln f_j(\boldsymbol{y_i}|\boldsymbol{\vartheta_j})\right). \tag{4.15}$$

In the first step for the algorithm, the 'E-step', the expectation is taken of the complete data log-likelihood function with respect to $\boldsymbol{s}|\boldsymbol{y}$. It can be shown that this expectation can be obtained by replacing the indicators $I[s_i = j]$ by the *ex post* probabilities $p_{i,j}$:

$$\begin{aligned} \ln \mathcal{L}_c(\boldsymbol{y}|\boldsymbol{\theta}) &\equiv E_{\boldsymbol{s}|\boldsymbol{y}}[\ln L_c(\boldsymbol{y}, \boldsymbol{s}|\boldsymbol{\theta})] \\ &= \underbrace{\sum_{i=1}^{N}\sum_{j=1}^{J} p_{i,j}\ln \pi_j}_{\text{(a)}} + \underbrace{\sum_{i=1}^{N}\sum_{j=1}^{J} p_{i,j}\ln f_j(\boldsymbol{y_i}|\boldsymbol{\vartheta_j})}_{\text{(b)}} \end{aligned} \tag{4.16}$$

In the second step, the 'M-step', this expectation is maximized with respect to the parameters $\boldsymbol{\theta}$. As the parameters π_j appear only in part (a) of (4.16), whereas the parameters

$\boldsymbol{\vartheta_j}$ appear only in (b), we can maximize two parts of this equation separately:

$$\max_{\pi_j} \quad \sum_{i=1}^{N}\sum_{j=1}^{J} p_{i,j} \ln \pi_j \qquad \text{for } j = 1, ..., J \tag{4.17}$$

$$\max_{\boldsymbol{\vartheta_j}} \quad \sum_{i=1}^{N}\sum_{j=1}^{J} p_{i,j} \ln f_j(\boldsymbol{y_i}|\boldsymbol{\vartheta_j}) \qquad \text{for } j = 1, ..., J \tag{4.18}$$

The maximization problem in (4.17) can be solved analytically. It can be written as:

$$\max_{\pi_j} \quad \sum_{j=1}^{J-1} \ln \pi_j \sum_{i=1}^{N} p_{i,j} + \ln\left(1 - \sum_{l=1}^{J-1} \pi_l\right) \sum_{i=1}^{N} p_{i,j} \qquad \text{for } j = 1, ..., J-1 \tag{4.19}$$

where the restrictions on the *ex ante* probabilities are used, which are given in (4.9). The first order conditions of this problem are written as:

$$\frac{\partial \ln \mathcal{L}_c(\boldsymbol{y}|\boldsymbol{\theta})}{\partial \pi_j} = \frac{\sum_{i=1}^{N} p_{i,j}}{\pi_j} - \frac{\sum_{i=1}^{N} p_{i,J}}{1 - \sum_{l=1}^{J-1} \pi_l} = 0 \qquad \text{for } j = 1, ..., J-1 \tag{4.20}$$

It is now easily verified that the solution to the latter system of equations is given by:

$$\pi_j = \frac{1}{N} \sum_{i=1}^{N} p_{i,j} \qquad \text{for } j = 1, ..., J-1 \tag{4.21}$$

Whether a solution to the problem given by (4.18) exist in closed form depends on the shape of the distribution functions $f_j(\cdot)$. The first order conditions are written as:

$$\frac{\partial \ln \mathcal{L}_c(\boldsymbol{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\vartheta_j}} = \sum_{i=1}^{N} p_{i,j} \frac{\partial \ln f_j(\boldsymbol{y_i}|\boldsymbol{\vartheta_j})}{\partial \boldsymbol{\vartheta_j}} = 0 \qquad \text{for } j = 1, ..., J \tag{4.22}$$

In case it is not possible to obtain closed-form expressions, numerical optimization techniques, such as the Newton-Raphson method, can be used.

The E-step and the M-step are repeated until convergence has been achieved. As a convergence criterion, one may use the (relative) change in the value of the log-likelihood function or the (relative) change in the parameter values. Schematically, the resulting estimation algorithm can be summarized as follows:

1. Fix the number of clusters J and generate starting values for the cluster-specific parameters $\boldsymbol{\vartheta_j}$ and the *ex ante* probabilities π_j, for $j = 1, ..., J$[1].

2. Given the current estimates of π_j and $\boldsymbol{\vartheta_j}$, compute the *ex post* probabilities $p_{i,j}$ using (4.12).

[1] As starting values for the *ex ante* probabilities, π_j, it is convenient to use random numbers on the unit-interval, satisfying $\sum_{j=1}^{J} \pi_j = 1$.

3. Given the current estimates of $p_{i,j}$, update the estimates of $\boldsymbol{\vartheta_j}$, for $j = 1, ..., J$.

4. Compute the *ex ante* probabilities, π_j, for $j = 1, ..., J$, using (4.21).

5. Compute the value of the log-likelihood function, $\ln L(\boldsymbol{y}|\boldsymbol{\theta})$, as given in (4.11).

6. Stop if convergence has been achieved, otherwise repeat steps 2-5.

The parameter estimates obtained have the same asymptotic properties as Maximum Likelihood estimates. Standard errors can be computed as discussed in Section 3.5.

Convergence of the EM algorithm

In contrast to the Newton-Raphson method, the EM algorithm always converges to a local maximum of the log-likelihood function if the value of the log-likelihood function increases in each iteration. This is true regardless of the shape of this function or the accuracy of the initial values for $\boldsymbol{\vartheta_j}$ and π_j (Titterington *et al.*, 1985; McLachlan and Basford, 1988). Although the algorithm may require many iterations, these iterations are computationally very attractive compared to Newton-Raphson iterations, as small parts of the likelihood function are maximized separately. Probably the most convenient way to estimate the parameters of a mixture panel data model is therefore to combine both techniques. Using random starting values, the EM algorithm can be used to find a solution near a local maximum of the likelihood function. After a fixed number of iterations, the parameter estimates of the solution with the maximum value of the likelihood function can be used as the starting values of the Newton-Raphson method, which will probably converge very quickly to the exact local maximum. To verify whether this procedure yields the global maximum of a particular problem, one could carry out a large number of runs, starting from a wide range of random starting values. It can be shown that if the number of runs tends to infinity, then the probability that the global maximum is found converges to one.

4.5 Model Selection

Using a latent class panel data model as defined in Section 4.3, we obtain an endogenous classification of our time series. However, the number of clusters has to be identified *a priori*. Within the classical statistical framework, this problem is still without a satisfactory solution[2] (Wedel and Kamakura, 1999). Let us discuss the main difficulty of the problem. Suppose one wishes to test the null hypothesis of J clusters against the alternative hypothesis of $J+1$ clusters. One may be inclined to use the likelihood ratio statistic for this test, which is simply the difference between the maximized values of the log likelihood of the two models. However, this test statistic is not asymptotically χ^2 distributed, as two of the regularity condition of the likelihood ratio test are violated. Firstly, the cluster

[2]See, however, Gourieroux *et al.*, 1982, for some first promising results.

specific parameters that correspond to cluster $J+1$ are unidentified under the null hypothesis. This causes the likelihood function to be flat with respect to these parameters at the optimum. As a result, the likelihood surface is not locally quadratic in this region. Secondly, the null hypothesis yields a local optimum of the likelihood function. Therefore, under the null hypothesis, the gradient vector of the likelihood function is identically zero (See Davies, 1977, McLachlan and Basford, 1988, and Hansen, 1992, among others).

Alternatively, one could apply bootstrapping methods. However, these methods are computationally very demanding, as the model under consideration has to be estimated for many synthetic data sets. Therefore, in practice one usually relies on information criteria. In the context of latent class models, information criteria strike a balance between the increase in fit obtained by imposing an extra cluster against the additional number of parameters required. The various information criteria differ in how to strike this balance. Commonly used criteria are the Akaike Information Criterion (AIC), which is based on the use of generalized entropy as a goodness-of-fit measure, defined as:

$$\text{AIC} = -\frac{2\ln L(\boldsymbol{y}|\boldsymbol{\theta})}{NT} + 2\frac{P}{NT}, \tag{4.23}$$

and the Schwartz/Bayesian Information Criterion (BIC), an asymptotic approximation to the Bayes factor:

$$\text{BIC} = -\frac{2\ln L(\boldsymbol{y}|\boldsymbol{\theta})}{NT} + \ln(NT)\frac{P}{NT}, \tag{4.24}$$

where P is the total number of parameters to be estimated. The BIC differs from the AIC only in that the number of parameters P is multiplied by $\ln(NT)$ instead of 2. Clearly, the BIC penalizes additional parameters more if the number of observations is larger than eight, which is usually the case. Due to the relatively severe penalty, a decision based on the BIC may result in too few clusters, whereas, in practice, the AIC sometimes tends to overestimate the number of clusters. In general, however, the performance of information criteria in the context of latent class models is unknown. In a recent study, Andrews and Currim (2003) compared the performance of seven information criteria to retain the number of clusters for a finite mixture logit model, manipulating many data characteristics. For this model, they found that the AIC, and especially a modified version of this criterion, performed best. For the modified AIC, referred to as AIC3, the penalty factor is three instead of two[3]:

$$\text{AIC3} = -\frac{2\ln L(\boldsymbol{y}|\boldsymbol{\theta})}{NT} + 3\frac{P}{NT}. \tag{4.25}$$

Their results, however, are not completely consistent with the results of previous studies, where different mixture models were used (See, for instance, Culter and Windham, 1994).

[3]The factor three shows up if one assumes the likelihood ratio test statistic to be asymptotically distributed as a noncentral χ^2 random variable with non-centrality parameter δ and $2(P_1 - P_0)$ degrees of freedom, where P_0 and P_1 denote the number of parameters estimated under the null and the alternative hypothesis, respectively (Bozdogan, 1994).

As a consequence, it is likely that no one information criterion is best for all types of models in all situations. To decide upon the number of clusters, it therefore appears to be wise to consider several criteria, and establish which criterion produces the most plausible results in each particular case.

4.6 Conclusions

Our data sets of real per capita GDP levels and growth rates are typical examples of panel data sets, as these data sets contain a unit, or cross-sectional dimension and a time dimension. In this chapter, the theoretical fundamentals have been discussed to model panel data sets. To specify a panel data model, one has to make two important decisions. Firstly, one should decide upon the functional form of the model, that incorporates the features of the time series under consideration. Secondly, one has to impose an appropriate structure on the parameters of the model. To illustrate this, for some parameters it is plausible to assume constancy over the cross-sectional dimension. Others, however, should differ between groups, or perhaps all, of the units. We have provided an overview of five parametric structures, which are widely used in practice.

A structure that seems relevant for the models we want to develop for our data sets is the discrete random effects structure. Models based on this parametric structure are called latent class models, or mixture models. Using discrete random effects, one assumes that the units, in our case countries, can be classified into J different clusters. Within each cluster of countries, one or more model parameters take the same value. These parameters might, for example, represent the average economic growth rate. To estimate the parameters of a latent class panel data model by Maximum Likelihood it is convenient to use the EM algorithm of Dempster *et al.* (1977). This method is computationally more attractive than the Newton-Raphson method, as small parts of the likelihood function are maximized separately. However, one of its drawbacks is that it converges very slowly to a maximum of the likelihood function, once it is near this solution. Therefore, one usually applies the Newton-Raphson method after a large fixed number of iterations of the EM algorithm. Apart from convergence, another important issue in latent class modelling is the decision upon the number of clusters. In practice, one usually relies on the verdict of information criteria. A problem with this approach is that it is likely that no one information criterion is best for all types of latent class models in all situations. Though strictly speaking this is subjective, one could therefore consider several criteria, and establish which criterium produces the most plausible results for the model and data set at hand.

Chapter 5

Multivariate Subperiod Analysis Using Static Models

5.1 Introduction

To become familiar with the estimation, model selection and evaluation of latent class panel data models in practice, we first consider two very basic static models and estimate their parameters for our panel of growth rates. The models are formulated such as to formalize our subperiod analysis of Section 2.3. It is interesting to compare the results of this approach to the analysis of correlations using MDS. For the models used in this chapter, we will establish that is it possible to solve the maximization problems in the M-step of the EM algorithm analytically. Consequently, estimation of these models is computationally very attractive.

In the next section, a model is introduced to describe clusters of countries having equal average growth rates and growth rate volatility. The model is estimated separately for the five subperiods, introduced in Section 2.3. As a consequence, the clustering obtained may be different for each subperiod. In contrast to this, Section 5.3 proceeds with discussing a model that can be used to obtain a constant clustering over time, by incorporating shifts in the average growth rate and growth rate volatility. Section 5.4 concludes the chapter.

5.2 A Model for Constant Growth and Growth Rate Volatility

Let us first assume that for each country i in our panel data set, the growth rates $\Delta y_{i,t}$, for $t = 1, ..., T$, are drawings from a normal distribution with mean μ_i and variance σ_i^2. These parameters are to be interpreted as the average growth rate and the growth rate volatility. Furthermore, assume that the countries can be classified into J clusters. For all countries classified to cluster $j \in J$, the mean and variance, denoted by μ_j and σ_j^2, respectively, are the same. Note that these means and variances are assumed to be constant over time.

The corresponding latent class panel data model is given by:

$$\begin{aligned} \Delta y_{i,t} = \mu_{s_i} + \varepsilon_{i,t}, \qquad & \varepsilon_{i,t} \sim N(0, \sigma^2_{s_i}), \\ \Pr[s_i = j] = \pi_j, \qquad \textstyle\sum_{j=1}^{J} \pi_j = 1, \qquad & \pi_j \geq 0. \end{aligned} \tag{5.1}$$

Given that country i belongs to cluster j, the conditional distribution of a series $\boldsymbol{\Delta y_i}$ is defined as:

$$f_j(\boldsymbol{\Delta y_i}|\mu_j, \sigma_j^2) = \prod_{t=1}^{T} \phi(\Delta y_{i,t}|\mu_j, \sigma_j^2), \tag{5.2}$$

where $\phi(\cdot|\mu, \sigma^2)$ denotes the probability density function of a normal random variable with mean μ and variance σ^2. The likelihood function can now be written as:

$$L(\boldsymbol{\Delta y}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\sigma^2}) = \prod_{i=1}^{N} \sum_{j=1}^{J} \pi_j f_j(\boldsymbol{\Delta y_i}|\mu_j, \sigma_j^2) = \prod_{i=1}^{N} \sum_{j=1}^{J} \pi_j \prod_{t=1}^{T} \phi(\Delta y_{i,t}|\mu_j, \sigma_j^2). \tag{5.3}$$

Estimation

The parameters of the model can be estimated using the EM algorithm. In Appendix A.6 it is verified that the *ex post* probabilities can be computed as:

$$p_{i,j} \equiv \Pr[s_i = j|\boldsymbol{\Delta y}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\sigma^2}] = \frac{\pi_j \prod_{t=1}^{T} \phi(\Delta y_{i,t}|\mu_j, \sigma_j^2)}{\sum_{l=1}^{J} \pi_l \prod_{t=1}^{T} \phi(\Delta y_{i,t}|\mu_l, \sigma_l^2)}. \tag{5.4}$$

For this model, it is also possible to derive closed-form expressions for the distribution-specific parameters μ_j and σ_j^2:

$$\mu_j = \frac{\sum_{i=1}^{N} \left(p_{i,j} \sum_{t=1}^{T} \Delta y_{i,t}\right)}{T \sum_{i=1}^{N} p_{i,j}} \qquad \text{for } j = 1, ..., J, \tag{5.5}$$

$$\sigma_j^2 = \frac{\sum_{i=1}^{N} \left(p_{i,j} \sum_{t=1}^{T} (\Delta y_{i,t} - \mu_j)^2\right)}{T \sum_{i=1}^{N} p_{i,j}} \qquad \text{for } j = 1, ..., J. \tag{5.6}$$

Given the data and the current estimates of $p_{i,j}$, these expressions will be used to update the estimates of μ_j and σ_j^2. Using the complete algorithm specified in Section 4.4, parameter estimation is now straightforward.

Results

The data used to estimate the parameters of this model could span the entire 20th century. However, it is very implausible to assume that growth rates and growth rate volatility were constant over such a long period of time. We will therefore estimate the model separately for five subperiods of the century. These subperiods, corresponding to different phases of economic development, were introduced in Section 2.3. Obviously, the assumption of parameter constancy is still questionable, but the results of this model are a good starting point for a more elaborate analysis.

Table 5.1: Selecting the number of clusters for model (5.1).

Subperiod	Criterion	Number of clusters J					
		2	3	4	5	6	Selected
1901-1918	AIC	-3.116	-3.106	-3.089	-3.070	-3.051	2
	AIC3	-3.096	-3.076	-3.050	-3.021	-2.992	2
	BIC	-3.059	-3.012	-2.959	-2.903	-2.847	2
1919-1939	AIC	-2.773	-2.765	-2.751	-2.735	-2.718	2
	AIC3	-2.756	-2.740	-2.718	-2.693	-2.668	2
	BIC	-2.722	-2.681	-2.635	-2.586	-2.536	2
1940-1945	AIC	-1.300	-1.311	-1.277	-1.226	-1.167	3
	AIC3	-1.241	-1.223	-1.159	-1.079	-0.990	2
	BIC	-1.181	-1.115	-1.003	-0.875	-0.739	2
1946-1973	AIC	-3.395	-3.465	-3.488	-3.505	-3.504	5
	AIC3	-3.382	-3.446	-3.463	-3.474	-3.467	5
	BIC	-3.353	-3.397	-3.394	-3.385	-3.358	3
1974-2000	AIC	-4.998	-4.992	-4.982	-4.969	-4.956	2
	AIC3	-4.985	-4.973	-4.956	-4.936	-4.917	2
	BIC	-4.955	-4.922	-4.885	-4.845	-4.805	2

To verify that the EM algorithm has converged to a maximum of the likelihood function, the parameter estimates obtained were used as the starting values for the Newton-Raphson method, which is run after 200 iterations of the EM algorithm. As a convergence criterion for the Newton-Raphson method, the change in the value of the log-likelihood function is used, which should be smaller than 10^{-4}. Each country i is assigned to the cluster j for which the *ex post* probability $\hat{p}_{i,j}$ is highest. This procedure is repeated 1000 times, starting from random starting values. It turned out that for $J = 2$, 80% of the solutions obtained were identical, whereas for $J = 5$ this was approximately 5%. We should therefore treat the estimation results with some care in case the number of clusters is large.

We base the decision upon the number of clusters on the values of information criteria, as discussed in Section 4.5. These values are shown in Table 5.1. In the right column, the number of clusters with the minimum value of the criterion is reported. For each subperiod, the AIC3 favours the same number of clusters as the AIC, except for the period of the Second World War. Looking in more detail, we observe, however, that the verdict of the AIC tends to be more unclear than the AIC3. Quite often, the values of AIC do not differ more than 0.01, for choices of J close to the number of clusters selected. In contrast to the AIC and AIC3, the BIC consistently selects only two clusters, due to its very high penalty on the number of parameters. Given the above consideration, we comply with the verdict of the AIC, giving it the benefit of the doubt as far as the period of the Second World War is concerned. The parameter estimates are shown in Table 5.2. For each subperiod, the clusters, represented by rectangles, are ordered vertically from high growth to low or even negative growth. Horizontally, the columns are ordered chronologically. Note that the standard deviations $\boldsymbol{\sigma}$ of the growth rates are shown instead of the variances $\boldsymbol{\sigma^2}$, for ease of comparison to Table 2.2.

For the period from 1901-1918, two groups can be distinguished, although only by a difference in standard deviation. A group of European countries, including the Netherlands, the UK and three Scandinavian countries, showed more stable growth than the other countries. While growth increases in the period from 1919-1939, this group stays together and is complemented by a few other countries. In the other group, characterized by many large economies, growth is substantially lower and more volatile. Obviously, this is mainly due to the 1929 stock market crash and the Great Depression. In the Second World War, the US and Canadian economies flourish as a result of the weapons industry. Most severely damaged were the Austrian and the Japanese economy. During the Golden Age the economies behaved very differently. This was also one of the conclusions of our MDS analysis in Section 2.4. Most economies that were severely damaged by the war, ended up in high growth clusters during the Golden Age. Especially Japan passed through a strong catch-up phase. Since this period, Germany and the Netherlands grew closely together due to their strong trading partnership. Remarkably, this group has very high volatility during the Golden Age. Finally, in the final quarter of the century, growth was very stable in most countries at a rate of two percent. In Finland, Sweden and New Zealand it was a bit lower and more volatile, but not exceptionally.

We end this section with two general remarks. Firstly, the standard errors are typically large. The size of the standard errors obviously depends on the length of the particular subperiod. This explains why the confidence intervals of the parameter estimates do not overlap for the two most recent subperiods, which are longer than the others, whereas they do for a short subperiod like the Second World War. Apart from that, the standard errors are relatively large for the first two periods, as there is more variation in the data of the first half of the century, as we concluded in Chapter 2. Secondly, one may argue that model solutions which imply one or more clusters to consist of a single country, should be excluded from further analysis. In that case, we should have estimated a model with four clusters for the period of the Golden Age, in which case Japan would have been classified to the current cluster of Austria, France and Italy. Apart from this matter, there is a significant difference, both in average growth and volatility, between Japan and the latter cluster.

Table 5.2: Parameter estimates of the *ex ante* probabilities $\boldsymbol{\pi}$, the growth rates $\boldsymbol{\mu}$ and the standard deviations $\boldsymbol{\sigma}$ for the model (5.1). The number of clusters J was selected using AIC. Estimated standard errors are in parenthesis.

1901-1918			1919-1939			1940-1945			1946-1973			1974-2000		
$\hat{\pi}_j$	$\hat{\mu}_j$	$\hat{\sigma}_j$	$\hat{\pi}_j$	$\hat{\mu}_j$	$\hat{\sigma}_j$	$\hat{\pi}_j$	$\hat{\mu}_j$	$\hat{\sigma}_j$	$\hat{\pi}_j$	$\hat{\mu}_j$	$\hat{\sigma}_j$	$\hat{\pi}_j$	$\hat{\mu}_j$	$\hat{\sigma}_j$
									JP					
									0.059	7.780	2.402			
									(0.057)	(0.454)	(0.621)			
AU, AT, BE, CA, FI, FR GE, IT, JP, NZ, SE, US			AU, BE, DM, FI, JP NL, NO, SE, SW, UK			AU, BE, CA, FI, NZ NO, SE, SW, UK, US			AT, FR, IT			AU, AT, BE, CA, DM FR, GE, IT, JP, NL NO, SE, UK, US		
0.713[1]	0.578	5.641	0.542[1]	2.412	4.869	0.581[1]	2.241	7.479	0.172	6.246	5.622	0.803[1]	1.961	1.798
	(0.387)	(0.644)		(0.422)	(0.737)		(1.118)	(1.528)	(0.093)	(0.623)	(0.932)		(0.095)	(0.131)
DM, NL, NO SW, UK			AT, CA, FR, GE IT, NZ, US			DM, FR, GE IT, NL			GE, NL			FI, NZ, SW		
0.287	0.539	3.414	0.458	1.671	7.218	0.287	-7.617	11.979	0.122	4.726	10.427	0.197	1.335	2.604
(0.146)	(0.393)	(0.797)	(0.184)	(0.622)	(1.134)	(0.138)	(2.774)	(3.445)	(0.081)	(1.380)	(2.108)	(0.106)	(0.291)	(0.426)
						AT, JP			AU, BE, CA, DM, FI NO, SE, SW, UK					
						0.131	-12.739	28.236	0.529[1]	3.223	2.678			
						(0.089)	(7.912)	(12.081)		(0.169)	(0.233)			
									NZ, US					
									0.118	1.678	5.505			
									(0.078)	(0.737)	(1.041)			

[1] $\hat{p}_J \equiv 1 - \sum_{j=1}^{J-1} \hat{p}_j$, where cluster J is the cluster with the largest number of countries.

5.3 A Model for Shifts in Growth and Growth Rate Volatility

In the previous section, we have studied growth in subperiods separately. The number of clusters and the classification varied over time. We observed however, that some countries were classified to the same cluster in more than one subperiod. Examples are Canada and the USA, and France and Italy. This indicates that groups of countries show common growth patterns for prolonged periods of time, perhaps even over the entire 20th century, which may be due to close geographical proximity, or, for example, because of historical (trading) partnerships. In this section we propose a model that allows for a different average growth rate and growth rate volatility in each of K predefined subperiods, while the cluster membership is assumed to be constant over time. The model that we consider is formulated as:

$$\begin{aligned} \Delta y_{i,t} &= \mu_{s_i,\kappa(t)} + \varepsilon_{i,t}, \qquad \varepsilon_{i,t} \sim N(0, \sigma^2_{s_i,\kappa(t)}), \\ \Pr[s_i = j] &= \pi_j, \qquad \textstyle\sum_{j=1}^J \pi_j = 1, \qquad \pi_j \geq 0, \end{aligned} \tag{5.7}$$

where the function $\kappa(\cdot)$ returns the index k of the subperiod the year t is linked to, with $k = 1, ..., K$:

$$\kappa(t) = \{k : T_{k-1} < t \leq T_k\}. \tag{5.8}$$

The parameter T_k denotes the final year of subperiod k, where $T_0 = 0$, $T_K = T$, and $T_{k-1} < T_k$, for all k, by definition. Given that country i belongs to cluster j, the conditional distribution of a series $\boldsymbol{\Delta y_i}$ is defined as:

$$f_j(\boldsymbol{\Delta y_i}|\mu_j, \sigma_j^2) = \prod_{t=1}^{T} \phi(\Delta y_{i,t}|\mu_{j,\kappa(t)}, \sigma^2_{j,\kappa(t)}). \tag{5.9}$$

Consequently, we can write the likelihood function as:

$$L(\boldsymbol{\Delta y}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\sigma^2}) = \prod_{i=1}^{N} \sum_{j=1}^{J} \pi_j \prod_{t=1}^{T} \phi(\Delta y_{i,t}|\mu_{j,\kappa(t)}, \sigma^2_{j,\kappa(t)}). \tag{5.10}$$

Estimation

Estimation of the parameters goes along exactly the same lines as for model (5.1). It is easily verified that the *ex post* probabilities are computed as:

$$p_{i,j} \equiv \Pr[s_i = j|\boldsymbol{\Delta y}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\sigma^2}] = \frac{\pi_j \prod_{t=1}^T \phi(\Delta y_{i,t}|\mu_{j,\kappa(t)}, \sigma^2_{j,\kappa(t)})}{\sum_{l=1}^J \pi_l \prod_{t=1}^T \phi(\Delta y_{i,t}|\mu_{l,\kappa(t)}, \sigma^2_{l,\kappa(t)})}. \tag{5.11}$$

Furthermore, $\mu_{j,k}$ and $\sigma^2_{j,k}$ are updated by:

$$\mu_{j,k} = \frac{\sum_{i=1}^N \left(p_{i,j} \sum_{t=T_{k-1}+1}^{T_k} \Delta y_{i,t}\right)}{(T_k - T_{k-1}) \sum_{i=1}^N p_{i,j}} \qquad \text{for } j = 1, ..., J,\ k = 1, ..., K, \tag{5.12}$$

$$\sigma^2_{j,k} = \frac{\sum_{i=1}^N \left(p_{i,j} \sum_{t=T_{k-1}+1}^{T_k} (\Delta y_{i,t} - \mu_{j,k})^2\right)}{(T_k - T_{k-1}) \sum_{i=1}^N p_{i,j}} \qquad \text{for } j = 1, ..., J,\ k = 1, ..., K. \tag{5.13}$$

Table 5.3: Selecting the subperiods and the number of clusters for model (5.7).

K	Subperiods	Criterion	Number of clusters J					
			2	3	4	5	6	Selected
3	1901-1939, 1940-1945,	AIC	-3.277	-3.298	-3.318	-3.327	-3.328	6
	1946-2000	AIC3	-3.269	-3.286	-3.302	-3.307	-3.303	5
		BIC	-3.236	-3.235	-3.232	-3.219	-3.197	2
4	1901-1939, 1940-1945,	AIC	-3.511	-3.536	-3.552	-3.561	-3.560	5
	1946-1973, 1974-2000	AIC3	-3.500	-3.520	-3.531	-3.535	-3.528	5
		BIC	-3.457	-3.453	-3.441	-3.421	-3.391	2
4	1901-1918, 1919-1939,	AIC	-3.286	-3.307	-3.328	-3.337	-3.336	5
	1940-1945, 1945-2000	AIC3	-3.275	-3.291	-3.307	-3.310	-3.304	5
		BIC	-3.232	-3.225	-3.217	-3.197	-3.167	2
5	1901-1918, 1919-1939,	AIC	-3.520	-3.543	-3.564	-3.572	-3.570	5
	1940-1945, 1945-1973,	AIC3	-3.507	-3.524	-3.538	-3.540	-3.531	5
	1974-2000	BIC	-3.453	-3.442	-3.427	-3.400	-3.362	2

Again, we use both the EM algorithm as specified in Section 4.4 and the Newton-Raphson method. The same convergence criteria are used.

Results

Similar to the previous section, we have estimated the model for different choices of the number of clusters. Moveover, for this model we varied the number of subperiods and the composition of these subperiods[1]. For each relevant combination, the AIC, AIC3 and BIC were computed, which are shown in Table 5.3. The mutual differences between the selections of the criteria are similar to our earlier findings. The BIC always favours the model with only two clusters, whereas the AIC3 and AIC tend to favour models with more clusters. Regrettably, only very small differences between the values of the criteria can be observed. Looking at the AIC, it is ambiguous whether five of six clusters are to be preferred. In contrast to the results for the previous model, the AIC3 does not provide more clarity. Combining the verdicts of both criteria, we slightly favour the model with five clusters and five subperiods to its equivalent with four subperiods, where the first two subperiods are combined. The estimation results are shown in Table 5.4.

In general, the clustering obtained is very natural. The first cluster consists of France and Germany, the engine of the European Union, complemented by the Netherlands. Growth in the first subperiod was particularly low, but it increased rapidly during the period before the Second World War. Despite average growth rates of -9% during this war, the group recovered very soon after. The second cluster consists of the countries that

[1]Models with more than five subperiods were also estimated. These estimates, however, proved to be very instable due to the large number of parameters.

Table 5.4: Parameter estimates of the *ex ante* probabilities $\boldsymbol{\pi}$, the growth rates $\boldsymbol{\mu}$ and the standard deviations $\boldsymbol{\sigma}$ for model (5.7) with $J = 5$ clusters and $K = 5$ subperiods. Estimated standard errors are in parenthesis.

			1901-1918		1919-1939		1940-1945		1946-1973		1974-2000	
j	Countries	$\hat{\pi}_j$	$\hat{\mu}_{j1}$	$\hat{\sigma}_{j1}$	$\hat{\mu}_{j2}$	$\hat{\sigma}_{j2}$	$\hat{\mu}_{j3}$	$\hat{\sigma}_{j3}$	$\hat{\mu}_{j4}$	$\hat{\sigma}_{j4}$	$\hat{\mu}_{j5}$	$\hat{\sigma}_{j5}$
1	FR, GE, NL	0.176	−0.471	5.587	2.798	7.487	−8.864	13.152	5.076	9.500	1.733	1.554
		(0.092)	(0.760)	(1.074)	(0.943)	(1.334)	(3.100)	(4.384)	(1.036)	(1.466)	(0.173)	(0.198)
2	AT, IT, JP	0.176	1.607	5.117	1.640	6.205	−12.269	24.668	6.880	4.263	2.176	1.704
		(0.092)	(0.696)	(0.983)	(0.782)	(1.105)	(5.814)	(8.222)	(0.465)	(0.656)	(0.189)	(0.233)
3	CA, NZ, US	0.176	1.630	5.612	0.783	7.248	5.844	7.410	1.972	4.811	1.541	2.272
		(0.092)	(0.764)	(1.079)	(0.913)	(1.291)	(1.747)	(2.469)	(0.525)	(0.741)	(0.252)	(0.342)
4	BE, FI, NO	0.177	−0.550	5.582	3.460	5.336	−1.524	6.390	4.013	2.533	2.367	2.270
		(0.093)	(0.760)	(1.076)	(0.672)	(0.952)	(1.507)	(2.134)	(0.277)	(0.382)	(0.253)	(0.342)
5	AU, DM, SE,	0.294[1]	0.600	3.712	1.838	4.163	1.398	7.573	2.880	2.610	1.558	1.955
	SW, UK		(0.392)	(0.551)	(0.407)	(0.574)	(1.384)	(1.958)	(0.221)	(0.306)	(0.169)	(0.220)

[1] $\hat{\pi}_5 \equiv 1 - \sum_{j=1}^{4} \hat{\pi}_j$.

bore the largest shocks during the Second World War, but passed through a very strong catch-up period after the war. Cluster 3 consists of the core of the Western Offshoots. Growth was typically low during the Great Depression and high during the Second World War. The growth patterns of the countries classified to the fourth cluster are similar to those of the first. The countries in this cluster, however, did not plunge into war, and ended up with very high growth after the Golden Age. Finally, the fifth cluster contains the most stable countries in terms of volatility. Growth in these countries was not harmed by the war, nor by any other events.

Note that the number of clusters selected for this model is relatively large compared to the number of clusters we selected for subperiods in our previous model. Comparing the results of the two models, we observe that the composition of the clusters obtained using model (5.7) is similar to that obtained for the Golden Age using model (5.1). Apparently, the divergence during the Golden Age has determined the classification of our new model to a large extent. Compared to model (5.1), the standard errors of model (5.7) are larger, as more parameters had to be estimated for the same number of observations. As for some periods the number of clusters is clearly too large, there is quite some overlap in the confidence intervals of the parameter estimates.

Comparing the time series analysis to the MDS analysis

We end this section with a comparison between the results of our time series analysis using latent class panel data models and our MDS analysis. Let us first discuss some important differences between both approaches. Firstly, using the latent class models discussed above, we forced each country to be classified unambiguously to a cluster. In contrast, our MDS analysis provided distances, or similarities, between countries, leaving the classification to the interpreter. Secondly, the classification of our latent class models is based on both average growth and growth rate volatility. For the MDS solutions, we only considered correlations between growth rates as a measure of similarity between countries. Additionally, we could have studied correlations between absolute growth rates, which supposedly measure volatility. Thirdly, an advantage of MDS is that the dynamic solutions provided a continuous picture of similarities in growth between countries, whereas we only considered predefined subperiods in our time series analysis. Finally, we have estimated standard errors for the parameters of our latent class models only. Note, however, that standard errors could have been obtained for the MDS solutions as well, by applying the moving blocks bootstrap for time series (See Groenen and Franses, 2000).

Comparing our static MDS solution to the classification obtained using model (5.7), we observe that the results of both approaches suggest strong similarities in growth within the cluster of Austria, Italy and Japan, and the cluster of Belgium, Finland and Norway. Furthermore, there are similarities in growth between Canada and the USA, and France and the Netherlands. Let us now discuss some of the differences of the results from the two approaches. Germany, which clearly has a strong trading partnership with the Netherlands and France, would not be interrelated to these countries based on the MDS results.

We mentioned, however, that the MDS solutions show similarities in growth only. Therefore, the classification of the latent class model may be based on similarities in growth volatility of these three countries. New Zealand, which we considered to be an outsider based on the MDS results, is added to the group of Canada and the USA by the latent class model. Finally, discussing the results of model (5.7), we concluded that the countries classified to the fifth cluster clearly belong together, because of their low volatility. In contrast, looking at the MDS solutions we would classify the two Scandinavian countries and Switzerland to the fourth cluster, Australia to the second cluster, and the UK to the third cluster. For geographical and historical reasons, one may prefer the MDS solutions for these countries.

5.4 Conclusions

In this chapter we have discussed two models to describe the variation in our panel data set of real per capita GDP growth rates. For the first model we assumed average growth and growth rate volatility of groups of countries to be constant over time. Its parameters were estimated separately for five subperiods of the 20th century. This procedure implied that the cluster membership of countries could vary over time. The results suggested that the 17 economies have diverged radically during the Golden Age. Before the Second World War, there was a large variation in the growth rate volatility of countries, whereas, in the final 25 years countries seem to have converged in terms of both growth and growth rate volatility. Finally, we observed that some countries were classified to the same cluster in more than one subperiod.

In contrast to this modelling approach, we estimated a model in which it is assumed that the cluster membership of countries is constant over the entire 20th century. Using this model, the clustering obtained is very natural in the sense that countries with a common history, or close geographical proximity are classified to the same cluster. The number of clusters obtained, however, is relatively large. Only for the subperiod of the Golden Age, we observed significant differences in the parameter estimates for the majority of the clusters. As far as the variability of cluster membership over time is concerned, our two models clearly reflected the extremes. Most likely, the truth is somewhere in between both extremes.

Finally, we discussed some differences between our time series analysis and the MDS analysis of Section 2.4. Perhaps the most important difference is that the classification of our latent class models is based on both average growth and growth rate volatility, whereas the MDS analysis is based on correlations in growth only. This might explain why we obtained a larger number of clusters using our latent class models, compared to the number of clusters we obtained studying the MDS solutions.

Chapter 6

Multivariate Analysis Using a Dynamic Model

6.1 Introduction

The two latent class panel data models discussed in Chapter 5 allowed us to study possible similarities in economic growth in a flexible way, in the sense that growth, growth volatility, and the cluster membership of countries could be varied over time. This approach provided useful insights into our data. Note, however, that the results of these models should not be interpreted as evidence for the similarities in the long-run growth paths of countries, which we established in Section 2.3. After all, by analyzing a panel of growth rates, we implicitly ignored possible differences in the absolute level of real per capita GDP across countries. Moreover, we assumed growth to be constant over subperiods, where our choice of subperiods was entirely subjective. Ideally, one should model similarities in the levels of the series, whilst treating possible changes in growth endogenously. The aim of this chapter is to develop such a model.

For this purpose, let us first elaborate on the nature of the similarities in long-run growth we just mentioned. The graphs shown in Appendix A.3 reveal that, for some countries, the path of long-run economic growth is close to linear, whereas, for others, this path is typically curved. However, also within the latter group, we established that there are strong similarities between the shapes of this curve. As we concluded that changes in long-run growth are mostly due to dramatic events that affected many or even all economies in our study, this may imply that economies recover from dramatic events in a similar way. The latter implication leads to the more general hypothesis that the dynamics of long-run economic growth are similar across countries.

To evaluate this hypothesis, we proceed as follows. Firstly, to model the dynamics of long-run growth, we use the knowledge gained in Chapter 3. In that chapter we concluded that our STAAR model may well describe long-run growth, provided that sufficient observations are available. We will extend this model to a panel data model.

To allow the dynamics of long-run growth to be equal across groups of countries, we treat the parameters summarizing the dynamics of long-run growth as discrete random effects. Again, to decide upon the number of clusters, information criteria will be used. Evidence for the above hypothesis is gained if we find that the appropriate number of clusters is considerably smaller than the number of countries in our study. Note that if this is the case, the above mentioned parameters are estimated over a larger number of observations, as desired. This increases the degrees of freedom and hence improves the efficiency of the estimates. As a consequence, it may well occur that the estimated trends obtained using this model better represent the long-run economic growth paths of our time series than the trends obtained using the univariate STAAR model.

The chapter is organized as follows. In the next section, we will formulate the latent class STAAR model, as briefly described above. As noted in Chapter 3, an important concern of the STAAR model is a possible bias in the parameter estimates in presence of heteroscedasticity in the error process. Therefore, in this chapter, attention is paid to modelling the time variation in volatility. Similar to the models in Chapter 5, the parameters of the model will be estimated using the EM algorithm. However, in this case, it is not possible to solve the maximization problems in the M-step of the algorithm analytically. Therefore, we have to resolve to numerical approximations of this solution. This is discussed in Section 6.3. Finally, in Section 6.4, the estimation results of the model are presented, and these are compared to our univariate results. This enables us to address the question whether our estimates of long-run economic growth are improved by exploiting cross-country similarities. Section 6.5 concludes this chapter.

6.2 A Multivariate Latent Class STAAR Model

To extend the univariate STAAR model of Chapter 3 to a panel data model, the most important step is to impose a structure on each group of parameters. As argued in the previous section, we want to treat the structural parameters characterizing the dynamics of long-run growth as discrete random effects. Recall that, in the STAAR model, this concerns the α, β, and γ parameters. The parameters α and β together determine the cutoffs of the interval in which long-run growth may vary over time, whereas γ measures the abruptness of changes in growth. Treating α, β, and γ as discrete random effects implies that we restrict these parameters to be the same within each of the J clusters of countries. As we want the classification of countries to be solely based on similarities in long-run growth, the other model parameters are treated as country fixed effects. We can now write the trend component as:

$$\begin{aligned} n_{i,t} &= n_{i,t-1} + \alpha_{s_i} + \beta_{s_i} F\Big((y_{i,t-1} - n_{i,t-1})/\sigma_{i,t-1}\Big|\gamma_{s_i}\Big), \\ \Pr[s_i = j] &= \pi_j, \qquad \textstyle\sum_{j=1}^{J} \pi_j = 1, \qquad \pi_j \geq 0, \end{aligned} \tag{6.1}$$

where $F(\cdot)$ is the logistic transition function as specified in (3.13). The initial values $n_{i,1}$, for $i = 1, ..., N$, are assumed to be unknown parameters. Note that although these parameters are part of the trend component, they are treated as country fixed effects, as we do not want the classification to be (partly) based on initial welfare.

Similar to our univariate model, the cycle is modelled as an ARMA process. Recall that we assumed the volatility of the cycle to be constant over time. However, this may have caused the estimates of β to be very high. Therefore, in this model, we allow volatility to change over time. Undoubtedly the most prominent change in volatility was the structural break located around 1940-1950. In Chapter 2 we observed that volatility was very high before this period, whereas it was low afterwards. Ideally, one should allow for smooth changes in volatility over time and an endogenous determination of the location of this structural break. However, the choice of our volatility component is severely restricted by data limitations. Therefore we assume that volatility is constant during the first and second half of the century, while we allow for a structural break in 1950[1]:

$$\sigma_{i,t} = \begin{cases} \xi_i^0 & \text{for } t = 1900, ..., 1950 \\ \xi_i^1 & \text{for } t = 1951, ..., 2000 \end{cases} \tag{6.2}$$

In summary, the latent class STAAR model is formulated as:

$$\begin{aligned} y_{i,t} &= n_{i,t} + z_{i,t} \\ n_{i,t} &= n_{i,t-1} + \alpha_{s_i} + \beta_{s_i} F\Big((y_{i,t-1} - n_{i,t-1})/\sigma_{i,t-1} \Big| \gamma_{s_i}\Big) \\ \phi_i(L) z_{i,t} &= \theta_i(L) \varepsilon_{i,t} \\ \varepsilon_{i,t} &\sim N(0, \sigma_{i,t}^2) \end{aligned} \tag{6.3}$$

where $\phi_i(L) = 1 - \phi_i^1 L - ... - \phi_i^p L^p$ and $\theta_i(L) = 1 - \theta_i^1 L - ... - \theta_i^q L^q$. Finally, $\sigma_{i,t}$ is defined as in (6.2), s_i satisfies (4.9), and $n_{i,1}$, for $i = 1, ..., N$, are unknown parameters.

6.3 Estimation

In order to facilitate the estimation of the model, let us impose a final restriction. Consistent with our approach in Chapter 3, we assume that the cycle $z_{i,t}$ is an AR process with a lag order of one[2]:

$$z_{i,t} = \rho_i z_{i,t-1} + \varepsilon_{i,t} \tag{6.4}$$

[1]Our decision upon the location of the break is based on a detailed analysis of the graphs in Appendix A.1. Note that, in general, this location differs across countries. For example, in Germany, the break is observed relatively late. Nevertheless, we consider the year 1950 to be a reasonably good overall estimate.

[2]Similar to the decision upon the number of clusters J, one could also decide upon the lag orders p and q using information criteria. However, this requires us to estimate the model for all plausible combinations of J, p and q, which is computationally very intensive.

For the latent class STAAR model, it is convenient to write the conditional distribution of a series $\boldsymbol{y_i}$, given that country i belongs to cluster j, as:

$$f_j(\boldsymbol{y_i}|\boldsymbol{\vartheta_j}) = \prod_{t=1}^{T} \phi(\varepsilon_{i,t}^{j}|0, \sigma_{i,t}^2). \tag{6.5}$$

Consequently, the likelihood-function of the model can be written as:

$$L(\boldsymbol{y}|\boldsymbol{\theta}) = \prod_{i=1}^{N}\sum_{j=1}^{J} \pi_j f_j(\boldsymbol{y_i}|\boldsymbol{\vartheta_j}) = \prod_{i=1}^{N}\sum_{j=1}^{J} \pi_j \prod_{t=1}^{T} \phi(\varepsilon_{i,t}^{j}|0, \sigma_{i,t}^2), \tag{6.6}$$

Note that we use the same notation here as we did in Section 4.4. The vector $\boldsymbol{\vartheta_j}$ denotes all country- and cluster-specific parameters associated with the distribution $f_j(\cdot)$, and $\boldsymbol{\theta}$ summarizes the model parameters. More specifically, $\boldsymbol{\theta} = (\boldsymbol{\pi}, \boldsymbol{\vartheta})'$. Finally, the term $\varepsilon_{i,t}^{j}$ denotes the residual of country i at time t in cluster j. Given that the cycle is an AR(1) process, this residual can be computed from (6.3) as:

$$\begin{aligned}
\varepsilon_{i,t}^{j} &= (1-\rho_i L)(y_{i,t} - n_{i,t}) \\
&= (y_{i,t} - \rho_i y_{i,t-1}) - \Big(\alpha_j + \beta_j F(z_{i,t-1}/\sigma_{i,t-1}|\gamma_j)\Big) \\
&\quad -(1-\rho_i)\Big(n_{i,1} + \alpha_j(t-2) + \beta_j \sum_{l=2}^{t-1} F(z_{i,t-l}/\sigma_{i,t-l}|\gamma_j)\Big).
\end{aligned} \tag{6.7}$$

Recall that, to maximize the expectation of the complete data log-likelihood function with respect to $\boldsymbol{s}|\boldsymbol{y}$, given by (4.16), we can consider two parts separately:

$$\max_{\pi_j} \quad \sum_{i=1}^{N}\sum_{j=1}^{J} p_{i,j} \ln \pi_j \qquad \text{for } j = 1, ..., J \tag{6.8}$$

$$\max_{\boldsymbol{\vartheta_j}} \quad \sum_{i=1}^{N}\sum_{j=1}^{J} p_{i,j} \ln \phi(\varepsilon_{i,t}^{j}|0, \sigma_{i,t}^2) \qquad \text{for } j = 1, ..., J \tag{6.9}$$

The solution to the first system of equations is given in (4.21). In contrast to the models discussed in Chapter 5, for the latent class STAAR model it is not possible to obtain closed-form solutions to the second system of equations. This can be understood by noting that the residuals $\varepsilon_{i,t}^{j}$ depend on the model parameters in a nonlinear way. Therefore we have to resolve to numerical approximations of this solution. Following Paap *et al.* (2004), we maximize the country- and cluster-specific parameters separately in an iterative manner. The estimation algorithm can be summarized as follows:

1. Fix the number of clusters J and generate starting values for the country- and cluster-specific parameters $\boldsymbol{\vartheta_j}$, and the *ex ante* probabilities π_j, for $j = 1, ..., J$.

2. Given the current estimates of π_j and $\boldsymbol{\vartheta_j}$, compute the *ex post* probabilities $p_{i,j}$ using (4.12).

3. Repeat M times:

   - For every country i, update the estimates of the country-specific parameters $n_{i,1}$, ρ_i, ξ_i^0 and ξ_i^1, given the current estimates of α_j, β_j, γ_j and $p_{i,j}$, for $j = 1, ..., J$.
   - For every cluster j, update the estimates of the cluster-specific parameters α_j, β_j and γ_j, given the current estimates of $n_{i,1}$, ρ_i, ξ_i^0, ξ_i^1 and $p_{i,j}$, for $i = 1, ..., N$.

4. Update the *ex ante* probabilities π_j, for $j = 1, ..., J$, using (4.21).

5. Compute the value of the log-likelihood function, $\ln L(\boldsymbol{y}|\boldsymbol{\theta})$.

6. Stop if convergence has been achieved, otherwise repeat steps 2-5.

To update the estimates of the country- and cluster-specific parameters in step 3, it is convenient to perform a fixed number of Newton-Raphson iterations, maximizing the expectation of the complete data log-likelihood function over the parameters considered. Note that, for the algorithm to convergence to a maximum of the log-likelihood function, it is not necessary to perform a full maximization in this step. An improvement of the complete data log-likelihood function suffices (Meng and Rubin, 1993).

It is often worthwhile to calibrate the algorithm carefully. A proper choice of the number of repetitions M, and the number of Newton-Raphson iterations may speed up convergence of the algorithm considerably. For our model and data set, it proved to be efficient to set the number of repetitions M equal to one, and the number of Newton-Raphson iterations equal to three. As a convergence criterion for the EM algorithm, we used the change in the value of the log-likelihood function, which should be smaller than 10^{-4}. Once the improvement of the log-likelihood was of order 10^{-2}, we increased M to five and the number of Newton-Raphson iterations to ten.

6.4 Results

The model is estimated with three, four and five clusters. Again, each country i is assigned to the cluster j for which the *ex post* probability $\hat{p}_{i,j}$ is highest. We select the appropriate number of clusters using the information criteria discussed in Section 4.5. The values of these criteria are shown in Table 6.1. Furthermore, the values of the log-likelihood function at convergence are reported. Surprisingly, all criteria tend to favour the model with four clusters, which leaves no room for subjective judgement. We note, however, that for the AIC and AIC3 criteria there are only very small differences between the values for four and five clusters. In all fairness, it is ambigious whether four or five clusters are to be preferred if we base our decision on one of these two criteria. The parameter estimates and the highest *ex post* probabilities of the latent class STAAR model with four clusters are shown in Table 6.2 and Table 6.3. In Appendix A.7, the estimated trend

Table 6.1: Selecting the number of clusters for the latent class STAAR model.

Criterion	Number of clusters J			
	3	4	5	Selected
AIC	-3.493	-3.522	-3.521	4
AIC3	-3.447	-3.475	-3.472	4
BIC	-3.246	-3.266	-3.255	4
Likelihood	3045.7	3073.5	3075.7	

components are presented graphically, together with the original series and the output series of the transition function. To construct the graphs, for each particular country, the cluster specific parameter estimates are used that correspond to the cluster with the highest *ex post* probability of this country.

Let us first make some general remarks. Firstly, as expected, the estimated standard errors of the parameters are much lower than those of our univariate models, due to the higher number of observations. For all clusters, the parameter β appears to be significant[3], which again indicates that the trend adjustment mechanism of the model is important. However, although perhaps significant, trend adjustments are marginal for countries assigned to the third cluster. For these countries, the trend of the latent class STAAR model practically reduces to a linear deterministic trend. Compared to our univariate results, the estimates of β are substantially lower for all countries under consideration, except for Denmark where β is only slightly higher. As a consequence, in general, the length of the intervals in which growth may vary is smaller. This implies that the trends evolve more slowly over time. A look at the graphs of the estimated trends and output series of the transition function confirms this finding. Moreover, the model does not seem to suffer from overfitting the trend component, in the sense that the direction of the trend tends not to be adjusted to short-run deviations. A drawback of a small value of β is that it limits the flexibility of the trend. Therefore the degree of persistence in our cycles is higher. Nevertheless, the estimates of ρ, reported in Appendix A.7, are still acceptably far from unity, exceeding 0.9 only in three cases. These three cases will be discussed below. Finally, given the high estimates of ξ_i^0 compared to ξ_i^1, volatility was indeed much higher during the first half of the century than during the prewar era. As a consequence, smaller deviations from the trend are tolerated during the latter period. This confirms our earlier observation that a careful treatment of changes in volatility over time is needed in order to obtain reliable estimates of long-run growth using the STAAR model.

We will now compare the characteristics of the different clusters obtained. For this purpose, we plot the estimated trend adjustments for different values of the standardized

[3]Recall that we do not formally test for the significance of β. See Section 3.5.

Table 6.2: Estimates of the cluster-specific parameters of the latent class STAAR model with four clusters. Estimated standard errors are in parenthesis.

j	Countries	$\hat{\pi}_j$	$\hat{\alpha}_j$	$\hat{\beta}_j$	$\hat{\gamma}_j$
1	AT, IT, JP	0.180**	0.038***	0.054***	1.289*
		(0.095)	(0.002)	(0.003)	(0.881)
2	AU, CA, NZ	0.192**	0.013***	0.015***	3.016**
		(0.108)	(0.001)	(0.002)	(1.457)
3	FR, GE, NL, US	0.219**	0.017***	−0.005***	1.689***
		(0.107)	(0.000)	(0.000)	(0.047)
4	BE, DM, FI, NO, SE, SW, UK	0.409[1]	0.018***	0.022***	1.786**
			(0.001)	(0.002)	(0.995)

*** Significant at the 1% level, ** at the 5% level, * at the 10% level.
[1] $\hat{\pi}_4 \equiv 1 - \sum_{j=1}^{3} \hat{\pi}_j$.

trend deviation. The resultant curves, shown in Figure 6.1, each provide a brief overview of the dynamics of 20th century long-run growth in countries classified to a particular cluster. The first cluster, which consists of Austria, Italy and Japan, will probably be no stranger to the reader. Relatively immature at the beginning of the century, these economies went through a strong catch-up period, as suggested by the relatively high estimate of α. As the estimate of β exceeds α, periods of negative long-run growth may also occur. This was the case, for example, during the Second World War. The second cluster covers Australia, Canada and New Zealand. Growth in this cluster is typically low, whereas the estimate of γ is somewhat higher compared to the other clusters. The latter is to allow for several more abrupt changes in the slope of the trend of these countries, of which the most prominent example is the change around 1940. Regrettably, the series of Canada deviates from the trend for a very long time, and appears to require a higher estimate of β. As a consequence, the degree of persistence in the cycle is high. The third cluster consists of France, Germany, the Netherlands and the US, where, considering the results of our earlier analyzes, the US is a newcomer in this group. The trends of these countries are estimated as linear lines, and it is clear that this result is most plausible for the US series. That the other economies are also classified to this cluster is not completely unexpected either, considering the fact that, after the Second World War, these economies seem to have returned to their prewar growth paths, which are approximately linear. Only near the end of the century a slight change in the slope of the trend can be observed for France and Germany. Note that the highest *ex post* probability of France is 0.72, which is relatively low. With probability 0.28, it would have been assigned to the fourth cluster. Interestingly, we estimated a fairly good, though more flexible trend for France

Table 6.3: Estimates of the country-specific parameters and the highest *ex post* probabilities of the latent class STAAR model with four clusters. Estimated standard errors are in parenthesis.

Country	$\hat{p}_{i,\hat{s}_i}$	$\hat{n}_{i,1}$	$\hat{\rho}_i$	$\hat{\xi}_i^0$	$\hat{\xi}_i^1$
Australia	0.964	8.377	0.834	0.045	0.021
		(0.044)	(0.062)	(0.001)	(0.001)
Austria	1.000	7.954	0.691	0.148	0.026
		(0.146)	(0.062)	(0.008)	(0.001)
Belgium	1.000	8.208	0.844	0.058	0.020
		(0.057)	(0.057)	(0.002)	(0.001)
Canada	0.969	7.896	0.952	0.071	0.025
		(0.063)	(0.031)	(0.003)	(0.001)
Denmark	1.000	7.971	0.724	0.054	0.024
		(0.054)	(0.074)	(0.002)	(0.001)
Finland	0.937	7.400	0.874	0.060	0.030
		(0.058)	(0.057)	(0.002)	(0.001)
France	0.720	7.932	0.940	0.102	0.014
		(0.052)	(0.013)	(0.001)	(0.000)
Germany	1.000	7.934	0.899	0.124	0.017
		(0.012)	(0.011)	(0.006)	(0.000)
Italy	1.000	7.469	0.748	0.084	0.024
		(0.084)	(0.073)	(0.004)	(0.001)
Japan	1.000	7.039	0.735	0.115	0.024
		(0.114)	(0.034)	(0.005)	(0.001)
Netherlands	0.996	8.043	0.937	0.108	0.019
		(0.028)	(0.020)	(0.004)	(0.000)
New Zealand	0.947	8.369	0.641	0.059	0.036
		(0.056)	(0.088)	(0.002)	(0.001)
Norway	1.000	7.478	0.897	0.053	0.023
		(0.051)	(0.049)	(0.002)	(0.001)
Sweden	1.000	7.860	0.791	0.039	0.017
		(0.039)	(0.067)	(0.001)	(0.000)
Switzerland	1.000	8.170	0.873	0.049	0.022
		(0.048)	(0.045)	(0.002)	(0.001)
UK	0.616	8.442	0.820	0.043	0.020
		(0.042)	(0.063)	(0.001)	(0.001)
USA	1.000	8.326	0.895	0.074	0.023
		(0.021)	(0.048)	(0.002)	(0.001)

Note: All parameters are significant at the 1% level.

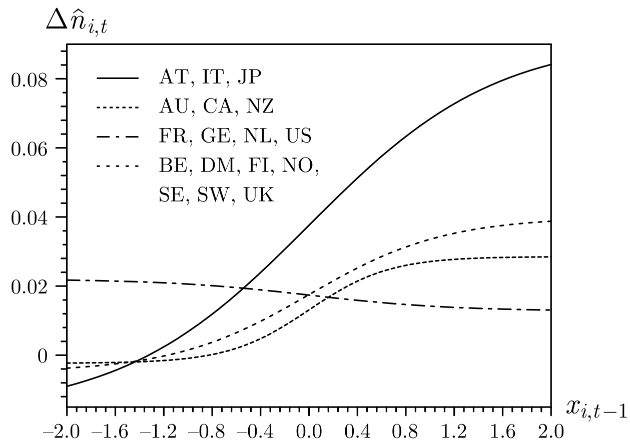


Figure 6.1: Estimated trend adjustments $\Delta\hat{n}_{i,t}$ as a result of a standardized trend deviation $x_{i,t-1} = (y_{i,t-1} - n_{i,t-1})/\sigma_{i,t-1}$.

using our univariate model. Comparing the parameter estimates of the two models, we observe, for example, that the gain in slowness of evolvement of the trend obtained by the multivariate model comes at the price of an increase in ρ from 0.82 to 0.94. Whether this is an acceptable price depends on the purpose of ones analysis. Finally, we discuss the fourth cluster, which consists of Belgium, Scandinavia, Switzerland and the UK. The growth characteristics of countries classified to this cluster differ only slightly from the second cluster. A closer look at Figure 6.1 reveals that, for negative deviations, the trend adjustment curves of the two clusters nearly overlap. However, for large, positive deviations, trend adjustments are substantially higher in the fourth cluster. This can be explained by the fact that economies classified to the latter cluster experienced somewhat higher growth spurts during the Golden Age. An exception to this rule is the UK. The fact that the UK does not entirely fit into this cluster is again apparent when looking at the highest *ex post* probabilities. After observing the data, the probability that the UK belongs to the fourth cluster is only 0.62, whereas, it belongs to the second cluster with probability 0.38.

6.5 Conclusions

In this chapter we hypothesized that the similarities in the long-run growth paths of economies, established in Section 2.3, are due to cross-country similarities in the dynamics of long-run growth. To evaluate this hypothesis, we extended the univariate STAAR model of Chapter 3 to a latent class panel data model. In this model, the structural parameters characterizing the dynamics of long-run growth are treated as discrete random effects. This implies that we restrict these parameters to be the same within J clusters of countries. The presence of strong similarities in the dynamics of long-run growth would coincide with a number of clusters J that is substantially smaller than the number of countries in our data set.

A technical concern of the STAAR model is that the parameter estimates appear to be biased in presence of time-varying volatility. To reduce the severity of this problem, we allowed a structural break in volatility to occur in 1950, as this break is generally thought to be the most prominent change in volatility during the 20th century. Once again, the parameters of the model were estimated using Expectation-Maximization. However, for this model, we presented a slightly different estimation algorithm, compared to the algorithm introduced in Section 4.4. In this algorithm, estimates of the country- and cluster-specific parameters are approximated numerically. It is efficient to update these estimates separately in an iterative manner. We estimated the model parameters for our panel data set of real per capita GDP levels. On the basis of information criteria, we established that the appropriate number of clusters is four. This number is considerably smaller than the number of countries considered, hence empirical evidence is gained for the above hypothesis.

Our next step was to compare the results of the latent class STAAR model to the results of the univariate model. We concluded that the trends obtained using the latent class model seem to better represent the long-run growth paths of the series than those obtained using the univariate model for the following reasons. Firstly, the trends evolve more slowly over time, as the intervals in which growth may vary are smaller. Secondly, the latent class STAAR model does not seem to suffer from overfitting the trend components, in the sense that the directions of the trends tend not to be adjusted to short-run deviations. Thirdly, although the degree of persistence in the cycles is higher, the estimates of the autoregressive parameters ρ are still far from unity, considering our small sample size. On balance, we have obtained better estimates of economic growth by exploiting cross-country similarities in the dynamics of long-run growth.

Finally, we discussed the characteristics of the different groups obtained. In Austria, Italy and Japan, the adjustments to long-run growth were typically high, as these economies went through a strong catch-up period during the 20th century. In contrast, for France, Germany, the Netherlands and the US, long-run growth is estimated to be constant. After the Second World War, these economies seem to have returned to their prewar growth paths, which are approximately linear. In between these two extremes are two clusters representing economies that showed minor changes in long-run growth over time. The difference between the countries classified to either of these clusters is that countries classified to one of the two experienced higher growth during the Golden Age.

Chapter 7

Summary and Conclusions

The economic policy-maker should be provided with reliable estimates of long-run economic growth in order for him or her to identify structural changes in economic performance. For this purpose, one often aims at distinguishing long-run growth from short-run fluctuations using time series models. In this thesis we took a two-dimensional view on this problem. Guided by the results of an exploratory data analysis, we first studied the dynamics of economic growth over time using univariate time series models. Secondly, we studied similarities in long-run growth across a panel of time series. These similarities were modelled using the latent class methodology, as put forward in this context by Paap *et al.* (2004). This approach ultimately lead to the formulation of a dynamic latent class panel data model in which the building blocks of these two studies were brought together.

Our methodology is applied to modelling 20th century economic growth, in terms of real per capita GDP, in 17 industrialized countries. An exploratory data analysis confirmed that, at least within this group of economies, growth tends to fluctuate around a long-run growth path. Typically, this path is nonlinear. Secondly, volatility in growth has changed over time. In particular, summary statistics computed over subperiods of the 20th century indicated that it has decreased sharply during the first decade after the Second World War. Thirdly, growth is subject to unanticipated shocks. Some of these shocks appear to have an impact on long-run growth. Generally, such shocks are due to key political or economic events that concern many countries. Examples are the Great Depression and the Second World War. Finally, we observed that the long-run economic growth paths of our countries show strong similarities over prolonged periods of time.

Let us first highlight the methodology as outlined above. We will then discuss the empirical results of our models. Finally, we provide some directions for further research.

Methodology

To model possible nonlinearities in the dynamics of economic growth, we discussed several univariate time series models. A basic premise of the models considered is that an annual macroeconomic time series can be decomposed as the sum of a trend and a cyclical component. For the trend component to represent the long-run growth path of the series it should evolve slowly over time whilst it is flexible enough to describe the time variation in long-run growth. We put forward a novel nonlinear trend model, which is referred to as the smooth trend adjustments autoregressive (STAAR) model. It is conjectured that this model strikes a good balance between flexibility and slowness of evolvement of the long-run growth process for a wide class of trending time series. The slope of our trend is allowed to vary within a fixed continuous interval, where the boundaries of the interval are to be estimated. If the length of this interval is zero, the model reduces to the linear deterministic trend model, whereas, if the interval is wide, the trend component of the model acts like a random walk with drift. It therefore embeds these two models as limiting cases. In each year, the value of the trend slope is determined by the standardized trend deviation in the previous year. A logistic transition function is used to map this deviation to the estimated interval mentioned above.

The second step was to study similarities in growth across the 17 countries under consideration. Countries were clustered based on common average growth rates and growth volatility over five subperiods of the 20th century. Using two static latent class panel data models, we allowed the number and the composition of the clusters either to be different in each subperiod, or to be constant over the entire century. The models are estimated using the EM algorithm by Dempster *et al.* (1977), and the decision upon the number of clusters is based on the values of information criteria.

Finally, we proposed a latent class STAAR model, which is based on a finite mixture of univariate STAAR models. By treating the structural model parameters that characterize the dynamics of long-run growth as discrete random effects, countries are clustered based on common long-run growth dynamics.

Empirical findings

In general, the results of our multivariate latent class models seem to demonstrate that estimates of economic growth can be improved by exploiting cross-country similarities in long-run economic growth. In particular, for most countries in our study, the trend obtained using the latent class STAAR model appears to better represent the long-run growth path of the series than the linear deterministic trend, the random walk, or the trend obtained using the univariate STAAR model. Apart from that, for the multivariate model, the standard errors of the parameter estimates are lower, due to the gain in efficiency.

Secondly, from a macro-, as well as a socioeconomic perspective, it is interesting to study the composition of the clusters obtained. For the latent class models discussed, this clustering is generally very natural in the sense that it roughly coincides with a clustering that is based on partnerships due to common history, culture and/or geographical proximity. For example, the Scandinavian countries are usually classified to the same cluster, as well as pairs of countries like Australia and New Zealand, and Germany and the Netherlands. The multivariate subperiods analysis of our panel of growth rates indicated that a clustering solely based on average growth and growth volatility tends to vary over time. Also the number of clusters is time varying. For example, we obtained a relatively large number of clusters for the Golden Age period, due to the great divergence in growth during this period. For the subperiods before and after this period, two or three clusters are obtained. Comparing the composition of the clusters for these two periods, we concluded that some economies that were initially classified to a low growth cluster have switched to a high growth cluster during the Golden Age. This indicates that these economies have caught up with the group of established economies. Examples are Austria, Italy and Japan.

Ultimately, let us touch upon the results of the latent class STAAR model, where countries are classified based on common long-run growth dynamics. For this model, we obtained four groups of countries. In line with earlier findings, one of these groups is Austria, Italy and Japan, where long-run growth, as well as adjustments to long-run growth are typically high, possibly as a result of the catch-up phase these economies went through. In contrast, for France, Germany, the Netherlands and the US, long-run growth is estimated to be constant. After the Second World War, these economies seem to have returned to their prewar growth paths, which are approximately linear. In between these two extremes are two clusters representing economies that showed minor changes in long-run growth over time. One of these clusters covers Australia, Canada and New Zealand. These economies appear to have benefited less from the flourishing Golden Age, compared to the other cluster, which includes Belgium, Scandinavia, Switzerland and the UK.

Further research

The nature of our modelling approach is exploratory in the sense that we did not formally test for the adequacy of our models. In further research, misspecification tests could be performed, such as tests for neglected serial correlation or nonlinearity. Moreover, it would be useful to derive a test for the null hypothesis of the linear deterministic trend model against the alternative of STAAR nonlinearity. Note, however, that this test is likely to have a nonstandard distribution, as it suffers from the problem of unidentified nuisance parameters under the null hypothesis (See Davies, 1977, and Hansen, 1996, among others).

Secondly, a limitation of the latent class STAAR model is that the composition and the number of clusters is assumed to be constant over time. Instead, the countries that seem to have caught up with others might have switched to a lower growth cluster after the Golden Age. It would be interesting to allow such changes in the composition over time to occur within a similar nonlinear modelling framework as considered in this thesis.

Finally, from an empirical point of view, it would be particularly relevant to evaluate the forecast accuracy of our models, and to compare this to the accuracy of competitive models. Obviously, it would also be interesting to apply our methodology to modelling other panels of (macroeconomic) time series, where the number of observations over the time- and/or the cross-sectional dimension is possibly higher.