Estimating and Forecasting Trend Output and Modelling Labor Input:

Application of Panel Techniques and Factor Models

Inaugural-Dissertation
zur Erlangung des Doktorgrades
des Fachbereichs Wirtschaftswissenschaften
der Johann Wolfgang Goethe-Universität
Frankfurt am Main

vorgelegt von
Dipl.-Volksw. Marcus Heribert Kappler
aus Stuttgart

2007

Acknowledgements

First and foremost I am indebted to my supervisor Prof. Dr. Uwe Hassler for finding interest in my research topics and for giving me considerable freedom to write my doctoral thesis. I am also very grateful to Prof. Dr. Dieter Nautz, who kindly consented to be my second supervisor. Furthermore, I would like to thank Prof. Dr. Dr. h.c. mult. Wolfgang Franz, President of the Centre for European Economic Research (ZEW), for continuously supporting this dissertation and for providing excellent working and research conditions. My thanks also go to all my colleagues from the ZEW Research Department Corporate Taxation and Public Finance and to the colleagues from the Research Department International Finance and Financial Management for the pleasant and constructive working environment and beneficial discussions. The many helpful comments from Dr. Friedrich Heinemann, Dr. Michael Schröder, Katrin Ullrich, Steffen Osterloh, Andreas Schrimpf and from several anonymous referees and from participants in international conferences are highly appreciated. I am also very grateful to the student assistants, notably to Qianying Chen and Heidi Hellerich, who provided able help in collecting and processing data and in editing this thesis.

I also thank Fritz Böhringer and the Böhringer-Ilsfeld Foundation for providing financial support for essential parts of this work within the framework of the project "ZEW Growth Monitor".

Finally, I would like to express my deep gratitude to my family, and in particular to my wife Juliane for unquestioningly supporting me throughout the work of this doctoral thesis and for believing in me.

Table of Contents

List of Figures

List of Tables

Preface

The policy relevance of obtaining reliable estimates of trend output and its evolution as well as being able to compare actual and trend growth rates between and across countries is obvious. Trend output—also often referred to as potential output—is the key concept for assessing a country's standard of living. Defined as the productive capacity that would be feasible under full or normal utilization of all production factors, trend output is a reference variable for the analysis of the business cycle, which is regularly employed by economic research institutes and economic advisory committees of national and international organizations. Furthermore, it is an important guideline for separating structural problems from business cycle phenomena, for deriving cyclically adjusted budget balances and for providing a variable of orientation for monetary policy. If carefully estimated, trend output contains valuable information about the medium-term economic outlook of a country which is particularly important for the planning of fiscal budgets.

Trend output is unobservable in practice and has to be estimated. Various competing methodologies to derive these estimates exist. While the in-sample performance of these estimates are often studied, an evaluation of the predictive accuracy of trend output growth from an out-of sample forecasting perspective is rarely carried out. The first chapter of this thesis seeks to fill the void by setting up a framework for the evaluation of these methods in terms of predictive accuracy. Among the many techniques, the production function approaches which explicitly relate trend output to capital and labor input as well as to technology are by far most commonly used. Due to its practical relevance, the analysis of accuracy focuses on the output growth projections from this approach. There is a considerable literature on the evaluation of growth forecasts over periods ranging from one to two years. However, there are only very few investigations of growth predictions over longer periods. In the analysis of this thesis I take a closer look at GDP growth predictions three to five years ahead and analyze whether existing approaches produce a reliable view about future economic developments. Thus, the focus is shifted to the longer oriented assessment of future economic performance.

A basic insight from the new growth literature reveals that sustainable long-run output and growth is determined by more factors than those the conventional empirical approaches to measure trend or potential output incorporate. New growth theories and in particular the voluminous empirical growth literature point to a host of additional determinants such as the accumulation of human capital, research and development and international trade as well as institutions and macroeconomic policies. However, these concepts are hardly integrated into time series based methods to estimate potential output and its corresponding growth path so far. Apart from that, empirical growth research is still conducted mainly in terms of studying differences in variation of output growth across countries. However, by construction, cross-country studies fail to consider the fact that growth factors such as the rates of physical and human capital accumulation vary over time. In the second chapter of this thesis, a panel data approach for identifying growth determinants is carried out which looks at both the cross-section and the time series variation of possible growth factors.

When dealing with a multi-country data set which comprises macroeconomic time series, the issues of non-stationarity, cointegration, as well as cross country dependence have to be taken into account in econometric estimation. Furthermore, assumptions about the degree of homogeneity and heterogeneity of parameters of econometric relationships have to be made in order to employ the most efficient estimators. Natural techniques to consider for parameter estimation and hypothesis testing are panel unit root and panel cointegration methods, which have advanced considerably in the last years. In contrast to the first generation of these tests which built on the assumption of independent units, the recently proposed tests of the second generation take fairly general patterns of cross-section dependence into account. In particular factor models have proven valuable in modelling dependencies across countries due to unobserved common effects. The assumption that economic time series of different countries share common global factors which, however, influence the single series quite individually, can often be justified on economic grounds. A much-cited example for a common global factor is that of technological progress.

In the last chapter of this thesis, factor modelling in the context of panel unit root testing to account for cross-section dependence is examined in greater detail. The interest in this chapter is the proper empirical modelling of the aggregate labor supply in terms of average hours worked. The central question is whether hours worked contain a unit root, which has important implications for the empirical validity of the Real Business Cycle and the New Keynesian models. Existing literature has not reached a definite conclusion on the time series properties of hours worked and it is demonstrated how recent methods for cross-dependent macroeconomic panel data can help to shed light on this controversial question.

Overall, this thesis takes a detailed view on empirical concepts—many of which are ubiquitous in practice— but provides no comprehensive theoretical perspective. For instance, the empirical difficulties arising from attempting to separate cycle from trend partly trace back to theoretical difficulties to clearly separate these components. However, these topics will only be touched on briefly whenever they help to provide clarification of empirical ideas.

Chapter 1

Projecting the Medium-Term: Methods, Outcomes and Errors for GDP Growth

1.1 Introduction

Realistic assessments of the medium-term growth capabilities of an economy are important for many purposes. Medium-term GDP forecasts are particularly vital for the planning of public budgets under the objective of a sustained budget policy, they build a basis for monetary policies and are relevant for firms with regard to making correct investment decisions in order to avoid inefficient resource allocations.

For the Member States of the European Monetary Union, medium-term projections carry special weight. Within the Stability and Growth Pact, the Member States are obliged to provide information about medium-term economic developments to the European Council and the European Commission in the form of a stability programme for the purpose of multilateral surveillance. These stability programmes include a regular presentation of how the medium-term objective for the budgetary position of close to balance or in surplus can be achieved and how the path of the general government debt ratio is expected to evolve.[1]

Medium-term projections are not only prepared by official governmental bodies,

[1] A Council regulation adopted in 1997 provides details. According to this regulation, each Member State has to deliver a report on the assumed development of government investment expenditure, real gross domestic (GDP) growth, employment and inflation. In particular, assumptions about medium-term GDP growth are of key interest in this respect since they provide a basis for deriving budget balances, government investment capabilities, employment growth and inflationary pressure. See European Commission (1997). Furthermore, in the year 2005 the ECOFIN Council released a Code of Conduct which incorporates elements of the Council regulation into guidelines which emphasize that Stability and Convergence programmes should be based on realistic and cautious macroeconomic forecasts, cf. European Commission (2005).

but also central banks and international institutions like the OECD and IMF regularly provide medium-term economic outlooks to analyze the potential development of the world economy, to deliver a guide for future monetary policies and to make a reference scenario available against which alternative assumptions can be studied. For instance, such tools can be utilized to see how various imbalances (e.g. current accounts, sectoral balances, debt stocks, etc.) identified in the short-term forecasts might evolve or be resolved as the economy progresses in the medium to long-run and how policies might need to change.

There is a considerable literature on the evaluation of GDP forecasts over shorter periods (1 to 24 months ahead). Important contributions for the UK and USA were made by Davies and Lahiri (1995), Granger (1996), Harvey et al. (2001), Fildes and Stekler (2002) and Stekler and Petrei (2003). The performance of forecasts by various national institutions in European countries is examined by Öller and Barot (2000). Holden et al. (1987) and Ash et al. (1998) focus on OECD forecasts, while Pons (2000) and Ashiya (2006) look at short-term predictions released by the OECD and IMF. Döpke and Fritsche (2006) as well as Kirchgässner and Müller (2006) provide studies for Germany.

In contrast to the evaluation of business cycle forecasts, the examination of forecasts of the economic development over the medium- or long-term hardly receives any attention in economic literature although the assessment of the latter is at least as important as performance checks of short-run oriented outlooks from a policy point of view. Notable exceptions among the few papers that investigate GDP growth predictions from a medium-term perspective are Lindh (2004) and Batista and Zaluendo (2004).

Against this background, this chapter first provides a survey of methods of medium-term forecasting that are used by governmental bodies in the G7 industrial countries and international institutions. As it turns out, the New Classical growth model with its assumptions about the supply-side functioning of an economy and conditional steady-state convergence plays a predominant role for medium-term forecasting. Therefore, the discussion of these procedures which are usually referred to as production function approaches (PFA) will receive special emphasis in the subsequent illustration.

In the second, empirical part of this chapter, it will be analyzed whether the production function procedures do produce reliable predictions of actual GDP growth over the medium-term. To this end, an out-of-sample forecast exercise based on quarterly data from National Accounts for the G7 countries is conducted and an evaluation of forecast errors is carried out. The formal evaluation of actual projections from official institutions, however, is difficult since these projections are usually published with a low frequency or have been prepared only recently and therefore exhibit a lack of time series observations which limits the application of statistical tests considerably. Despite this restriction, available projections from national and international sources are also included in the analysis below, however, these projections are compared to the actual medium-term development of GDP and the pseudo projections from the out-of-sample analysis in a more stylized fashion.

The producers of medium-term forecasts are aware of the limits to precision of predictions beyond the usual business cycle frequencies and denote such forecast "projections", rather than definite forecast (e.g. Carnot et al., 2005). The term projection is used since predicting is usually conducted by extrapolating from past observations and these projections mainly serve to illustrate broad trends in the sense of providing a baseline-scenario for the assessment of alternative case scenarios. Medium-term projections typically abstract from the prediction of future cyclical developments and therefore do not claim to have rich information value in terms of correlation with actual outcomes.

Nevertheless, in order to be a reliable tool for policy analysis the methods typically employed should at least yield projections that do not systematically over- or underestimate actual GDP development over the medium-term. Tests for unbiasedness are therefore a central issue of the present chapter since this is the same as testing if projections are weak rational and consistent and hence meet basic optimality requirements. Even if projections are unbiased they may nevertheless be very inaccurate. Therefore, the results of tests for forecast accuracy are also reported, although accuracy in terms of correlation with outcomes is not a primary claim of such more longer-oriented forecasts.[2]

Due to the design of the out-of-sample analysis, the corresponding multi-step forecasts result in forecast errors that are serially correlated. In that case efficiency of projections does not rule out serial correlation of the forecast-errors. In order to explicitly account for serial correlation in error processes and to perform consistent tests for unbiasedness and accuracy, a simple model of forecast errors is employed to analytically derive the exact covariance matrix of forecast errors and appropriate test statistics. We use a framework for testing forecast unbiasedness which is inspired by the work of Brown and Maital (1981), Keane and Runkel (1990), Davies and Lahiri (1995) and Clements et al. (2007), while the accuracy test draws on the contributions of Diebold and Mariano (1995) and Harvey at al. (1997). It is shown that this particular framework has advantages in small samples over the approaches usually employed to inference in forecast error analysis. Empirical implementation

[2] A note on terminology: In the following sections, the terms "projections", "forecasts" and "predictions" are used synonymously for the medium-term forecasts considered in this chapter, whereas the most appropriate understanding of these forecasts is that of projections as they are meant to indicate likely future developments based on extrapolation of past trends, rather than deliver precise point forecasts of GDP growth.

of these tests is straightforward and conducted for three to five year cumulative forecasts of GDP growth based on the production function approach for Canada, France, Germany, Italy, Japan, the United Kingdom and the USA.

The rest of the chapter proceeds as follows. Section 1.2 discusses commonly used approaches for producing medium-term predictions and reviews the relevant literature. Section 1.3 is extensively devoted to the implementation and analysis of the PFA and explains the testing strategy in detail. Results of the forecast evaluation are presented in section 1.4 while section 1.5 summarizes and concludes.

1.2 Approaches for predicting medium-term growth

The aim of this section is to give a brief overview of the mainstream approaches for the preparation of medium-term GDP projections which are currently in use by governmental bodies and policy-oriented international institutions, to highlight the key features of the conventionally employed methods and to motivate the practical relevance of the subsequent empirical analysis.[3]

Besides yielding a key reference variable for the medium-term planning of public budgets, projections of the main economic development that go beyond the typical business cycle forecast horizons have become an increasingly important tool for the policy analysis conducted by national authorities and international institutions.

A key element of all applied methods is the concept of potential output. In a nutshell, potential output denotes the level of real GDP attainable without raising inflation when the economy is operating at a high rate of resource use. The original definition goes back to Okun (1962). The importance of the concept of potential output for the preparation of predictions originates from the assumption that in the medium- to long-run the economy evolves according to its potential growth rate. This assumption also implies that output always shows a tendency to return to its potential path and that deviations of actual output from the potential level are only temporary and can not be sustained for long periods. Output growth will tend to be less than potential growth when output is above potential output and more than potential growth when it is below the potential level.

The theoretical underpinning for such an understanding of the behavior of the economy is twofold: First, the existence of a long-run growth path is delivered

[3]The illustrations in this section draw on technical reports and working papers by the OECD, the IMF, the European Commission and Central Banks but also on an extensive report conducted by the ZEW in cooperation with CEPS, Brussels, on behalf of the German Ministry of Economics and Labor with the title "Methods of Medium-Term Economic Forecasting". For this purpose, information on the approaches and methods used by governmental bodies in Germany, France, the United Kingdom, Italy, the Netherlands and the USA was gathered with the aid of a questionnaire which was sent to the persons responsible for the official projections by governments or administrations in the respective country. More detailed references are given in the subsequent sections.

by macroeconomic growth theory, which either specifies the long-run growth of an economy as being solely determined by exogenous forces (New Classical theory, the Solow-Swan model, for instance) or by endogenezing long-run economic growth by modelling important determinants more as functions of economic decisions.[4] Usually, these theories ignore cyclical fluctuations. Secondly, the existence of output gaps can be justified and explained from theories of business cycle fluctuations, which give insight into the causes of cyclical output movements around its potential or trend level. Although several theoretical approaches that analyze the interaction between cyclical movements and long-run growth have been brought up (e.g. Stadler, 1990), the conceptual separation between "growth" and "business cycle" is still prevailing particularly in applied work although this dichotomy is somehow artificial. While theories of fluctuations play an important role for the preparation of business cycle forecasts, they are of minor significance for assessing the medium- to long-term outlook.

The potential output of a country can not be observed and must therefore be estimated. A variety of methods have been developed for these purposes which can be categorized into several broad classes: *Production function approaches* (PFA), *statistical filters*, *system approaches* and *multivariate time series models*.[5] The PFA are the main concern of this chapter and will be reviewed in greater detail below. *Statistical filters* such as bandpass filters or the Hodrick-Prescott (HP) filter extract trends from GDP directly without explicit reference to economic theory. As illustrated below, these filters often serve as an auxiliary tool for the implementation of more theory-oriented methods.

The *system approaches* build on the full specification of simultaneous models which describe the interlink between key variables such as output, inflation and unemployment. Usually potential output is modeled as latent variable and the parameters of the model and potential output are estimated within the Kalman filter framework.[6]

Structural vector autoregressions (SVAR) are the most widely used models in the class of the *multivariate time series models*. Blanchard and Quah (1989) introduced this methodology which aims to identify different demand and supply innovations in a vector autoregressive (VAR) model with the aid of long-run neutrality restrictions on the various types of innovations. In this framework, a measure of potential output

[4] Aghion and Howitt (1998), Chapter 1, provide a comprehensive illustration of various new growth theories which endogenize technology as a driver of long-term economic growth.

[5] It is not the aim of the present chapter to provide a comprehensive survey and comparison of the many methods to estimate potential output. These can be found, for instance, in Bjørnland et al. (2005), Chagny and Döpke (2001), Cerra and Saxena (2000) or Dupasquier et al. (1997).

[6] Apel and Jansson (1999) illustrate the system approach in detail and apply it to Swedish data. Further applications of this methodology can be found in Fabiani and Mestre (2004), Ögünç and Ece (2004) or Benes and N'Diaye (2004).

is derived by the identified supply-side innovations since by assumption these are the only components that have a permanent effect on output.[7] Gosselin and Lalonde (2006) recently proposed an Eclectic Approach (EA) that combines the Hodrick-Prescott smoothing method with an equilibrium path generated by an SVAR on which the estimation of potential output in an augmented HP estimation setup is conditioned. The EA overcomes some of the shortcomings of the plain HP-filter and enriches it with information of a structural economic relationship.[8]

The measures of potential output arising from the various methods rarely yield a unified view and therefore policy-oriented institutions typically base their analysis on a mixture of methods. However, for a forward-looking assessment of potential production capacities and for the derivation of medium-term projections production function or growth accounting approaches are most widely-used. The OECD[9], the IMF[10] and the European Commission[11] employ a PFA. The German government uses a PFA for projections of GDP within the annual medium-term fiscal outlook. Besides the European Central Bank itself, many national central banks in Europe also base part of their assessment of the current situation of the business cycle and the estimation of the future macroeconomic performance on production function approaches.[12] Concepts that are closely related to the PFA are growth accounting methods which decompose trend output growth into components such as growth of labor productivity, growth in average hours worked, growth in employment rates and growth in population of working age. The advantage of these methods is that they do not rely on measures of the capital stock or capital services and some practitioners regard the preparation of forward projections of the individual components of the growth accounting methods as easier than the preparation of input projections for the PFA. The Congressional Budget Office in the USA[13] and the HM Treasury in the United Kingdom[14], for instance, use a growth accounting framework to derive medium-term projections.

[7]The SVAR methodology is a workhorse for many empirical problems. Examples of applications to estimate potential output and the output gap are provided by Gerlach and Smets (1999), Fritsche and Logeay (2002), Scacciaviavillani and Swagel (2002) or Claus (2003).

[8]See chapter 2 of this thesis for an application of the EA.

[9]A full documentation of the OECD method to compute potential output with the PFA and to prepare medium-term scenarios and projections is given by Beffy et al. (2006).

[10]The IMF's production function approach for the industrial countries is documented in De Masi (1997).

[11]Röger (2006) and Denis et al. (2002) describe the European Commission approach in detail.

[12]A full description of the recent research activity of the German *Bundesbank*, *Banque de France* and *Banca d'Italia* with respect to the analysis of growth and business cycles is given by Baghli et al. (2006) and Bassanetti et al. (2006). The contributions of these authors document well that production function approaches play an important role in modelling the supply side of European economies for policy analysis.

[13]A Background Paper of the Congressional Budget Office provides a summary of this growth accounting approach which is based on a textbook Solow growth model. See CBO (2004) for details.

[14]See HM Treasury (2002).

Although macroeconomic theory and particularly growth theory have developed new and more comprehensive insights into growth processes of economies than the PFA with its standard neoclassical frame of reference is capable of capturing, it is still very popular in practice.[15] The appeal of using a production function for estimating potential output and projecting its path unquestionably comes from its economic underpinning and the fact that projections for key input variables are either readily available or can be constructed by extrapolating from past trends. One distinct merit of the PFA over univariate methods is the use of population data for which projections are relatively reliable several years ahead. Perhaps the most significant advantage of the PFA is that it is based on a comprehensive economic framework which links potential output to its fundamental determinants. This in turn facilitates the assessment of the impact of policy changes or structural shifts of the economy on potential output. The key determinants of production also provide many channels through which adjustments can enter the assessment of future potential output growth. The underlying trends can easily be adjusted on judgemental grounds, when necessary, if the forecaster has additional information on the evolution of these inputs from outside the PFA framework.[16]

Obviously, the PFA is also subject to several caveats. Most importantly, it relies on data that—in addition to the target variable itself—must be estimated and therefore brings in additional sources of uncertainty surrounding the resulting potential output measures. This problem concerns the capital stock data and the non-accelerating inflation rate of unemployment (NAWRU), since both are also unobserved and have to be estimated adequately. A further problem is that the PFA builds on production function parameters which are usually imposed rather than econometrically estimated, thereby necessitating the setting of further assumptions about the economy. Since the PFA relies on trend measures of the various inputs, the question arises how to derive plausible trend values of, for instance, the potential labor input. The subsequent sections which are devoted to the implementation of the PFA demonstrate and discuss these problems in greater detail.

The assumption that an open output gap closes is an integral part of all PFA and growth accounting based projection methods. As mentioned above, the hypothesis that the output gap closes sooner or later refers directly to the neoclassical growth model in which the economy always tends towards a steady-state where output of effective labor is constant due to diminishing returns to scale with regard to factor inputs. Diminishing returns to scale also imply that the speed of convergence to the

[15]For instance, the numerous contributions to the Handbook of Economic Growth edited by Aghion and Durlauf (2005) clearly illustrate the many factors that are expected to influence the production potential of an economy and long-run growth.

[16]See Butler (1996), pp. 15 for more on the role of judgement on potential output estimates and policy-analysis.

steady-state condition positively depends on how far the economy deviates from its steady state. Even if the assumption of steady-state convergence can be sustained based on empirical evidence, as will be shown below, the critical feature of the practical implementation is the fixed period assumption during which the output gap is closed. For five year GDP growth projections, for example, it is typically assumed that the output gap closes over the five year horizon. Section 1.3.2 provides a more detailed discussion of this proceeding.

Other methods than the above described are employed or have been proposed to compute trend output and to derive projections, notably large macroeconometric models, dynamic stochastic general equilibrium models (DSGE) and cointegrating VAR models. See Garrat et al. (2006) for a review. However, in particular the latter approaches are typically designed and used for the evaluation of system responses to macroeconomic shocks and the preparation of short-term forecasts and play only a minor role for the production of longer-term outlooks. Besides, most of these models incorporate a New Classical production function with long-run restrictions that are in line with predictions of the PFA. Recently, de la Croix et al. (2006), Lindh (2004) and Lindh and Malmberg (1999) have developed models to estimate medium- and long-run GDP growth that are mainly based on demographic data and these models have been proven to perform well for the Swedish economy. However, in the light of the outstanding practical relevance and its straightforward replicability, the rest of the chapter will focus on analyzing the forecast performance of the PFA based methods.

1.3 Analysis of the production function approach

Among forecasters it is widely accepted that forecasts beyond the usual business cycle frequencies of 1 to 2 years tend to have few or zero information content (eg. Isiklar and Lahiri, 2007, for evidence from cross-country surveys). Given these insights, the obvious question arises why one should conduct an analysis of forecasts that far exceed horizons which are typically regarded as the limits for which present information can be used in shaping a view of future developments. Although it certainly can not be expected that growth projections 3 to 5 years ahead show a close connection to movements of actual growth, however, suitable medium-term projections should at least meet minimum requirements in order to be of any use for policymakers.

Principal requirements of such projections are unbiasedness and improved accuracy vis-à-vis naïve forecasts. Unbiasedness is a prerequisite for rational forecasts and implies that medium-term growth projections of GDP are on average in line with actual trend developments and therefore show no tendency to systematically over- or underestimate GDP growth. For example, this is particularly important for the medium-term planning of public budgets in order to avoid deficits in the medium and long-run.

Even if projections are unbiased, they may nevertheless be very inaccurate and lead to large forecast errors. Accuracy is an important criteria for judging forecasts quality. However, as it has been pointed out, correlation with actual outcomes is not a primary concern of medium-term projections as they are rather meant to illustrate broad trends. However, if forecasts from simple models show a tighter linkage to actual developments than predicted trends that are prepared with the aid of the PFA, which incorporates a more elaborated view of the economy, then the efficiency of the latter approach is seriously called into question.

After an extensive presentation of the empirical implantation of the PFA, the issues of bias and accuracy are explored in greater detail.

1.3.1 Implementing the production function approach

The PFA builds on a standard growth accounting framework which is depicted in many research papers and textbooks. A further formal description of this concept may not contribute much to theoretical insights, but is necessary for the demonstration of the specification of the projection analysis below. In the following, a formulation is adopted which is most closely related to descriptions in Giorno et al. (1995), McMorrow and Roeger (2001), Carnot et al. (2005), Cotis et al. (2004) or Beffy et al. (2006).

The starting point is the specification of potential supply of the economy. The total output of the economy is produced according to a standard New Classical Cobb-Douglas production function with capital and labor input:

$$Y_t = (E_t N_t)^{\alpha} K_t^{(1-\alpha)} \tag{1.1}$$

Y_t denotes output , N_t labor input, K_t capital input and E_t the Harrod-neutral labor augmenting Total Factor Productivity (TFP).[17] Labor input comprises several key variables of the labor market and enters the production function on a hours

[17] In applications, the specification of the Cobb-Douglas function and the assumption of Harrod-neutral technological progress is typically not motivated on theoretical grounds but rather used ad hoc. However, there are also profound arguments based on micro theory to use Cobb-Douglas technology. Jones (2005) shows that models which incorporate steady-state growth—a key assumption of the PFA—lead to global production which takes the Cobb-Douglas form and produces a setup where technological change in the local production is entirely labor-augmenting in the long-run. This result is derived with a microfounded growth model that builds on the distribution of ideas, a popular approach of new growth theories. Acemoglu (2003) also derives a micro-framework for the standard neoclassical growth model with labor-augmenting technical change.

worked basis rather than on number of employed:

$$N_t = H_t L_t \tag{1.2}$$
$$L_t = PW_t PR_t (1 - U_t) \tag{1.3}$$

In the above equation , H_t is the annual amount of hours worked per employee that is multiplied by the total employment of the economy to yield a measure of total labor input. Employment in turn is determined by the working age population PW_t, the participation rate PR_t and the level of unemployment U_t. The TFP as the Solow residual, which captures all the factors that affect output but are not directly included in labor, such as technology, results from equation (1.1):

$$E_t = Y_t^{-\alpha} K_t^{-(1-\alpha)/\alpha} N_t^{-1} \tag{1.4}$$

In order to obtain a measure of potential output of the economy, several trend variables (indicated with an asterisk) are substituted in equation (1.1):

$$Y_t^* = (E_t^* N_t^*)^{\alpha} K_t^{(1-\alpha)} \tag{1.5}$$
$$N_t^* = H_t^* L_t^* \tag{1.6}$$
$$L_t^* = PW_t PR_t^* (1 - U_t^*) \tag{1.7}$$

Obviously, the tricky part of implementing the production function approach is the use of adequate and reasonable trend values for the input variables. Typically several trend variables are generated by smoothing the series with the aid of statistical filters, whereas the time series filter of Hodrick and Prescott (1997, HP) is by far the most frequently utilized tool for this purpose. In the implementation below, for instance, the HP filter with its standard smoothing parameter $\lambda = 1600$ for quarterly data is used to filter the data for *hours worked*, the *participation rate* and the *TFP*. Since the application of the HP filter results in cyclical components of the filtered series that fluctuate around zero, such a procedure always defines potential output as being generated with a “normal” level of hours worked, labor force participation and TFP.

In order to derive the total contribution of labor, the notion of a “natural” rate of unemployment generally enters the calculation of N_t^* through the concept of the NAWRU (Non-Accelerating Wage Rate of Unemployment). The NAWRU is an estimate of the unemployment rate that results in employment levels which are consistent with stable wage inflation and lead to a sustainable level of potential output that does not raise inflationary pressure. While the use of filter techniques for the computation of trend values for hours worked, participation and TFP represents rather an ad hoc approach, the NAWRU estimates for U_t^*, however, bring in a complete theoretical labor market framework into the estimation of potential output. Furthermore, the degree of sophistication for empirically deriving NAWRU estimates usually far exceeds the data treatment of the remaining input variables and parameters of the production function approach.

Typically, data for the capital stock enters equation (1.5) directly. Such a procedure computes potential output as the contribution of capital services at maximum utilization since the existing stock of fixed assets always constitutes its maximal contribution to production. Due to data limitations, consideration of a "normal" or average level of capital services in the computation of potential output is hardly feasible. Therefore, one has to keep in mind that such a treatment implies a certain inconsistency regarding the assumptions about the degree of factor utilization, since capital is assumed to operate at maximum capacity while for labor input a normal level of factor utilization is assumed instead.

Estimating the partial elasticities

Besides trend variables of the inputs to production, knowledge of the partial elasticities of output with respect to labor and capital is required to determine the TFP and the level of potential output. The common approach to derive figures for these parameters merits further in-depth discussion as this is another source where concrete assumptions about the workings of the economy enter the procedure to estimate potential output. Moreover, data measurement issues play an important role for estimating these elasticities.

Key assumptions for deriving empirical counterparts for the partial elasticities are perfect competition in the factor and product markets as well as constant returns to scale of the production technology in the long run. The first assumption justifies the use of labor compensation numbers from National Accounts data as a measure for the labor elasticity of output (α) since under perfect competition in equilibrium factor prices equal marginal productivities.[18] The assumption of constant returns to scale in turn allows one to obtain the capital elasticity of output as one minus the labor share, i.e. labor compensation as a fraction of output.

The above mentioned proceeding constitutes the most popular method for estimating α in growth accounting. Although very popular, the National Accounts approach is subject to some caveats (Musso and Westermann, 2005). For example,

[18] As is well known, factor prices correspond to the partial elasticities in the Cobb-Douglas production function.

if firms earn rents from temporary monopolies due to innovation, the contribution of capital is overestimated in such a growth accounting framework since the imposed capital share $(1-\alpha)$ includes these rents. As a consequence, the contribution from TFP is underestimated. Furthermore, computing the capital contribution to production with the aid of the residual elasticity $(1-\alpha)$ attributes the net indirect taxes which are a component of GDP all to capital although a large part of the value added to finance these taxes has been generated by labor. Therefore, neglecting indirect taxes as a labor contribution also overestimates the capital share of production. In addition, the figures of the capital share include payments accruing to both reproducible and non-reproducible capital such as land and natural resources. For this reason capital share estimates derived from capital stock data, which are usually calculated using the perpetual inventory method form investment flows, will be lower than those derived from labor compensation data (Caselli and Feyrer, 2007). Lastly, one has to add to the compensation of employees the income of the self-employed. This component, however, can not be observed as it is a part of the gross operating surplus and gross mixed income. A typical approach is to assume labor income of the self-employed to be equivalent to the average compensation per employee. Under this assumption the adjusted labor share is simply the sum of the unadjusted labor share and the unadjusted labor share times the fraction of the self-employed over the employees. Table 1.1 shows the averages of the unadjusted and adjusted labor share of the G7 countries computed from annual National Accounts data.

Table 1.1: Labor shares from National Accounts Data

	Unadjusted	Adjusted
Canada	0.540	0.628
France	0.528	0.600
Germany	0.553	0.624
Italy	0.449	0.673
Japan	0.537	0.705
United Kingdom	0.570	0.647
USA	0.583	0.639
G7	0.537	0.645

Notes: Labor shares correspond to the ratio of the compensation of employees over GDP taken from the OECD Economic Outlook database. The adjusted labor share takes into account the imputed labor income of the self-employed: *Adjusted labor share=Unadjusted labor share · (No. of employees + No. of self-employed)/No. of employees.* Entries are averages of annual data from 1972 to 2005.

Column 2 of table 1.1 contains the figures for the adjusted labor share which is the measure generally used in growth accounting. These values fluctuate between 0.6 and 0.7 for the G7 countries. The average for α over all countries yields a value of 0.64 which comes very close to the popular rule of thumb value of 2/3.[19] Taking the fraction of self-employed into account can raise the labor share significantly as can be seen from the case of Italy. For this country, the adjusted labor share is more than twenty percentage points higher than the unadjusted labor share. What can be learned from table 1.1 is that adjusted labor shares do not vary much across countries and a simple rule of thumb value is at least broadly in accordance with cross-country averages of adjusted labor share data.

A further and more interesting question is whether it is possible to retrieve econometric estimates of α that match the figures calculated from National Accounts data and if there is statistical support for the assumption of constant returns to scale of the Cobb-Douglas technology. Econometric estimates of factor shares are regularly criticized and, as it turns out below, not without reason. Temple (2006) provides a recent survey on this matter.[20]

In the following, estimating the parameters of the Cobb-Douglas function is carried out in a dynamic framework by assuming that the logarithm of output y_t follows an Autoregressive Distributed Lag (ARDL) model. For estimation and identification of the structural Cobb-Douglas parameters, the ARDL is re-parameterized into an Error Correction Model (ECM) and the estimation techniques of Pesaran et al. (1999) are employed.[21] If output follows a Cobb-Douglas technology, the logarithms of output, capital and labor input are cointegrated and an ECM model is an appropriate empirical specification. The estimators proposed by Pesaran et al. (1999) allow a balanced degree of homogeneity and heterogeneity assumptions concerning long-run and short-run coefficients and therefore constitute a suitable ground for comparing econometric estimates and averages from National Accounts sources. Table 1.2 shows the estimation results and provides more detailed information on the estimation.

First, it stands out that single country OLS estimates of the ARDL models yield implausible coefficient estimates (see table 1.2). The magnitudes of individual estimates of the labor share do not match the figures computed from National Accounts

[19] E.g. King and Rebelo (1999), p. 954.

[20] A typical argument is that the level or growth rate of technical efficiency constitutes an omitted variable since it is usually not included in estimated equations but highly relevant and likely to be correlated with growth rates of input factors. Therefore, estimated parameters are biased and the contribution of factor accumulation is probably overestimated. In order to defuse the omitted variable problem, the growth rate of technical efficiency (growth rate of total factor productivity) is assumed to follow a linear trend in the estimations below.

[21] See chapter 2 of this thesis for more details on these techniques.

Table 1.2: Factor share estimates from the Cobb-Douglas function

	$\widehat{\alpha}$: Labor elasticity		$\widehat{\beta}$: Capital elasticity	
Individual estimates				
Canada	0.401	(0.354)	0.492	(0.548)
France	0.723***	(0.190)	0.373***	(0.057)
Germany	1.215***	(0.505)	0.579**	(0.257)
Italy	-0.759**	(0.421)	0.782***	(0.323)
Japan	1.011***	(0.403)	0.155	(0.226)
United Kingdom	0.522***	(0.075)	0.330**	(0.177)
USA	0.419**	(0.183)	0.658***	(0.189)
MGE	0.504**	(0.240)	0.481***	(0.081)
PMGE	0.528***	(0.051)	0.343***	(0.037)

Notes: */**/*** denotes significance to the 1%/5%/10% level according to quantiles from the standard normal distribution. Figures in brackets are the standard errors. Mean Group (MG) estimates are average coefficients of individual estimates from Error Correction Models (ECM) corresponding to the following long-run relationship: $y_{it} = a_{it} + \tau_i t + \alpha_i n_{it} + \beta_i k_{it}, \quad i = 1, ..., N, T = 1, ..., T$. Lower case letters denote logarithms. See text for definitions of variables. The Pooled Mean Group (PMG) maximum likelihood estimates are based on heterogeneous short-run dynamics but restrict all the long-run coefficients to be the same across countries. Selection of the lag orders of short-run dynamics of each country is based on the Schwarz Bayesian information criteria with a maximum lag order of four. A likelihood ratio test does not reject the hypothesis of equal long-run coefficients across countries. The seasonally adjusted observations cover the period from the first quarter of 1972 to the last quarter of 2005.
Post estimation diagnostic tests of residuals from PMG estimation do not indicate serial correlation except for Italy where the null of no serial correlation of 4^{th}-order can not be rejected. Ramsey's RESET test using the square of the fitted values is significant for France and insignificant for the other countries. Non-normality is rejected for the residuals of the Pooled Mean Group error correction equations for the United Kingdom and Italy. Italy is also the only country for which residuals are not homoscedastic according to White's heteroscedasticity test. In general, test diagnostics for the residuals of Italy in Pooled Mean Group estimation are poor and do not recommend adopting such an empirical specification for this country whereas for the remaining G7 countries the diagnostics support this kind of model specification.

data and this holds for all countries.[22] In fact, correspondence between econometric estimates and available information on the labor share from National Accounts is

[22] Note that the total amount of labor input enters the estimation equation of the Cobb-Douglas function and therefore the estimate $\widehat{\alpha}$ should be a measure of the adjusted labor share. The estimated capital elasticity $\widehat{\beta}$ refers to the reproducible capital stock only whereas the estimated capitals share from the labor compensation data represents both, reproducible and non-reproducible capital.

almost achieved if coefficients are restricted to be the same across countries. The mean group (MG) estimates and the pooled mean group (PMG) estimates are comparable to the measure of the unadjusted labor share from labor compensation data from National Accounts and somewhat lower than the corresponding adjusted labor share figures.

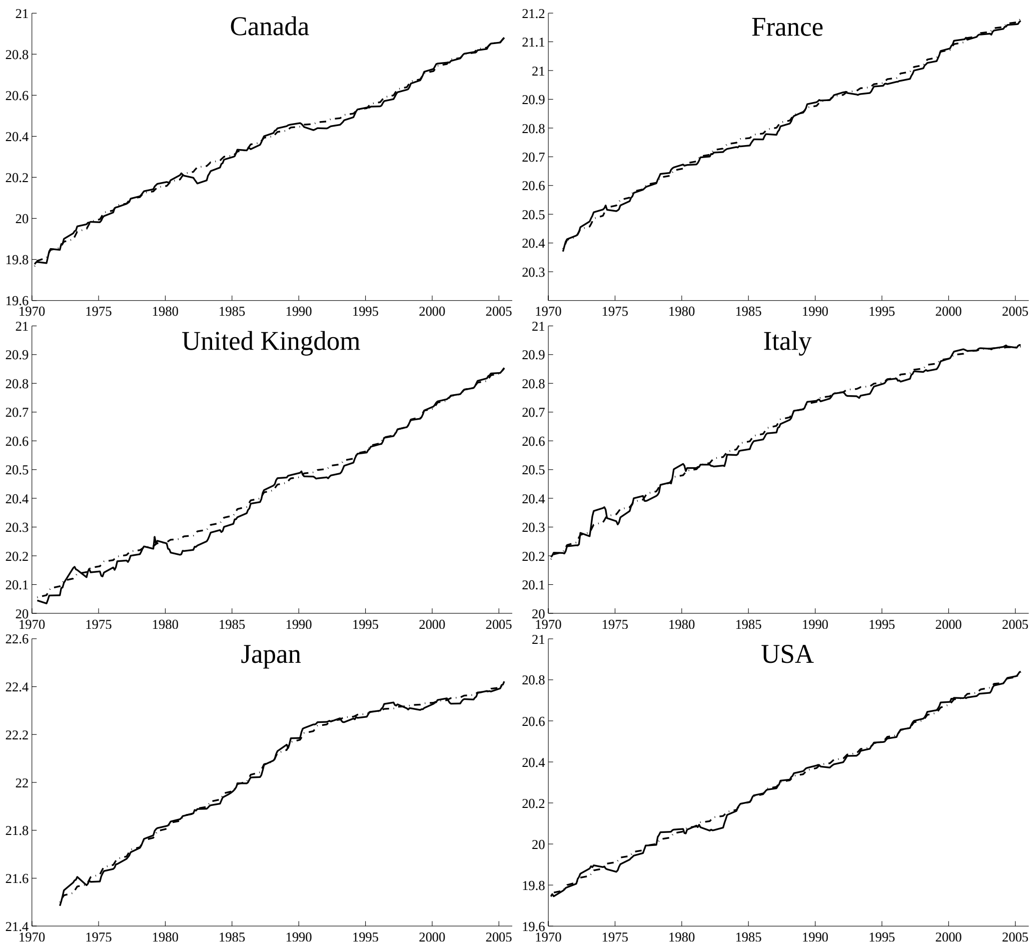
The possibility to test rather than to impose constant returns to scale is an advantage of the econometric approach. Testing constant returns to scale of the MG estimates and the PMG estimates amounts to a test if the respective coefficient estimates of α and β add up to one. Testing these restrictions with the aid of a Wald tests results in a test statistic of 0.01 for the MG estimates and 0.83 for the PMG estimates. According to the critical values from the chi-squared distribution with one degree of freedom, neither of both tests is able to reject the null that the sum of the estimated coefficients of the labor and capital share is one. Consequently, the assumption of constant returns to scale is supported by econometric estimates within the MG and PMG estimation framework. The factor share estimates for Germany, Italy and Japan, however, highlight the difficulties to test the constant returns to scale restriction on the individual country level.

An overall conclusion from the preceding exercise is the following: Econometric support for the usual assumptions of the growth accounting procedures can be provided and econometrically estimated parameters which are broadly in line with the National Accounts data approach can be obtained. However, this works only if one imposes restrictions regarding cross-country parameter homogeneity either by simply averaging individual estimates of long-run coefficients or by imposing the restriction that long-run parameters are the same across countries while short-run parameters are allowed to vary. Single country estimates, however, can yield very implausible parameter estimates (Germany, Japan) or are not able to statistically support Cobb-Douglas technology at all (Italy). The estimates of α according to table 1.1 and table 1.2 demonstrate that not for a single country do both approaches to measure the labor share coincide so it remains a matter of choice which method to use. The National Accounts approach needs assumptions of perfect competition and constant returns to scale while the econometric approach does not rely on these assumptions but needs to impose restrictions with regard to parameter homogeneity across countries in order to produce significant and reasonable results.

For the implementation below, the average adjusted labor share from the National Accounts approach is used for every country, mainly for two reasons. First, using the same value of $\widehat{\alpha} = 0.65$ for the G7 countries seems reasonable since individual estimates do not vary much around the average value. Secondly, the National Accounts approach is the most common proceeding to estimate partial factor elasticities in implementation of the production function approach and the value used here is even in accordance with an often employed rule of thumb.[23]

1.3.2 In-sample estimates of potential output

Figure 1.1 shows the results of the in-sample computation of potential output corresponding to the above outlined production function method in logarithmic form for Canada, France, the United Kingdom, Italy, Japan and the USA. The seasonally adjusted quarterly data is taken from the OECD Economic Outlook database.[24] As can be seen, actual GDP fluctuates more or less symmetrically around its potential level over time.

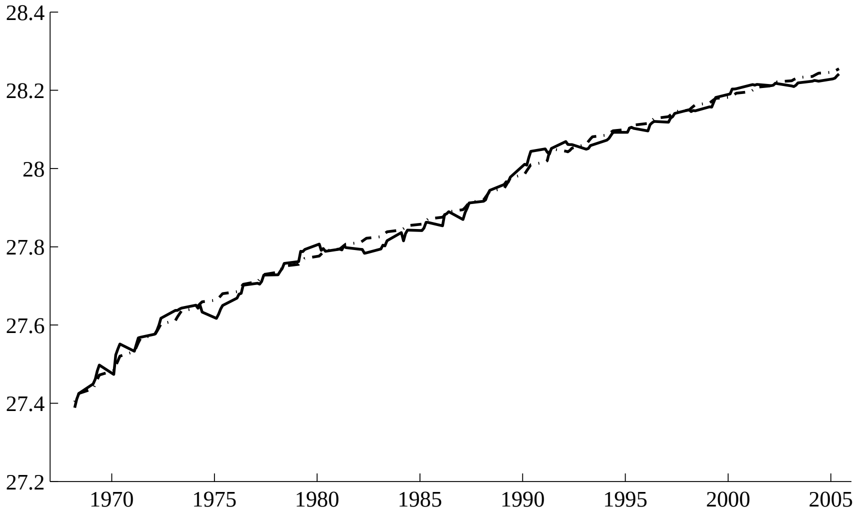


Notes: Dotted lines denote actual GDP while solid lines represent potential GDP.

Figure 1.1: Potential and actual GDP

[23]Another good reason to rely on these National Accounts estimates is the slightly better forecast performance. Using the econometrically estimated value of α according to the PMG estimate ($\widehat{\alpha} = 0.53$) in the out-of-sample forecasts analysis below results in forecasts which are in general worse than forecasts employing $\widehat{\alpha} = 0.65$ with respect to Root Mean Squared Error.

[24]Section 1.5 in the appendix provides more detailed information about the data set.



Notes: See figure 1.1

Figure 1.2: Potential and actual GDP for Germany

A special case: German data and the treatment of the unification break

Two characteristics of the available data for Germany demand a special treatment of the application of the PFA to compute potential output. First, a lack of time series observations for Germany for the periods before 1991 due to the territorial separation within Germany requires a linking of West-German and all-German data which, however, induces a level-break at the time of the German unification. For reasons that become clear below, the out-of-sample analysis is tremendously distorted if potential output is marked by a sizeable level shift. In order to eliminate the reunification break, the first differences of the affected variables have been regressed on an impulse dummy and the level series have been recalculated by integration of the residuals from the dummy regression.[25]

Secondly, data for the German capital stock for the total economy is only available from 1991 onwards whereas data for the capital stock of the private sector is available for West-Germany and Germany over the period from 1960 to 2005. In contrast to the computation of potential output for the other G7 countries, the production function version of Giorno et al. (1995) is used to estimate Germany's

[25] Fritsche and Logeay (2002) use this technique to remove the unification outlier in German data of quarterly GDP growth. Stock and Watson (2003) propose to remove such an outlier by replacing it by the median of the three observations on either side of the observations. Since the results are not very sensitive to the choice between both approaches, the impulse dummy method has been selected and level series have been recalculated with the first observation of the variable in question as starting values. For this reason the resulting artificial level series is the extension of West-German GDP after the unification based on growth rates for all-German data. In this case, economic interpretation of the level of potential output after the first quarter of 1991 is hardly meaningful, however, the proceeding does not constitute a shortcoming for the out-of-sample analysis which focuses on growth rates.

potential output. This alternative computation is identical to the above outlined proceeding with the only difference that it builds on a business-sector production function instead of a total-economy Cobb-Douglas production technology. Within this approach, potential output for the total economy is obtained by adding actual value added in the government sector to potential output of the business sector. Obviously, this implies that output of the government sector equals its potential level throughout. Figure 1.2 shows the path of potential output for Germany.

Output Gap closing assumption and implementation

A concept which is directly linked to potential output is that of the output gap. The output gap is defined as the positive or negative deviation of actual output from potential output and plays an important role for the derivation of medium-term growth projections. A common assumption which draws on mainstream macroeconomic theory is that, in the long run, the path of actual output coincides with the path of potential output. Therefore, sooner or later output will return to potential once deviated from that path. In this regard, the output gap is a measure of how far the economy is currently away of its potential and determines the growth rate that is needed in order to close the output gap over a given period.[26] In practice, this idea is implemented in a rather ad hoc fashion and it is typically assumed that output gradually approaches potential output over the medium-term projection period. Figure 1.3 illustrates these points by stylizing the derivation of projections over the period from T_0 to T_1.

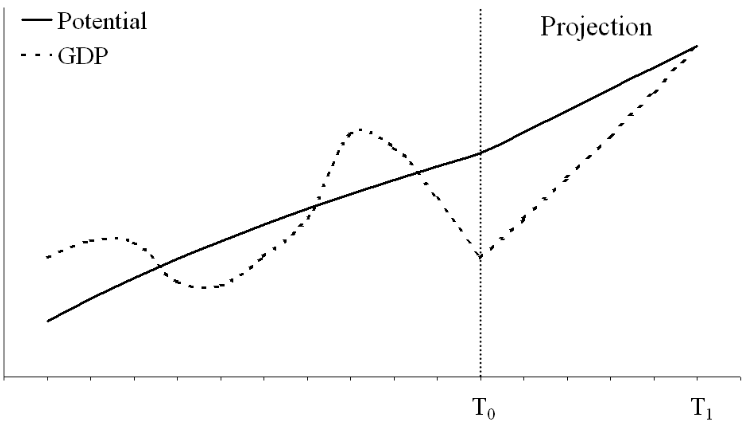


Figure 1.3: Potential output and the output gap

In the beginning period of the projection T_0 , the economy faces a negative

[26]Formally, this assumption implies that the average quarterly growth rate of GDP, necessary to close the Gap over h horizons, is $\overline{g}^h = g^h_{t+1} = \ldots = g^h_{t+h} = (Y^*_{t+h}/Y_t)^{1/h} - 1 \simeq \frac{1}{h}(\ln Y^*_{t+h} - \ln Y_t)$. Y^*_{t+h} is the *level* of potential output after h quarters.

output gap that is closed until the end of the projection period T_1 as actual output converges to potential output. If the starting point of a projection is a negative output gap, it is clear that the resulting growth rates of GDP need to be above the potential growth rate for a prolonged period in order to catch-up with potential growth. GDP evolves in an analogous manner if the output gap is positive at the beginning of the projection period in which case projected growth needs to be beyond potential growth for consecutive periods in order to close the gap from above.

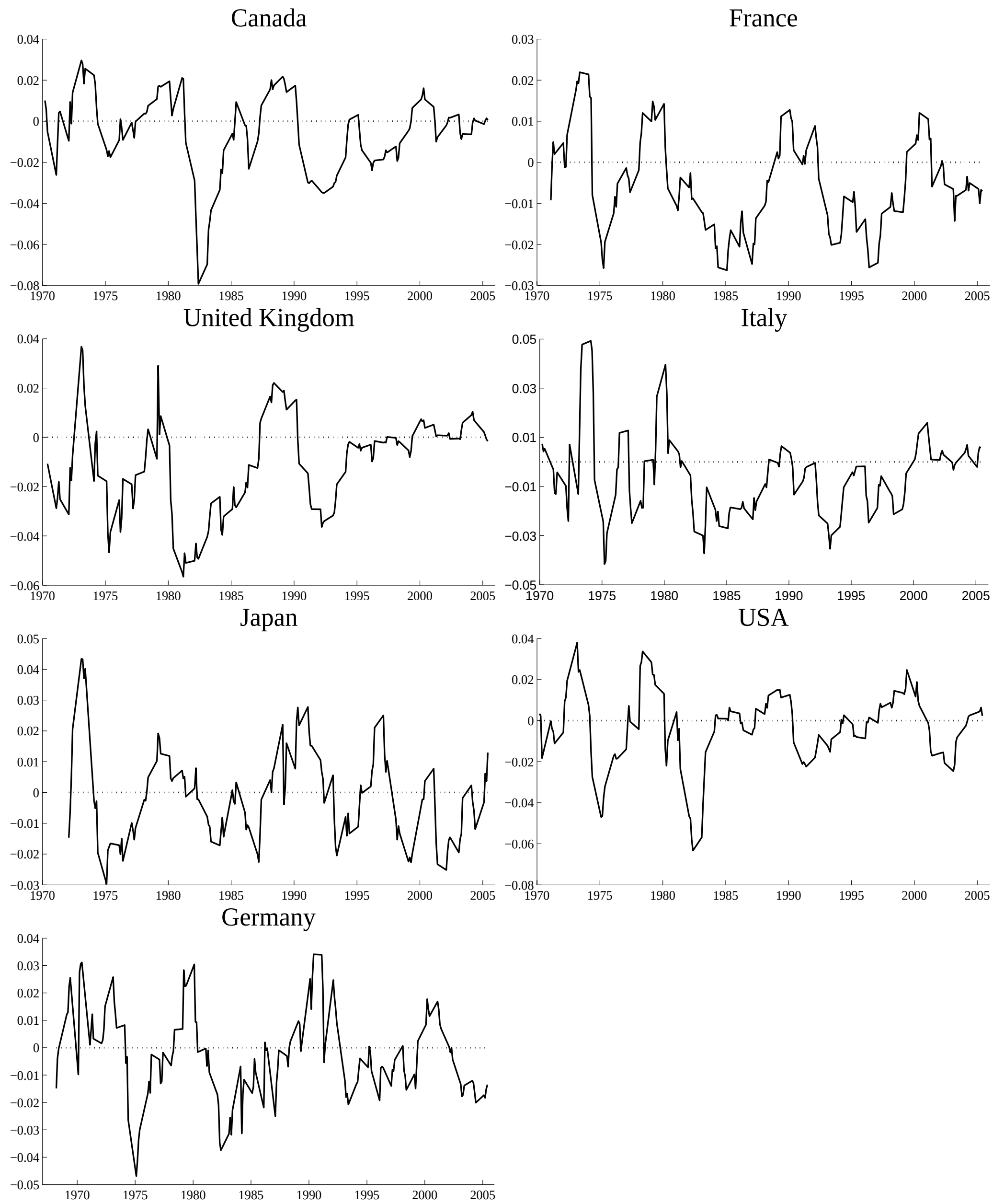


Figure 1.4: Output gaps from the PFA

Obviously, in many respects, such a simplified and stylized scenario of economic dynamics seems to be questionable (Carnot et. al, 2005). The assumption that the catching up process precisely starts at the moment the projection is being prepared, for example, is highly artificial and only by chance will real time dynamics match with such a growth prediction. The output gap may still increase after the beginning period of the projection and close later as assumed resulting in growth dynamics which are fairly different from the predicted ones. Furthermore, it might be more realistic to assume that negative output gaps follow positive output gaps, and vice versa, instead of expecting growth to settle at the potential rate after the gap is vanished. From figure 1.4, which displays the output gaps for the G7 countries corresponding to the PFA method, it can be seen that positive and negative gaps alternate quite frequently in real economies.

There would be more points of criticism to mention, however, given that medium-term projections are not intended for forecasting the cyclical output fluctuations several years ahead, such a simple approach may nevertheless be useful for the prediction of broad future trends. Naturally, the growth rates for the individual years that are derived from the gap closing scenario are not interpretable from a business cycle perspective. In this case, the development of GDP over the entire projection period which results when the economy gradually returns to potential output is the focus of interest and should be referred to for the evaluation of predictive accuracy.

The back-to-trend approach imposes some basic requirements on the output gap that can be readily checked on the basis of an analysis of the gap estimates. Zero mean and stationarity are the most important requirements in order to give empirical support for the assumption that the gap closes automatically. If the gap is non-stationary, there is no guaranty that imbalances unwind and the occurrence of permanent gaps would be possible, although such a behavior would be difficult to justify on theoretical grounds. Table 1.3 displays descriptive statistics of the output gap derived from the PFA. Evidence for stationarity is reported with the aid of standard Augmented Dickey-Fuller tests.[27]

The entries in table 1.3 clearly show that the gap measures for the G7 countries meet this requirement. Apart from the United Kingdom, the estimated gaps are on average very close to zero. Furthermore, the ADF t-statistic is highly significant for all countries and implies stationarity.

Another crucial assumption related to the gap closing scenario concerns the period within which the gap is closed. Usually, this time span is determined by the ending period of the projection and justified rather on practical than on empirical

[27] Elliott et al. (1996) have developed more powerful unit root tests than the standard ADF tests. However, the authors also show that in the case where there is no deterministic component—as is the case in the present test setup—there is no room for improving the power of the Dickey-Fuller t test.

Table 1.3: Properties of the PFA output gap estimates

	CAN	DEU	FRA	GBR	ITA	JPN	USA
Starting period	70Q2	68Q2	71Q1	70Q4	70Q2	72Q1	70Q2
Mean	−0.006	−0.003	−0.005	−0.011	−0.006	−0.002	−0.004
Std. Dev.	0.019	0.016	0.011	0.020	0.016	0.015	0.018
ADF t-statistic	-3.29^{***}	-3.98^{***}	-2.62^{***}	-2.48^{**}	-2.62^{***}	-3.50^{***}	-3.46^{***}
No. of lagged diff.	1	0	4	0	6	0	1
$\widehat{\rho}$	0.90	0.81	0.91	0.92	0.86	0.83	0.88
Av. duration of gap (in years)	2.61	2.77	3.84	3.84	3.50	2.55	2.48

Notes: All observations end in the last quarter of 2005. The starting quarters vary between countries as indicated in the table. The Augmented Dickey-Fuller (ADF) tests have been conducted without deterministic terms in the estimation equations. The number of lagged difference terms of the ADF test were chosen with the aid of the modified Akaike information criterion and the maximum lag length has been set to 12 throughout. $^{*}/^{**}/^{***}$ denotes significance to the 1%/5%/10% level according to MacKinnon's (1996) one-sided p-values. $\widehat{\rho}$ is the estimate of the autoregressive coefficient from the ADF regression. The average duration of the output gap is the number of consecutive quarters in which the output gap was either positive or negative whereas durations less or equal to 4 quarters have been excluded from the calculation.

grounds. Since one is interested in the growth projection over the entire period, results do not change if the actual output returns sooner than assumed to its potential level and subsequently evolves with the potential growth rate. However, if the gap typically closes later than assumed, the back-to-trend scenario yields a predicted overall growth rate which is no longer in line with the actual development. The question whether it is realistic to assume periods of 3 to 5 years for closing the gaps should also be answered empirically.

Two statistics in table 1.3 assess the typical duration of a negative or positive output gap. The first statistic is the estimated autoregressive coefficient $\widehat{\rho}$ from the ADF test regressions. This coefficient informs about the persistence of the output gap time series. The second statistic is a measure of the average duration of the output gap and is based on a simple counting of the number of consecutive quarters in which the gap estimate does not change its sign. The autoregressive coefficients are in the range of 0.81 to 0.92 and point to rather persistent output gaps. This impression is also conveyed by graphical inspection of the historical evolution of the gap measures (see figure 1.4). The implication of the $\widehat{\rho}$ estimates can be illustrated with the aid of the following example: Consider an AR(1) model for the output gap of Germany and assume that the economy is hit by a positive shock which leads to a deviation of actual output from potential output. In the absence of other shocks, an autoregressive coefficient of $\widehat{\rho} = 0.81$ implies that more than 95% of the gap will be closed after 16 quarters. While such a hypothetical example helps to illustrate the dynamics inherent to the gap estimates, however, past output gaps exhibited rather individual patterns and varying duration times.

From the counting exercise follows that, on average, the duration of the gaps for the seven countries was from 2.48 years (USA) to 3.84 years (France and Great Britain). At the same time, the series depicted in figure 1.4 also show that output gaps can last for several years. Marked examples are the pronounced negative output gaps at the beginning of the eighties for France and the United Kingdom, which had lengths of 8.5 and 7.5 years, respectively. However, these periods are exceptional cases and the overall conclusion from the duration analysis is that, although artificial, the restriction that a gap is closed after 3 or 5 years (depending on the projection horizon) is not too far from reality and may serve as an acceptable scenario in the absence of alternatives.

Forward-looking assessment of potential output

The production function specifies the main components that determine potential output. In order to derive GDP projections, the future prospects of potential output have to be assessed. Typically, this task is accomplished by extrapolating the key variables from past trends, however, it is also the stage of the projection process where judgemental adjustments usually enter the quantitative estimation by deciding whether historical trends can be sustained over the projection period, or whether they should be adjusted on the grounds of additional information coming from outside the PFA framework. A neutral scenario (baseline scenario), which incorporates a no-change assumption of the evolution of the key components builds a natural starting point for alternative scenarios in order to illustrate the range of possible outcomes and to demonstrate the uncertainties inherent to the projection.

In the out-of sample experiment of section 1.3.3 a neutral scenario for the projection of potential output has been chosen. The following list explains which assumptions have been made and how forecasts for the individual inputs to the computation of a forward projection of potential output have been generated (recall equations (1.1) to (1.7) from above).[28] Note that such an analysis has to take account of the real-time characteristic of the sample data, i.e only information that could have been known to the forecaster at the time the pseudo-forecast is produced should be employed for the prediction of subsequent potential output.

- The **Total Factor Productivity** is estimated as the Solow residual corre-

[28]These assumptions mainly follow the proceedings documented in Carnot et al. (2005), p. 163-64 and Denis et al. (2002), p. 22-23.

sponding to equation (1.4) and extended over the projection horizon with the aid of ARIMA-model forecasts. The HP-filter is applied afterwards in order to obtain a trend value of TFP that can be fed into the Cobb-Douglas production function.[29]

- The interdependence between GDP growth and capital investment makes it difficult to derive projections for the **capital stock** from a theoretical point of view. However, given the smooth trending behavior of the capital stock data one typically observes, predicting this input variable econometrically is straightforward. Also ARIMA-model forecasts that are smoothed with the HP-filter are employed for a forward projection of this component.

- Extending the number of **working age population** over the projection horizon is done with the aid of actual population data. No forecast is used for this variable since reliable projections of population data over medium-term horizons are typically readily available from demographic surveys to the forecaster.[30]

- The extrapolation of the trend **participation rate** and the trend in **hours worked** is also carried out with the aid of ARIMA-model forecasts and the HP-filter. In practice, projecting the future evolution of these variables is typically based on extra information about whether past trends are maintained over the projection horizon or whether trend changes are likely. However, such a proceeding is not feasible in the recursive out-of sample analysis.

- The **NAWRU**, which is taken from OECD sources, is assumed to evolve unchanged from its last value at the period when the projection starts. For lack of alternative information, a flat extrapolation of the NAWRU seems to be most consistent with the notion of a stable long-run unemployment rate.

This section finishes the description of the implementation of the PFA. Again, it should be stressed at this point that it is not the aim of this chapter to investigate the general theoretical suitability of the PFA for estimating potential output, but to check the predictive performance of a method that is so ubiquitous in policy analysis.

[29]The lag selection of the ARIMA models have been specified by means of the Schwarz's Bayesian information criterion throughout. The maximum lag length was 4 quarters for all series. The models have been estimated with the aid of the MATLAB function *armaxfilter* from Kevin K. Sheppard's GARCH toolbox.

[30]Lindh (2004) explains in more detail the uncertainties related to demographic projections which essentially concern mortality, fertility and migration. All in all he concludes that the first 5 or 10 years of a demographic projection are fairly reliable with respect to forecast error compared to standard projections of economic variables.

1.3.3 Multi-step forecasts and analysis of errors

For the analysis of forecast errors from the out-of sample experiments, a framework inspired by the work of Brown and Maital (1981), Keane and Runkle (1990), Davies and Lahiri (1995) and Clements et al. (2007) is employed to derive the covariance structure of cumulative forecast errors. It is shown that this particular framework has advantages in small samples over the approaches usually employed to inference in forecast error analysis.

The analysis of forecast errors is based on cumulative forecasts of quarterly differences of the logarithm of GDP and the corresponding realized log-differences. The design of the forward looking analysis is as follows:

- The total number of observations is T. An initial sample of observations is chosen, say, from the first observation to t^* with $t^* < T$. The PFA is employed to produce h forecasts for the growth rate of GDP based on this sample. These multi-step forecasts over the periods from $t^* + 1$ to $t^* + h$, given information available at time t^*, are denoted as $\Delta y_{t^*+1|t^*}, \Delta y_{t^*+2|t^*}, ..., \Delta y_{t^*+h|t^*}$.
- Next, the h multi-step forecasts are cumulated to $F^h_{t^*} = \Delta y_{t^*+1|t^*} + \Delta y_{t^*+2|t^*} + ... + \Delta y_{t^*+h|t^*} = \sum_{i=1}^{h} \Delta y_{t^*+i|t^*}$ to yield medium-term forecasts of GDP growth. Also, the quarterly growth rates of actual GDP, $\Delta y_{t^*+1}, \Delta y_{t^*+2}, .., \Delta y_{t^*+h}$ are summed up to $A^h_{t^*} = \sum_{i=1}^{h} \Delta y_{t^*+i}$.
- Forecast errors are computed:

$$e^h_{t^*} = A^h_{t^*} - F^h_{t^*} = \sum_{i=1}^{h} \Delta y_{t^*+i} - \sum_{i=1}^{h} \Delta y_{t^*+i|t^*} \tag{1.8}$$

- The sample is expanded by one quarter, i.e. the next forecasts are conducted as $F^h_{t^*+1} = \sum_{i=1}^{h} \Delta y_{t^*+1+i|t^*+1}$ and errors are obtained as $e^h_{t^*+1} = A^h_{t^*+1} - F^h_{t^*+1}$.
- The procedure is iterated until $t^* + j = T - h$, $j = 0, 1, \ldots, T^*$, $T^* = T - t^* - h$.

In order to execute a test for forecast unbiasedness, the correlation structure of the forecast errors induced by the overlapping nature of the forecasting procedure needs to be derived. The following error components model will therefore be helpful. It is assumed that the errors as depicted in equation (1.8) have the following structure:[31]

[31] Davies and Lahiri (1995) use such a model to analyze forecast errors in a panel data setting using professional forecasts. Clements et al. (2007) build on this model to test whether forecasts of the Federal Reserve are systematically biased and efficient. The framework allows them to pool information over horizons and represents an analogue application to the forecast errors analysis in the present chapter.

$$e_t^h = A_t^h - F_t^h = \sum_{i=1}^{h} u_{t+i} + \phi = \nu_t^h + \phi, \quad t = t^*, \dots, T-h \tag{1.9}$$

According to this model, the forecast errors of GDP growth over h horizons are the sum of the cumulative effect of all disturbances to the growth rate that occurred between period t and $t+h$ and a bias term which is given by ϕ. This error model is consistent with rational forecasts if the bias term is omitted since from that it follows that $E[e_t^h] = 0$. Thus, a test for unbiased forecasts employs the null hypothesis that $\phi = 0$ in a regression based on equation (1.9).

Assuming rationality of forecasts and i.i.d. disturbances gives $E[u_t] = 0$, $E[u_t^2] = \sigma_u^2$ and $E[\nu_t^h] = 0$. The cumulative forecasts are overlapping and therefore induce serial correlation among forecast errors in different periods since adjacent forecasts share a common subrange, determined by the difference in time of the two errors in which they share the same disturbances (cf. Davies and Lahiri, 1995 or Brown and Maital, 1981). From equation (1.9) it follows that

$$\begin{aligned} E[(\nu_t^h)^2] &= h\sigma_u^2 \\ E[\nu_t^h \nu_{t+k}^h] &= \begin{cases} (h-|k|)\sigma_u^2 & \text{for} \quad k = -(h-1), \dots, 1, \dots, h-1 \\ & \text{and} \quad t+h > t+k > t-h \\ 0 & \text{else} \end{cases} \end{aligned}$$

Therefore, rather than being diagonal, the variance matrix $E[\nu^h \nu^{h\prime}] = \Sigma_\nu$ takes the following block diagonal form: [32]

$$\underset{(T^* \times T^*)}{\Sigma_\nu} = \sigma_u^2 A \tag{1.10}$$

with

$$A = \begin{pmatrix} a^{(0)} & a^{(1)} & \cdots & a^{(h-1)} & 0 & \cdots & & & \cdots & 0 \\ a^{(1)} & a^{(0)} & a^{(1)} & \cdots & a^{(h-1)} & 0 & \cdots & & & \vdots \\ \vdots & a^{(1)} & a^{(0)} & a^{(1)} & \cdots & a^{(h-1)} & 0 & \cdots & & \\ a^{(h-1)} & \cdots & a^{(1)} & a^{(0)} & a^{(1)} & \cdots & a^{(h-1)} & 0 & \cdots & \\ 0 & a^{(h-1)} & \cdots & a^{(1)} & a^{(0)} & a^{(1)} & \cdots & a^{(h-1)} & 0 & \\ \vdots & 0 & a^{(h-1)} & \cdots & a^{(1)} & a^{(0)} & a^{(1)} & \cdots & a^{(h-1)} & 0 \\ & & 0 & a^{(h-1)} & \cdots & a^{(1)} & a^{(0)} & a^{(1)} & \cdots & a^{(h-1)} \\ & & & 0 & a^{(h-1)} & \cdots & a^{(1)} & a^{(0)} & a^{(1)} & \vdots \\ \vdots & & & & 0 & a^{(h-1)} & \cdots & a^{(1)} & a^{(0)} & a^{(1)} \\ 0 & \cdots & & & & 0 & a^{(h-1)} & \cdots & a^{(1)} & a^{(0)} \end{pmatrix} \tag{1.11}$$

[32] $\nu^h = (\nu_t^h, \nu_{t+1}^h, \dots, \nu_{t+T^*}^h)$ is the vector that contains the stacked cumulative shocks.

$$a^{(k)} = (h-k), \quad k = 0, \cdots, h-1$$

From (1.10) it is apparent that only in the case of a one-step ahead forecast ($h = 1$) are the errors ν_t^h serially uncorrelated. The variance-covariance specification is very parsimonious since it depends only on one unknown parameter, σ_u^2, which can be estimated as shown below.

Test of bias in cumulative forecasts

The following test of unbiasedness has its origins in the work of Mincer and Zarnowitz (1969) and Holden and Peel (1990). A test of weak rationality amounts to a test of forecast unbiasedness in (1.9), where

$$H_0: \quad \phi = 0 \tag{1.12}$$

The test statistic of interest is

$$t_\phi = \frac{\widehat{\phi}}{\widehat{\sigma}_\phi} \tag{1.13}$$

with

$$\widehat{\phi} = \frac{1}{T^*} \sum_{t=t^*}^{T-h} e_t^h \tag{1.14}$$

and the consistent covariance matrix estimator

$$\widehat{\sigma}_\phi^2 = (X'X)^{-1} X' \widehat{\Sigma}_\nu X (X'X)^{-1} = \frac{1}{T^{*2}} i'_{T^*} \widehat{\Sigma}_\nu i_{T^*} \tag{1.15}$$

and $X = i_{T^*}$ with i_{T^*} as a vector of ones with dimension T^*.[33] The expressions (1.14) and (1.15) constitute a feasible estimation since the covariance matrix Σ_ν depends only on one unknown parameter which can readily be obtained. $\widehat{\Sigma}_\nu$ is constructed according to (1.10) with an estimate of the average quarterly disturbance variance. This can be obtained in the following way. Let $\widehat{\nu}^h = (\widehat{\nu}_t^h, \widehat{\nu}_{t+1}^h, \ldots, \widehat{\nu}_{t+T^*}^h)$ be a vector that encloses estimates of ν_t^h which are the computed deviations of each forecast error from the bias estimate $\widehat{\phi}$. Since $E[\nu^h \nu^{h\prime}] = \sigma_u^2 A$, an estimate of the disturbance variance is given by[34]

$$\widehat{\sigma}_u^2 = \frac{1}{T^*} \widehat{\nu}^{h\prime} A^{-1} \widehat{\nu}^h \tag{1.16}$$

We refer to the above outlined approach as generalized least squares (GLS)

[33] Cf. Clements et al. (2007).

[34] This result uses the fact that the trace tr of a scalar is the scalar. It holds that $tr(\sigma_u^2 I) = \sigma_u^2 T = E[tr(\nu\nu' A^{-1})] = E[tr(\nu' A^{-1} \nu)] = E[\nu' A^{-1} \nu]$, whereas I is the identity matrix. Replacing population moments with sample moments gives equation (1.16).

framework although simple averaging (OLS) is used to estimate the bias term ϕ.[35] The focus of interest is rather on the GLS standard errors as given by equation (1.15).

Table 1.4: Size properties of Newey-West based tests of forecast unbiasedness

	$T = 120$			$T = 100$			$T = 80$		
h	1%	5%	10%	1%	5%	10%	1%	5%	10%
4	4.68	12.43	19.39	4.95	13.01	19.94	5.45	13.58	20.63
8	6.74	15.08	22.21	7.37	16.06	23.29	8.58	17.60	24.97
12	8.85	18.11	25.36	10.27	19.45	26.78	12.57	22.25	29.49
16	11.19	20.61	27.76	13.18	23.00	30.41	16.87	26.88	34.34
20	13.89	23.69	30.98	16.94	27.24	34.62	22.17	32.67	39.77

Notes: The effective sample size is $T - h$. For each forecast step and sample size, 100000 replications of experimental data following the stochastic process as given by equation (1.9) have been generated. The disturbances u_t are individually distributed $\mathcal{N}(0, 1)$ and $\phi = 0$ has been set throughout in order to obtain data that represent unbiased forecasts. The HAC estimator is based on Bartlett kernel weights and a truncation lag of $h - 1$. Entries denote rejection frequencies at nominal significance levels of 1%, 5% and 10%. Computational work was performed in MATLAB.

A common approach to take serial correlation in a test of unbiasedness into account is to apply the standard errors of Newey and West (1987) which correct for autocorrelation and heteroscedasticity. Bartlett weights in the formula for the Newey-West covariance matrix ensure that the matrix is positive definite but are also meant to model the declining influence of autocorrelations as the separation of observation pairs in time grows. The decline of the autocorrelations of forecast errors as the distance between them grows larger is the key feature of the overlapping nature of the forecast error analysis.[36] The matrix in (1.10) clearly illustrates this. However, the appendix demonstrates that the use of kernel weights in the HAC estimator is not appropriate in a test of weak rationality when the forecast errors follow (1.9). This estimator has difficulties in capturing the correct standard errors in finite samples. Table 1.4 shows results of the size properties of the Newey-West t-statistic in a test under the null hypothesis of unbiased multi-step forecasts provided

[35] Both the OLS and GLS estimators are known to be consistent, however, the latter is more efficient than the former. Yet we prefer to compute the bias estimate with the aid of OLS since the GLS estimator in fact minimizes a *weighted* sum of squared errors, which in contrast to the simple average sum of squared errors has the disadvantage that it does not possess an intuitive and straight interpretation in the present application.

[36] Cf. Clements (2005), p. 7-9, for an illustration of the application of Newey-West covariance matrix estimation techniques in the context of rationality tests of multi-step forecasts.

by a Monte Carlo experiment.[37] The entries of table 1.4 display the empirical sizes of Newey-West's t-statistics for a test of $\widehat{\phi} = 0$ for various forecast steps h and sample sizes T, whereas the chosen quantities for h and T are of the same magnitude as the forecast horizons and the observation numbers in the subsequent out-of sample analysis. The experimental data is generated under $\phi = 0$. Empirical sizes of the GLS based test for unbiasedness are not reported since these appeared to be identical to the nominal sizes throughout. The entries in the table make it clear that the Newey-West based test is heavily oversized as soon as $h > 0$ and the size distortion increases with h and declining T.

Test of forecast accuracy

A convenient framework to test for forecast accuracy was introduced by Diebold and Mariano (1995, DM) while Harvey et al. (1997) enhanced it to improve the test performance in small samples. The DM-test is based on a forecast error loss differential. Following a usual convention, a quadratic loss differential is used below in order to test whether the forecasts from the production function model and the forecasts from the random walk model have equal accuracy. Medium-term projections of GDP growth have positive value if they predict the economic development better than naïve forecasts. Besides, using a quadratic loss function in the present context is adequate since negative and positive forecast errors should be given the same weight while larger forecast errors in absolute value should be given higher weight than smaller errors for the purpose of evaluating the accuracy.

The motivation and derivation of the test of forecast accuracy is as follows. Consider two forecast error series $\tilde{e}_t^h$ and $\overline{e}_t^h$ originating from two different forecast models that share the same target. In this case, the average of the quadratic loss differential is given by:

$$d = \frac{1}{T^*} \sum_{t=1}^{T} d_t^h, \tag{1.17}$$

and

$$d_t^h = (\tilde{e}_t^h)^2 - (\overline{e}_t^h)^2 \tag{1.18}$$

whereas it is assumed throughout that the errors individually follow the component model introduced above:

[37] Cooper and Priestley (2006) and Ang and Baekert (2006), for example, show in a similar test-setup that Newey-West t-statistics can lead to size distortions of tests for stock return predictability when using overlapping observations.

$$\tilde{e}_t^h = \sum_{i=1}^{h} \tilde{u}_{t+i} + \tilde{\phi} = \tilde{\nu}_t^h + \tilde{\phi}, \quad E[\tilde{e}_t^h] = \tilde{\phi}, \quad V[\tilde{e}_t^h] = h\sigma_{\tilde{u}}^2 \tag{1.19}$$

$$\overline{e}_t^h = \sum_{i=1}^{h} \overline{u}_{t+i} + \overline{\phi} = \overline{\nu}_t^h + \overline{\phi}, \quad E[\overline{e}_t^h] = \overline{\phi}, \quad V[\overline{e}_t^h] = h\sigma_{\overline{u}}^2 \tag{1.20}$$

The test statistic of interest is given by

$$DM = \frac{d}{\sqrt{\widehat{V}(d)}} \tag{1.21}$$

$\widehat{V}(d)$ is the estimated variance of d, including any autocovariances $\widehat{\gamma}_d(k)$ of d at displacement k. Following DM, the variance of d in the presence of overlapping forecasts over h horizons is given by:[38]

$$\widehat{V}(d) = \frac{1}{T^*} \sum_{k=-(h-1)}^{(h-1)} \widehat{\gamma}_d(k) \tag{1.22}$$

and $\widehat{\gamma}_d(k)$ is the estimated autocovariance of d. DM propose to estimate (1.22) with the aid of a weighted sum of sample autocovariances as in the work of Newey and West (1987). In applied work, this is the most conventional approach to obtain an estimate of $V(d)$.[39] However, having stated an explicit model for the forecast errors of interest, derivation of the exact variances and covariances is straightforward and should help to improve the small sample problems inherent to the latter method. Consider the error models (1.19) and (1.20) with bias terms $\tilde{\phi}$ and $\overline{\phi}$. In the case that $Cov(\tilde{u}_t, \overline{u}_t) = 0$ and under the assumption that quarterly shocks $\tilde{u}_t, \overline{u}_t$ are normally distributed, the following expression for the variance of d_t^h results:[40]

$$\gamma_d(0) = V[(d_t^h)] = V[(\tilde{e}_t^h)^2] + V[(\overline{e}_t^h)^2] = 2h\sigma_{\tilde{u}}^2(h\sigma_{\tilde{u}}^2 + 2\tilde{\phi}^2) + 2h\sigma_{\overline{u}}^2(h\sigma_{\overline{u}}^2 + 2\overline{\phi}^2) \tag{1.23}$$

However, the assumption of uncorrelated disturbances resulting from two forecast models that have the same target is not realistic. Dependence arises since the

[38] Cf. Diebold and Mariano (1995), p. 135.

[39] Since the test statistic is known to be oversized in small samples, Harvey et al. (1997) propose to augment the Diebold-Mariano test with a corrective factor, which is given by $K = \sqrt{(T^* + 1 - 2h + h(h-1)/T^*)/T^*}$ which leads to the modified DM test $mDM = K \cdot DM$. The authors also demonstrate that the power of the test is improved when critical values of the Student t distribution are used.

[40] If $a \sim \mathcal{N}(\mu, \sigma^2)$, then $\left(\frac{a-\mu}{\sigma}\right)^2 \sim \chi^2(1)$. Since a Chi-squared distributed random variable with one degree of freedom has an expected value of 2, it follows that $V\left[\left(a^2 - 2a\mu + \mu^2\right)\right] = 2\sigma^4$. From the properties of the variance of sums it is apparent that $V[a^2] = 2\sigma^4 + 4\sigma^2\mu^2 = 2\sigma^2(\sigma^2 + 2\mu^2)$.

forecast errors share macroeconomic shocks that are in general not predictable. In order to account for the presence of quarterly disturbances that are common to both forecast errors, the covariance of $\tilde{u}_t$ and $\overline{u}_t$ needs to be included in equation (1.23). Taking $Cov(\tilde{u}_t, \overline{u}_t) = \sigma_{\tilde{u},\overline{u}}$ into account leads to:

$$Cov\left[(\tilde{e}_t^h)^2, (\overline{e}_t^h)^2\right] = 2h\sigma_{\tilde{u},\overline{u}}(h\sigma_{\tilde{u},\overline{u}} + 2\tilde{\phi}\overline{\phi}) \tag{1.24}$$

Combining these results and rearranging expressions produces the following formula for the variances and autocovariances of the quadratic loss differential:[41]

$$\begin{aligned} \gamma_d(k) &= E[d_t^h d_{t-k}^h] - E[d_t^h]E[d_{t-k}^h] = \\ &= 2\check{h}\left(\sigma_{\tilde{u}}^2(\check{h}\sigma_{\tilde{u}}^2 + 2\tilde{\phi}^2) + \sigma_{\overline{u}}^2(\check{h}\sigma_{\overline{u}}^2 + 2\overline{\phi}^2) - 2\sigma_{\tilde{u},\overline{u}}(\check{h}\sigma_{\tilde{u},\overline{u}} + 2\tilde{\phi}\overline{\phi})\right) \\ & \check{h} = h - |k| \end{aligned} \tag{1.25}$$

Replacing population moments with sample moments in equation (1.25) yields an applicable expression for the variance estimate of d. The variances of $\tilde{u}_t$ and $\overline{u}_t$ can be estimated like in equation (1.16) while the covariance is estimated analogously as follows

$$\widehat{\sigma}_{\tilde{u},\overline{u}} = \frac{1}{T^*}\widehat{\tilde{\nu}^{h}}' A^{-1}\widehat{\overline{\nu}^h} \tag{1.26}$$

Estimates of $\tilde{\phi}$ and $\overline{\phi}$ can be obtained by following (1.14).[42]

The analogy to the test of forecast unbiasedness is obvious: Performing the DM test is identical to running the regression $d_t^h = \alpha + \varepsilon_t$ and to computing the consistent t-statistic of $\widehat{\alpha}$. Furthermore, computing $\widehat{V}(d)$ after equation (1.22) is the same as computing $\widehat{V}(d) = \frac{1}{T^*} i'_{T^*}\widehat{A}i_{T^*}$ with $\widehat{A}$ being of the form as shown by equation (1.11), whereas in this case the individual elements of A, $a^{(k)}$, are replaced with estimates of the sample autocovariances $\widehat{\gamma_d}(k)$.

In the following, the finite sample size of the test statistic for equal forecast accuracy vis-à-vis the conventional modified DM test which estimates $\widehat{V}(d)$ with the aid of Newey-West HAC covariances is assessed on the grounds of a Monte Carlo analysis. Size distortions of various tests for forecast accuracy based on HAC estimators in small samples are well documented in the work of Clark (1999). This study, however, considers only one- and two-step ahead forecasts while forecast horizons are much larger in the present out-of-sample exercise. The designs of the subsequent experiment under the null hypothesis of equal forecast accuracy is as

[41] This result is established more rigorously in appendix 1.5, page 59.

[42] Note that it is not appropriate to perform the test of accuracy with the aid of bias-removed forecasts. The consideration of both elements— forecast bias and error variance—is just the central feature of this test.

follows: First, two unbiased forecast error series are drawn from a bivariate standard normal distribution and the desired degree of contemporaneous correlation among the two error series is imposed.[43] Then these forecast errors are cumulated over various horizons and afterwards the modified DM test and the test as described by equations (1.21),(1.22) and (1.25) are performed for sample sizes of $T = 80, 100$ and 120 as well as contemporaneous correlations of $\rho = 0.5$ and 0.9.[44] The test statistic of the latter is computed with sample estimates of the variances and covariance $\widehat{\sigma}^2_{\tilde{u}}, \widehat{\sigma}^2_{\overline{u}}$ and $\widehat{\sigma}_{\tilde{u},\overline{u}}$.

In view of the entries of table 1.5 it is apparent that the HAC covariance based test is oversized and the size distortion has the same magnitude for all sample sizes and horizons. In contrast to that, the GLS based tests seem to have good size properties, but tend to be slightly undersized for tests at the 10% level. Note that the effective sample size depends on h, i.e. the number of observations which are actually feasible for computing the estimates is $T - h$.

The overall impression from the experiment is that, on balance, the GLS based tests appear to have the best properties. In absolute value, the size distortions of the GLS test are smaller than the distortions of the HAC based test, even if the small sample adjustment of Harvey et al. (1997) is taken into account. Again, it is worth emphasizing that the GLS test outlined above only relies on estimates of the variance of the error components σ^2_u and on an estimate of the bias term ϕ for the respective error series and on the covariance between the two series. Thus, these test procedures build on very parsimonious parameter specifications, and according to the Monte Carlo evidence, come up with favorable characteristics in small samples.

Although the PFA to produce medium-term forecasts is model driven and exact variances of forecast errors would in principle be feasible, the tests for forecast unbiasedness and accuracy outlined above have advantages for several reasons. First, derivation of exact forecast-error variances for the production function approach which involves separately estimated variables like the NAWRU seems to be difficult if not impossible. Building around the outlined model of forecast errors can circumvent the difficult task of delivering exact analytical error covariances. Secondly, the approach is parsimonious in terms of parameters involved, simple to compute and takes the exact structure of the error correlation from overlapping forecasts into account. Finally, it seems to be a good alternative to the usually employed non-parametric heteroscedasticity and autocorrelation consistent (HAC) estimators which are known to suffer from size distortions in small samples.

[43] The desired correlation is achieved by premultiplication of the original error series with the Choleski factor of the required correlation matrix. Cf. Diebold and Mariano (1995), p.138, for details.

[44] It is worth mentioning that the case of $\rho = 0.9$ is particularly relevant for the present analysis of GDP growth which exhibits strong correlations among errors from different models.

Table 1.5: Size properties of tests for equal forecasts accuracy

		$\rho = 0.5$								
		$T = 120$			$T = 100$			$T = 80$		
	h	1%	5%	10%	1%	5%	10%	1%	5%	10%
HAC	4	2.04	8.33	15.04	1.99	8.39	15.20	2.07	8.34	15.13
	8	2.32	8.91	15.91	2.21	8.72	15.77	2.13	8.72	15.73
	12	2.43	8.99	16.07	2.43	8.91	15.93	2.51	8.66	15.56
	16	2.55	8.96	15.90	2.55	8.92	15.72	2.66	8.43	14.98
	20	2.61	8.79	15.73	2.67	8.43	14.85	2.82	7.99	13.77
GLS	4	0.88	4.48	9.33	0.85	4.47	9.23	0.83	4.45	9.11
	8	1.10	4.70	9.19	1.12	4.76	9.07	1.08	4.51	8.72
	12	1.20	4.74	9.04	1.26	4.74	8.97	1.29	4.69	8.56
	16	1.29	4.83	8.92	1.41	4.85	8.76	1.37	4.56	8.25
	20	1.44	4.85	8.68	1.44	4.76	8.35	1.51	4.71	8.14
		$\rho = 0.9$								
		$T = 120$			$T = 100$			$T = 80$		
	h	1%	5%	10%	1%	5%	10%	1%	5%	10%
HAC	4	2.10	8.30	14.86	2.04	8.21	14.94	2.09	8.37	15.16
	8	2.31	8.91	15.97	2.42	8.99	15.90	2.23	8.78	15.85
	12	2.40	8.94	16.05	2.39	8.85	15.92	2.45	8.62	15.70
	16	2.53	8.94	15.96	2.53	8.77	15.61	2.62	8.49	14.95
	20	2.68	8.81	15.54	2.80	8.66	15.11	2.85	8.03	13.83
GLS	4	0.87	4.49	9.22	0.88	4.43	9.09	0.87	4.44	9.01
	8	1.09	4.80	9.34	1.11	4.80	9.22	1.07	4.61	8.84
	12	1.24	4.78	9.09	1.24	4.72	8.92	1.28	4.64	8.61
	16	1.32	4.81	8.89	1.41	4.76	8.67	1.38	4.77	8.52
	20	1.49	4.88	8.79	1.58	4.97	8.67	1.57	4.71	8.24

Notes: *HAC* denominates the tests that are based on non-parametric HAC estimates of the modified test statistic mDM and *GLS* denotes the corresponding estimates that build on the covariance estimator according to equation (1.25). The effective sample size is $T - h$. For each forecast step and sample size, 100000 replications of experimental data following the stochastic process as given by equations (1.19) and (1.20) with $\tilde{\phi} = \overline{\phi} = 0$ have been generated. The disturbances $\tilde{u}_t$ and $\overline{u}_t$ are first drawn from a bivariate standard normal distribution and then the contemporaneous correlation of ρ has been imposed. These experimental data represent forecasts of same accuracy. The HAC estimator is based on Bartlett kernel weights and a truncation lag of $h-1$. Entries denote rejection frequencies at nominal significance levels of 1%, 5% and 10%. Computational work was performed in MATLAB.

1.4 Out-of-sample results

This section presents the empirical results of the out-of-sample analysis and compares the pseudo forecasts with corresponding projections from official institutions for the respective country. The data used to implement the PFA as well as the projections from official sources are explained in section 1.5 in the appendix.

All country tables shown below have an identical structure. For each of the three, four and five year forecast horizons, these tables show the key measures of forecast performance of the different forecast models. In addition to the forecasts that arise from the gap-closing scenario as outlined above (PFA, gap closing), two other forecasts are considered: The first is a random walk forecast (RW) which is based on the average growth rate over the respective sample period of each forecast step. The second forecast is the growth rate derived from directly extrapolating potential output without considering the transitional dynamics originating from closing the output gap (PFA, direct). While the RW forecast represents a typical naïve forecast, the latter is meant to capture whether the consideration of transitional dynamics towards potential output as employed in the PFA gap closing version helps to improve forecast precision.

The first three rows of each block in the tables report the number of cumulative forecasts available for evaluation as well as the mean forecast and actual forecasts expressed as average annual growth rates which are derived from the underlying quarterly growth rates. The next rows contain the average forecast error (bias) which is the difference between the mean of the actual growth rate and the mean of the forecast. The indented rows following the bias estimate report the heteroscedasticity and autocorrelation consistent (HAC) t-statistics and two-sided p-values according to the Newey-West formula and the GLS t-value and two-sided p-value from the bias test as described in section 1.3.3.

Root mean squared errors (RMSE) and mean absolute errors (MAE) are also reported, which can both be regarded as a combination of bias and variance measures. The ratio of the RMSE from two different models gives Theils's U index of inequality which measures the degree to which the PFA forecasts differ from the RW forecasts. A value greater than one implies that the random walk forecasts have better accuracy than the PFA forecast. However, this measure does not indicate whether the difference in accuracy is statistically significant. In order to close this gap, the remaining rows of each block in the tables display the results of the forecast accuracy tests as outlined in section 1.3.3. Once again, both the conventional HAC test statistics and the GLS test statistics are reported. P-values refer to two-sided tests of the null hypothesis.

For the sake of completeness, the results of the three, four and five year ahead forecasts are reported although in most cases test outcomes for a country hold equally for all years. This means, for instance, that if a significant bias of the three year forecast is detected, this bias will typically be also significant for the four and five years ahead forecast. Forecast performance is in general not specific to a certain horizon but rather dependent on the forecast method (PFA (Gap closing), PFA (GA) or RW) and, needless to say, the considered country.

Subsequent to each table, a graph is shown which depicts the pattern of the pseudo forecast from the PFA (Gap closing), the actual GDP development as well as available projections from governments or official authorities. For Germany and the USA, these predictions refer to a five year horizon whereas for the remaining countries the three years ahead predictions are shown.

Comparable official projections are limited with respect to covered time periods. Furthermore, the preparation periods and announcement dates of these official projections do not show a one-to-one correspondence to the respective beginning periods of the pseudo forecasts and official statistics of economic data known to the forecasters at the time the projections were produced are slightly different from the figures used here due to data revisions.[45] Hence, this comparison is rather sketchy than strictly formal. Yet for the US data, for example, the differences in medium-term growth rates between the real time data („first announcements“) and the final data from OECD sources are not substantial.

The comparison of official projections and the forecasts from the above exercise should help explain the workings of the production function approach in a practical setting. Typically, the various determinants of potential output and its medium-term development are not extrapolated in a mechanistic fashion but enriched with expert opinion and a whole series of qualitative assumptions. In particular, projections from official governmental authorities or institutions that are closely tied to governments or even projections from supranational institutions are often accused of being over-optimistic.[46] If a neutral scenario per se produces a biased forecast, these forecasts might be improved by judgmental add-factors that restore efficiency, however, in the case that neutral scenarios are already unbiased, there might be little scope for improving these forecasts and judgemental adjustments eventually lead to non-rational predictions.

The subsequent sections show the tables country by country and briefly comment on the individual outcomes.

[45] In addition, the projections from official sources considered below are produced at annual frequency whereas the pseudo forecasts are conducted at quarterly frequency. For this reason, the out-of-sample exercise offers four possible forecasts each year that could in principle be used for comparison with the annual official projections. The pseudo forecast made in the last quarter of the respective years have been chosen for this purpose. This choice assures a comparable reference period for the medium-term predictions although the publication dates of the official forecasts do not imply a strictly comparable level of information at the date both types of forecasts are produced. However, since the information difference concerns only one or two quarters, we do not regard it as a problem that considerably limits the comparisons of forecast performances.

[46] For example, cf. Batista and Zaluendo (2004) for a concise literature review of IMF's medium-term growth projections and the discussion of over-optimism.

1.4.1 Germany

Table 1.6 contains the out-of-sample results for the growth forecasts for Germany. The table reveals that for all three forecast horizons both PFA versions result in forecasts that are unbiased. The mean of the forecasted and actual values are almost identical and the bias test does not reject the hypothesis of zero mean forecast error, irrespective of whether t-statistic is consulted. In contrast, the random walk forecast with a three year horizon shows an average forecast error of -0.727 which is significant at the 5%-level according to the HAC t-values and significant at the 8%-level according to the GLS t-values. It is also biased for the other forecast horizons with respect to the HAC statistics, however, p-values of the GLS statistics show significance beyond the 10%-level.

Overall, extrapolating from past GDP growth trends resulted in systematically upward biased three, four and five years ahead growth predictions for Germany. In addition, the PFA (GA) and the PFA (gap closing) forecasts have Theil's U values that are strictly less than one over all horizons with the lowest values measured for the gap closing version. Although forecast accuracy seems to be in favor of the production function approach, the accuracy tests also demonstrate that the differences in squared forecast errors are never significant. Another interesting insight is that RMSE decrease with increasing forecast horizons, i.e. the forecasts become more and more accurate with rising span. In anticipation of the upcoming sections, this result also holds for the forecasts for the other G7 countries. One reason that longer horizon forecasts might be more precise than shorter horizon ones is that GDP growth trends predicted by the PFA are more valid for longer periods and that over shorter periods some cyclical effects still prevail which are captured less accurately by a forecasting framework that solely builds on the production-side of the economy.

Figure 3.2 illustrates the degree to which the PFA (gap closing) forecasts and the projections from the German government are able to explain the actual GDP development. A conspicuous finding is the loose correspondence of both projections with the actual GDP development over the considered five-year span. Only in the period from 1997 to 2002 does the pseudo PFA-forecast display a close connection to the actual growth rates. For the remaining years, neither the pseudo forecasts nor the official projections predict a GDP development in advance that retrospectively matches with the course of the actual growth rates. This failure is particularly apparent for the period from 1989 to 1993 where the German economy enjoyed an economic boom whose pervasion did not seem to be predictable. The preceding error analysis has shown that PFA yields unbiased forecasts. However, the prediction error for the government projection is on average -0.486 and implies an upward bias.

Table 1.6: Results of forecast evaluation for Germany

	PFA(direct)	PFA(gap closing)	RW
Horizon = 3 years			
Number of cumulative forecasts:	92	92	92
Mean forecast:[1]	1.839	1.918	2.598
Mean actual:[1]	1.872	1.872	1.872
Average forecast error (Bias):[1]	0.033	-0.046	-0.727
HAC t-value (p-val.):	0.112 (0.91)	-0.183 (0.86)	-2.294 (0.02)
GLS t-value (p-val.):	0.051 (0.96)	-0.061 (0.95)	-1.787 (0.08)
Root Mean Squared Error (RMSE):	1.190	1.085	1.354
Mean Absolute Error (MAE):[1]	0.956	0.827	1.097
Theil's U:	0.879	0.801	-
Average loss differential (Accuracy):	-3.766	-5.906	-
HAC t-value (p-val.):	-0.907 (0.37)	-1.364 (0.18)	-
GLS t-value (p-val.):	-0.407 (0.68)	-0.599 (0.55)	-
Horizon = 4 years			
Number of cumulative forecasts:	88	88	88
Mean forecast:[1]	1.844	1.907	2.609
Mean actual:[1]	1.935	1.935	1.935
Average forecast error (Bias):[1]	0.090	0.027	-0.675
HAC t-value (p-val.):	0.328 (0.74)	0.116 (0.91)	-2.236 (0.03)
GLS t-value (p-val.):	0.120 (0.90)	0.033 (0.97)	-1.501 (0.14)
Root Mean Squared Error (RMSE):	0.998	0.893	1.178
Mean Absolute Error (MAE):[1]	0.801	0.693	0.992
Theil's U:	0.847	0.758	-
Average loss differential (Accuracy):	-6.255	-9.446	-
HAC t-value (p-val.):	-0.833 (0.41)	-1.266 (0.21)	-
GLS t-value (p-val.):	-0.321 (0.75)	-0.475 (0.64)	-
Horizon =5 years			
Number of cumulative forecasts:	84	84	84
Mean forecast:[1]	1.849	1.910	2.618
Mean actual:[1]	1.985	1.985	1.985
Average forecast error (Bias):[1]	0.137	0.075	-0.633
HAC t-value (p-val.):	0.533 (0.60)	0.345 (0.73)	-2.245 (0.03)
GLS t-value (p-val.):	0.193 (0.85)	0.101 (0.92)	-1.378 (0.17)
Root Mean Squared Error (RMSE):	0.878	0.781	1.057
Mean Absolute Error (MAE):[1]	0.686	0.596	0.894
Theil's U:	0.831	0.739	-
Average loss differential (Accuracy):	-8.654	-12.674	-
HAC t-value (p-val.):	-0.719 (0.47)	-1.148 (0.25)	-
GLS t-value (p-val.):	-0.299 (0.77)	-0.451 (0.65)	-

Notes: [1]: Annual averages in percentage, Sample period: 1968:2 to 2005:4, Forecast period: 1980:1 to 2005:4, Computational work was performed in MATLAB.

Indeed , the figure 3.2 shows that the pattern of the official projections runs parallel to the course of the pseudo forecast which conveys a “neutral” or baseline scenario. Thus, a systematical deviation from neutral assumptions and an overly optimistic view can be stated for the official government projections which, we bear in mind, constitute an important figure for budget planning.

Notes: The date always refers to the last year of the projection. See section 1.5 in the appendix for details. Average error of government projections: -0.486, RMSE of government projections: 1.036

Figure 1.5: BMWA projections for 5-year GDP growth in Germany

1.4.2 USA

In contrast to the outcomes for Germany, for the USA we find that the random walk model demonstrates better forecast performance than the PFA based forecast. As depicted in table 1.7, the forecasts produced with both PFA versions exhibit positive bias for all horizons but at the same time these estimates are not significant according to the GLS t-values. In contrast, these bias estimates are highly significant for all forecast horizons with respect to the HAC t-statistics.

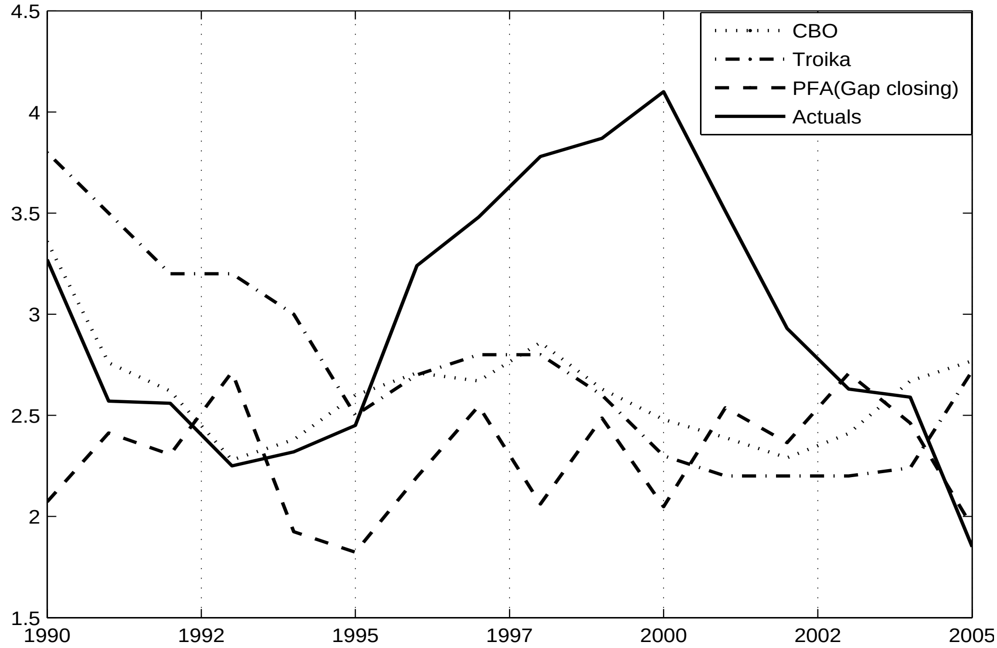
However, discrepancy of inference does not hold for the tests of forecast accuracy. Here we do not reject the null hypothesis of equality between the squared prediction errors of the PFA (GA) and the RW forecast at the 10%-level when looking at the GLS and HAC t-values. So the choice of which method to use for the calculation of robust standard errors does not influence the test decision. In general, similar to the results of the Monte Carlo experiments of section 1.3.3, t-values associated with the GLS procedure are smaller than the HAC based t-values. Overall, the performance measures for the USA are clearly to the credit of the random walk model.

Table 1.7: Results of forecast evaluation for the USA

	PFA(direct)	PFA(gap closing)	RW
Horizon = 3 years			
Number of cumulative forecasts:	72	72	72
Mean forecast:[1]	2.287	2.348	3.103
Mean actual:[1]	2.993	2.993	2.993
Average forecast error (Bias):[1]	0.706	0.645	-0.110
HAC t-value (p-val.):	2.385 (0.02)	2.399 (0.02)	-0.376 (0.71)
GLS t-value (p-val.):	1.494 (0.14)	1.388 (0.17)	-0.446 (0.66)
Root Mean Squared Error (RMSE):	1.254	1.145	0.939
Mean Absolute Error (MAE):[1]	1.127	1.013	0.773
Theil's U:	1.335	1.220	-
Average loss differential (Accuracy):	6.210	3.872	-
HAC t-value (p-val.):	1.383 (0.17)	0.907 (0.37)	-
GLS t-value (p-val.):	0.840 (0.40)	0.576 (0.57)	-
Horizon = 4 years			
Number of cumulative forecasts:	68	68	68
Mean forecast:[1]	2.280	2.313	3.106
Mean actual:[1]	2.964	2.964	2.964
Average forecast error (Bias):[1]	0.684	0.650	-0.142
HAC t-value (p-val.):	2.501 (0.01)	2.638 (0.01)	-0.517 (0.61)
GLS t-value (p-val.):	1.085 (0.28)	1.038 (0.30)	-0.568 (0.57)
Root Mean Squared Error (RMSE):	1.096	1.033	0.783
Mean Absolute Error (MAE):[1]	0.916	0.855	0.665
Theil's U:	1.400	1.320	-
Average loss differential (Accuracy):	9.408	7.267	-
HAC t-value (p-val.):	1.051 (0.30)	0.893 (0.38)	-
GLS t-value (p-val.):	0.528 (0.60)	0.428 (0.67)	-
Horizon =5 years			
Number of cumulative forecasts:	64	64	64
Mean forecast:[1]	2.269	2.295	3.106
Mean actual:[1]	2.955	2.955	2.955
Average forecast error (Bias):[1]	0.686	0.660	-0.151
HAC t-value (p-val.):	2.701 (0.01)	2.955 (0.00)	-0.587 (0.56)
GLS t-value (p-val.):	1.309 (0.20)	1.235 (0.22)	-0.578 (0.57)
Root Mean Squared Error (RMSE):	0.989	0.932	0.670
Mean Absolute Error (MAE):[1]	0.767	0.719	0.611
Theil's U:	1.477	1.391	-
Average loss differential (Accuracy):	13.243	10.493	-
HAC t-value (p-val.):	0.855 (0.40)	0.783 (0.44)	-
GLS t-value (p-val.):	0.557 (0.58)	0.446 (0.66)	-

Notes: [1]: Annual averages in percentage, Sample period: 1970:2 to 2005:4, Forecast period: 1985:1 to 2005:4, Computational work was performed in MATLAB.

Figure 1.6 provides a graphical assessment of the performance of the Troika's, the CBO's and the PFA forecasts. The hump-shaped behavior of five-year average GDP growth which begins in the mid-nineties and ends in the year 2002 is the most eye-catching element in this figure. Another remarkable fact is that none of the projections follow this pattern. Before 1995, the CBO's projection was almost



Notes: The date always refers to the last year of the projection. See section 1.5 in the appendix for details. Average error of Troika projection: 0.215, RMSE of Troika projection: 0.895, Average error of CBO projections: 0.345, RMSE of CBO projections: 0.733

Figure 1.6: CBO and Troika projections for 5-year GDP growth in the USA

in line with the actual movement of GDP growth whereas the projections released by the Troika were apparently upward biased. The unsteady fluctuation and the cautious level of the pseudo forecast stands out, which is a graphical confirmation of the outcomes reported in table 1.7.

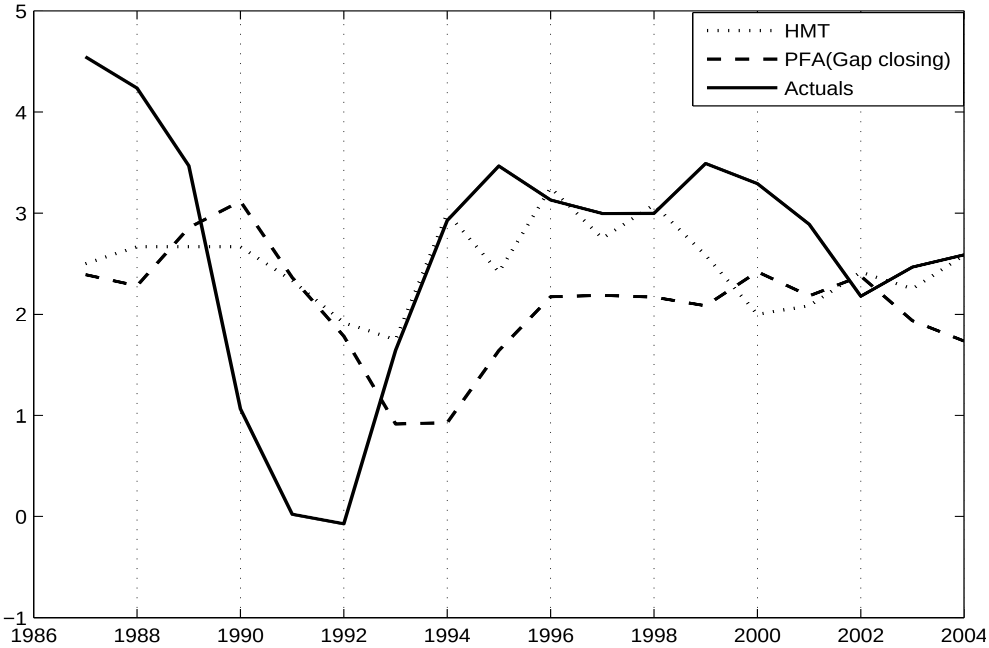
1.4.3 United Kingdom

The outcomes of the forecast performance tests for the United Kingdom are displayed in table 1.8. The bias estimates for the PFA (GA) forecasts, which arise from directly projecting potential growth, amount to values of around 0.5%, however, these estimates are not significant according to both test statistics. Similarly, the PFA (Gap closing) and RW models produce unbiased forecasts over all considered spans. In terms of accuracy, the RW model clearly wins the race: Only for the three year horizon are Theil's U values in favor of the PFA (Gap closing) model. The remaining test outcomes imply that the RW forecasts have a closer tie to the final outcomes than the other predictions. The average loss differential is positive and significant for the four and five year spans when looking at the GLS t-statistics throughout. For the PFA (gap closing) forecasts, in particular, fairly substantial loss differentials are observed: For the five years ahead forecast, the difference between the squared errors of the PFA (direct) and the RW forecasts is 11.7 percentage points and purports that the RW is on average quite a few percentage points closer to the true value than the former forecast.

Table 1.8: Results of forecast evaluation for the United Kingdom

	PFA(direct)	PFA(gap closing)	RW
Horizon = 3 years			
Number of cumulative forecasts:	72	72	72
Mean forecast:[1]	2.063	2.323	2.170
Mean actual:[1]	2.536	2.536	2.536
Average forecast error (Bias):[1]	0.474	0.214	0.366
HAC t-value (p-val.):	1.112 (0.27)	0.530 (0.60)	0.908 (0.37)
GLS t-value (p-val.):	0.969 (0.34)	0.433 (0.67)	1.358 (0.18)
Root Mean Squared Error (RMSE):	1.474	1.304	1.326
Mean Absolute Error (MAE):[1]	1.264	1.076	1.092
Theil's U:	1.111	0.983	-
Average loss differential (Accuracy):	3.710	-0.521	-
HAC t-value (p-val.):	1.081 (0.28)	-0.127 (0.90)	-
GLS t-value (p-val.):	3.677 (0.00)	-0.359 (0.72)	-
Horizon = 4 years			
Number of cumulative forecasts:	68	68	68
Mean forecast:[1]	1.994	2.203	2.163
Mean actual:[1]	2.476	2.476	2.476
Average forecast error (Bias):[1]	0.482	0.273	0.313
HAC t-value (p-val.):	1.226 (0.22)	0.709 (0.48)	0.882 (0.38)
GLS t-value (p-val.):	1.045 (0.30)	0.599 (0.55)	1.113 (0.27)
Root Mean Squared Error (RMSE):	1.304	1.169	1.100
Mean Absolute Error (MAE):[1]	1.138	0.992	0.925
Theil's U:	1.185	1.063	-
Average loss differential (Accuracy):	7.848	2.498	-
HAC t-value (p-val.):	1.840 (0.07)	0.569 (0.57)	-
GLS t-value (p-val.):	2.951 (0.00)	4.001 (0.00)	-
Horizon =5 years			
Number of cumulative forecasts:	64	64	64
Mean forecast:[1]	1.935	2.116	2.154
Mean actual:[1]	2.446	2.446	2.446
Average forecast error (Bias):[1]	0.511	0.330	0.292
HAC t-value (p-val.):	1.408 (0.16)	0.903 (0.37)	0.907 (0.37)
GLS t-value (p-val.):	0.925 (0.36)	0.616 (0.54)	0.994 (0.32)
Root Mean Squared Error (RMSE):	1.137	1.019	0.908
Mean Absolute Error (MAE):[1]	1.004	0.894	0.801
Theil's U:	1.252	1.122	-
Average loss differential (Accuracy):	11.702	5.352	-
HAC t-value (p-val.):	1.879 (0.06)	1.389 (0.17)	-
GLS t-value (p-val.):	1.790 (0.08)	4.887 (0.00)	-

Notes: [1]: Annual averages in percentage, Sample period: 1970:4 to 2005:4, Forecast period: 1985:1 to 2005:4, Computational work was performed in MATLAB.



Notes: The date always refers to the last year of the projection. See section 1.5 in the appendix for details. Average error of HMT projections: 0.134, RMSE of HMT projections: 1.145

Figure 1.7: HMT projections for 3-year GDP growth in the United Kingdom

The HAC t-values imply insignificant loss differentials for the PFA (gap closing) predictions at the four and five year horizon, however, we regard the GLS statistics as being more reliable in the light of the experimental outcomes reported in section 1.3.3 and therefore conclude that the RW generated more accurate forecasts than the other models for horizons beyond three years.

Figure 1.7 shows the three years ahead growth projections from the HMT and the PFA vis-à-vis the actual GDP development. A prolonged period of underestimation of growth by the PFA forecasts during the second half of the nineties is visible and also that these forecasts adjust too late to a changing growth trend. A further negative point would be that, after 2002 when average growth caught up, the PFA forecasts still indicated a decline of growth. On the positive side, the HMT projections stand out with a remarkably good forecast performance record in the period from 1993 to 1998. Before the year 1993, the HMT and pseudo forecast nearly coincide but are over-optimistic. In the years 1991 and 1992, the bias for the annual average growth rate over the three year forecast horizons amounts to 2 percentage points for both of these predictions, which leads to substantial forecast errors. On balance, however, the HMT projections display good forecast performance.

1.4.4 France

Results for France are given in table 1.9. The average forecast of the PFA (gap closing) and the average realized growth rates are nearly identical. Average forecast errors for all horizons are therefore not significantly different from zero.

Table 1.9: Results of forecast evaluation for France

	PFA(Direct)	PFA(gap closing)	RW
Horizon = 3 years			
Number of cumulative forecasts:	72	72	72
Mean forecast:[1]	1.789	2.032	2.502
Mean actual:[1]	2.106	2.106	2.106
Average forecast error (Bias):[1]	0.317	0.074	-0.397
HAC t-value (p-val.):	0.923 (0.36)	0.256 (0.80)	-1.249 (0.22)
GLS t-value (p-val.):	0.838 (0.40)	0.197 (0.84)	-1.734 (0.09)
Root Mean Squared Error (RMSE):	1.092	0.903	1.015
Mean Absolute Error (MAE):[1]	0.926	0.741	0.851
Theil's U:	1.075	0.889	-
Average loss differential (Accuracy):	1.452	-1.941	-
HAC t-value (p-val.):	0.294 (0.77)	-0.567 (0.57)	-
GLS t-value (p-val.):	0.277 (0.78)	-0.567 (0.57)	-
Horizon = 4 years			
Number of cumulative forecasts:	68	68	68
Mean forecast:[1]	1.803	1.993	2.509
Mean actual:[1]	2.104	2.104	2.104
Average forecast error (Bias):[1]	0.301	0.112	-0.404
HAC t-value (p-val.):	0.875 (0.38)	0.377 (0.71)	-1.309 (0.20)
GLS t-value (p-val.):	0.681 (0.50)	0.250 (0.80)	-1.614 (0.11)
Root Mean Squared Error (RMSE):	1.026	0.867	0.939
Mean Absolute Error (MAE):[1]	0.901	0.741	0.764
Theil's U:	1.093	0.923	-
Average loss differential (Accuracy):	2.741	-2.097	-
HAC t-value (p-val.):	0.290 (0.77)	-0.291 (0.77)	-
GLS t-value (p-val.):	0.255 (0.80)	-0.263 (0.79)	-
Horizon =5 years			
Number of cumulative forecasts:	64	64	64
Mean forecast:[1]	1.820	1.982	2.513
Mean actual:[1]	2.092	2.092	2.092
Average forecast error (Bias):[1]	0.272	0.110	-0.421
HAC t-value (p-val.):	0.863 (0.39)	0.395 (0.69)	-1.508 (0.14)
GLS t-value (p-val.):	0.639 (0.52)	0.256 (0.80)	-1.656 (0.10)
Root Mean Squared Error (RMSE):	0.919	0.788	0.846
Mean Absolute Error (MAE):[1]	0.829	0.704	0.657
Theil's U:	1.085	0.931	-
Average loss differential (Accuracy):	3.192	-2.390	-
HAC t-value (p-val.):	0.220 (0.83)	-0.202 (0.84)	-
GLS t-value (p-val.):	0.199 (0.84)	-0.193 (0.85)	-

Notes: [1]: Annual averages in percentage, Sample period: 1971:1 to 2005:4, Forecast period: 1985:1 to 2005:4, Computational work was performed in MATLAB.

Yet unbiasedness is only one side of the coin. RMSE are large and the difference between the squared errors from naïve RW model forecasts and the squared errors from both PFA forecasts are not significant, irrespective of the t-statistic one looks at. The RW forecast is significantly biased at the 10%-level over most time spans according to the GLS test statistic. Overall, the PFA (gap closing) models predictions' stand out slightly with the most favorable outcomes.

There is no official projection from national sources for the medium-term growth available to us which could be used for an illustrative comparison. We therefore draw on the IMF's three years ahead projections for the French economy over the period from 1993 to 2005. The figure 1.8 shows the results when the pseudo forecasts are compared to the IMF's projections and the final outcomes. A lack of accuracy of both predictions is visible, but the heavily biased IMF projections are most salient. For most periods, the IMF's predictions are roughly one percentage point higher than the unbiased pseudo forecasts which can be regarded as incorporating a neutral scenario of future economic growth. Notice that the IMF projections are nearly parallel to the pseudo forecast, so there is clearly scope to improve the IMF projections. Once again, both predictions were not capable of capturing the hike of growth that occurred around the turn of the millennium.

Notes: The date always refers to the last year of the projection. See section 1.5 in the appendix for details. Average error of IMF projections: -0.914, RMSE of IMF projections: 1.462

Figure 1.8: IMF projections for 3-year GDP growth in the France

1.4.5 Italy

For Italy, both PFA forecasts are continuously unbiased. However, the PFA version which builds on the back-to-trend scenario generates average forecast errors which are larger in absolute value than the PFA (direct) forecasts (see table 1.10). The random walk model predictions deviate to a large extent from the actual values and test outcomes clearly imply biasedness. None of the accuracy tests in the table are significant, meaning that the PFA forecasts do not have better predictive value in terms of accuracy than the random walk forecast.

Notes: The date always refers to the last year of the projection. See section 1.5 in the appendix for details. Average error of IMF projections: -1.211, RMSE of IMF projections: 1.419

Figure 1.9: IMF projections for 3-year GDP growth in Italy

In figure 1.9, the pseudo forecasts as well as projections from IMF staff are contrasted with the final medium-term growth rates of GDP. Again, the reference horizon is three years. The track record of the PFA projections is fairly good for the period of 1995 to 2001, but rather poor for the beginning and ending of the period considered for this comparison.

As well as the IMF projections for France, the projections for Italy are also too optimistic in all periods. The bias estimate amounts to -1.2 and this further implies that a systematic deviation from a neutral scenario about the trend evolution of output can be assumed for the IMF projections.

Table 1.10: Results of forecast evaluation for Italy

	PFA(direct)	PFA(gap closing)	RW
Horizon = 3 years			
Number of cumulative forecasts:	72	72	72
Mean forecast:[1]	1.865	2.221	2.481
Mean actual:[1]	1.726	1.726	1.726
Average forecast error (Bias):[1]	-0.139	-0.495	-0.754
HAC t-value (p-val.):	-0.433 (0.67)	-1.553 (0.12)	-2.397 (0.02)
GLS t-value (p-val.):	-0.300 (0.77)	-1.098 (0.28)	-3.027 (0.00)
Root Mean Squared Error (RMSE):	1.095	1.173	1.241
Mean Absolute Error (MAE):[1]	0.948	0.983	1.020
Theil's U:	0.883	0.945	-
Average loss differential (Accuracy):	-3.064	-1.475	-
HAC t-value (p-val.):	-0.794 (0.43)	-0.577 (0.57)	-
GLS t-value (p-val.):	-0.545 (0.59)	-0.641 (0.52)	-
Horizon = 4 years			
Number of cumulative forecasts:	68	68	68
Mean forecast:[1]	1.825	2.102	2.494
Mean actual:[1]	1.712	1.712	1.712
Average forecast error (Bias):[1]	-0.113	-0.390	-0.782
HAC t-value (p-val.):	-0.375 (0.71)	-1.282 (0.20)	-2.888 (0.01)
GLS t-value (p-val.):	-0.226 (0.82)	-0.780 (0.44)	-2.830 (0.01)
Root Mean Squared Error (RMSE):	1.034	1.088	1.152
Mean Absolute Error (MAE):[1]	0.836	0.901	0.968
Theil's U:	0.897	0.944	-
Average loss differential (Accuracy):	-4.142	-2.325	-
HAC t-value (p-val.):	-0.604 (0.55)	-0.498 (0.62)	-
GLS t-value (p-val.):	-0.353 (0.73)	-0.338 (0.74)	-
Horizon =5 years			
Number of cumulative forecasts:	64	64	64
Mean forecast:[1]	1.763	1.998	2.505
Mean actual:[1]	1.699	1.699	1.699
Average forecast error (Bias):[1]	-0.064	-0.299	-0.807
HAC t-value (p-val.):	-0.263 (0.79)	-1.181 (0.24)	-3.894 (0.00)
GLS t-value (p-val.):	-0.121 (0.90)	-0.563 (0.58)	-2.833 (0.01)
Root Mean Squared Error (RMSE):	0.862	0.923	1.043
Mean Absolute Error (MAE):[1]	0.713	0.773	0.899
Theil's U:	0.826	0.885	-
Average loss differential (Accuracy):	-8.663	-5.906	-
HAC t-value (p-val.):	-0.780 (0.44)	-0.713 (0.48)	-
GLS t-value (p-val.):	-0.400 (0.69)	-0.397 (0.69)	-

Notes: [1]: Annual averages in percentage, Sample period: 1970:2 to 2005:4, Forecast period: 1985:1 to 2005:4, Computational work was performed in MATLAB.

1.4.6 Japan

Average GDP growth in Japan amounted to roughly two percent each year during the period from 1985 to 2005 on which the forecast evaluation indices shown in table 1.11 are based.

Table 1.11: Results of forecast evaluation for Japan

	PFA(direct)	PFA(gap closing)	RW
Horizon = 3 years			
Number of cumulative forecasts:	72	72	72
Mean forecast:[1]	2.151	2.201	3.470
Mean actual:[1]	2.068	2.068	2.068
Average forecast error (Bias):[1]	-0.082	-0.133	-1.402
HAC t-value (p-val.):	-0.161 (0.87)	-0.254 (0.80)	-2.434 (0.02)
GLS t-value (p-val.):	-0.111 (0.91)	-0.181 (0.86)	-3.270 (0.00)
Root Mean Squared Error (RMSE):	1.628	1.613	2.161
Mean Absolute Error (MAE):[1]	1.329	1.344	1.941
Theil's U:	0.754	0.746	-
Average loss differential (Accuracy):	-18.155	-18.604	-
HAC t-value (p-val.):	-1.315 (0.19)	-1.283 (0.20)	-
GLS t-value (p-val.):	-0.948 (0.35)	-1.020 (0.31)	-
Horizon = 4 years			
Number of cumulative forecasts:	68	68	68
Mean forecast:[1]	2.237	2.264	3.509
Mean actual:[1]	1.982	1.982	1.982
Average forecast error (Bias):[1]	-0.255	-0.282	-1.527
HAC t-value (p-val.):	-0.494 (0.62)	-0.526 (0.60)	-2.599 (0.01)
GLS t-value (p-val.):	-0.356 (0.72)	-0.402 (0.69)	-3.464 (0.00)
Root Mean Squared Error (RMSE):	1.506	1.506	2.137
Mean Absolute Error (MAE):[1]	1.231	1.268	1.980
Theil's U:	0.704	0.705	-
Average loss differential (Accuracy):	-36.822	-36.786	-
HAC t-value (p-val.):	-1.228 (0.22)	-1.207 (0.23)	-
GLS t-value (p-val.):	-1.129 (0.26)	-1.173 (0.25)	-
Horizon =5 years			
Number of cumulative forecasts:	64	64	64
Mean forecast:[1]	2.311	2.334	3.546
Mean actual:[1]	1.907	1.907	1.907
Average forecast error (Bias):[1]	-0.404	-0.427	-1.640
HAC t-value (p-val.):	-0.804 (0.42)	-0.812 (0.42)	-2.879 (0.01)
GLS t-value (p-val.):	-0.519 (0.61)	-0.559 (0.58)	-3.344 (0.00)
Root Mean Squared Error (RMSE):	1.405	1.411	2.131
Mean Absolute Error (MAE):[1]	1.201	1.234	1.979
Theil's U:	0.659	0.662	-
Average loss differential (Accuracy):	-64.142	-63.717	-
HAC t-value (p-val.):	-1.216 (0.23)	-1.215 (0.23)	-
GLS t-value (p-val.):	-1.184 (0.24)	-1.220 (0.23)	-

Notes: [1]: Annual averages in percentage, Sample period: 1972:1 to 2005:4, Forecast period: 1985:1 to 2005:4, Computational work was performed in MATLAB.

The PFA (direct) and PFA (gap closing) models are able to predict GDP growth rates that approximately match with this development: The mean forecasts are only slightly above two percent over all horizons and bias estimates are not significant even once. The opposite holds for the random walk forecast. Here, the average predicted growth rates are much too high and thus bias estimates deviate significantly from zero throughout. Compared to the other country results, the random walk model does the worst job since Japan, particularly during the nineties, was not able to sustain the dynamic growth rates from past years. Using the example of Japan, the benefit from employing a production function approach that incorporates various trend indices as opposed to a simple univariate trend extrapolation of GDP shows up noticeably.

Notes: The date always refers to the last year of the projection. See section 1.5 in the appendix for details. Average error of IMF projections: -1.528, RMSE of IMF projections: 1.836

Figure 1.10: IMF projections for 3-year GDP growth in Japan

An effect of the bad performance of the RW model's forecasts emerges in the Theil's U ratios. These are between 0.75 and 0.65. The loss differentials are large at all horizons, however, the hypothesis of equal forecast accuracy between the PFA forecasts and the RW forecasts is never rejected by the GLS and the HAC statistics.

A good impression of forecast performance is also provided by figure 3.1. The pseudo forecasts are plotted against the three years ahead projections from IMF's forecasting staff. Again, the course of actual GDP growth is not caught by the predictions. However, the PFA (gap closing) forecasts do at least fluctuate at an appropriate level of growth while the IMF projections are once more clearly oversized. The IMF should have had cognizance of the assumptions that lead to neutral predictions as conveyed by the PFA forecasts since these employ only information that would have been available at the time the IMF released its projections. Hence, these projections seem to have been built on intended optimism rather than on an neutral or cautious assumption about the likely future prospects of the Japanese economy.

1.4.7 Canada

The last outcomes to discuss are those for Canada. Table 1.12 shows the corresponding results. It can be seen from the estimates that the PFA models tend to underestimate realized growth while the RW tends to overshoot. The PFA (gap closing) forecasts are mostly in conformance with average true growth. However, bias estimates are insignificant for all model forecasts and horizons. To sum up, due to a lack of significance, the key forecast performance figures in table 1.12 do not provide clear guidance as to which model to put more confidence in when preparing medium-term growth forecasts.

Notes: The date always refers to the last year of the projection. See section 1.5 in the appendix for details. Average error of IMF projections: -0.419, RMSE of IMF projections: 1.013

Figure 1.11: IMF projections for 3-year GDP growth in Canada

The last figure displays three years ahead PFA forecasts, comparable projections from the IMF and actual growth outcomes.

Table 1.12: Results of forecast evaluation for Canada

	PFA(direct)	PFA(gap closing)	RW
Horizon = 3 years			
Number of cumulative forecasts:	72	72	72
Mean forecast:[1]	2.262	2.540	3.225
Mean actual:[1]	2.760	2.760	2.760
Average forecast error (Bias):[1]	0.498	0.220	-0.465
HAC t-value (p-val.):	0.999 (0.32)	0.476 (0.64)	-0.885 (0.38)
GLS t-value (p-val.):	1.074 (0.29)	0.479 (0.63)	-1.388 (0.17)
Root Mean Squared Error (RMSE):	1.661	1.466	1.618
Mean Absolute Error (MAE):[1]	1.418	1.177	1.112
Theil's U:	1.027	0.906	-
Average loss differential (Accuracy):	1.268	-4.227	-
HAC t-value (p-val.):	0.123 (0.90)	-0.491 (0.62)	-
GLS t-value (p-val.):	0.148 (0.88)	-0.699 (0.49)	-
Horizon = 4 years			
Number of cumulative forecasts:	68	68	68
Mean forecast:[1]	2.288	2.499	3.229
Mean actual:[1]	2.727	2.727	2.727
Average forecast error (Bias):[1]	0.438	0.228	-0.502
HAC t-value (p-val.):	0.848 (0.40)	0.464 (0.64)	-0.921 (0.36)
GLS t-value (p-val.):	0.784 (0.44)	0.407 (0.69)	-1.383 (0.17)
Root Mean Squared Error (RMSE):	1.487	1.356	1.469
Mean Absolute Error (MAE):[1]	1.264	1.096	1.072
Theil's U:	1.012	0.923	-
Average loss differential (Accuracy):	0.819	-5.120	-
HAC t-value (p-val.):	0.042 (0.97)	-0.316 (0.75)	-
GLS t-value (p-val.):	0.045 (0.96)	-0.365 (0.72)	-
Horizon =5 years			
Number of cumulative forecasts:	64	64	64
Mean forecast:[1]	2.306	2.487	3.233
Mean actual:[1]	2.702	2.702	2.702
Average forecast error (Bias):[1]	0.396	0.215	-0.531
HAC t-value (p-val.):	0.722 (0.47)	0.410 (0.68)	-0.925 (0.36)
GLS t-value (p-val.):	0.722 (0.47)	0.393 (0.70)	-1.546 (0.13)
Root Mean Squared Error (RMSE):	1.324	1.214	1.355
Mean Absolute Error (MAE):[1]	1.156	1.030	1.087
Theil's U:	0.977	0.896	-
Average loss differential (Accuracy):	-2.095	-9.050	-
HAC t-value (p-val.):	-0.059 (0.95)	-0.299 (0.77)	-
GLS t-value (p-val.):	-0.075 (0.94)	-0.403 (0.69)	-

Notes: [1]: Annual averages in percentage, Sample period: 1970:2 to 2005:4, Forecast period: 1985:1 to 2005:4, Computational work was performed in MATLAB.

For the period from 1993 to 1998, the by now familiar diagnosis also stands out here: the IMF's projections are visibly too high. However, after 1998, the same projections tend to result in underestimations of true growth but return to an over-optimistic path towards the end of the sample. By contrast, the pseudo forecasts are located too low in most periods, confirming the finding that the average forecast error is positive. The calculated bias of the IMF's projections amounts to -0.42 using the 13 available observations, which is roughly the size of the random walk's model forecast bias. Naturally, such a stylized assessment can not replace a more rigorous statistical analysis of forecast precision and the established results might not hold in general.

1.5 Summary and conclusion

Realistic projections of the medium-term growth capabilities are important for many purposes, however, in contrast to the evaluation of business cycle forecasts, the examination of forecasting approaches and actual predictions of the economic development over the medium- or long-term hardly receives any attention in the economic literature.

This chapter begins with a survey of methods for medium-term predictions that are used by governmental bodies in the major industrial countries and international institutions. It turns out that the production function approach with its assumptions about the supply-side functioning of the economy and conditional steady-state convergence plays a pre-dominant role for the preparation of medium-term projections of output growth three to five years ahead.

Against this background, the aim of the present chapter is to check the predictive value of the PFA as a mainstream approach to estimate potential output and to derive forecasts from it. There have been a number of studies that have analyzed the outcomes of the various methods to estimate potential output in-sample. This chapter follows a different path by evaluating the value of the production function approach with the aid of an explicitly forward-looking analysis. Due to the design of the out-of-sample analysis, the corresponding multi-step forecasts result in forecast errors that are highly serial correlated. In order to account for serial correlation in error processes and to perform consistent tests for unbiasedness and accuracy, a simple model of forecast errors is employed to analytically derive the exact covariance matrix of forecast errors. Empirical implementation of these tests is straightforward and it has been shown that they have good size properties in small samples.

The evaluation of the forecast errors of the out-of-sample analysis for the observation period from 1985 to 2005 highlights the following: The production function approach yields unbiased projections of real GDP growth for three, four and five year horizons for most countries, but misses other important features of actual GDP developments. Root mean squared errors and mean absolute errors are large and the predictions often only capture a small fraction of the time variation of actual GDP growth. For most countries, projections from the PFA are at least capable of beating naïve forecasts in terms of root mean squared errors, however, differences in accuracy are not statistically significant in the majority of cases. All in all, these are noteworthy results in view of the large forecast horizons. However, the analysis also shows that a simple random walk model produces better predictions for the future economic growth in the USA and the United Kingdom.

More importantly, the PFA predictions do not overshoot as opposed to some official projections, however, they underestimated the trends in the USA and the United Kingdom, two economies that experienced exceptionally strong growth during the nineties. At the same time, the example of Japan shows that the PFA forecasts were in some respect able to capture the decline in growth which marked the Japanese economy during the last decade. In general, however, forecasts are typically flat compared to actual growth rates and prolonged periods of boom or economic decline do not seem to be predictable. This is an analogue to the results typically found in the evaluation of business cycle forecasts (Fildes and Stekler, 2002). A common outcome in this literature is that business cycle turning points are hardly detected in advance, the same seems to hold for more longer-oriented predictions. However, in contrast to this literature, we found that forecast accuracy increases with forecast horizon. One reason that longer horizon forecasts might be more precise than shorter horizon ones is that GDP growth trends predicted by the PFA are more valid for longer periods and that over shorter periods some short-run fluctuations still prevail which are captured less precisely by such a forecasting framework. To sum up, the PFA seems to be suitable for delivering cautious predictions which are particularly useful for a sound planning of public expenditures in the medium-run.

The pseudo forecasts serve as a sort of "status quo" or neutral benchmark which incorporate an assessment of the future economic outlook if factor contributions and total factor productivity follow regular trends. The comparisons of these pseudo forecasts with projections from official authorities have shown that the German government's and the IMF's future assessments of economic developments, in particular, tend to deviate systematically from neutral assumptions and result in a systematic overestimation of actual GDP evolutions over the medium-run. These findings suggest that there is still room for improving the rationality of several officially released medium-term predictions.

Appendix

Data definitions and sources

The data source for the production function based forecast error analysis is the OECD Economic Outlook No. 79, published in June 2006. The seasonally adjusted quarterly data comprises the key variables necessary to compute potential output as shown by equation (1.5) to (1.7).
The reference variable on which the computation of cumulative growth rates and forecast errors are based on is the **real Gross Domestic Product**.

The **capital stock** is the sum of all fixed assets that provide continuous services by being employed repeatedly for output production. The data are based on a recently revised method that takes a differentiated account of the flow of productive use of different capital assets with differing age and efficiency profiles. Particularly, these new capital stock estimates feature a more precise treatment of ICT equipment in terms of price and efficiency trends.[47]

The **working age population** is the number of people in the age group of 15 to 64.

The **labor force participation rate** is defined as the number of persons in the labor force (persons employed or unemployed) as a fraction of the working age population.

The NAWRU estimates are also taken from OECD calculations. This variable is an estimate of the rate of unemployment consistent with constant wage inflation and is denoted as **non-accelerating wage rate of unemployment (NAWRU)**. The OECD uses an Kalman-Filter technique to obtain time-varying NAWRU estimates.[48]

The number of paid **hours worked per employee** on an annual basis includes paid overtime but excludes paid hours that are not worked due to vacations, sickness, etc.

Derivation of the TFP requires the use of **total employment** data, which includes all employees and self-employed persons.

As explained in section 1.3.2, **capital stock** data for **Germany** refer to the **business sector** instead of the total economy. In addition, data on **employment in the business sector** and **real GDP of the business sector** is used to calculate the TFP for Germany. Data on **employment of the government sector** is needed to adjust potential employment of the total economy.

The **Labor shares** are based on annual data from the OECD Economic Out-

[47] Beffy et al. (2006) provide a technical description of the capital stock estimation procedure.

[48] Details of the estimation design can be found in Richardson et al. (2000). In addition, this paper gives an extensive review of empirical studies of the NAWRU and empirical procedures to estimate it.

look database. The Labor share corresponds to the ratio of the **compensation of employees** over **GDP**. The adjusted labor share takes the ratio of the **total employment** over the number of **self-employed** into account.

Governmental bodies and IMF projections

Besides analyzing the pseudo forecasts of the PFA, a look is also taken at the projections published by international institutions and government bodies. For **Germany**, official projections issued by the government are taken from the medium-term fiscal outlook *Finanzplan des Bundes* which are usually published in summer and refer to a five-year prediction horizon. The data cover the period from 1985 to 2005, whereas the date always refers to the last year of the projection, e.g. the value for 2004 defines the average growth rate over the period from 2001 to the end of 2005.[49]

Official growth projections for the **USA** are taken from two sources: The first series of official projections is from the so-called *Troika*, which comprises selected staff members from the *President's Council of Economic Advisers (CEA)* and from the *U.S. Treasury and Office of Management and Budget (OMB)*. These figures are published in the *Economic Report of the President* around early February each year. The second series of projections stems from the *Congressional Budget Office (CBO)*, which unveils its future economic assessment every January in *The Economic and Budget Outlook*. Both of these 5 years ahead predictions cover the periods from 1990 to 2005.

Her Majesty's Treasury (HMT) emits medium-term projections for the **United Kingdom** in the context of the *Pre-Budget* and *Budget Report* every November and March. Here, we take the 3 years ahead projection released in March of each year. These projections range from 1987 to 2004.

For lack of suitable data from national authorities, annual predictions from the *International Monetary Fund (IMF)* are used for **France, Italy, Japan** and **Canada** to base the comparison of predictive accuracy on. The *IMF* issues medium-term projections within the biannual *World Economic Outlook* every spring and autumn. The 3 years ahead projections shown in the graphs are taken from the spring edition of each year from 1993 to 2005.

Inconsistency of the kernel-based HAC estimator

Assume that the forecast errors follow the data generating process as given by equation (1.9), which is repeated for convenience:

[49] Heinemann 2006, p. 253-254 describes the procedure of the medium-term fiscal outlook, which has remained unchanged since 1968, in greater detail.

$$e_t^h = A_t^h - F_t^h = \sum_{i=1}^{h} u_{t+i} + \phi = \nu_t^h + \phi \tag{1.27}$$

The error components u_t are iid with $E[u_t] = 0$, $E[u_t^2] = \sigma_u^2$ and $E[u_{t+i}u_{t+j}] = 0\,\forall i \neq j$. As shown in the text, this error model leads to the variances and covariances of e_t^h being $E[(\nu_t^h)^2] = h\sigma_u^2$ and $E[\nu_t^h \nu_{t+k}^h] = E[\nu_t^h \nu_{t-k}^h] = (h - |k|)\sigma_u^2$ with the symmetric covariances being solely determined by the time distance between errors and being cut off when the distance exceeds the forecast horizon. With the aid of $\widehat{\sigma}_u^2$, a consistent covariance matrix estimator, $\widehat{\Sigma}_\nu$, of Σ_ν is readily constructed (see equation 1.10) and the variance estimator for a test of $\widehat{\phi} = 0$ in an OLS regression is directly given by equation (1.15). Expanding this expression by applying matrix algebra (and skipping asterisks) results in

$$\widehat{\sigma}_\phi^2 = \frac{1}{T^2} i_T' \widehat{\Sigma}_\nu i_T = \frac{1}{T^2}\left(Th\widehat{\sigma}_u^2 + \sum_{k=1}^{h-1} 2(T-k)(h-k)\widehat{\sigma}_u^2) \right) \tag{1.28}$$

The term in brackets on the left hand side represents the sum of all elements of the block diagonal matrix $\widehat{\Sigma}_\nu$. The notation in (1.28) facilitates the subsequent comparison of the GLS covariance estimator with the non-parametric kernel-based estimator of Newey and West (1987). Since $\underset{T\to\infty}{plim}\,\widehat{\sigma}_u^2 = \sigma_u^2$, it follows that also $\underset{T\to\infty}{plim}\,\widehat{\sigma}_\phi^2 = \sigma_\phi^2$. The textbook formula for the Newey-West HAC estimator corresponding to the regression model (1.27) with an intercept as sole regressor becomes:[50]

$$\widehat{\sigma}_{\phi,NW}^2 = (X'X)^{-1}\widehat{\Sigma}_{NW}(X'X)^{-1} = \frac{1}{T^2}\widehat{\Sigma}_{NW} \tag{1.29}$$

where:

$$\widehat{\Sigma}_{NW} = \sum_{t=1}^{T} (\widehat{\nu}_t^h)^2 + \sum_{k=1}^{h-1} w_k \sum_{t=k+1}^{T} 2(\widehat{\nu}_t^h \widehat{\nu}_{t-k}^h) \tag{1.30}$$

and w_k denote kernel weights that serve the purpose of weighting down disturbance correlations as the separation in time grows and ensuring that the estimate of the covariance is positive definite.[51] Often a Bartlett kernel in the form of $1 - \frac{k}{h}$ is used for w_k. Replacing the sample disturbance moments in (1.30) with the estimates of the corresponding parameterized expressions results in the following expression for the HAC variance estimator for $\widehat{\phi}$:

[50] In the present application, the usually unknown truncation lag in Newey-West formula is completely determined by the forecast horizon h.

[51] Cf. Clements (2005), p. 8-9.

$$\begin{aligned}\widehat{\sigma}^2_{\phi,NW} &= \frac{1}{T^2}\left(\sum_{t=1}^{T} h\widehat{\sigma}^2_u + \sum_{k=1}^{h-1} w_k \sum_{t=k+1}^{T} 2(h-k)\widehat{\sigma}^2_u\right) \\ &= \frac{1}{T^2}\left(Th\widehat{\sigma}^2_u + \sum_{k=1}^{h-1} w_k 2(T-k)(h-k)\widehat{\sigma}^2_u\right) \qquad (1.31)\end{aligned}$$

On comparing the formula for the consistent GLS estimator of $\widehat{\sigma}^2_\phi$ (see equation 1.28) with the Newey-West estimator as shown in equation (1.31), it becomes clear that—unless $w_k = 1 \forall k$—$\underset{T\to\infty}{plim}\, \widehat{\sigma}^2_{\phi,NM} \neq \sigma^2_\phi$. Setting $w_k = 1$ allows consistent estimation in the Newey-West framework, however, the properties of the Newey-West estimator and the GLS estimator still differ in finite samples since the former builds on sample moment estimates of $\widehat{\nu}^h_t$ while the latter relies only on an estimate of $\widehat{\sigma}_u$. The outcomes of the experimental study demonstrate the size distortion of the non-parametric HAC estimator (see table 1.4 in the text).

Derivation of the variance and covariances of the quadratic loss-differential

Assume that $\tilde{u}_t \sim \mathcal{N}(0, \sigma^2_{\tilde{u}})$ and $\overline{u}_t \sim \mathcal{N}(0, \sigma^2_{\overline{u}})$. Furthermore, let $\tilde{u}_t$ and $\overline{u}_t$ be uncorrelated over time but contemporaneously correlated with $Cov(\tilde{u}_t, \overline{u}_t) = \sigma_{\tilde{u},\overline{u}}$. Given these assumptions, the following results for the expectations of the product of squared sums of two contemporaneously correlated random variables will be useful for the subsequent derivation:

$$E\left[(\sum_{i=1}^{h}\tilde{u}_t)^2(\sum_{i=1}^{h}\tilde{u}_{t-k})^2\right] = \begin{cases}(3h^2-4h|k|+2|k|^2)\sigma^4_{\tilde{u}} & \text{for} \quad h>|k| \\ h^2\sigma^4_{\tilde{u}} & \text{for} \quad h\leq |k|\end{cases} \qquad (1.32)$$

$$\begin{aligned}E\left[(\sum_{i=1}^{h}\tilde{u}_t)^2(\sum_{i=1}^{h}\overline{u}_{t-k})^2\right] &= \\ E\left[(\sum_{i=1}^{h}\overline{u}_t)^2(\sum_{i=1}^{h}\tilde{u}_{t-k})^2\right] &= \begin{cases}h^2\sigma^2_{\tilde{u}}\sigma^2_{\overline{u}} + 2(h-|k|)^2\sigma^2_{\tilde{u},\overline{u}} & \text{for} \quad h>|k| \\ h^2\sigma^2_{\tilde{u}}\sigma^2_{\overline{u}} & \text{for} \quad h\leq |k|\end{cases} \qquad (1.33)\end{aligned}$$

The aim, however, is to show that the covariance between the loss differential d^h_t and its lagged values with displacement k takes the following form:

$$\gamma_d(k) = \begin{cases}2\check{h}\left(\sigma^2_{\tilde{u}}(\check{h}\sigma^2_{\tilde{u}} + 2\tilde{\phi}^2) + \sigma^2_{\overline{u}}(\check{h}\sigma^2_{\overline{u}} + 2\overline{\phi}^2) - 2\sigma_{\tilde{u},\overline{u}}(\check{h}\sigma_{\tilde{u},\overline{u}} + 2\tilde{\phi}\overline{\phi})\right) & \text{for} \quad h>|k| \\ 0 & \text{for} \quad h\leq|k|\end{cases}$$

whereas $\check{h} = h - |k|$.

Given the forecast error models of equation (1.19) and (1.20), the quadratic loss differential becomes

$$d_t^h = (\tilde{e}_t^h)^2 - (\overline{e}_t^h)^2 = (\sum_{i=1}^{h} \tilde{u}_{t+i} + \tilde{\phi})^2 - (\sum_{i=1}^{h} \overline{u}_{t+i} + \overline{\phi})^2 \tag{1.34}$$

The covariance is computable as $\gamma_d(k) = E[d_t^h d_{t-k}^h] - E[d_t^h]E[d_{t-k}^h]$. The parametric solution for the product of the two expected values in this expression can readily be obtained.

Since

$$\begin{aligned} E[d_{t-k}^h] &= E\left[(\sum_{i=1}^{h} \tilde{u}_{t-k+i})^2 + 2\tilde{\phi}\sum_{i=1}^{h} \tilde{u}_{t-k+i} + \tilde{\phi}^2\right] \\ &- E\left[(\sum_{i=1}^{h} \overline{u}_{t-k+i})^2 + 2\overline{\phi}\sum_{i=1}^{h} \overline{u}_{t-k+i} + \overline{\phi}^2\right] &(1.35) \\ &= h\sigma_{\tilde{u}}^2 - h\sigma_{\overline{u}}^2 + \tilde{\phi}^2 - \overline{\phi}^2 \qquad \forall\, k &(1.36) \end{aligned}$$

it follows that

$$E[d_{t-k}^h]E[d_t^h] = \left(\tilde{\phi}^2 - \overline{\phi}^2 + h(\sigma_{\tilde{u}}^2 - \sigma_{\overline{u}}^2)\right)^2 \tag{1.37}$$

In contrast, deriving the solution for the expected value of the product of the loss differential and its lagged value is more cumbersome. Expanding $E[d_t^h d_{t-k}^h]$ and omitting expressions with an expected value of zero leads to

$$
\begin{aligned}
E[d_t^h d_{t-k}^h] &= E[\tilde{\phi}^4] + E[\overline{\phi}^4] - 2E[\tilde{\phi}^2\overline{\phi}^2] + E\left[\tilde{\phi}^2(\sum_{i=1}^{h}\tilde{u}_{t+i})^2\right] - E\left[\overline{\phi}^2(\sum_{i=1}^{h}\tilde{u}_{t+i})^2\right] \\
&+ 4E\left[\tilde{\phi}^2\sum_{i=1}^{h}\tilde{u}_{t+i}\sum_{i=1}^{h}\tilde{u}_{t-k+i}\right] + E\left[\tilde{\phi}^2(\sum_{i=1}^{h}\tilde{u}_{t-k+i})^2\right] - E\left[\overline{\phi}^2(\sum_{i=1}^{h}\tilde{u}_{t-k+i})^2\right] \\
&+ E\left[(\sum_{i=1}^{h}\tilde{u}_{t+i})^2(\sum_{i=1}^{h}\tilde{u}_{t-k+i})^2\right] - 4E\left[\tilde{\phi}\overline{\phi}\sum_{i=1}^{h}\tilde{u}_{t-k+i}\sum_{i=1}^{h}\overline{u}_{t+i}\right] \\
&- E\left[\tilde{\phi}^2(\sum_{i=1}^{h}\overline{u}_{t+i})^2\right] + E\left[\overline{\phi}^2(\sum_{i=1}^{h}\overline{u}_{t+i})^2\right] - E\left[(\sum_{i=1}^{h}\tilde{u}_{t-k+i})^2(\sum_{i=1}^{h}\overline{u}_{t+i})^2\right] \\
&- 4E\left[\tilde{\phi}\overline{\phi}\sum_{i=1}^{h}\tilde{u}_{t+i}\sum_{i=1}^{h}\overline{u}_{t-k+i}\right] + 4E\left[\overline{\phi}^2\sum_{i=1}^{h}\overline{u}_{t+i}\sum_{i=1}^{h}\overline{u}_{t-k+i}\right] \\
&- E\left[\tilde{\phi}^2(\sum_{i=1}^{h}\overline{u}_{t-k+i})^2\right] + E\left[\overline{\phi}^2(\sum_{i=1}^{h}\overline{u}_{t-k+i})^2\right] - E\left[(\sum_{i=1}^{h}\tilde{u}_{t+i})^2(\sum_{i=1}^{h}\overline{u}_{t-k+i})^2\right] \\
&+ E\left[(\sum_{i=1}^{h}\overline{u}_{t+i})^2(\sum_{i=1}^{h}\overline{u}_{t-k+i})^2\right] \qquad (1.38)
\end{aligned}
$$

To find the parametric solution of (1.38), the cases $h > |k|$ and $h \leq |k|$ need to be differentiated.

From corollary (1.32) and (1.33), the following results:

For $h > |k|$,

$$
\begin{aligned}
E[d_t^h d_{t-k}^h] &= \tilde{\phi}^4 + \overline{\phi}^4 - 2h\overline{\phi}^2\sigma_{\tilde{u}}^2 - 3h^2\sigma_{\tilde{u}}^2 - 4hk\sigma_{\tilde{u}}^2 + 2k^2\sigma_{\tilde{u}}^2 \\
&- 2\sigma_{\overline{u}}^2(h^2\sigma_{\tilde{u}}^2 + 2k\overline{\phi}^2 - 3h\overline{\phi}^2) + \sigma_{\overline{u}}^2(3h^2 - 4hk + 2k^2) \\
&- 2\tilde{\phi}^2(\overline{\phi}^2 - 3h\sigma_{\tilde{u}}^2 + 2k\sigma_{\tilde{u}}^2 + h\sigma_{\overline{u}}^2) - 4\sigma_{\tilde{u},\overline{u}}^2(h-k)^2 \\
&- 8\tilde{\phi}\overline{\phi}(h-k)\sigma_{\tilde{u},\overline{u}} \qquad (1.39)
\end{aligned}
$$

and subtracting (1.37) from (1.39) gives the non-zero autocovariance formula as shown above and in equation (1.25) in the text.

For $h \leq |k|$, the term in (1.38) collapses to

$$
E[d_t^h d_{t-k}^h] = \left(\tilde{\phi}^2 - \overline{\phi}^2 + h(\sigma_{\tilde{u}}^2 - \sigma_{\overline{u}}^2)\right)^2 \qquad (1.40)
$$

which is the same as $E[d_{t-k}^h]E[d_t^h]$. Therefore, in this case the autocovariance of d_t is zero.

Chapter 2

Estimating Trend Growth Using Panel Techniques

2.1 Introduction

Given that explaining cross-country growth differentials is of major interest in economics, it is hardly surprising that there is an extensive literature of theoretical and empirical research on economic growth. The weak growth rates of the larger EU-countries during the second half of the nineties, in particular, have revived public interest in this topic.

This chapter does not aim to introduce a new set of potentially growth enhancing or growth impeding variables, neither will it give a comprehensive survey of the vast growth literature. Rather, the existing evidence and suggestions are used and a new approach of estimating trend growth of advanced economies is proposed. The suggestion seeks to combine econometric methods that have been used to test and estimate the implications of the extended Solow growth model in a cross sectional time series setting with an application of multivariate time series filter techniques. Filter techniques are used in order to "smooth" estimates of trend growth which result from fitted values of structural econometric relationships. Dealing with a cross country time series data set involves issues of panel integration and panel cointegration under cross sectional dependence as well as assumptions about the degree of homogeneity and heterogeneity of model parameters. Moreover, the problem of robustness of correlations between growth and the potential determinants that has been raised by the literature needs to be accounted for.

The econometric specification in the present study is derived from an augmented neoclassical growth model allowing for a non-diminishing returns to scale production function, which is a standard approach in empirical research (e.g. Mankiw et al., 1992, Islam, 1995, Barro, 1997). A straightforward extension of this standard model is pursued that goes back to Cellini (1997), Sarno (1999) and to Bassanini et al. (2001). The latter authors focus on the role of policies and institutions for growth of OECD countries. Panel econometric techniques proposed by Pesaran et al. (1999) and Pesaran (2006) are used in order to estimate panel error correction models (ECM) which account for cross sectional dependence through a factor model. Furthermore, panel integration and cointegration methods suggested by Demetrescu et al. (2006) and Pedroni (1999, 2004) are employed. Estimated equilibrium paths are smoothed with multivariate filter techniques following the approach of Gosselin and Lalonde (2006).

This chapter is divided into 6 sections. Section 2.2 concisely reviews some key concepts of the theoretical and empirical growth literature in order to motivate the analysis and in particular the variables considered in the empirical part of this chapter. The following section outlines the theoretical embedding and empirical strategy while section 2.4 is devoted to explaining and discussing the chosen econometric approach and strategy for the model specification search. Data and results are presented in section 2.5 while section 2.6 summarizes and concludes. A literature survey of the various approaches to measure trend or potential output, which would complete the following discussion, is provided in section 1.2 of chapter 1 of this thesis and will be omitted here.

2.2 The sources of economic growth

Extensive surveys of variables used in empirical growth studies can be found in Barro and Sala-i-Martin (2001), Durlauf et al. (2005) and Durlauf and Quah (1999). Subsequently, only a brief overview of the most significant concepts will be given. Temple (1999) introduced a helpful classification of variables typically considered in the growth literature which will also be adopted here.

2.2.1 The proximate sources of growth

The current understanding of economic growth is greatly influenced by the new (endogenous) growth theory. Aghion and Howitt (1998) provide a concise summary of many ideas of this strand of theory. However, empirical work still dates back to the neoclassical growth model of Solow (1956), although it has undergone many extensions and changes in interpretation particularly in the light of the new growth theories. The proximate sources of growth basically comprise the variables considered in the original Solow-model, which is why they are sometimes denoted as "Solow variables". Besides the variables that will be discussed in more detail below, the baseline Solow model usually comprises a term that includes the population growth, the depreciation rate and the growth rate of the level of technology. A formal presentation of the extended neoclassical growth model will be given in section 2.3.1. In the following paragraphs a more narrative overview of the basic ideas will be given.

Physical capital

The neoclassical model considers a production function with physical capital as the only reproducible input into the production process. The critical assumption of this model is that returns to scale to physical capital are diminishing. Therefore, investment in capital influences the level of aggregated output rather than the growth rate. Hence, in the long-run the growth rate of the economy is a function of the exogenous rate of population growth, the exogenous rate of technological change and the natural rate of capital depreciation. Even though the neoclassical model does not explain growth sustaining determinants based on economic decisions, it does at least point to the most important factor: technological change.

Endogenous growth models support a broader view of capital and relax the assumption of diminishing returns to capital. The models of endogenous technological change developed by Grossman and Helpman (1991) and Aghion and Howitt (1992), for instance, incorporate knowledge as additional input to production.[1] On the firm-level, the utilization of knowledge in the production process still features diminishing returns to scale but generates economy-wide externalities through spill-over effects. Externalities also apply to physical capital if technological innovation is embodied in new capital and improves the economy-wide adoption of new technologies. De Long and Summers (1991) point to the importance of investment in equipment as a source of externalities. For empirical applications, the accumulation of physical capital remains one of the key variables. It is usually measured as the investment share of GDP.

Human capital

Another way to introduce externalities into growth models with a broad view of capital is to consider the role of human capital. Classical references for growth models of human capital are Romer (1986) and Lucas (1988). Since the work of Mankiw et al. (1992) was published, it has became a standard approach of empirical growth models to include a measure of human capital stock or a measure of its accumulation together with physical capital. This usually improves the fit of empirical models and is in line with theories emphasizing the importance of education and training

[1] A formal discussion of these types of endogenous growth models can be found in Durlauf and Quah (1999).

for growth.

The question whether the stock of human capital or its rate of accumulation matters for economic growth has recently aroused interest and its answer mainly depends on the theoretical approach. The Lucas (1988) model clearly stresses the importance of the accumulation of human capital while models of endogenous technological progress usually account for the stock of human capital. The empirical results obtained by Teles (2005) indicate that the Lucas (1988) model satisfactorily explains the growth rate of "rich" countries. Against the background of these findings the empirical part of this chapter (see section 5) also uses a measure of human capital investment rather than a measure of the human capital stock since the attention is directed to growth determinants of advanced industrial countries. Due to data limitations the focus of the human capital variable will be on schooling rather than training. Therefore, just one dimension of the much broader concept of human capital will be considered.

2.2.2 The wider sources of growth

The potential determinants of growth considered in the next sections are basically a subset of the many candidates considered by new growth series that have attracted a great deal of attention in the context of advanced countries. Much of these determinants can be seen to provide a deeper understanding of the central component of long-run growth where the Solow model leaves a void, namely the evolution of Total Factor Productivity (TFP). The technological level of an economy determines the productivity of the inputs to production. Besides the technological efficiency under which an economy operates there is also the political and institutional setting that affects the overall productivity. In fact, most variables in recent empirical growth studies that go beyond the analysis of the proximate sources of growth belong to this category. Two issues in particular have attracted attention: Fiscal and monetary policy as well as financial market development. In addition to the above, the following section will examine the role of international trade and demography.

Research and development

Theories of endogenous technological growth in the spirit of Romer (1986) naturally emphasize the influence of research and development (R&D) on economic growth. However, as Temple (1999) points out, even though there is already an overwhelming microeconomic evidence for high private returns to R&D, there are some well-known problems in measuring the contribution of research to productivity growth. One reason for the difficulty in resolving research driven growth models is that the underlying concept of R&D-models (the concept of knowledge and ideas) is so hard to pin down.[2] Nevertheless, in empirical models one can try to proxy for R&D through private and public expenditures on R&D.

Inflation

Though monetary policy is usually referred to through its impact on business-cycles, there could also be an impact on the long-run growth path via investment and investment uncertainty. Traditionally, the influence of inflation on growth is analyzed by means of considering the influence of monetary growth changes on the level of the physical capital stock in the context of a neoclassical growth model. Tobin (1965) argues that inflation increases the opportunity cost of holding money and therefore encourages people to invest in physical capital. Temple (2000) points out that this should not be an important consideration since money balances are usually only a small fraction of the physical capital stock and therefore this effect of inflation on capital accumulation seems negligible. The effect mentioned by Tobin (1965) can even be reversed when altering assumptions: If money has to be held prior to the purchase of capital goods, inflation is expected to lower the steady-state capital stock (Stockman, 1981). The effects of inflation on investment are becoming increasingly significant when considering endogenous growth theories in which returns to broad capital are constant.

A possibly stronger influence of inflation on investment is exerted through the tax system: An imperfectly indexed tax system increases the user cost of capital when inflation rises, since the value of depreciation allowances falls simultaneously. A higher user cost of capital increases the profitability required to undertake an investment project. Therefore, there should be an overall negative impact through inflation on the accumulation of physical capital and possibly steady-state growth (Feldstein, 1983). Besides the level of inflation its variability could also be connected to investment and growth through the impact on uncertainty (Anh and Hemmings, 2000). Arguments brought forward in this spirit point to the fact that inflation increases uncertainty and therefore introduces noise into the workings of markets. Both the level of inflation and its variation will be considered in the empirical section of this chapter.

Government activity

In terms of fiscal policy the role of the government in setting the economic framework in which economic growth takes place is of major concern. Many publicly financed activities are not aimed at improving economic growth in the first place. The levy of social contributions in order to redistribute resources or the public stimulation of the

[2] See Temple (1999), p.140-141 for a full discussion.

demand side of the economy in times of weak overall demand are just two examples of government activities that do not aim at raising long-run growth, but do follow wider policy objectives. Countries with huge public sectors may extend activities into areas that might be more efficiently carried out by private agents leading to productivity losses on the aggregate level. On the other hand there is Wagner's law which implies that the income elasticity of demand for government services is larger than unity since the scope of government increases with the level of income. Fölster and Henrekson (2001, 2006) provide a deeper discussion of Wagner's law and the role of the government with regard to economic growth. Therefore, in an empirical investigation the observed correlation between income and government size may be positive. A priori, the expected sign of the relationship between government size and economic growth seems unclear. The variables used in the econometric analysis are real government consumption as a proportion to GDP as well as the sum of direct and indirect taxes and social contributions received by government as a proportion to GDP. Real government consumption is more than just a proxy for a special component of public spending. This variable is often perceived as an indicator of government size and the importance of the public sector in the economy. Moreover, the government deficit will be considered.

Taxes

In terms of the composition of public revenues it is useful to distinguish between distortionary taxes (direct taxes, e.g. taxes on income and firm profits) and taxes that are generally regarded as less distortionary (indirect taxes, taxation of goods and services). Distortionary taxes could affect the investment decision in physical and human capital of agents and hence reduce growth, while non-distortionary taxes do not influence the incentives of economic agents (Widmalm, 2001, Padovano and Galli, 2002 and Milesi-Ferretti and Roubini, 1998.) For example, if labor supply is inelastic, the intertemporal consumption path of an agent is not influenced by a consumption tax or a flat tax on labor income. Since not all taxes may be equally distorting, the tax mix is possibly an important growth determinant. In this chapter it will be accounted for via the tax-ratio variable (the proportion of total indirect taxes to total direct taxes).

Labor markets

Another source of influences on economic growth is the institutional design of national labor markets. Labor market rigidities in particular are likely to affect the growth rate negatively because they lead to under-utilization of the human capital stock. De La Fuente (2003) notes that since the underlying theoretical models of empirical growth-studies describe the evolution of labor productivity (i.e. output per employed worker), it should be expected that the growth of income *per capita* will also depend on the behavior of participation and unemployment rates. Hence the empirical model of this chapter which also uses GDP per capita as dependent variable considers a measure for the non-cyclical unemployment rate in the form of the NAWRU (Non-Accelerating Wage Rate of Unemployment).

Financial markets

A special strand of the growth literature focuses on the relationship between financial market development and economic growth. Levine (1997, 2005) gives a comprehensive overview of this issue. These theories emphasize the efficiency with regard to information processing of highly developed financial markets. Financial markets collect and distribute information concerning investment projects efficiently and allow investors to pool risks through the allocation of various financial assets. Particularly with regard to the funding of new technologies, stock markets seem to be superior to credit funding due to their efficient allocation of information (Allen and Gale, 2000). Besides the gains in overall efficiency, developed financial markets may increase the level of investments through the provision of attractive assets. The empirical literature is somewhat ambiguous about the growth enhancing properties of financial markets when considering industrial countries (Rousseau and Wachtel, 2005). In this chapter the traditional set of indicators to proxy for the influence of financial markets is used: The stock market capitalization of listed companies to GDP as an indicator of relative development and size of a stock market. Furthermore, the turnover which is defined as the ratio of stocks traded to market capitalization is considered as a proxy for market liquidity. In order to account for the influence of credit markets the variable of total credits to the private sector over GDP is also included.

International trade

A further sphere of influence on growth to be considered is international trade. The economic literature indicates several arguments as to why engagement in international trade could be beneficial for an economy. Traditional reasoning is based on the idea that trade is promotive because comparative advantages are exploited. However, other arguments are also brought forward. The exposure to competition through openness or the diffusion of technology through trade can stimulate the economic growth of a country. Coe and Helpman (1995) emphasize the role of trade for technological diffusion. Trade, however, can also simply be a reflection of growth patterns in the sense that trade is endogenous to the growth progress (Baldwin, 2000). To proxy for the trade openness of a country the share of exports and imports to GDP is used in the empirical analysis. Furthermore, to capture trade effects as a result of price advantages or disadvantages, an indicator of the terms of trade (export prices over import prices) is accounted for.

Demography

The last issue to be considered is demography and the consequences of an aging population on growth. There are different arguments as to why aging could influence the growth path. On the hand there are influences on the input factors to production. The relative provision of inputs is likely to change under an aging society since the relative labor force decreases, which leads to an increased capital-labor ratio. This, in turn, may lead to a disinvestment of physical capital since the ratio of capital to labor needs to be adjusted, leading to slower rates of GDP growth. There could also be influences of an aging population on the rate of technological progress if innovations and technological adaptation are exacerbated by an older population whose human capital is sufficiently depreciated. Siebert (2002) provides these arguments. However, empirical evidence for these theories is flawed by the fact that no country has undergone the whole process of aging so far. Nevertheless, these ideas are incorporated into the empirical section by means of the variable of age dependency and the variable of the proportion of the population over the age of 65 to total population.

By no means do the above described wider sources of growth constitute a complete set of growth-enhancing or growth-impeding variables. Instead, this set of variables is intended to bundle the most evident record of the theoretical and empirical growth literature. A complete list of the variables used in this chapter and the corresponding data sources can be found in the appendix. The next section explains the empirical model and the econometric approach which is used to shed light on the empirical relevance of the above discussed concepts.

2.3 Theoretical embedding and empirical strategy

The empirical analysis refers to an extended New Classical Solow growth model and in particular to its empirical implications raised by Cellini (1997), Sarno (1999), Binder and Pesaran (1999) and Pesaran (2004b). These authors basically show that the Solow growth model implies a stochastic steady state if the variables that determine equilibrium output per capita are stochastic variables. Traditionally, the implications of the Solow model have been studied with the aid of pure cross-sectional data in the spirit of Barro's regressions and in that case, variation of variables over time is ruled out by construction. The assumption that variables are characterized by stochastic trends as soon as one deals with cross sectional *time series* seems obvious. As a consequence, error correction and cointegration techniques are the natural tools to test the implications of the Solow model. The specific empirical design of the subsequent analysis builds on Mankiw et al. (1992), Islam (1995) and in particular on Bassanini et al. (2001), who augment the standard New Classical growth model with a set of political and institutional variables. The approach is briefly reviewed and the dynamic equation employed in the empirical testing is derived. Subsequently, the econometric techniques used for estimating the dynamic models and for panel cointegration testing are outlined.

2.3.1 Derivation of steady-state dynamics

The model is derived from a textbook New Classical approach that builds around a constant returns to scale production function with three inputs (labor, physical and human capital). In this section we will look at the deterministic version of the model in order to set out the general theoretical embedding. Considerations of the stochastic nature of the variables and the steady-state condition are provided in the subsections where the econometric formulation of the model is given.

The aggregate production function for country i in period t is

$$Y_i(t) = K_i(t)^{\alpha} H_i(t)^{\beta} (A_i(t) L_i(t))^{1-\alpha-\beta} \qquad \alpha, \beta > 0,\ \alpha + \beta < 1 \tag{2.1}$$

$Y_i(t)$ is aggregate production, $K_i(t)$ the stock of physical capital, $L_i(t)$ labor input, $H_i(t)$ the stock of human capital and $A_i(t)$ the Harrod-neutral level of economic and technological efficiency. The partial elasticities of output with regard to its inputs are α and β, respectively, and are assumed to be identical for all countries. Following standard assumptions, the dynamics of physical capital, human capital and labor are:[3]

$$\dot{K}_i(t) = s_i^K(t) Y_i(t) - \delta K_i(t) \tag{2.2}$$

$$\dot{H}_i(t) = s_i^H(t) Y_i(t) - \delta H_i(t) \tag{2.3}$$

$$\dot{L}_i(t) = n_i(t) L_i(t) \tag{2.4}$$

$s_i^K(t)$ is the fraction of output that is invested in physical capital and $s_i^H(t)$ is the fraction that is invested in human capital. Population grows at the exogenous rate $n_i(t)$. Physical and human capital depreciate at the same rate of δ. The level of economic and technological efficiency $A_i(t)$ consists of two components: economic

[3] Dotted variables represent derivatives with respect to time.

efficiency $I_i(t)$ and the level of technological progress $\Omega_i(t)$:

$$A_i(t) = I_i(t)\Omega_i(t) \tag{2.5}$$

The level of technological progress grows at the exogenous rate $g_i(t)$:

$$\dot{\Omega}_i(t) = g_i(t)\Omega_i(t) \tag{2.6}$$

whereas the level of economic efficiency is a log-linear function of institutional and policy variables $V_i^j(t)$ (the variables described in the previous section):[4]

$$\ln I_i(t) = p_i(0) + \sum_{j=1}^{M} p_j \ln V_i^j(t) \tag{2.7}$$

Let $k_i(t) = K_i(t)/A_i(t)L_i(t)$ and $h_i(t) = H_i(t)/A_i(t)L_i(t)$ be quantities per effective unit of labor. Since $\alpha + \beta < 1$, the economy converges to a steady state defined by

$$\overset{*}{k}_i(t) = \left(\frac{(s_i(t)^K)^{1-\beta}(s_i(t)^H)^{\beta}}{n_i(t) + g_i(t) + \delta}\right)^{\frac{1}{1-\alpha-\beta}} \tag{2.8}$$

and[5]

$$\overset{*}{h}_i(t) = \left(\frac{(s_i(t)^K)^{\alpha}(s_i(t)^H)^{1-\alpha}}{n_i(t) + g_i(t) + \delta}\right)^{\frac{1}{1-\alpha-\beta}} \tag{2.9}$$

Note that unlike in the textbook version of the Solow model, the steady-state values in (2.8) and (2.9) are not constant but vary both across countries and over time since the determining variables are not constant, either. Substituting these expressions into the production function and taking logarithms produces the following equation for output per capita in steady state:

$$\begin{aligned}\ln\left[\frac{\overset{*}{Y_i}(t)}{L_i(t)}\right] = \ln \Omega_i(t) + p_i(0) + \sum_{j=1}^{M} p_j \ln V_i^j(t) + \frac{\alpha}{1-\alpha-\beta} \ln s_i^K(t) + \\ \frac{\beta}{1-\alpha-\beta} \ln s_i^H(t) - \frac{\alpha+\beta}{1-\alpha-\beta} \ln\left(g_i(t) + n_i(t) + \delta\right)\end{aligned} \tag{2.10}$$

Output per capita in steady-state depends on the accumulation of physical and

[4]See Bassanini et al. (2001). The number of institutional factors M will in principle be very large. However, in order to keep the analysis empirical operational, only a small subset of these variables can be considered for which, however, it is assumed that they comprise the most significant influences of economic efficiency.

[5]Star superscripts denote steady state values. For a more rigorous derivation see Mankiw et al. (1992) and Cellini (1997).

human capital, population growth, the rate of technological progress, the rate of depreciation and the level of institutional and political efficiency. Equation (2.10) describes the evolution of the level of output per capita in the long-run. Mankiw et al. (1992) derive the dynamics of output per capita in the neighborhood of the steady state. Cellini (1997) shows that these dynamics imply that the movements of output per capita in a given country follow an error correction mechanism, an implication which will be of major interest in the following sections.

2.4 Econometric approach and specification search

This section outlines the employed econometric techniques to estimate the steady-state equation (2.10) in a dynamic panel framework which allow for a certain degree of heterogeneity and cross section dependence among countries. Subsequently, the pursued strategy for model specification will be explained. The presence of such a large number of possible growth determinants has raised questions of model identification and uncertainty in the empirical growth literature. It will be discussed how the strategy of the present chapter is related to this literature which is typically concerned with the specification search of empirical growth models in the context of pure cross section surveys.

2.4.1 Panel integration and cointegration tests

The time series of the present survey are regularly characterized by deterministic and stochastic trends. These properties have to be accounted for in econometric estimation. In order to test the non-stationarity of the variables, panel unit root tests are considered. An extensive survey and application of several panel unit root tests in the presence of heterogeneous and cross correlated countries is given in chapter 3 of this thesis, hence they will only be referred to briefly in this section. Chapter 3 shows that panel unit root tests increase power against univariate procedures. However, as soon as the data is characterized by cross section dependencies these tests can be severely biased if this kind of dependence is not accounted for in the testing procedure. The assumption that the OECD countries are independent is obviously not a suitable hypothesis. Therefore, in order to use robust techniques, the second generation panel unit root test proposed by Demetrescu et al. (2006, DHT) is considered. DHT show that this test is fairly reliable when applied to ADF tests in correlated panels, even when the cross-section and time dimension are of moderate sizes. The DHT test builds on the p-values obtained from univariate unit root tests and combines them to the single test statistic $t(\widehat{\rho}^*, \kappa)$ with the aid of the modified inverse normal method (see section 3.4.3 in chapter 3 for details). Under the null hypothesis that all the time series are nonstationary, the test statistic results in an approximate standard normal distribution.

After having detected which variables are potential candidates for establishing a long-run relationship by means of panel unit root tests, cointegration analysis is the next consequential step. Consider the following static regression

$$y_{it} = \varphi_i d_{it} + \mathbf{x}_{it}\beta_i + u_{it} \tag{2.11}$$

in which $\varphi_i d_{it}$ represents a deterministic part including any individual intercepts or time trends or both, and $\mathbf{x}_{it}$ is a $k \times 1$ vector of $I(1)$ regressors, i.e. these regressors are integrated of order one. Given that our focus is on establishing an empirical counterpart of the steady-state condition for the log of GDP per capita, $\mathbf{x}_{it}$ comprises all variables entering the right hand side of (2.10), i.e. in discrete time notation $\mathbf{x}_{it} = (\ln s_{it}^K, \ln s_{it}^H, n_{it}, \ln V_{it}^5, \ldots, \ln V_{it}^m)'$. Here it is assumed that the growth rate of technological progress g_{it} can partly be modeled as a deterministic time trend, while the time-invariant depreciation rate δ is a component of the intercept term.

Entorf (1997) and Kao (1999) demonstrate that the problem of establishing a spurious regression result is more likely to occur in panel data regressions than in pure time series studies. For this reason it is important to check whether the errors u_{it} in a panel data regression as given by equation (2.11) are stationary.

In the following we will consider single-equation approaches to test for panel cointegration and to estimate long-run relationships. System approaches for estimating the number and parameters of the cointegration vectors in panel data models have been developed by Larsson et al. (2001) and Groen and Kleibergen (2003). However, Breitung and Pesaran (2005) point to the poor small sample performance of the ML estimators on which Larsson et al. (2001) base their strategy and the nonlinear GMM techniques underlying Groen and Kleibergen's (2003) procedure do not seem to be appropriate for consideration of small samples, either. Recently, a Global VAR approach has been proposed by Pesaran et al. (2004) and applied by Dees et al. (2005) which likewise requires fairly large data sets in order to produce reasonable results. In general, the demand for data of the system approaches is quite high. Therefore, we will adhere to the single-equation framework, given the moderate sample size of the present study, despite the fact that it imposes the rather strong restriction that either zero or one cointegration vectors are permitted. However, the theory of long-run growth suggests one cointegration relationship and this shall be our main concern for empirical testing in the remainder of the chapter.

The determination of the order of integration for the variables is important for setting up the cointegration analysis. If there is a linear combination of two or more non-stationary series that is stationary, the non-stationary time series are said to be cointegrated. This stationary combination can be interpreted as long-run equilibrium. In this chapter, cointegration tests are carried out by applying the unit root test from DHT to the residuals of the estimated static cointegration regression following equation (2.11) under the assumption of slope homogeneity ($\beta_i = \beta \, \forall i$). Basically, this is a panel residual based approach similar to the cointegration tests in the single country analysis of Engle and Granger (1987). As before, cross-sectional dependency is accommodated in the DHT framework by taking the correlation of the underlying probits into account. Testing whether homogeneous slopes can be maintained in the long-run relationship is of primary concern of the present study, which assumes that the considered industrial countries have similar steady-states and in particular seeks to exploit the cross section information on steady-state parameters. Pooling the long-run parameters may fail to deliver a viable cointegration vector. However, cointegration may still exist if slopes are permitted to be country-specific. To test for this possibility, we will also look at Pedroni's (1999, 2004) method to test for cointegration in heterogeneous panels with multiple regressors.

The Pedroni (1999) test builds on the residuals u_{it} from cointegration regressions with country specific cointegration vectors. Pedroni (1999) derives test statistics which converge to standard normal distributions under the null hypothesis. The Pedroni (1999) test is based on testing the order of integration of the residuals of the cointegration regression such as (2.11). The author offers seven statistics to test whether the autoregressive coefficient of the residual based regression is unity. Four test-statistics pool the autoregressive coefficient (ρ_i) along the so called within-dimension, while three statistics pool along the between-dimension by taking the average of the coefficients. The four statistics that refer to the within-dimension are based on the following hypotheses

$$H_0(\text{No cointegration}) : \rho_i = 1 \quad \forall i, \quad H_1 : \rho_i = \rho < 1 \quad \forall i$$

Both under H_0 as well as under H_1 these statistics assume a common value for the autoregressive coefficient ρ of the underlying residual regression. By contrast, the test statistics of the between-dimension allow for an additional source of potential heterogeneity, since they do not assume a common autoregressive coefficient on the residual test. The hypotheses are

$$H_0(\text{No cointegration}) : \rho_i = 1 \quad \forall i, \quad H_1 : \rho_i < 1$$

Pedroni (1999) refers to the four within test statistics as panel cointegration statistics and to the three between test statistics as group mean cointegration statistics. The panel cointegration statistics are a variance bounds test (v-statistic), an analogue to the Phillips-Perron (1988) ρ test (panel rho-statistic) and nonparametric Phillips-Perron (1988) test (panel PP-statistic) and a parametric ADF statistics (panel ADF-statistic). The group tests are the group rho-statistics, the group PP-statistic and the group ADF-statistic. These one-sided test statistics are distributed asymptotically standard normal after appropriate standardization.[6] The Pedroni (2004) test is not explicitly designed to deal with cross-section correlation but can take common time effects into account in order to accommodate some forms of cross-sectional dependence. For this purposes it is supposed that the disturbances for each country can be decomposed into common disturbances that are shared among all countries and independent idiosyncratic disturbances that are specific to each country (Pedroni, 2004). In order to derive tests statistics that are robust to such a form of dependence, the variables are demeaned before they enter the regression (2.11), i.e. each variable is expressed as the derivation from its time series average in the form of $\tilde{z}_{it} = z_{it} - 1/T \sum_{t=1}^{T} z_{it}$. In the next section we will consider a factor model which is much more flexible with respect to the accommodation of cross section dependence, in that it allows the countries to be influenced by the common factor fairly individually.

When faced with seven test statistics, the question arises which one is the most suitable for finite samples. Fortunately, Pedroni (2004) provides extensive Monte Carlo studies with regard to size and power of the various panel cointegration statistics he offers to the researcher. In conclusion, in very small samples the group-rho statistics are somewhat undersized and constitute the most conservative of the seven tests. Pedroni (2004) further concludes that the panel-v statistic tends to have the greatest power in large panels relative to the other statistics and can be most useful when the alternative is very close to the null hypothesis. The other test statistics lie somewhere in between these two cases. In Pedroni (1997) more detailed Monte Carlo results with respect to the ADF versions of the tests are reported. In these experiments the group-ADF statistic appears to be the most powerful in smaller samples (when T is smaller than 20), followed by the panel v-statistic and panel rho-statistic. In the light of these results and given the small dimension of the panel data at hand, we chose to base test decision mainly on the outcomes of the group-ADF tests.

Other residual based panel cointegration tests have also been proposed, notably by Kao (1999) and Westerlund (2005). However, we regard the tests outlined above as being sufficient for testing for spurious correlations in the present framework and therefore restrict the panel cointegration testing to these procedures.

[6] See Pedroni (1999) for exact calculation and required moment adjustments of these statistics.

2.4.2 Panel Error Correction Model (ECM) and estimation

This section outlines the single-equation panel ECM used for estimation and explains the Pooled Mean Group (PMG) estimator proposed by Pesaran, Shin and Smith (1999, PSS hereafter). If the variables in question are cointegrated they can subsequently be formulated as a ECM model. In principle, the coverage of the data allows us to estimate N separate regressions or N separate ECMs. However, the aim here is to exploit the cross-sectional dimension of the data to gain more precise estimates of the long-run relationship and to increase power of applied tests statistics.

Denoting $y_{it} = \ln(Y_{it}/L_{it}),\ i = 1, ..., N, t = 1, ..., T$ as the logarithm of output per capita for country i, the following ECM corresponding to the steady-stage condition in (2.10) can be considered for estimation[7]

$$\begin{aligned}\Delta y_{it} = -\phi_i &\left(y_{it-1} - \varphi_{0i} - \varphi_{1i} t - \theta_{1i} \ln s_{it}^K - \theta_{2i} \ln s_{it}^H - \theta_{3i} n_{it} - \sum_{j=4}^{m} \theta_{ij} \ln V_{it}^j \right) \\ &+ \sum_{k=1}^{p_i} b_{0ik} \Delta y_{it-k} + \sum_{k=0}^{p_i} b_{1ik} \Delta \ln s_{it-k}^K + \sum_{k=0}^{p_i} b_{2ik} \Delta \ln s_{it-k}^H \\ &+ \sum_{k=0}^{p_i} b_{3ik} \Delta \ln n_{it-k} + \sum_{j=4}^{m} \sum_{k=0}^{p_i} b_{ik}^j \Delta \ln V_{it-k}^j + \epsilon_{it}\end{aligned} \tag{2.12}$$

The first term in brackets on the right hand side of equation (2.12) is the long-run equilibrium of the logarithm of GDP per capita. Note that the error correction equations are formulated in terms of current rather than lagged levels of the regressors. This follows the original work of PSS and allows an autoregressive distributed lag (ARDL) model of order $(1, 0, \ldots, 0)$ as a special case of a data generating process of (2.12). A deterministic time trend (t) is included to account for the exogenous growth of technological progress. Inclusion of variables in the long-run equilibrium depends on the time series properties of the variable in question. Stationary variables will enter as exogenous variable while only integrated variables potentially enter the long-run relationship. Lagged differences of the endogenous variables are included to capture short-run adjustment dynamics (remaining terms of equation 2.12). The ϵ_{it} are cross-section specific error terms. The error correction coefficients ϕ_i are country-specific measures of the speed of adjustment to equilibrium. These parameters can be interpreted as the velocity at which fast an economy returns

[7]For simplicity of notation only, in (2.12) it is assumed that all differenced variables for a country i enter with the same lag order, namely p_i. Naturally, in estimations, lag orders are allowed to vary across variables and countries.

to its long-run growth once it has deviated from that path due to business-cycle shocks, for instance. Note that this steady-state convergence is not to be confused with the "catching-up convergence" (Bernard and Durlauf, 1996) employed in pure cross-section surveys and which is concerned with the transition of countries to new, possibly common steady states. Typically this kind of research looks at the correlation between initial per capita output levels and subsequent growth rates for a group of countries comprising advanced industrial countries *and* less developed economies.

In growth empirics it is common to analyze only the cross-section dimension. In addition, assumptions about common growth factors—at least in country-samples that share similar characteristics (advanced industrial countries with market economies for example)—seem to be appropriate. Given that the OECD countries considered in the present analysis have access to common technologies, active trade relations and dynamic capital flows, it seems plausible to assume that they tend to have similar long-run production parameters. In terms of short-run dynamics, this assumption seems rather implausible. PSS propose an estimator that pools the coefficients of the long-run relationship whilst allowing the short-run dynamics to be heterogeneous. They call this procedure Pooled Mean Group (PMG) estimation and the corresponding set of estimated parameters PMG estimates.

In terms of equation (2.12), applying the PMG estimation imposes the following homogeneity restrictions on the long-run parameters:

$$\theta_{1i} = \theta_1, \theta_{2i} = \theta_2, ..., \theta_{mi} = \theta_m \quad \forall i, i = 1, ..., N \tag{2.13}$$

While the homogeneity restrictions are imposed for the steady-state equation, short-run dynamics are estimated heterogeneously to allow for different dynamic adjustment patterns across countries.

The PMG estimates are obtained by maximizing the concentrated log-likelihood function corresponding to the ECM specification (2.12) under the restriction (2.13) and the assumption that the disturbances ϵ_{it} are independently distributed across i and t with zero means and variances σ_i^2. Furthermore, the model needs to be stable in the sense that the roots of the characteristic equation of the steady-state fall outside the unit circle. If these properties hold, PPS have shown that this approach yields consistent and asymptotically normal estimates of the short-run and long-run coefficients irrespective of whether the underlying regressors are $I(1)$ or $I(0)$.

The standard PMG estimation framework assumes that countries are totally independent. A more reasonable assumption is that countries are cross-correlated due to international linkages and common influences such as common technology shocks. Following Pesaran (2006), it is supposed that cross-correlation can be captured by a factor error structure. Under this assumption, the errors of equation (2.12) are given by

$$\epsilon_{it} = \gamma_i f_t + e_{it} \tag{2.14}$$

in which f_t is a unobserved common effect and e_{it} are independently distributed country-specific errors. Such a specification of an empirical model seems to be more in line with a production function featuring technology as a country-specific unobservable variable that may comprise common components across countries (Pesaran, 2004b).

Pesaran (2006) shows that augmenting the panel ECM with a set of cross-sectional averages of all variables featuring distinct weighting schemes can capture the correlated error component. In order to account for cross section dependence in estimating the parameters of the ECM, equation (2.12) may be augmented with the following variables: $\overline{y}_{t-1}$, $\sum_{k=1}^{p_i} \Delta\overline{y}_{t-k}$, $\overline{\mathbf{x}}_t$, $\sum_{k=0}^{p_i} \Delta\overline{\mathbf{x}}_{t-k}$. Overlined variables denote cross-country averages in the form of $\overline{z}_t = \sum_{i=1}^{N} z_{it}$ and $\mathbf{x}_{it}$ comprises all level series entering the right hand side of (2.12).

However, considering the large time series dimension of this approach, following Binder and Bröck (2006) we pursue a more parsimonious specification which results in conducting a two-step procedure. The authors denote it "two-step correlated effects augmentation (TS-CEA)". It can be applied as follows. The basic insight that lies behind the Common Correlated Effects estimators developed in Pesaran (2006) is that a proxy for the unobserved common factor can be obtained as

$$\hat{f}_t = \Delta\overline{y}_t - \widehat{\overline{\theta}}_0 - \widehat{\overline{\theta}}_4 t - \widehat{\overline{\phi}}\overline{y}_{t-1} - \widehat{\overline{\theta}}' \overline{\mathbf{x}}_{it} - \sum_{k=1}^{p_i} \widehat{\overline{b}}_0 \Delta\overline{y}_{t-k} - \sum_{k=0}^{p_i} \widehat{\overline{\mathbf{b}}}'_k \Delta\overline{\mathbf{x}}_{t-k} \tag{2.15}$$

in which hatted coefficients are from a first step estimation of

$$\Delta\overline{y}_t = \overline{\theta}_0 + \overline{\theta}_4 t + \overline{\phi}\overline{y}_{t-1} + \overline{\theta}' \overline{\mathbf{x}}_{it} + \sum_{k=1}^{p_i} \overline{b}_0 \Delta\overline{y}_{t-k} + \sum_{k=0}^{p_i} \overline{\mathbf{b}}'_k \Delta\overline{\mathbf{x}}_{t-k} + \epsilon_t \tag{2.16}$$

In a second step, we replace f_t from (2.14) with $\hat{f}_t$ from (2.15) and estimate the ECM as shown by equation (2.12) and (2.14) with the help of this factor estimate.

Homogeneity of long-run parameters should not be imposed but rather be tested. Whether long-run parameters are homogeneous as imposed by the PMG estimator can be tested with the help of a Hausman (1978) test. Consistent estimates of the long-run coefficients can be obtained from the Mean Group (MG) estimator, which simply computes the average of the individual OLS estimates of the the long-run coefficients from the ECM. These, however, will be inefficient if long-run coefficients are indeed homogeneous. Under this condition, the PMG estimates are consistent and efficient. Therefore, the effect of heterogeneity on the means of the parameters can be determined by a Hausman-type test between the MG and PMG estimates. The test statistic is given by

$$H = \hat{o}' \left[var(\hat{o}) \right]^{-1} \hat{o} \tag{2.17}$$

where $\hat{o}$ is a (m x 1) vector of the differences between the MG and PMG coefficient estimates and $var(\hat{o})$ is the corresponding covariance matrix. Under the homogeneity hypothesis of the long-run equilibrium, the Hausman statistic is asymptotically distributed as χ^2 variate with m degrees of freedom, the number of estimated long-run parameters. Since $var(\hat{o})$ need not to be positive definite, in some cases the test may not be applicable. In addition, tests of the equality of the long-run slope coefficients can be carried out using likelihood ratio or other classical statistical procedures.

There have been other single-equation procedure to estimate the cointegration relations proposed, notably the "fully-modified OLS" approach of Pedroni (2000) and Phillips and Moon (1999) and the "dynamic OLS" procedures due to Saikkonen (1991). The latter is based on a static OLS regression such as (2.11) that is augmented with leads and lags of the first differenced regressors. However, both of these methods rely on kernel based nuisance parameter estimates to adjust the relevant test statistics which may perform poorly in small samples (Breitung and Pesaran, 2005). Therefore, we do not pursue this branch of panel cointegration estimation techniques given the data set at hand.

2.4.3 Deriving trend output estimates: an EMVF approach

After having estimated a suitable dynamic panel data model, the established long-run relationship can be used to derive country-specific estimates of the trend value of GDP per capita and its according growth rate. One can formalize this idea as follows. Collect the cross section specific variables of the long-run relationship in the $T \times k$ matrix $X_i = (x_{i1}, \ldots, x_{iT})'$. Furthermore, denote the $T \times 2$ matrix of deterministic regressors $D_i = ([1, \ldots, 1]', [1, 2, \ldots, T]')$ and let $\widehat{\theta}$ be the $k \times 1$ vector of PMG estimates of the long-run coefficients and $\widehat{\varphi}_i$ the 2×1 vector of country specific PMG estimates of the deterministic regressors.

It follows that the cross section specific vector of fitted values for the level of GDP per capita $\widehat{y}_i = (\widehat{y}_{i1}, \ldots, \widehat{y}_{iT})$, is given by

$$\widehat{y}_i = X_i \widehat{\theta} + D \widehat{\varphi}_i \tag{2.18}$$

In principle, equation (2.18) could be directly used to compute the steady-state output per capita and the steady-state growth rates for a country. Unfortunately, such a proceeding would yield unsteady measures of trend output since the explanatory variables of the X_i matrix are unlikely to be measured at their steady-state values. These variables may depict undue volatility which in turn translates into volatile trend output measures. A promising approach of using econometrically derived equilibrium paths in combination with time series smoothing techniques in order for the former to be suitable as policy tools can be found in the applied literature of multivariate time filters. In essence, this research deals with the linkage of economic theory such as a Phillips curve relation with time series filter techniques and is particularly popular among practitioners.

The primary aim of the multivariate filters is to improve the efficiency of the univariate time series filter of Hodrick and Prescott (1997, HP) and to increase the economic content of the resulting estimates by adding long-run relationships grounded in an econometric approach to the minimization problem.[8] In particular among researchers at the *Bank of Canada* multivariate filters are actual working tools. Typically, economic theory is embedded in the form of residuals from a Phillips curve equation or in the form of equilibrium paths generated by structural vector autoregressions (SVAR). We do not go into detail here but refer to the comprehensive studies of Laxton and Tetlow (1992), Butler (1996), Rennison (2003) and Gosselin and Lalonde (2006). The latter contribution is the one which we will follow most closely below. Instead of employing a structural output estimate derived from a SVAR as Gosselin and Lalonde (2006) do we will use the estimated long-run relation from the panel ECM as given by equation (2.18).

The Extended Multivariate HP-filter (EMVF) of Gosselin and Lalonde (2006) is given by the following minimization problem

$$\min_{\tau_i} F(\tau_i) = (y_i - \tau_i)' W_y (y_i - \tau_i) + (\widehat{y}_i - \tau_i)' W_{\widehat{y}_i} (\widehat{y}_i - \tau_i) + \lambda \tau_i' S' S \tau_i \tag{2.19}$$

The vector $y_i = (y_{1i}, \dots, y_{iT})'$ comprises the observed values of the logarithm of GDP per capita for economy i and $\tau_i = (\tau_{i1}, \dots, \tau_{iT})'$ contains the smoothed or trend values that minimize (2.19). The parameter λ is the usual smoothing parameter of the HP-Filter, which is typically equal to 100 for annual data, and the matrix S takes the second differences of τ_{it}. The two diagonal matrices W_y and $W_{\widehat{y}_i}$ assign weights to the output gap and to the deviation of the steady-state value of the ECM from the trend series, respectively. In the present application, these weight matrices are set equal to the identity matrix, but in general they allow a fairly flexible penalizing of the restriction terms in equation (2.19). For instance,

[8] In particular, the multivariate filters help to reduce the familiar end-sample problems of the HP-filter which essentially consists of a centered moving average in mid-sample that becomes one-sided closer to the start and the end of the sample.

by setting individual diagonal elements equal to zero, period specific output gaps or steady-state deviations can be excluded from the computation of the filter. Such an exclusion may improve the trend estimate if the researcher has additional information which, say, recommends to down-weight, up-weight or exclude individual steady-state observations or if outliers ought to be handled.

Based on experimental grounds, Rennison (2003) provides evidence supporting the efficiency and reliability of the EMFV in the case that structural output estimates from a SVAR are used as extensions to the filter. Naturally, Rennison's (2003) Monte Carlo experiments do not allow us to draw conclusions with regard to the present study. However, provided that the estimations of the panel ECM result in statistically and economically sensible steady-state specifications, we belief that similar properties of the present version of the EMVF may hold.

2.4.4 Search strategy

Specification search in growth econometrics is a notorious problem. Economic growth theory proposes a huge number of variables that potentially explain economic growth. The survey of the literature at the beginning of this chapter gives a rudimentary overview of the many suggestions. The mere dimension of possible regressors implies that one simply faces a small-sample problem in cross-country growth estimations. The available cross-country sample data typically does not leave sufficient degrees of freedom to estimate a general model and then drop regressors whose coefficients would converge to zero in (theoretical) large samples (Sala-i-Martin et al., 2004). To put it differently, a general-to-specific approach for model selection is virtually not feasible. Therefore, investigators commonly consider only a small number of explanatory variables in their attempt to establish a statistically significant linkage between economic growth and a particular variable of concern. However, a typical problem encountered in cross-country growth regressions is that the significance, sign and magnitude of estimated coefficients of a certain variable can change considerably when altering the set of explanatory variables in the regression specification.

In the context of cross-country growth regressions a sensitivity analysis designed to challenge this problem was first conducted by Levine and Renelt (1992), who surveyed the robustness of the many variables that are considered to be correlated with economic growth based on Leamer's (1983, 1985) "extreme bounds analysis". Ever since, this paper has stimulated a growing literature that is concerned with model uncertainty in growth econometrics. Examples include Sala-i-Martin (1997), Sala-i-Martin et al. (2004), Bleaney and Nishiyama (2002), Florax et al. (2002) and Fernandez et al. (2001). A usual strategy in this strand of research is to estimate as many models as (computer power renders) possible and then to compute summary statistics of interest such as extreme bounds, the fraction of regressions in which a particular variable is significant or posteriori distributions of regression coefficients in studies following the Bayesian approach. All of these methods have been applied to pure cross-section data sets and—at least not to our knowledge—not to a cross-sectional time series framework yet.

Given the complexities involved in specifying, estimating and testing a cointegrated panel ECM, a mechanistic econometric evaluation and summing-up of few parameter estimates does not seem to be a sensible approach as opposed to the robustness analysis in the pure cross-section framework. In some cases the various statistical tests considered are inconclusive and the results need to be carefully interpreted with respect to the theoretical predictions of the long-run model. In addition, the purpose of the present analysis is to test the growth relevance of the variables but also to search for a model with economically meaningful long-run properties in order that it can be employed for an theory-guided estimation of the trend output and trend growth of an economy.

In the light of these considerations, the search strategy for variables that are correlated with GPD per capita in the long-run is as follows. As in Levine and Renelt (1992) and Sala-i-Martin (1997), we restrict the set of right hand side variables of the long-run GDP per capita equation to always contain certain key variables and then add one variable from the wider sources of growth after another. In particular, we always keep the Solow variables in the ECM and add one variable from the wider growth candidates, estimate this model and in a next step estimate a second model which includes another variable from the pool of wider growth sources and so on. In doing so we end up with a total of 16 empirical models. In contrast to the robustness checks from pure cross section studies, this approach looks only at a very small number of models. Limiting the number of estimated models has the advantage that one can conduct detailed analysis for each individual estimation and base judgement on whether an empirical specification supports a reasonable long-run relationship on the outcomes of several tests jointly. Obviously, generalization of the proceeding would be straightforward by, say, drawing two or more variables from the pool simultaneously that are added to the fixed Solow variables in every step. However, this comes at the cost of an increasing number of models to evaluate since the number of possible combinations of variables increases rapidly and may be difficult to deal with in a cross-sectional time series framework.

The main guidelines to assess the validity of the specifications in the present analysis can be summarized by the following issues: Are the variables homogenously and/or heterogeneously cointegrated? Do negative and significant error correction coefficients confirm a long-run relationship? Do Hausman-tests support homogeneous long-run slope parameters? Are PMG and MG estimates significant? Do the coefficients of the Solow variables not vary with model specifications? An estimated panel model which is in line with the theoretical steady-state predictions outlined in section 2.3 will give a positive answer to each of these questions.

Subsequent to the specification search, an extended model comprising the variables for which the most supportive evidence of explaining GDP per capita is estimated and employed for the application of the EMVF approach.

2.5 Data and results

The used panel data set is compiled from various sources, notably from international institutions such as the OECD, the World Bank and the IMF. The cross section coverage varies with the time series under consideration. For most variables the study comprises 23 countries over 30 years, however, in particular for the time series of the subject areas "fiscal policy" and "labor markets" the cross section is reduced to 12 countries due to missing values. In addition, for few variables, time series observations are not available for the entire time span from 1971 to 2000. Table 2.10 in the appendix provides an overview of the data sources and coverage. For the unit root tests we use the maximum number of countries for each respective variable whereas in the case of the ECM's, estimations build on the 12 countries that have observations for *all* variables under consideration. A lack of time series observations for Germany for the periods before 1991 due to the territorial separation within Germany requires a linking of West-German data for the period from 1971 to 1991 and all-German data for the period thereafter. Due to this linking several time series for Germany exhibit outliers in the year 1991. Following Stock and Watson (2003) such observations have been replaced with the median of the three observations on either side of the data points in question.

We are aware that the present sample size is rather short in order to draw strong conclusions from the subsequent test outcomes since the econometric test procedures outlined above generally build on asymptotic properties. While the lengths of the time series may be sufficient to study features of economic growth, their low frequency of observation leaves only few degrees of freedom for estimation in a rather ambitious panel framework. However, it is a general problem that many economically interesting candidate variables for explaining long-run output and growth are measured only at an annual frequency for which, however, long and consistent cross country data is typically rare. Keeping these caveats in mind, the following exercise is not intended to give definite answers but rather to demonstrate a general strategy to analyze economic growth within a cross-sectional time series data-set.

In the interest of maintaining a parsimonious specification, all estimations are limited to include only a small number of lagged variables. The univariate tests are less affected by this limitation, however, typically for the estimations of the ECMs and the TS-CEAs a maximum lag order of one or two is considered.

2.5.1 Results of the panel unit root tests

Before the outcomes of the panel unit root tests are discussed, we want to present some preliminary considerations. A question that sometimes arises among researchers is whether it is sensible to model theoretically bounded variables such as the investment rate in human capital or the long-term unemployment rate as $I(1)$ variables. Here we follow the stance of the empirical literature that is concerned with a data-coherent modelling of economic relations and for this reason we prefer to base the decision on whether to treat a variable as $I(1)$ or $I(0)$ in the estimations of economic relations on the outcomes of econometric tests. Naturally, unit root tests based on longer available samples may produce different results than the ones presented below and may reflect the theoretical limitation of many considered variables.

Table 2.1 provides an overview of the DHT test outcomes for the variables of the data set. Test results for both the level and the first difference of the respective variable are reported. For the individual ADF tests which underly the DHT panel unit root procedure, the decision whether to include deterministic regressors such as an intercept or trend is crucial, since the asymptotic distribution of the test statistics is influenced by that choice. Given the high number of individual ADF regressions involved in this exercise a rather pragmatic proceeding was chosen in dealing with the inclusion of deterministic components. Since none of the variables variate around a zero mean, the inclusion of an intercept is always appropriate. Whenever it seemed adequate to consider a trend stationary model as an alternative, a time trend was also included. The ADF regressions for the first differenced series exclude a deterministic time trend throughout but always include an intercept. With regard to the number of included lagged differences we let the modified Schwarz Bayesian information criteria (SBC) chose.[9] The maximum lag length was set using the criterion proposed by Schwert (1989). By way of robustness check, we also took a look at the panel unit root results based on model selection with the help of the modified Akaike criterion. In general, however, we did not observe conflicting results concerning the decision whether to regard a variable as either $I(1)$ or $I(0)$.

The panel unit root statistics summarized in table 2.1 confirm the familiar result that most economic times series contain a unit root and become stationary when transformed into the first difference form. Only for the *tax quota*, the *standard deviation of inflation* and the *age dependency ratio* did we find evidence that the level

[9]See Ng and Perron (2001) for a discussion of modified information criteria.

Table 2.1: Summary of panel unit root tests

Variable	CS #	Level	First diff.
		$t(\widehat{\rho}^*, \kappa)$	$t(\widehat{\rho}^*, \kappa)$
GDP per capita, log	23	-0.86 (0.20)	**-5.53** (0.00)
Investment ratio, log	23	-0.25 (0.40)	**-6.04** (0.00)
Human capital investment, log	23	-0.86 (0.19)	**-5.82** (0.00)
Population growth	23	-0.90 (0.18)	**-13.21** (0.00)
Government consumption (% of GDP), log	23	1.19 (0.88)	**-5.20** (0.00)
Tax quota, log	12	**-1.99** (0.00)	-
Tax ratio, log	12	-0.45 (0.33)	**-8.42** (0.00)
Net lending government (% of GDP)	12	0.68 (0.75)	**-6.10** (0.00)
Inflation (CPI)	23	-0.35 (0.36)	**-10.17** (0.00)
Standard deviation of inflation	23	**-4.17** (0.00)	-
R&D expenditure (% of GDP), log	20	0.14 (0.56)	**-6.23** (0.00)
Openness (imports + exports/GDP), log	23	5.30 (0.99)	**-2.44** (0.01)
Terms of trade, log	12	-1.03 (0.15)	**-4.19** (0.00)
Stock market capitalization (% of GDP), log	16	4.96 (0.99)	**-5.88** (0.00)
Turnover ratio, log	23	4.33 (0.99)	**-8.98** (0.00)
Credits to private sector (% of GDP), log	21	1.33 (0.91)	**-5.89** (0.00)
NAWRU, log	12	4.05 (0.99)	**-2.55** (0.01)
Age dependency ratio, log	23	**-3.29** (0.00)	-
Population over 65 (% total population), log	23	4.27 (0.99)	-0.99 (0.16)

Notes: CS # denotes the number of included cross sections. Panel unit root test were conducted with the help of the modified inverse normal method to account for cross-section dependence as proposed by Demetrescu et al. (2006). The parameter κ was set to 0.2. See text for a description of this approach. The underlying p-values were derived from Augmented Dickey-Fuller (ADF) tests which include a constant term throughout. Whenever it seemed appropriate to consider a trend stationary model as an alternative, a time trend was also included. The number of lagged difference terms of the ADF test was chosen with the aid of the modified Schwarz Bayesian information criterion and the maximum lag length was set according to the rule $k_{max} = \text{int}(12(T/100)^{1/4})$ of Schwert (1989). One-sided p-values according to the Standard Normal distribution in brackets. Computation of the one-sided p-values for the ADF t-statistics draws on MacKinnon (1996). Computational work was performed in *Eviews* and MATLAB.

of these series is already stationary. The outcomes of the demographic variables require some deeper discussion. The *age dependency ratio*, which is the ratio of the combined child population (0-14 years) and the aged population (65 years and over) to the economically active population (age 15-65 years), is found to be stationary. However, the *ratio of the population over 65 to the total population*, a variable which seems quite similarly constructed to the age dependency ratio at first glance, shows rather different time series properties. The DHT panel unit root test provides relatively strong support for the view that the level of these series are non-stationary. In addition, a test on the first difference of these series also indicates non-stationarity when the order of the lag length is selected with the help of the modified Schwarz Bayesian criterion. In the case that the modified Akaike information criterion is allowed to select the lag order of first differences, the corresponding $t(\widehat{\rho}^*, \kappa)$ statistic amounts to -0.75 with a p-value of 23% (not reported in table 2.1). When the lag order is selected manually, the hypothesis that there is a unit root in the first difference of the population over 65 variable is not rejected for lag orders of one and two, but rejected at the 10% significance level for orders of three and five. If five lags of first differences are included in the ADF regressions, the DHT tests are insignificant once again. Overall, the available data does not lead to a clear conclusion as to whether population over 65 variables are integrated once or twice. To avoid the problem of mixing $I(1)$ and $I(2)$ variables we use these variables in the first differenced form throughout the subsequent estimations.[10]

The establishment of the unit root properties is an important preliminary step for the following estimation of the long-run relationship between GDP per capita and its potential determinants. Whenever a variable is found to be $I(0)$ it can not form a cointegration relation with an $I(1)$ variable by definition. For the estimations of the ECMs below, an $I(0)$ diagnosis implies that the corresponding series will enter the error correction equation in levels as an exogenous explanatory variable and not the long-run relation of the error correction component. In contrast, the $I(1)$ variables will enter the long-run relationship of the ECM estimations and for these variables panel cointegration procedures are performed which test for homogeneous and heterogeneous cointegration.

2.5.2 Results of the panel ECMs

After having determined the order of integration for the variables, models that always include real GDP per capita, the investment share, human capital and population growth and which are sequentially augmented with variables from the pool of institutional and political indicators are estimated along the lines of PSS. In addition, for each of the 16 ECMs, panel cointegration tests are conducted in order to check whether the PMG estimates formulate a tenable long-run relation. In a next step, an extended model comprising variables for which the evidence from the sequential estimation procedure gives most conclusive evidence of being related to long-run output per capita is estimated. This helps to get a more comprehensive picture of the determinants of output. Moreover, the derived long-run relationship provides a basis to compute the country-specific estimates of the trend value of GDP per capita and its according growth rate with the aid of the EMVF.

[10] Garratt et al. (2006) discuss the ambiguity of unit root test results and the problem of how to best deal with variables that are on the borderline of being $I(0)/I(1)$ or $I(1)/I(2)$. The authors refer to Haldrup's (1998) review of the econometric analysis of $I(2)$ variables which points to the dangers of inappropriate application of econometric methods designed for use with I(1) variables. He suggests that it is often useful to transform time series a priori to obtain variables that are unambiguously $I(1)$ rather than dealing with mixtures of $I(1)$ and $I(2)$ variables directly.

The strategy to add one variable at a time may lead to an omitted variable problem if the included variable is one of several correlated and significant variables. In that case this variable will capture some part of the correlation of the omitted variable with the left hand side variable, i.e. one has misspecification bias. When dealing with an omitted variables problem, it is a basic insight of regression analysis to be cautious in interpreting the coefficient magnitudes of the separately augmented models, as well as in selecting from these variables to form an extended model. The correlation matrix of the regressors shown in table 2.11 in the appendix helps in assessing how severe such a problem may be for the subsequent estimations. Fairly substantial correlation coefficients can be observed for few pairs of regressors. In particular, *government consumption* and the *tax quota* have a correlation coefficient of 0.84. Furthermore, *government consumption* is correlated with *trade openness*. The corresponding coefficient is 0.64. The *tax quota* is also highly correlated with *trade openness*. For these two regressors the computed correlation coefficient amounts to 0.81. Not surprisingly, the variables that measure the development of the *financial markets* are correlated among each other. Lastly, higher correlations can be observed between the *expenditure on research and development* and the *financial market* indicator *stock market capitalization*. In general, table 2.11 provides evidence for only moderate forms of correlation among the complete set of regressors. However, for estimations including correlated regressors the correlation with the omitted variables needs to be kept in mind in interpreting the subsequent results.

The PMG estimation and panel cointegration results for the ECMs are reported in tables 2.2 and 2.3. Before the estimation outcomes are discussed in more detail, some general remarks about the structure and organization of these tables are necessary. The first column of tables 2.2 and 2.3 shows the variable identifiers of the estimated model. The Solow variables $\ln s^K$, $\ln s^H$ and n are kept constant and appear in all ECMs. The second column contains the PMG coefficient estimates together with the according standard errors and an indication of significance based on two-sided critical values from the standard normal distribution.

The third column shows Hausman test statistics which compare the PMG and MG coefficient estimates individually. In the next column, the Hausman statistic that tests the coefficients jointly is shown. If the Hausman statistics are significant according to the critical values from the $\chi^2(m)$ distribution, the homogeneity hypothesis of long-run coefficients is rejected and country-specific long-run slope parameters may provide a more appropriate model specification.

Table 2.2: PMG estimates and cointegration test results for various models (I)

Explanatory variable	PMG estimates	Hausman test	Hausman test (joint)	LR test	Loading parameter	$t(\widehat{\rho}^*, \kappa)$ CI test	Group-ADF
Fiscal policy							
$\ln s^K$	0.28***(0.05)	0.51					
$\ln s^H$	0.11***(0.02)	1.42					
n	-0.06***(0.01)	1.30					
$\ln cg$	-0.49***(0.07)	0.57					
			3.58	240.67***	-0.17***(0.03)	-3.03***	-0.21
$\ln s^K$	0.36***(0.02)	2.50					
$\ln s^H$	0.12***(0.02)	0.11					
n	-0.04***(0.01)	3.13*					
$\ln taxq$	0.69***(0.14)	-					
			3.95	140.25***	-0.28***(0.06)	-4.46***	-2.30**
$\ln s^K$	0.48***(0.04)	3.72*					
$\ln s^H$	0.17***(0.02)	0.67					
n	-0.10***(0.01)	1.38					
$\ln tr$	0.11***(0.03)	2.14					
			10.83**	235.90***	-0.18***(0.06)	-5.12***	-1.91**
$\ln s^K$	0.23***(0.05)	2.61					
$\ln s^H$	0.04 (0.03)	0.26					
n	-0.08***(0.02)	0.11					
$nlgq$	0.00** (0.00)	0.23					
			4.45	121.70***	-0.22***(0.04)	-3.23***	-2.30**
Monetary policy							
$\ln s^K$	0.40***(0.05)	0.75					
$\ln s^H$	0.11** (0.05)	0.80					
n	-0.00 (0.01)	1.59					
i	-1.84***(0.25)	0.23					
			29.68***	238.28***	-0.20***(0.04)	-8.10***	-1.05
$\ln s^K$	0.34***(0.03)	0.48					
$\ln s^H$	0.19***(0.03)	0.18					
n	-0.08***(0.01)	0.60					
isd	0.67***(0.14)	-					
			0.61	159.21***	-0.21***(0.05)	-4.46***	-2.30**

Notes: $^{*}/^{**}/^{***}$ denotes significance to the 1%/5%/10% level. Figures in brackets are the standard errors. The cross-section of all estimations covers 12 countries except the specifications including the *financial market* variables. For these specifications the cross section number is reduced to 11 due to missing observations for Norway. See table 2.10 for detailed data coverage and variable identifiers.
For the **PMG estimations**, an intercept, a time trend and the common factor estimate given by equation (2.15) are included in all specifications. The error correction coefficient is the MG estimate, i.e. it is computed as the arithmetic mean of the individual error correction coefficient estimates $\widehat{\phi}_i$. The corresponding standard error is obtained with the help of the non-parametric variance estimator $V(\widehat{\overline{\phi}}) = [N(N-1)]^{-1}\sum_{i=1}^{N}(\widehat{\phi}_i - \widehat{\overline{\phi}})^2$. For stationary variables which enter the ECM in levels but not the cointegrating relationship, only the MG estimates are computed and therefore Hausman tests are not applicable. For these specifications, the joint Hausman test refers only to the basic (Solow) variables, which are $I(1)$. Selection of the lag orders of short-run dynamics of each country is based on the Akaike information criteria with a maximum lag order of two. The concentrated likelihood function has been maximized with the aid of the Newton-Raphson method. Computational work was performed in MATLAB.
The $\boldsymbol{t(\widehat{\rho}^*, \kappa)}$ statistic of the DHT test has been specified with $\kappa = 0.2$. See table 2.1 for further notes on specifications of these tests.
The **Group-ADF** statistic of the Pedroni (1999) cointegration test is based on a static regression according to equation (2.11) including heterogeneous intercepts and heterogeneous time trends. Common time effects have been subtracted out prior to estimations. The number of lagged difference terms of the underlying ADF regressions was selected using a step down procedure, starting from 4 lagged differences. Computational work was performed in WinRats.

Table 2.3: PMG estimates and cointegration test results for various models (II)

Explanatory variable	PMG estimates	Hausman test	Hausman test (joint)	LR test	Loading parameter	$t(\widehat{\rho}^*, \kappa)$ CI test	Group-ADF
Research and development							
$\ln s^K$	0.23^{***} (0.02)	0.48					
$\ln s^H$	-0.06^{***} (0.01)	0.75					
n	-0.40 (0.56)	1.35					
$\ln rd$	0.13^{***} (0.01)	4.29^{**}					
			18.83^{***}	241.92^{***}	-0.45^{***} (0.10)	-6.30^{***}	-1.66^{**}
International trade							
$\ln s^K$	0.18^{***} (0.03)	1.55					
$\ln s^H$	-0.00 (0.02)	0.03					
n	-0.05^{***} (0.01)	1.12					
$\ln open$	0.29^{***} (0.05)	1.04					
			3.61	220.90^{***}	-0.29^{***} (0.06)	-3.90^{***}	1.52^{*}
$\ln s^K$	0.31^{***} (0.02)	0.28					
$\ln s^H$	0.02 (0.02)	0.15					
n	-0.04^{***} (0.01)	1.05					
$\ln tot$	-0.18^{***} (0.03)	0.14					
			2.61	257.78^{***}	-0.33^{***} (0.07)	-3.54^{***}	-1.86^{**}
Financial markets							
$\ln s^K$	0.23^{***} (0.06)	0.16					
$\ln s^H$	0.17^{***} (0.03)	0.94					
n	-0.09^{***} (0.02)	6.44^{**}					
$\ln cap$	0.12^{***} (0.03)	0.28					
			24.21^{***}	199.05^{***}	-0.15^{***} (0.04)	-3.53^{***}	-1.33^{*}
$\ln s^K$	0.75^{***} (0.12)	5.95^{**}					
$\ln s^H$	-0.59^{***} (0.18)	6.97^{**}					
n	-0.08^{***} (0.03)	0.95					
$\ln turn$	-0.00 (0.00)	0.32					
			0.47	129.88^{***}	-0.09^{***} (0.03)	-3.62^{***}	-0.55
$\ln s^K$	0.34^{***} (0.03)	0.17					
$\ln s^H$	0.20^{***} (0.03)	0.87					
n	-0.09^{***} (0.02)	0.11					
$\ln credit$	0.01 (0.01)	0.06					
			2.00	378.28^{***}	-0.20^{***} (0.06)	-3.59^{***}	-0.23
Labour markets							
$\ln s^K$	0.29^{***} (0.02)	1.84					
$\ln s^H$	0.04^{***} (0.01)	0.28					
n	-0.06^{***} (0.01)	6.20^{**}					
$\ln nawru$	-0.06^{***} (0.01)	0.00					
			8.02^{*}	236.28^{***}	-0.34^{***} (0.09)	-3.80^{***}	-1.88^{**}
Demography							
$\ln s^K$	0.31^{***} (0.05)	1.57					
$\ln s^H$	-0.50^{***} (0.09)	4.28^{**}					
n	-0.03^{***} (0.01)	0.45					
$\ln adr$	0.85^{***} (0.14)	-					
			15.78^{***}	191.69^{***}	-0.16^{***} (0.05)	-4.46^{***}	-2.30^{**}
$\ln s^K$	0.32^{***} (0.04)	1.54					
$\ln s^H$	0.10^{***} (0.03)	1.24					
n	-0.07^{***} (0.02)	0.26					
$\Delta \ln pop65$	0.01^{***} (0.01)	0.90					
			1.73	133.71^{***}	-0.19^{***} (0.04)	-3.81^{***}	-2.82^{***}

Notes: See table 2.2

For each model, likelihood ratio (LR) test statistics are reported in column 5 of tables 2.2 and 2.3 which are based on comparing the log-likelihood of the unrestricted model with the log-likelihood of the model that restricts the long-run slope parameters to being the same across each group. Rejection implies that the panel model which comprises country-specific coefficient estimates of the long-run parameters is more likely to be supported by the data than the panel model that imposes homogeneous coefficients.

Furthermore, the result tables show the MG estimates of the error correction coefficient along with an indication of significance. A negative and significant parameter value is expected if the long-run slope estimates establish a cointegration relationship.

The last two columns of tables 2.2 and 2.3 report panel cointegration tests results. The $t(\widehat{\rho}^*, \kappa)$ statistic is derived by applying the DHT panel unit root test to the residuals of the static cointegration regression which imposes slope homogeneity on the explanatory variables but leaves the coefficients of the deterministic part unrestricted. The Group-ADF statistic of Pedroni (1999, 2004), which is based on residuals from a cointegration regression in which both the deterministic terms *and* slope coefficients are permitted to be heterogeneous across countries, is presented in the last column of the tables. If both the DHT and Pedroni (1999, 2004) tests reject the null hypothesis of no cointegration, we regard this as a strong indication that the according variables are actually forming an equilibrium relationship.

The ECM estimations include Austria, Belgium, Germany, Denmark, France, the United Kingdom, Italy, Japan, the Netherlands, Norway, Sweden and the USA. These countries have observations for all variables of the data set and constitute the largest possible intersection. Most estimations cover the period from 1971 to 2000, only the specifications including the *financial markets* and *research and development* variables cover shorter periods due to data availability. Further details on estimation techniques and model selection guidelines are given in the note to table 2.2. Before going into detail of the individual results reported in table 2.2 and 2.3, some basic findings are worth mentioning.

First, the PMG estimates are significant in most model specifications suggesting that many variables are indeed relevant for economic growth, at least at first glance. However, significance of coefficient estimates does not signify the existence of long-run relationships and PMG estimations need to be accompanied by cointegration tests.

Secondly, another conspicuous outcome of the econometric analysis is that the DHT test on the residuals from the static regression with homogeneity restrictions on the slope coefficients is significant for every model specification, while the Group-ADF test more frequently fails to reject the null hypothesis of no cointegration. We have discussed the small sample performance of Pedroni's (1999, 2004) cointegration test above, however, we have been silent on the performance of the DHT test in finite samples so far. In the context of panel unit root testing, with the aid of Monte Carlo experiments DHT show that the modified inverse normal method delivers good results for medium and strong cross-correlation and various sizes of T and N at the 5% level. For obvious reasons, the question whether the insights of these experimental outcomes can be conferred to cointegration tests with multiple regressors remains open. Therefore, we can not definitely rule out that the DHT test outcomes in the present analysis reflects a small sample bias in part. For these reasons, a careful and joint interpretation of both the Group-ADF and DHT cointegration test results seems advisable.

Thirdly, LR tests never fail to reject the null hypothesis of parameter homogeneity. Interestingly enough, PSS also encounter a general rejection of the LR test in both of their empirical applications. The authors discuss the interpretation of this feature to some extent and point to various sources of this problem, however, there seems to be no general answer to it unless one is willing to adhere to a single country estimation. The latter, howerver, is not in line with our the intents mentioned at the outset of the study. In contrast to the LR test, the Hausman test results provide a more differentiated picture for the question as to whether long-run slope parameters should be pooled or not. We will comment on the individual results of the Hausman test in due time.

Fourthly, the estimates of the error correction coefficients show that the long-run relation makes an important contribution to the equations explaining growth of GDP per capita. The magnitude of this coefficient estimate varies somewhat across model specifications but always appears with a significant and negative sign. According to these estimates, the average speed of equilibrium adjustment is fairly rapid. The error correction coefficient estimates are in the range of -0.45 to -0.09 with a median of the estimates of -0.20.

Now we turn to the interpretation of the coefficient estimates of the individual explanatory variables, which can be interpreted as output elasticities due to the logarithmic transformation. First, the outcomes of the Solow variables which are kept in all of the estimated model specifications are discussed, followed by an interpretation of results of the sequential augmentation.

The coefficient estimate of physical *capital investment* emerges as relatively robust in the sense that the sign and magnitude of the estimated coefficient of the investment variable does not change considerably when altering the set of explanatory variables. Usually growth regressions are plagued by this phenomenon as pointed out in section 2.4.4. However, we are able to confirm the findings of Levine and Renelt (1992), Sala-i-Martin (1997), Florax et al.(2002) and Fernandez et al. (2001). According to these studies, the investment in physical capital belongs to one of the few robust determinants of growth.

In contrast, the *human capital* variable appears with much more volatile coefficient estimates across PMG models. Furthermore, the PMG estimates are insignificant in several estimations. Similar to Mankiw et al. (1992) we proxy for the rate of human capital accumulation that measures the percentage of the total population attending secondary and tertiary school. Such enrollment ratios have also been used in the work of Barro (1991) and Levine and Renelt (1992) among many others and may be regarded as a "classical" indicator of the investment in human capital. However, the use of enrollment ratios as proxies to the flow of human capital investment has often been questioned (e.g. Wößmann, 2003) and we also regard the fragile outcome of the present analysis as more of a problem of the empirical implementation of a theoretical concept that is difficult to operationalize.

The coefficient estimate of the *population growth* variable hardly varies across models and has the expected negative sign which growth theory predicts. The coefficient estimates that refers to estimations in which the population growth variable is actually significant lie in the range of [-0.10, -0.03]. The median of the estimates amounts to -0.07.

The estimation and test outcomes of the sequentially augmented models are summarized in the following paragraphs. In the subject area of *fiscal policy*, *government consumption*, the overall *tax quota* and the *tax ratio* coefficients are significant in the estimations. *Government consumption* appears with a negative sign, while the PMG coefficient estimates of the other indicators are positive. As mentioned before, the DHT test rejects the hypothesis of no cointegration throughout. The Group-ADF test is unable to reject "no cointegration" for the estimation that includes *government consumption* but rejects for the remaining specifications. The PMG estimate of the *public deficit* is significantly positive but very limited in magnitude. The joint Hausman test is significant in the *tax ratio* specification. The observed rejection of overall poolability of the long-run slope coefficients might be due to the *human capital* variable for which the individual Hausman test fails to diagnose slope homogeneity in this estimation. Overall, the results suggest a significant impact of fiscal policy settings on output per capita across countries and over time. At the same time, the signs of the estimated correlations also show that fiscal activity may not be characterized by simple relations such as "government activity is bad for growth". Rather, a differentiated view on which kind of activity is pursued is recommended.[11]

In the subject field of *monetary policy*, *inflation* and the *variation of inflation*,

[11]See also Agell et al. (1997) for a survey of studies analyzing the public sector and fiscal activity. The authors claim that empirical studies do not allow a statement on whether the relation between the extent of the public sector and the economic growth is "positive, negative, or nonexistent".

both measured with the aid of consumer price indices, have negative and significant PMG estimates. However, in the *inflation* specification the joint Hausman test results in a rejection of the homogeneous slope hypothesis and the Group-ADF test does not imply a cointegration relationship. The stationary measure of the *variation of inflation* enters the panel ECM as a country specific variable, which by construction cannot form a homogeneous cointegration relation with the other $I(1)$ variables. In table 2.2, the MG estimate along with a non-parametric estimate of the standard error of these regressors is reported. This estimate is significant but positive which partly contradicts a priori expectations.

The outcomes of the empirical growth model that include the *expenditures on research and development*, in addition to the Solow variables are shown at the top of table 2.3. The Group-ADF test statistic amounts to -1.66 which is significant to the 5% level. The PMG estimate is positive and significant but equality across countries is rejected according to the Hausman test. Consequently, the PMG model specification should be discarded in favor of an estimation of country-specific equations. Note that the number of observations for this estimated ECM is limited since the data coverage stems from the period of 1981 to 2000 only and may also explain poor empirical results.

The variables that characterize *international trade* activity provide very satisfactory results both from econometrical and economical points of view. *Trade openness* is significant and positive in the PMG estimation and cointegration test results confirm the existence of a long-run link between trade openness and GDP per capita. Furthermore, the Hausman test does not reject the hypothesis of parameter homogeneity neither on the individual nor on the joint level. Very similar results are reported for the *terms of trade* estimation except that the sign of the PMG estimate is negative in this specification. A negative coefficient estimate is not surprising since the terms of trade variable might be regarded as an indicator of international trade competitiveness. A negative coefficient supports the notion that international price competitiveness is important for the exploitation of trade benefits and may raise the level of GDP per capita. However, inclusion of the trade variables renders the coefficient estimate of the human capital variable insignificant in both estimations, which may be due to the small positive correlation between the trade and the human capital measures.

Concerning the variables of the subject area of *financial markets*, only the coefficient of the *stock market capitalization* variable has a reasonably low standard error which indicates significance. "No Cointegration" is rejected at the 10% level of significance by the Group-ADF test. However, the estimations that include the *turnover ratio* and *credits to the private sector* result in insignificant PMG estimates and give no indication of cointegration according to the Group-ADF test statistic.

Given the extensive literature pointing to the economic benefits of developed financial markets, these poor results need some qualification. Rousseau and Wachtel (2005) report very similar results which are, however, derived from a different empirical estimation strategy. The authors find that the finance-growth relationship is not as strong according to more recent data as it was in the original studies with data for the period from 1960 to 1989. Rousseau and Wachtel (2005) offer two possible explanations which in particular can also help to understand the outcomes of the present study. First, financial depth may have had greater value as a shock absorber in the 1970s and 80s, decades characterized by worldwide nominal shocks. Furthermore, they also find that among poorer counties, the relationship is positive but imprecisely measured and among very rich countries it is absent. The general conclusion of their study is that the widely accepted effect of finance on growth is still present, but fragile. Given our own empirical results we have nothing to add to this reasoning.

The PMG estimation including the *NAWRU* supports the notion that a high structural unemployment rate may be growth-impeding. The PMG estimate is negative and significant and the Hausman test is highly indicative with respect to coefficient homogeneity across countries. Furthermore, both the DHT and Pedroni (1999, 2004) cointegration tests imply the existence of a long-run relation.

The last two PMG estimations refer to the demographic influence on output and growth. Panel unit root tests diagnose stationarity of the *age dependency ratio* which is why it enters the ECM as a regressor with country-specific coefficients instead of the long-run relation. The corresponding MG estimate of the *age dependency ratio* coefficient is positive and significant. Note that the cointegration tests and the PMG estimates refer to the long-run equation which comprises the Solow variables only. Recall that the *population over 65* variable is included in the first differences form in order to avoid the potential mixing of $I(1)$ with $I(2)$ variables. The diagnostic statistics with regard to parameter homogeneity and cointegration of the estimation are satisfactory, however, the significant and positive PMG coefficient estimate implies a rather counter-intuitive economic interpretation. Since there are fundamental problems in observing the impact of aging on economic development in empirical research, one of them being the difficulty in establishing the time series properties of the proxy variables, such estimation results should be regarded with caution.

2.5.3 Results of an extended panel ECM

In order to obtain a more comprehensive picture, an extended model comprising variables for which the evidence from the preceding estimations give most conclusive evidence of being related to long-run output per capita is estimated. The estimated cointegration relation from this model forms the basis of the trend output and trend growth computation with the help of the EMVF, which will be reported in the subsequent section.

Based on the results from the preceding analysis and given the limited degrees of freedom available, estimation of a model comprising *capital investment*, *population growth*, *trade openness*, *terms of trade* and the *NAWRU* is pursued. Note that we do not include the human capital variable due to poor outcomes which have been discussed above. The decision to use this model specification is based on the search criterions set out in section 2.4.4 but still reflects to some degree the subjective choice of the researcher. We also tried various different model estimations but found the following empirical model to be one that fits the data satisfactorily well while at the same time providing a reasonable economic interpretation.

In what follows we present the estimation and diagnostic results of the extended model in more detail. Table 2.4 shows the PMG estimate of the long-run slope coefficients vis-à-vis the according MG estimates. The MG estimate of the error correction coefficient is indicated below. In order to assess the short-run part of the ECM, the MG estimates of the coefficient estimates of the first differences are shown in the bottom part of table 2.4. As mentioned before, these short-run dynamics enter the panel ECM with country-specific coefficient estimates but are summarized in MG form in order to avoid excessive notation.

The PMG coefficient estimates are highly significant for all explanatory variables, while weak significance of MG estimates can be observed only for the *capital investment* and the *terms of trade* variable. The signs of the PMG estimates are in accordance with what theory predicts.

All variables pass the Hausman test and testing the coefficient estimates jointly also leads to an acceptance of the pooling restriction. Again, the LR test rejects the hypothesis of cross-country parameter equality of the long-run slope coefficients. The MG estimate of the *error correction coefficient* is significantly negative as well as the intrinsic individual estimates which are indicated in table 2.5. This table shows both the PMG estimates of the *error correction coefficients* and the corresponding estimates based on single country OLS estimations that do not impose long-run slope restrictions. Concerning the PMG results, an insignificant coefficient estimate is only found for Norway, while the OLS estimates are insignificant for Belgium and the USA. In fact, for the USA, the OLS coefficient estimate is even positive. In general, the *error correction coefficient* estimates imply a cointegration relation of GDP per capita and the considered variables.

Furthermore, given the panel cointegration test outcomes of the DHT $t(\widehat{\rho}^{*}, \kappa)$ CI test and the Pedroni (1999) Group-ADF test, which are shown in the bottom part

Table 2.4: PMG and MG estimation results for an extended ECM

Variable	PMG estimates		MG estimates		Hausman test[1]
			Long-run coefficients		
$\ln s_t^K$	0.262***	(0.016)	0.172*	(0.099)	0.863
$\ln n_t$	-0.044***	(0.006)	-0.036	(0.038)	0.048
$\ln open_t$	0.078***	(0.024)	0.133	(0.121)	0.211
$\ln tot_t$	-0.065**	(0.027)	-0.245*	(0.140)	1.730
$\ln nawru_t$	-0.041***	(0.006)	-0.097	(0.067)	0.694
			Error correction coefficient		
$\ln y_{t-1}$	-0.387***	(0.090)	-0.641***	(0.133)	
			Short-run coefficients		
$\Delta \ln y_{t-1}$	0.220***	(0.074)	2.960	(0.150)	
$\Delta \ln s_t^K$	0.104*	(0.057)	0.104*	(0.057)	
$\Delta \ln s_{t-1}^K$	-0.030	(0.022)	-0.019	(0.030)	
Δn_t	0.010*	(0.006)	0.024***	(0.008)	
Δn_{t-1}	0.010	(0.008)	0.008	(0.006)	
$\Delta \ln open_t$	0.058**	(0.028)	0.124**	(0.060)	
$\Delta \ln open_{t-1}$	-0.007	(0.019)	0.016	(0.036)	
$\Delta \ln tot_t$	0.011	(0.021)	0.024	(0.071)	
$\Delta \ln tot_{t-1}$	0.033	(0.041)	0.028	(0.059)	
$\Delta \ln nawru_t$	0.042	(0.032)	0.098***	(0.035)	
$\Delta \ln nawru_{t-1}$	0.011	(0.031)	-0.015	(0.047)	
			Factor and deterministic coefficients		
$\widehat{f}_t$	0.837***	(0.179)	0.513***	(0.187)	
Time trend	0.006***	(0.002)	0.013***	(0.004)	
Intercept	4.355***	(1.090)	7.290***	(1.564)	
			Panel cointegration tests		
$t(\widehat{\rho}^*, \kappa)$ CI test			-4.856***		
Group-ADF			-2.452***		

Notes: */**/*** denotes significance to the 1%/5%/10% level. Figures in brackets are the standard errors. [1]The joint Hausman test amounts to 3.432, which is insignificant according to the critical values from the $\chi^2(5)$ distribution. The *LR* test statistic is 411.114, which is highly significant with respect to the critical values of the $\chi^2(55)$ distribution. The *error correction coefficient*, *short-run coefficients* and *factor and deterministic coefficients* refer to averages of the individual estimates and corresponding standard errors are obtained with the aid of the non-parametric variance estimator. See table 2.2 for further details on the estimation process and specification techniques.

of table 2.4 and which both reject the "no cointegration" hypothesis, the estimated panel ECM provides strong evidence that a cointegration relationship between GDP per capita and the variables of interest is established.

Table 2.5: Estimates of the error correction coefficients ϕ_i

	PMG		OLS	
Austria	-0.295***	(0.109)	-1.172***	(0.235)
Belgium	-0.458***	(0.115)	-0.164	(0.250)
Germany	-0.392***	(0.060)	-0.628***	(0.126)
Denmark	-0.232***	(0.077)	-0.925***	(0.084)
France	-0.309***	(0.042)	-0.432***	(0.126)
UK	-0.802***	(0.141)	-0.767***	(0.160)
Italy	-0.354***	(0.045)	-0.479***	(0.068)
Japan	-0.079*	(0.049)	-0.429***	(0.116)
Netherlands	-0.298***	(0.076)	-0.885***	(0.126)
Norway	-0.098	(0.095)	-0.864***	(0.130)
Sweden	-1.159***	(0.118)	-1.330***	(0.210)
USA	-0.161*	(0.105)	0.379	(0.168)

Notes: */**/*** denotes significance to the 1%/5%/10% level according to the one-sided critical values of the standard normal distribution. Figures in brackets are the standard errors.

The average estimates of the *short-run coefficients* shown in the middle part of table 2.4 are significant only for few dynamic regressors. However, country-specific dynamic components can make important contributions in the individual equations. Table 2.6 provides an overview of the dispersion of lag orders of the variables across the individual country ARDLs that correspond to the country ECMs. Model selection was carried out with the help of the SBC. The SBC selects high orders of lags in particular for the equations for France, Norway, Sweden and the USA. For the other countries smaller models suffice.

Summary statistics that shed further light on the appropriateness of the PMG panel ECM at the individual country level are reported in table 2.7. For most country equations, the diagnostic statistics are generally satisfactory as far as tests on the residual serial correlation, functional form, normality and heteroscedasticity are concerned. Problems of serial correlation are observed only for the equations for Denmark and Norway whereas incorrect functional forms are indicated for the equations of France and Norway. The RESET test for functional form is weakly significant for Italy and the Netherlands. Normality of residuals is rejected only for the German equation and the hypothesis of no residual heteroscedasticity is refused at the 10% level of significance for Belgium. Due to small values of the *unadjusted* R^2 and a limited number of degrees of freedom, the *adjusted* $\overline{R}^2$ becomes negative for Austria and Norway. For the remaining countries the ECM seems to fit the historical series of the first difference of GDP per capita quite well. The positive $\overline{R}^2$

Table 2.6: Orders of lags in the ARDL model

	$\ln y_t$	$\ln s_t^K$	$\ln n_t$	$\ln open_t$	$\ln tot_t$	$\ln nawru_t$
Austria	1	2	0	0	2	1
Belgium	2	1	2	1	0	0
Germany	1	1	0	0	1	2
Denmark	2	0	0	1	2	0
France	2	1	2	2	2	2
UK	2	0	1	0	0	1
Italy	1	0	0	1	2	1
Japan	1	1	0	0	0	2
Netherlands	2	1	2	2	2	0
Norway	2	2	2	2	2	2
Sweden	2	2	2	2	2	2
USA	2	2	2	1	2	2

Notes: These lag orders where selected by the minimum of the Schwarz Bayesian information criterion (SBC). A maximum lag order of two was considered.

Table 2.7: Diagnostic statistics for the extended panel ECM

	$\chi^2_{SC}[4]$	$\chi^2_{FF}[1]$	$\chi^2_N[2]$	$\chi^2_H[1]$	$\overline{R}^2$
Austria	0.25	1.36	0.43	2.12	-0.04
Belgium	0.03	0.98	0.05	2.94*	0.60
Germany	6.92	0.11	6.86**	1.10	0.48
Denmark	40.21***	0.03	1.44	0.08	0.18
France	3.40	16.06***	0.20	1.95	0.86
UK	1.12	0.10	4.45	0.03	0.34
Italy	2.15	2.88*	3.68	0.52	0.83
Japan	0.07	1.58	1.42	1.92	0.73
Netherlands	2.66	7.78*	0.22	0.54	0.54
Norway	17.50***	13.58***	0.98	0.89	-0.39
Sweden	7.86*	0.20	3.02	0.00	0.92
USA	0.07	2.55	0.94	2.41	0.62

Notes: */**/*** denotes significance to the 1%/5%/10% level. The following χ^2 diagnostic statistics, which refer to the residuals that are based on the PMG estimates, are reported. $\chi^2_{SC}[4]$: Lagrange multiplier test of residual serial correlation. $\chi^2_{FF}[1]$: Ramsey's RESET test using the square of the fitted values. $\chi^2_N[2]$: Jarque-Bera test for normality based on a test of skewness and kurtosis of residuals. $\chi^2_H[1]$: Heteroscedasticity test based on the regression of squared residuals on squared fitted values.

lie in the interval of $[0.18, 0.92]$.

In terms of robustness of long-run parameter estimates, we also conducted a model selection with the aid of the Akaike (AIC) information criterion as well as with various fixed lag orders for all the variables of the model. Table 2.8 presents alternative PMG and MG estimates for four different ARDL specifications.

Table 2.8: Alternative PMG and MG estimation results of the extended ECM model for different ARDL specifications

ARDL order	Variable	PMG estimates		MG estimates		Hausman test[1]
		Long-run coefficients				
Chosen by the AIC ($k^{max}=2$)	$\ln s_t^K$	0.255***	(0.017)	0.166	(0.106)	0.730
	$\ln n_t$	-0.048***	(0.006)	-0.073	(0.054)	0.214
	$\ln open_t$	0.051*	(0.027)	0.260	(0.231)	0.831
	$\ln tot_t$	-0.053*	(0.030)	-0.307**	(0.157)	2.741*
	$\ln nawru_t$	-0.041***	(0.006)	-0.070	(0.074)	0.151
		Error correction coefficient				
	$\ln y_{t-1}$	-0.356***	(0.090)	-0.662***	(0.152)	
		Long-run coefficients				
(1,0,0,0,0,0)	$\ln s_t^K$	0.256***	(0.032)	0.224***	(0.068)	0.291
	$\ln n_t$	-0.015**	(0.008)	-0.029	(0.019)	0.653
	$\ln open_t$	0.247***	(0.041)	0.182**	(0.073)	1.181
	$\ln tot_t$	0.037	(0.038)	-0.101*	(0.061)	8.388***
	$\ln nawru_t$	-0.071***	(0.009)	-0.021	(0.052)	0.960
		Error correction coefficient				
	$\ln y_{t-1}$	-0.339***	(0.050)	-0.563***	(0.064)	
		Long-run coefficients				
(1,1,1,1,1,1)	$\ln s_t^K$	0.264***	(0.044)	0.080	(0.198)	0.915
	$\ln n_t$	-0.067***	(0.016)	-0.036	(0.043)	0.632
	$\ln open_t$	0.170**	(0.066)	0.338	(0.211)	0.703
	$\ln tot_t$	-0.043	(0.053)	-0.285	(0.219)	1.298
	$\ln nawru_t$	-0.046***	(0.012)	0.122	(0.159)	1.109
		Error correction coefficient				
	$\ln y_{t-1}$	-0.224***	(0.051)	-0.411***	(0.066)	
		Long-run coefficients				
(2,2,2,2,2,2)	$\ln s_t^K$	0.274***	(0.018)	0.114	(0.102)	2.578
	$\ln n_t$	-0.050***	(0.007)	-0.061	(0.045)	0.062
	$\ln open_t$	0.049*	(0.028)	0.317	(0.235)	1.324
	$\ln tot_t$	-0.086***	(0.031)	-0.233	(0.145)	1.084
	$\ln nawru_t$	-0.034***	(0.007)	-0.060	(0.070)	0.125
		Error correction coefficient				
	$\ln y_{t-1}$	-0.354***	(0.103)	-0.665***	(0.141)	

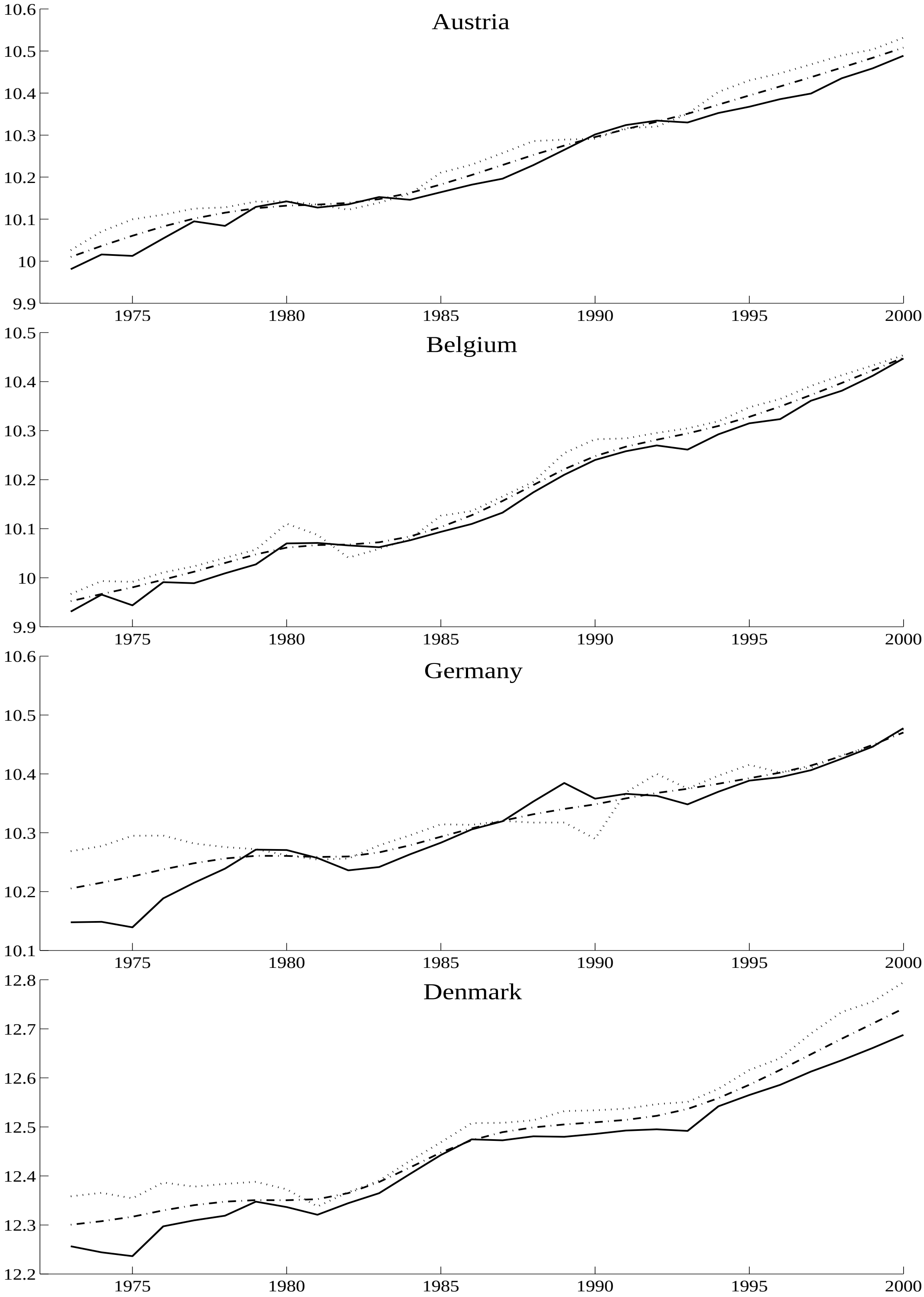
Notes: See table 2.4.

A general finding is that parameter estimates from the PMG approach do not vary largely across ARDL specifications and this in particular holds for the coefficient estimates of the *capital investment*, *population growth* and the *NAWRU* variables. Concerning *trade openness* and *terms of trade*, outcomes somehow depend on the dynamic specification and most precise parameter estimates are obtained in letting the SBC select the lag order (recall the model shown in table 2.4). A further result of the robustness analysis is that PMG estimates are less sensitive to the choice of orders of the ARDL model than the MG estimates. The MG coefficient estimates for the *capital investment* variable varies over the interval $[0.080, 0.224]$, for instance, while the PMG estimates lie in the narrow range of 0.255 to 0.274. The non-parametric variance estimates of the MG coefficients imply significance only in the ARDL (1,0,0,0,0,0) case whilst leading to insignificant coefficient estimates in almost all other specifications. The *error correction coefficient* estimates of the PMG are generally lower than the corresponding MG estimates. Hausman tests typically do not reject the restriction of equal slope parameters across countries, only for the *terms of trade* variable can a significant Hausman statistic be observed for two ARDL specifications.

In general, the results demonstrate that the PMG approach seems fairly robust to the choice of lag orders. Since the results of the model presented in table 2.4 provide the most precise PMG coefficient estimates, it will form the basis for the application of the EMVF that will be illustrated in the next section.

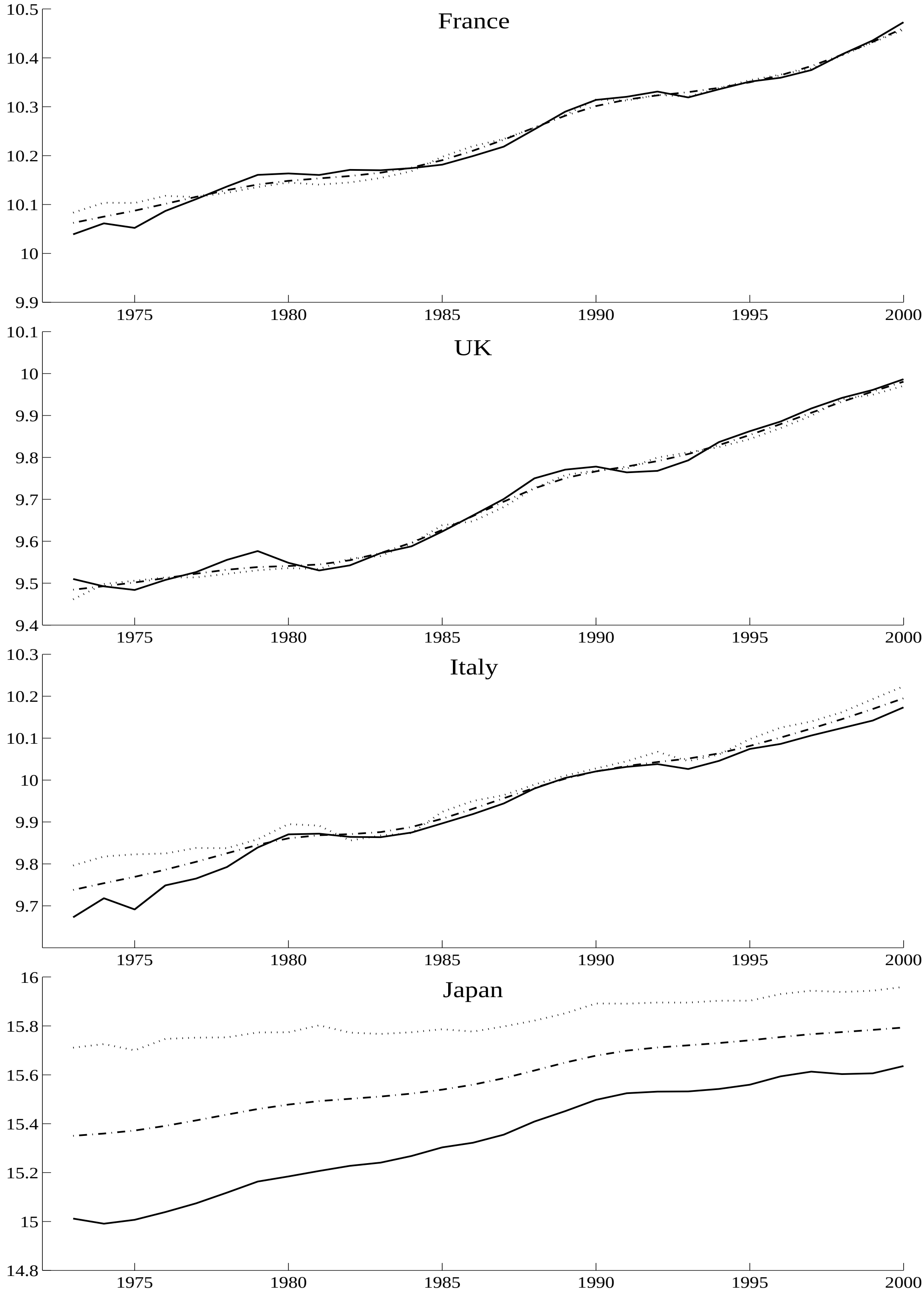
2.5.4 Results of the EMVF

With the help of the PMG estimates as reported in table 2.4 and the equations (2.18) and (2.19), trend estimates of the level of GDP per capita are readily computed. Figures 2.1 to 2.3 present the level of actual GDP per capita, the country-specific fitted values of the long-run relationship and the smoothed values that are obtained by the EMVF. Not surprisingly, the long-run component derived from the panel ECM is not as smooth as the trend estimates from the EMVF and for most countries it is also subject to some fairly significant downward and upward shifts at various points in the sample. Of course, the variability of the fitted values is inherited from the variability of their determinants. For most countries, the fitted values of the long-run relationship fluctuate around the actual series of GDP per capita. Exceptions are Denmark, Italy, Japan, the Netherlands and Norway, for which the fitted equilibrium levels lie above the actual levels in most periods.



Notes: — actual ··· fitted – · – trend

Figure 2.1: Trend, actual and fitted values of GDP per capita (I)



Notes: — actual ··· fitted – · – trend

Figure 2.2: Trend, actual and fitted values of GDP per capita (II)

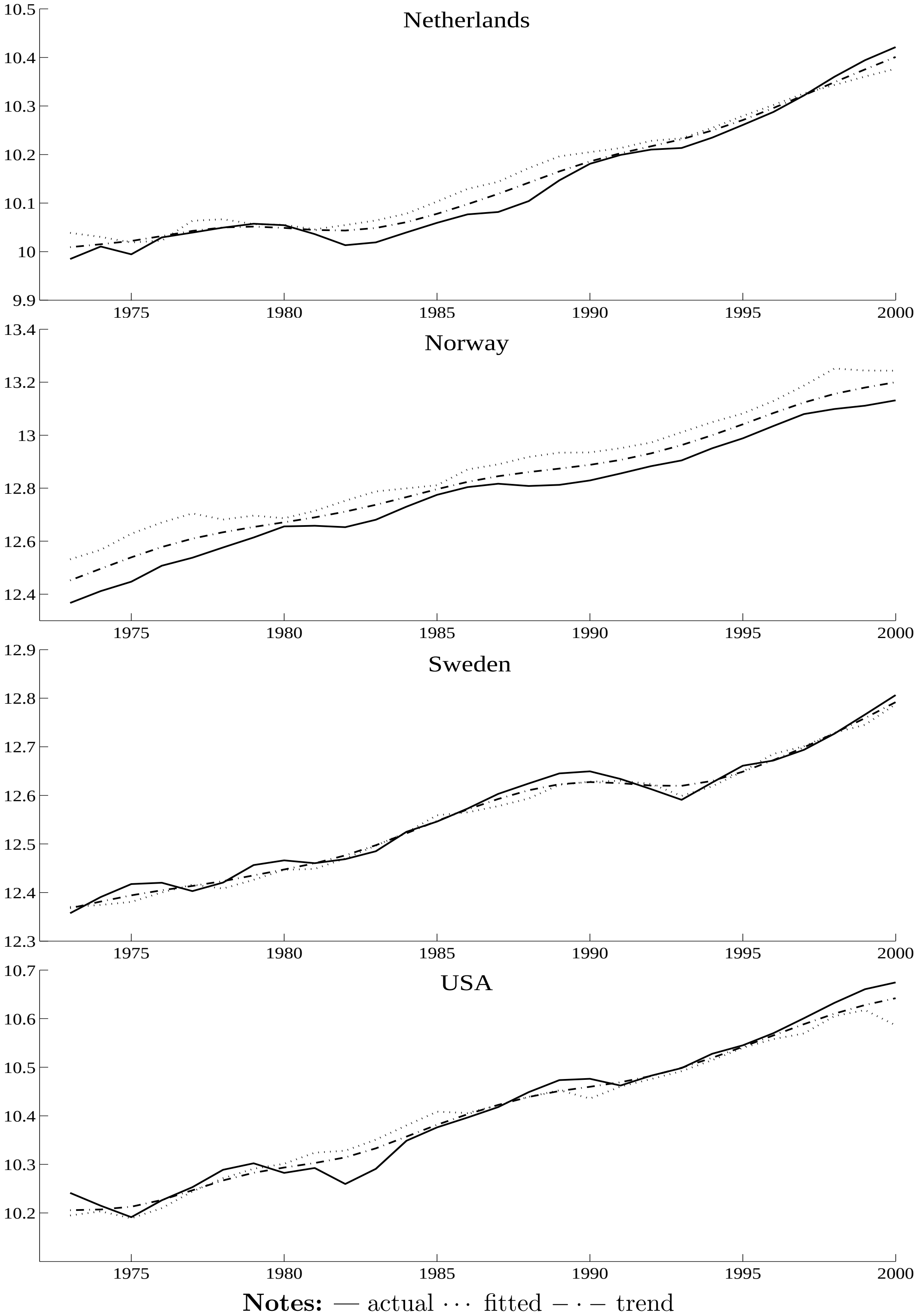
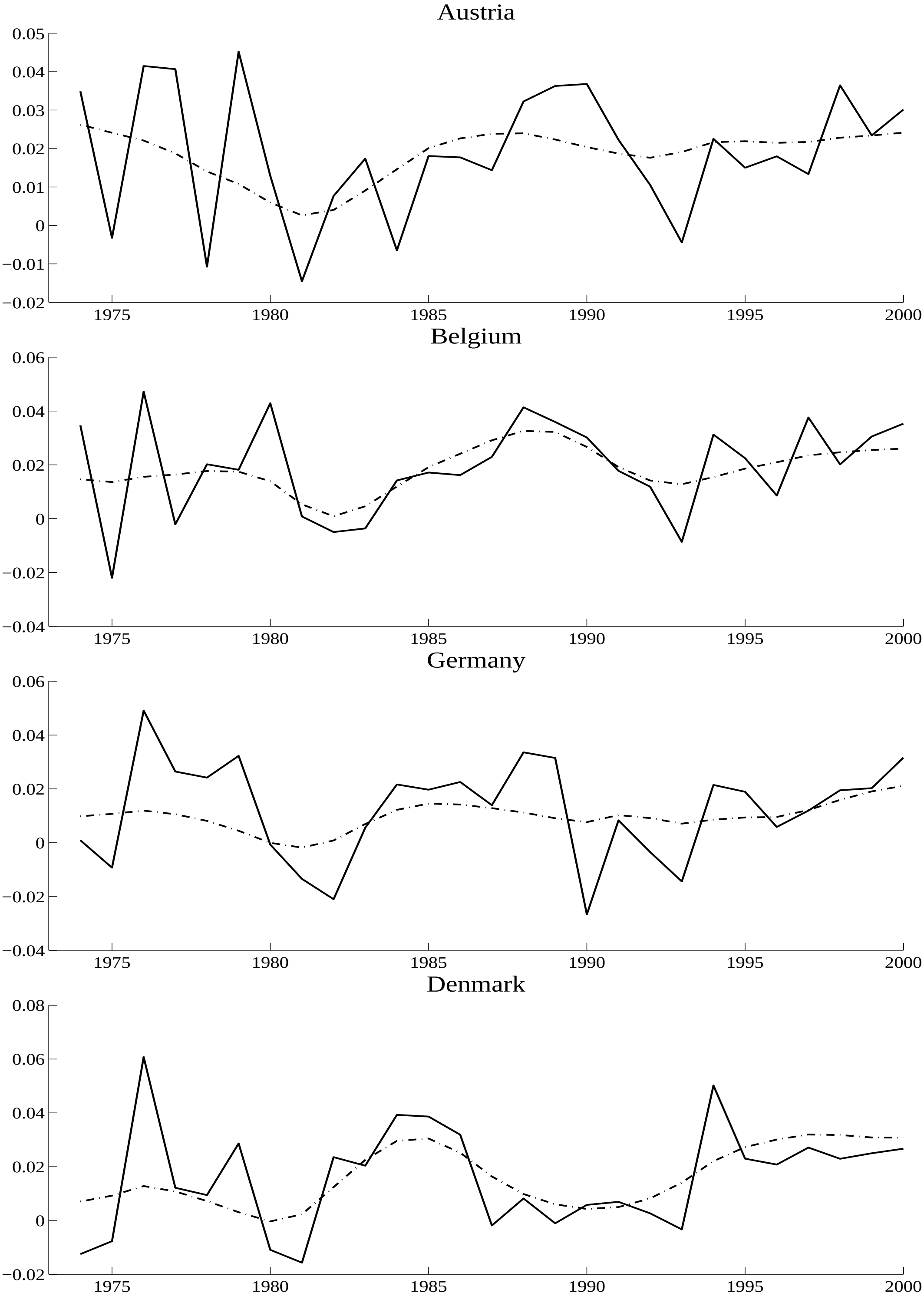
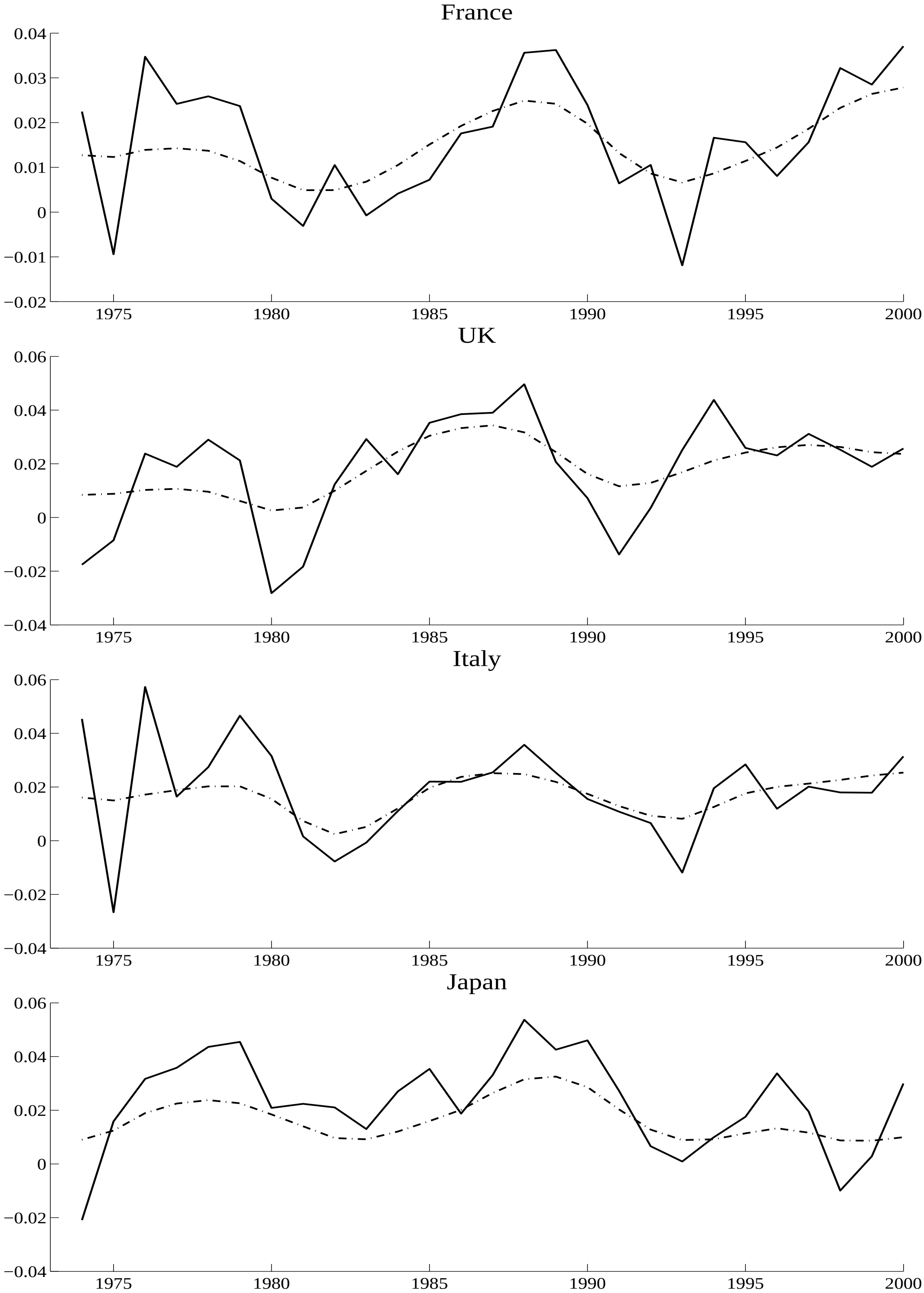


Figure 2.3: Trend, actual and fitted values of GDP per capita (III)



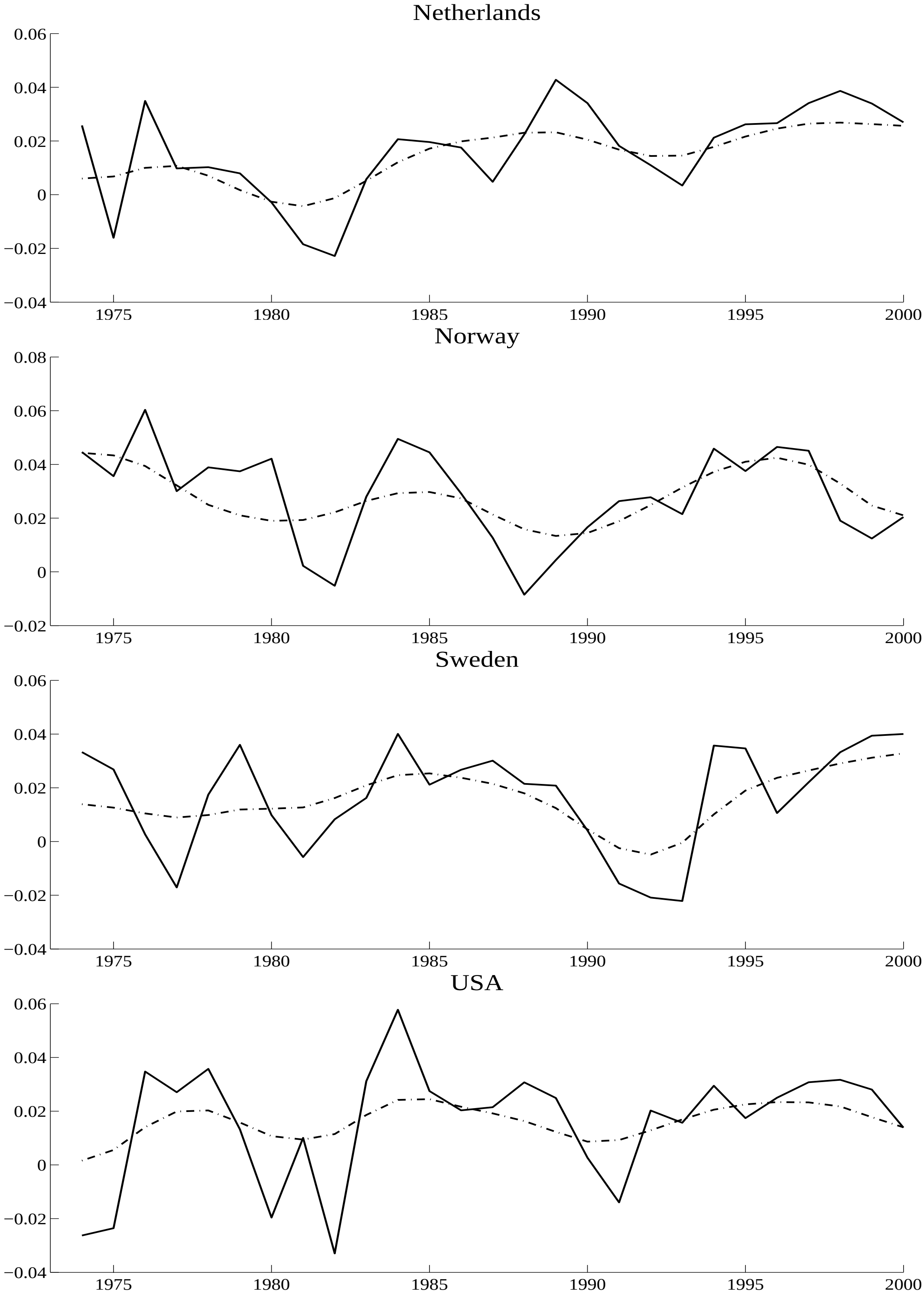
Notes: — actual − · − trend

Figure 2.4: Growth of trend and actual GDP per capita (I)



Notes: — actual − · − trend

Figure 2.5: Growth of trend and actual GDP per capita (II)



Notes: — actual − · − trend

Figure 2.6: Growth of trend and actual GDP per capita (III)

For Japan, in particular, the computed equilibrium path derived from the PMG estimates is located quite substantially above the actual level of GDP per capita. As a consequence, the smoothed series turns out to lie in between the actual and the fitted series due to the assignment of equal weights in the minimization problem of the EMVF. To complete the illustration, growth rates of the trend series of GPD per capita vis-à-vis the realized values are shown in figures 2.4 to 2.6. As expected, the trend growth record is very smooth. A typical pattern of trend growth which can be observed for nearly all countries is the decline at the beginning of the sample period, followed by a hump-shaped movement in the course of the eighties and continuing with an acceleration after a decline at the beginning of the nineties. Thus, the EMVF estimates provide a nice graphical summary of the global economic developments during the considered period from 1974 to 2000. The global economic downturns at the beginning of the eighties and nineties leave their marks not only in the realized per capita growth rates but also to a certain degree in the trend estimates. Such a partial attribution of cyclical movements to the trend estimate is a typical property of time series filters following the lines of Hodrick and Prescott (1997), in which a weighted sum comprising a component which determines the closeness of the trend to the actual series and a term that captures the variability of the trend is minimized. The weighting parameter λ (recall equation 2.19 in section 2.4.3) of the trend variability criteria thereby determines the smoothness of the resulting series.[12]

The EMFV estimates for the year 2000 are very close to the actual values and coincide with those in the United Kingdom, Norway and the USA. However, careful interpretation of the end-of-sample data points of the filtered series is particular important. These filter outcomes suffer to a certain degree from what is commonly known as the end-of-sample problem. Since the trend value for a certain data point is computed using data from both before and after that date, such filter methods have difficulties in identifying the trend at the end of the sample because fewer and fewer future values are available to include in the computation of such an average. The use of the “structural” information from the fitted values of the long-run relationship in the minimization problem of the EMFV may reduce the end-of-sample problem since by construction this component is less influenced by short-run fluctuations, but as mentioned before, still carries some degree of time variability. In practice, the end-of-sample problem is partly addressed by extending the actual series with a couple of forecasted data points. For the sake of completeness, averages of the estimated trend and actual growth rates over the sample period are provided by

[12]In the present application we keep to the familiar value of $\lambda = 100$ for annual series, however, other values have been proposed which may be used as well. Based on a number of different arguments , Ravn and Uhlig (2002), for instance, propose to use a value of 6.25 for λ in the case of the simple HP-Filter.

Table 2.9: Averages of trend growth and actual growth of GDP per capita

	Trend growth	actual growth
Austria	0.018	0.018
Belgium	0.018	0.018
Germany	0.011	0.012
Denmark	0.016	0.015
France	0.014	0.016
UK	0.018	0.017
Italy	0.017	0.018
Japan	0.019	0.022
Netherlands	0.015	0.016
Norway	0.027	0.027
Sweden	0.015	0.016
USA	0.016	0.016

Notes: The sample period is from 1974 to 2000.

table 2.9. From this table we see that averages for both values coincide and there is no tendency in the EMFV approach to systematically underestimate or overestimate observed growth rates.

2.6 Summary and conclusion

In this chapter, an approach for identifying growth factors with cross sectional time series data and an alternative statistical method for the determination of trend growth of GDP per capita is proposed. The considered growth factors can be traced back to suggestions and evidence of the theoretical and empirical growth literature, which is concisely reviewed at the beginning of this chapter. The used econometric techniques take the non-stationary nature and the heterogeneity of the data as well as cross section dependence across countries into account. The empirical outcomes suggest that many variables that have been prominently suggested in the literature indeed demonstrate a long-run correlation with economic growth. Furthermore, the estimation results of an extended ECM show that several of these variables taken together help to explain much of the historical growth patterns both across countries and over time. The trend output and growth paths that are derived from this extended ECM with the aid of a multivariate time series filter illustrate how such panel models can be used for policy applications. Naturally, due to the limited dimension of cross-sectional and time series data as well as possible data quality problems, the present analysis has limits in the extent to which it can draw generalized conclusions. A further issue is the potential simultaneity of the potential determinants of economic growth which demands investigations based on a full system approach. However, cross-country growth studies are often characterized by such shortcomings. A main concern of this chapter is to demonstrate which tests may be conducted and what estimators may be employed when searching for the determinants of trend growth with cross sectional time series data, rather than providing definite answers.

In contrast to the many ad-hoc methodologies for estimating trend growth, the approach discussed here which derives trend estimates from rigorous econometric evidence constitutes a transparent proceeding for comparing the different trend measures and the states of business cycles across countries.

Appendix

Table 2.10: Description of data

Variable	Coverage	CS #	Identifier
Real GDP per capita[1]	1971-2000	23	y
Basic variables			
Investment/GDP (public + private)[1]	1971-2000	23	s^K
Investment in human capital (Number of people enrolled in secondary and tertiary education/Total population)[3]	1971-2000	23	s^H
Population growth[1]	1971-2000	23	n
Fiscal Policy			
Government consumption/GDP[1]	1971-2000	23	cg
Tax quota (Indirect taxes + direct taxes + social contributions/GDP)[1]	1971-2000	12	$taxq$
Tax ratio (Direct taxes/Indirect taxes)[1]	1971-2000	12	tr
Public deficit[1]	1971-2000	12	$nlgq$
Monetary Policy			
Inflation (consumer prices)[1]	1971-2000	23	i
Standard deviation of inflation (past three years)[1]	1971-2000	23	isd
Research and development			
Expenditure on research and development/GDP (public + private)[1]	1981-2000	21	rd
International trade			
Trade openness (imports + exports/GDP)[1]	1971-2000	23	$open$
Terms of Trade (export prices/import prices)[1]	1971-2000	12	tot
Financial markets			
Stock market capitalization/GDP[2]	1976-2000	16	cap
Turnover ratio (stocks traded/stock market capitalization)[2]	1976-2000	16	$turn$
Credits to private sector/GDP[4]	1976-2000	21	$credit$
Labour markets			
NAWRU (non-accelerating wage rate of unemplyoment)[1]	1971-2000	12	$nawru$
Demography			
Age dependency ratio ([persons aged 0-14 years + persons over 65 years] /persons aged 15-64 years)[2]	1971-2000	12	adr
Population over 65/total population[2]	1971-2000	12	$pop65$

Sources:
[1]OECD Economic Outlook, various editions, [2]World Development Indicators of the World Bank, [3]Education Database of the UNESCO, [4]International Financial Statistics of the IMF

Cross section (see CS # above):

23 Austria, Australia, Belgium, Switzerland, Germany, Denmark, Spain, France, the United Kingdom, Greece, Ireland, Iceland, Italy, Japan, South-Korea, Luxembourg, the Netherlands, Norway, New Zealand,Portugal, Sweden, Turkey, the USA

21 Austria, Australia, Belgium, Switzerland, Germany, Denmark, Spain, France, the United Kingdom, Greece, Ireland, Italy, Japan, South-Korea, Luxembourg, the Netherlands, Norway, New Zealand, Portugal, Sweden, the USA

16 Austria, Australia, Belgium, Switzerland, Germany, Denmark, Spain, France, the United Kingdom, Greece, Italy, Japan, South-Korea, the Netherlands, Sweden, the USA

12 Austria, Belgium, Germany, Denmark, France, the United Kingdom, Italy, Japan, the Netherlands, Norway, Sweden, the USA

Table 2.11: Correlation matrix of the explanatory variables

	s^K	s^H	n	cg	$taxq$	tr	$nlgq$	i	isd	rd	$open$	tot	cap	$turn$	$credit$	$nawru$	adr	$\Delta pop65$
s^K	1.00																	
s^H	-0.19	1.00																
n	-0.04	0.03	1.00															
cg	-0.53	0.09	-0.18	1.00														
$taxq$	-0.29	0.17	-0.33	0.84	1.00													
tr	0.08	0.31	-0.10	0.14	0.28	1.00												
$nlgq$	0.28	0.08	0.01	-0.07	0.07	-0.00	1.00											
i	-0.36	-0.26	0.31	0.27	0.08	-0.02	-0.33	1.00										
isd	-0.29	-0.17	0.07	0.26	0.05	-0.05	-0.28	0.37	1.00									
rd	0.04	-0.09	0.05	-0.12	-0.26	-0.09	0.38	-0.32	0.00	1.00								
$open$	-0.22	0.28	-0.36	0.64	0.81	0.13	-0.02	-0.09	-0.07	-0.36	1.00							
tot	-0.40	0.36	-0.26	0.26	0.30	0.10	0.01	-0.22	-0.17	-0.11	0.36	1.00						
cap	-0.02	0.08	-0.22	-0.23	-0.30	-0.36	0.40	-0.47	-0.21	0.56	-0.12	0.18	1.00					
$turn$	0.17	0.13	-0.09	-0.37	-0.22	0.06	0.32	-0.48	-0.25	0.42	-0.15	0.09	0.45	1.00				
$credit$	0.32	0.11	-0.06	-0.27	-0.14	0.22	0.18	-0.23	-0.17	0.04	-0.05	0.01	0.06	0.21	1.00			
$nawru$	-0.40	0.34	-0.24	0.20	0.31	0.18	-0.30	-0.05	-0.15	-0.38	0.38	0.54	0.01	0.05	-0.07	1.00		
adr	-0.68	0.20	0.13	0.44	0.22	0.11	0.04	0.40	0.25	0.19	0.08	0.14	0.08	-0.24	-0.20	-0.01	1.00	
$\Delta pop65$	0.44	-0.21	-0.31	-0.42	-0.42	-0.15	-0.09	-0.30	-0.10	0.09	-0.38	-0.13	0.29	0.12	0.10	-0.12	-0.43	1.00

Notes: Correlation coefficients have been computed using the common sample period from 1981 to 2000. All variables are in logarithms, except n, $nlgq$, i and isd. See table 2.10 for further details.

Chapter 3

Panel Tests for Unit Roots in Hours Worked

3.1 Introduction

Employing the appropriate statistical model to measures of aggregate labor supply is important for several empirical applications. For example, whether aggregate hours worked are specified as a level or difference stationary time series can have far reaching consequences for the validity of predictions of Real Business Cycle (RBC) models, as the prominent debate initiated by Galí (1999) and taken up by Christiano et al. (2003) demonstrates. According to this controversy, the response of the labor market to technology shocks in a structural vector autoregression (SVAR) analysis crucially depends on the specification of hours worked. If hours worked are employed in levels, hours usually rise after a positive technology shock. If, on the other hand, hours worked are used in first differences, hours fall after the same shock. In the same manner that the first outcome is in line with the predictions of standard RBC models, the latter gives support for New Keynesian models of the macroeconomy assuming monopolistic competition, sticky prices and variable effort. However, in order to use SVAR models and impulse response functions to analyze dynamics of a system, the data must either conform or be transformed to conform to a tractable probability model so that inference can be drawn correctly. Therefore, careful inspection of the time series properties of hours worked is required before specifying such models.

Average hours worked is also a variable of interest in the discussion about the differences in work effort between Americans and Europeans. Important contributions to this field of activity are from Prescott (2004), Blanchard (2004) and Alesina et al. (2005). Among other things, the reasonings in those papers involve estimates of macro elasticities of labor supply, a theoretical and empirical assessments of the labor supply tax rate nexus and many possible explanations for the persistent behavior of the aggregate labor supply. Traditionally, macroeconomic labor supply elasticities have been estimated by simply evaluating the cross section dimension of the data due to a lack of time series. Meanwhile, data availability has improved and the comprehensive data set of Nickell and Nunziata (2001), for instance, allows estimation along the cross sectional and time series dimension. Appropriate transformations to maintain standard limiting theories or testing for cointegration to avoid spurious results is necessary if working with integrated variables.

It is well known that univariate tests for unit roots lack power if the variable is a stationary but highly persistent time series. The purpose of panel unit root tests is to increase power over univariate tests by combining information across units. Standard panel unit root tests, however, suffer from size distortions if the units are cross sectionally dependent, as it is likely in cross country studies.

The contribution of this chapter is to provide evidence of the non-stationarity of hours worked for OECD countries by applying several panel unit root tests that allow for cross country dependencies. A further contribution is to show that cross country dependence in hours worked can be empirically handled by allowing a factor structure to generate this dependency. The feasibility of estimating a common factor structure by analyzing the cross section variation in the data is also an advantage of panel methods over univariate procedures. Lastly, it is shown that the persistent behavior of hours worked originates both from a common factor and country specific sources.

The analysis starts with a short description of the employed data and the data source. Then, the results of standard univariate Augmented Dickey-Fuller (ADF) tests are reported and based on the residuals of these ADF regressions the cross section dependence inherent in the panel is assessed. Subsequently, a sequential testing strategy for unit root testing in cross sectionally dependent panels is accomplished and several tests of the so-called second generation are conducted. First, the panel unit root test of Demetrescu et al. (2006, DHT hereafter) is considered to illustrate the principle of meta-analysis in unit root testing and to find out whether there is a homogeneous unit root in the data. The principle of meta-analysis is the main thread running through the panel tests considered here, so first outlining the DHT test is a good starting point for the following. In addition, when discussing the PANIC procedure of Bai and Ng (2004, BN hereafter) it will be shown that the procedure from DHT offers an obvious improvement of the pooled test of BN. In contrast to the following approaches, the DHT test does not rely on a specific model of dependence structure. However, most of the panel unit root tests of the second generation build on the assumption that cross-section dependence can be captured through one or more common factors.

For a robustness check, the analysis continues with an application of Pesaran's

(2005), Pesaran hereafter) and Phillips and Sul's (2003, PS hereafter) testing methods which assume that cross section dependence originates from a single common factor. In order to examine if there is more than one common factor driving the evolution of hours worked, the method of Moon and Perron (2004, MP hereafter) is employed which addresses this problem adequately. Under the null hypothesis, the PS and MP tests assume the same order of integration for the idiosyncratic component and the common factor(s). In contrast, the PANIC procedure from BN allows the order of integration of these components to differ. Therefore, it is advisable not to stop the testing sequence with the results of PS and MP. In order to get a richer picture of the dynamics inherent to the data at hand, the BN procedure is used in a last step to show that the observed non-stationarity is due to both a common unobserved factor and country specific error components. Furthermore, this result indicates that the individual time series of hours worked are not cointegrated along the cross sectional dimension. The last section of the chapter offers summaries and conclusions on this matter.

3.2 Data

An important requirement for the subsequent estimations is the utilization of sound data which permit reliable cross country comparisons. Throughout this chapter, (average) hours worked refer to annual hours worked per employee. Hours worked on a per employee basis is the most comprehensive empirical counterpart for the labor input variable implied by most macroeconomic theories, e.g. general equilibrium business cycle models.[1]

The data for 30 OECD countries is taken from the Total Economy Database of The Conference Board and Groningen Growth and Development Centre from the University of Groningen. For most countries, the covered period is from 1950 to 2005.[2] The figures include paid overtime hours but exclude paid hours that are not worked due to vacation, sickness, etc. The University of Groningen compiles the figures from national labor force surveys and national establishment surveys as well as from national and international sources. International data sources include the OECD, the U.S. Bureau of Labor Statistics and the comprehensive studies of Angus Maddison.[3]

[1]Christiano et al. (2003) use total hours worked per capita for the U.S while Galí (1999) uses total hours worked and demonstrates the robustness of his results against per capita measures in subsequent papers. Alesina et al. (2005) base their estimations of the effect of tax rates on annual hours worked per person in the 15-64 age group.

[2]For the Czech Republic, Hungary, Poland, the Slovak Republic, the Republic of Korea and Mexico shorter periods are observed. See table 3.1 below and figures 3.4 to 3.6 in the appendix for further details on data coverage. Until 1990, the figures refer to West Germany and to Germany afterwards.

[3]A more detailed description of the data and adjustment methods can be found under

For interpretation of hours worked per employee as a labor supply quantity, it is important to notice that mainly three factors influence the evolution of this variable. The first influence comes from usual hours worked per week for full-time workers. Besides paid overtime hours, this component mainly reflects standard weekly hours which are the result of collective agreements between employer and employees or national legislation. The next factor affecting annual hours is the fraction of part-time workers. Obviously, an increase in the fraction of people who chose to work part-time decreases the aggregate measure of hours worked per employee. A further influence, of course, comes from days of paid vacations.

From the decomposition above one can conclude that deterministic or stochastic trends in hours worked can arise from various sources.

3.3 Single country analysis

Though the focus of the present chapter is on panel unit root tests, a natural starting point is the single country unit root testing. Individual Augmented Dickey-Fuller tests (ADF) for the logarithm of hours worked are presented below. This preliminary analysis serves several purposes: First, it gives a quick glance at the time series properties of the data at hand and at the possible diffusion of the number of integrated time series in the cross section. Second, in a next step, the residuals of these ADF regressions are utilized for estimating and testing the degree of cross section correlation in the panel. Furthermore, some of the subsequent tests for unit roots in panels with cross section dependence build on statistics of these univariate regressions.

When specifying ADF regressions, the decision about inclusion of appropriate deterministic components is important since the critical values for the ADF tests depend on that choice. As hours worked do not vary around zero, inclusion of an intercept is essential. However, a decision on inclusion of a linear time trend is not that clear-cut. Wolters and Hassler (2006) propose including a trend in the test regression whenever a series is suspicious of a linear trend upon visual inspection, because decision may not rely on the standard t-statistic of the estimated coefficient of the time regressor. Hamilton (1994) recommends fitting a specification that is a plausible description of the data under both the null hypothesis and the alternative, if the researcher does not have a specific null hypothesis. In addition to this he proposes including a linear trend as a regressor if there is an obvious trend in the data.

A downward trend in hours worked since the seventies is observable for many countries in the panel (see figures 3.4 to 3.6 in the appendix). However, this trend

http://www.ggdc.net

Table 3.1: Individual $ADF(l_i)$ test statistics

		Intercept			Intercept and trend		
Country	Obs.*	t-stat	p-value	l_i	t-stat	p-value	l_i
Australia	56	-2.20	0.21	0	-1.69	0.74	0
Austria	56	0.87	0.99	1	-2.82	0.20	1
Belgium	56	-1.95	0.31	1	0.17	1.00	0
Canada	56	-1.75	0.40	2	-1.14	0.91	2
Switzerland	56	-1.72	0.42	0	-1.11	0.92	0
Czech Republic	17	-0.87	0.77	0	-1.69	0.71	0
Germany	56	-1.58	0.49	0	-0.62	0.97	0
Denmark	56	-0.77	0.82	2	-0.84	0.96	1
Spain	56	-0.44	0.89	1	-1.73	0.72	1
Finland	56	-1.10	0.71	0	-1.47	0.83	0
France	56	-0.21	0.93	1	-1.79	0.70	0
United Kingdom	56	-0.74	0.83	3	-1.31	0.87	2
Greece	56	-1.58	0.49	0	-0.84	0.95	0
Hungary	26	-2.51	0.12	0	-2.09	0.52	0
Ireland	56	0.35	0.98	2	-2.03	0.57	0
Iceland	56	-1.38	0.59	3	-0.44	0.98	2
Italy	56	-0.27	0.92	0	-1.34	0.87	5
Japan	56	-0.26	0.92	1	-1.35	0.87	0
Republic of Korea	44	-2.01	0.28	0	-1.34	0.86	0
Luxembourg	56	-1.39	0.58	1	-0.80	0.96	2
Mexico	47	-1.63	0.46	2	0.54	1.00	9
Netherlands	56	-2.03	0.27	1	1.36	1.00	0
Norway	56	-0.95	0.76	1	-0.93	0.94	1
New Zealand	56	-2.71	0.08	0	-0.44	0.98	4
Poland	17	-1.17	0.66	0	-2.23	0.45	0
Portugal	56	-1.07	0.72	0	-0.91	0.95	8
Slovak Republic	17	-1.24	0.63	0	-1.18	0.88	0
Sweden	56	-1.86	0.35	1	-0.78	0.96	1
Turkey	56	-1.58	0.49	0	-0.84	0.95	0
USA	56	-0.91	0.78	1	-1.59	0.78	0

Notes: *total number of Observations. All tests were executed with the help of *Eviews*. MacKinnon (1996) p-values. Lag length l_i was chosen due to the minimum of the modified Schwarz Bayesian information criterion. Maximum lag length was 3, 5, 9 or 10, depending on the individual number of time series observations from the interval $[17, 56]$.

stopped for some countries during the eighties (Denmark, Spain, the United Kingdom, Iceland, Norway, New Zealand, Sweden and the USA) and still seems to continue for Germany, Ireland and Portugal. Standard RBC theory states that hours worked should rather be constant, hypothesizing hours worked being a stationary process fluctuating around a constant mean.[4] This would suggest using an intercept

[4]Constant behavior of hours worked per worker is a feature of the balanced growth path if the number of workers grows with the population in the long-run.

without deterministic trend specification for the ADF regressions. On the other hand, the increasing participation rates of women of who many chose to work part-time thereby reducing the aggregate measure of hours worked could be possibly approximated, at least locally, by a linear trend specification. Neither economic theory nor visual inspection of hours worked for most countries provides clear guidance on whether to include a linear trend or not in the regressions. Therefore, both specifications are considered below.

In summarizing table 3.1, the following can be observed: On the 10% level of significance for the ADF regressions including only an intercept, the null hypothesis of non-stationarity is rejected only for New Zealand. When concentrating on the outcomes of the ADF tests which employ an intercept and trend specification, the null hypothesis is not rejected for any of the countries in the cross section. Overall, regarding hours worked as non-stationary time series is favored over a trend stationary specification.

However, ADF tests lack power relative to the alternative that the series is a persistent, but stationary process. For example, this lack of discriminatory power is one of the reasons why Christiano et al. (2003) do not regard classical univariate unit root diagnostics as helpful in deciding whether to treat hours worked for the US as a level or difference stationary stochastic process.[5] Increasing power of unit root tests through the pooling of information across countries is the primary aim of panel unit root tests and a reason for the popularity of these tests. Therefore, testing the order of integration of hours worked with the help of panel data seems to offer an obvious solution to the power problem. The next section gives a brief outline of the panel assumptions and hypothesis employed in the remainder of the chapter.

3.4 Panel analysis

3.4.1 The panel unit root framework

Surveys of panel unit root tests are given by Breitung and Pesaran (2005), Choi (2004), Banerjee (1999) and with a special focus on second generation panel unit root tests by Gutierrez (2006), Jang and Shin (2005) and Gengenbach et al. (2004), among others. Only the basic framework is given below.

It is assumed that the time series for N cross sections evolve according to:

$$h_{it} = d_{it} + x_{it} \tag{3.1}$$

[5] Christiano et al. (2003) circumvent deciding on the basis of univariate unit root tests. Instead, they employ an encompassing criterion to select between the competing specifications. Cf. pp.8 for details.

$$x_{it} = \phi_i x_{it-1} + u_{it} \tag{3.2}$$

where $i = 1, ..., N$, $t = 1, ..., T_i$ and d_{it} represent deterministic components including any individual intercepts or individual time trends or both. The cross section specific autoregressive coefficient is ϕ_i. Equations (3.1) and (3.2) translate into an expression for the observable variables:

$$h_{it} = \phi_i h_{it-1} + d_{it} - \phi_i d_{it-1} + u_{it} \tag{3.3}$$

Panel unit root tests of the first generation assume independent units h_{it} and typically suppose that the idiosyncratic disturbances u_{it} are i.i.d. across i and t with $E(u_{it}) = 0$, $E(u_{it}^2) = \sigma_i^2$ and $E(u_{it}^4) < \infty$.[6] Examples of the modelling strategy of u_{it} in the presence of cross section dependence are given below.

Most panel unit root tests build their testing strategy around ADF type regressions corresponding to equation (3.3). A test for the presence of a unit root in the panel is represented by the null hypothesis $H_0 : \phi_1 = \cdots = \phi_N = \phi = 1$. Two types of tests can be distinguished, depending on the alternative hypothesis under consideration. The first type of test considers a homogeneous alternative, i.e. it takes the form $H_1 : \phi_1 = \cdots = \phi_N = \phi < 1$. Examples are the tests of Levin et al. (2002), Breitung (2000) and Hadri (2000). The second sort of tests employs a heterogeneous alternative hypothesis: $H_1 : \exists i$ with $\phi_i < 1, i = 1, ..., N$. This implies that there is a subgroup $N_0 \leq N$ for which $\phi_1 < 1, ..., \phi_{N_0} < 1$. The tests of Im et al. (2003), Maddala and Wu (1999) or Choi (2001) involve this alternative hypothesis.[7] Irrespective of the alternative under consideration, when the null hypothesis of a unit root is rejected, one can only conclude that a certain fraction of units in the panel is stationary. The panel unit root tests under cross section dependence outlined in the subsequent sections assume a heterogeneous alternative throughout.

As mentioned above, the advantage for testing the unit root hypothesis on the basis of cross sectional time series is the amplification of power. The gain in power by switching from univariate unit root tests to panel unit root tests is well documented for example in the papers of Levin and Lin (1992) and Levin et al. (2002).

However, if the panel features cross section dependence, classical panel unit root tests suffer from serious size distortions. As it is shown in the next section, the panel data of hours worked for OECD countries is characterized by significant cross section correlation that should not be neglected in unit root testing. Therefore, outcomes of first generation panel tests for unit roots in hours worked are not reported here.

The implication of cross section dependence is surveyed by several authors. Gen-

[6] Cf. for example, Breitung and Pesaran (2005).

[7] Breitung and Pesaran (2005) note, that despite the different treatment of the alternative hypothesis, both tests can be consistent against both types.

genbach et al. (2004) give a brief literature overview to simulation studies that assess the performance of panel unit root tests under the presence of cross correlation and cross section cointegration. Banerjee et al. (2005) demonstrate how panel unit root tests become oversized in the presence of long-run cross unit relationships. Hassler and Tarcolea (2005) also conclude by investigating nominal long-term interest rates for 12 OECD countries that ignoring or modelling cross-correlation in multi-country studies may heavily affect the outcome of non-stationarity panel analyzes. Pesaran (2005) demonstrates by means of Monte Carlo simulations that panel unit root tests that do not account for cross section dependence can be seriously biased if the degree of dependence is sufficiently large. Phillips and Sul (2003) show that OLS estimators provide little gain in precision compared with single equation OLS when cross section dependence is ignored in the panel regression. Furthermore, commonly used panel unit root tests are no longer asymptotically similar under the presence of cross section dependence. Strauss and Yigit (2003) demonstrate that the greater the extent of cross correlations and their variation, the higher is the size distortion of the Im et al. (2003) test.

3.4.2 Cross section dependence in the panel of hours worked

There are several potential causes for cross section dependence in the present panel: Common observed and unobserved factors or general residual correlation that remains after controlling for common influences. Examples for such factors affecting average hours worked are the above-mentioned technology shocks.

Pesaran (2004a) proposes a simple test for error cross section dependence that has the correct size and sufficient power even in small samples. To check if the OECD panel at hand is characterized by cross section dependence, the residuals of the individual $ADF(l_i)$ regressions from the preceding single country analysis are used to compute Pesaran's (2004a) test statistic. The test draws on the residuals of both the intercept only and the intercept and linear trend specifications. The test statistic of cross section dependence for an unbalanced panel is computed as[8]

$$CD = \sqrt{\frac{2}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \sqrt{T_{ij}} \hat{\rho_{ij}} \right), \tag{3.4}$$

where $\hat{\rho_{ij}}$ are the pairwise correlation coefficients from the residuals of the ADF regressions. The correlations are computed over the common set of observations T_{ij} for i and j, $i \neq j$. The CD statistic is distributed standard normal for $T_{ij} > 3$, if the number of country specific observations exceeds the number of regressors in the underlying equation and sufficiently large N. As Pesaran (2004a) demonstrates the

[8] Cf. Pesaran (2004a), p.17.

good performance of the CD test in small samples, it seems to be well suited for the present cross section of 30 countries with numbers of time observations ranging from 17 to 56.

Table 3.2: Test of cross section dependence within different regions

	OECD	European Union	Europe	North. Europe	Non Europe	G7
Residuals from $ADF(l_i)$ regression with intercept						
CD statistic	12.71	5.79	9.92	7.61	3.46	5.22
P-value	0.00	0.00	0.00	0.00	0.00	0.00
$\bar{\hat{\rho}}$	0.10	0.08	0.10	0.33	0.09	0.16
Residuals from $ADF(l_i)$ regression with intercept and linear trend						
CD statistic	12.96	6.22	10.41	7.96	2.74	6.09
P-value	0.00	0.00	0.00	0.00	0.01	0.00
$\bar{\hat{\rho}}$	0.10	0.08	0.11	0.34	0.08	0.18

Notes: CD test is based on the residuals of the individual $ADF(l_i)$ regressions, sample is unbalanced, i.e. $T_i \in [17, 56]$. The CD statistic is asymptotically normally distributed. P-values refer to a two-sided test. $\bar{\hat{\rho}}$ is the simple average of the pair-wise residual correlation coefficients.

OECD=Australia, Austria, Belgium, Canada, Switzerland, Czech Republic, Germany, Denmark, Spain, Finland, France, United Kingdom, Greece, Hungary, Ireland, Iceland, Italy, Japan, Republic of Korea, Luxembourg, Mexico, Netherlands, Norway, New Zealand, Poland, Portugal, Slovak Republic, Sweden, Turkey, Unites States

European Union=Austria, Belgium, Czech Republic, Germany, Denmark, Spain, France, United Kingdom, Greece, Hungary, Ireland, Italy, Luxembourg, Netherlands, Poland, Portugal, Slovak Republic, Sweden

Europe=Austria, Belgium, Switzerland, Czech Republic, Germany, Denmark, Spain, Finland, France, United Kingdom, Greece, Hungary, Ireland, Iceland, Italy, Luxembourg, Netherlands, Norway, Poland, Portugal, Slovak Republic, Sweden

Northern Europe=Denmark, Finland, Iceland, Norway, Sweden

Non Europe=Australia, Canada, Japan, Republic of Korea, Mexico, New Zealand, Turkey, USA

G7=Canada, Germany, France, United Kingdom, Italy, Japan, USA

Table 3.3: Cross section dependence across Europe and Non European countries

	ADF with intercept	ADF with intercept and trend
CD statistic	4.60	4.67
p-value	0.00	0.00
$\bar{\hat{\rho}}$	0.03	0.04

Notes: CD test is based on pair-wise residual correlations between each European and non European country. See table 3.2 for further notes.

Table 3.2 shows the CD statistics for countries within the OECD, the European

Union, Northern Europe, Non-European countries and the Big Seven western industrial countries. The upper part of table 3.2 contains CD statistics that employ residuals from ADF estimations with intercept only while the lower part displays the results that rely on ADF residuals from an intercept and linear trend regression. The hypothesis of zero cross section correlation is rejected for all regions and both ADF specifications at the 1%-level of significance. In both specifications, according to the average correlation coefficients, the highest degree of cross section dependence is found for the countries within the group of Northern Europe, followed by the countries within the G7. The group of Non-European countries shows about the same degree of dependence as the countries within the European Union and within geographical Europe.

The CD statistic can also be used to test for dependence across regions with distinct countries. Table 3.3 displays the CD statistic that builds on correlations which are computed for the ADF residuals of each European country with the residuals of each Non-European country.[9] By rejecting the null hypothesis of cross section independence at the 1%-level, these test statistics also indicate the presence of error dependence across the countries of Europe and the group of Non-European countries. However, the average residual correlation coefficient $\overline{\hat{\rho}}$ is rather low.

Overall, the outcomes of the preceding tests clearly indicate the presence of cross section dependence of hours worked in the panel of OECD countries. In addition, the estimates of the average correlation coefficients for different regions suggests that residual correlation is heterogeneous rather than homogeneous.

Tests for the presence of unit roots in hours worked should take these dependence into account in order to produce reliable results. The next section addresses this issue by applying second generation unit root tests for panel data.

3.4.3 Panel unit root tests for cross sectionally dependent panels

In this section, the panel unit root tests of DHT, Pesaran, PS, MP and BN which all allow for cross section dependence among units are illustrated.[10] We consider a sequence of tests mainly for robustness reasons, but also to demonstrate to the reader the many possibilities to conduct such unit root tests. Besides similarities, the considered tests differ in terms of strategy and test statistics employed, so using a series of tests should give a comprehensive picture of the dynamic characteristics of the data at hand. In order to assess the respective tests with respect to small

[9] For this version of the test, the CD statistic is calculated as $CD_{N_1N_2} = \sqrt{\frac{1}{N_1N_2}}\left(\sum_{i=1}^{N_1}\sum_{j=N_1+1}^{N_1+N_2}\sqrt{T_{ij}}\hat{\rho_{ij}}\right)$, whereas N_1 is the number of countries in region 1 and N_2 is the number of countries in region 2.

[10] Tests that build on a GLS approach are not considered in the present analysis as they rely on T being substantially larger than N which is not the case for the panel data at hand. Cf. Breitung and Das (2005), for instance.

sample properties and power properties against roots in the neighborhood of one, references to simulation outcomes in the literature are given.

More precisely, the testing order is as follows: In a first step, the test of DHT is conducted to see whether it indicates a unit root in the data or not. Starting from a single country analysis, the DHT test statistic is readily computed since it simply combines individual p-values. For robustness reasons and to allow the cross section dependence being caused by a common factor, the testing continues with an application of the procedures from Pesaran and PS. In order to check whether more than one common factor should be taken into account, in a next step the test of MP is used. If a unit root is detected at this stage and one is simply interested in this result, the testing sequence could in principle be stopped here. However, since both the MP and PS test assume the same order of integration of the common factor and idiosyncratic component under the null hypothesis, these tests do not help to assess the *source* of possible non-stationarity. The sum of two time series can have dynamic properties very different from the individual series themselves (Bai and Ng, 2004). Therefore, in a final step the BN procedure is employed to test for the presence of unit roots in the idiosyncratic components and the common factors *separately*. This gives a much richer insight into the dynamics of the individual time series. In addition, the BN approach amounts to a test for no cointegration among the individual time series of hours worked. To see the complementarities between the different tests more clearly, consider the following relation: Since the PS and MP tests effectively remove the common factor, it implies that if these tests both reject the null hypothesis of a unit root and the BN test rejects the unit root hypothesis for the idiosyncratic component but not for the common factor, this is a strong indication that the individual time series are cointegrated.[11] In addition, the DHT test can be used to confirm this result if it signifies non-stationarity.

As mentioned above, we start with an outline of the DHT approach since it offers a nice introduction into the idea of building meta statistics which is also shared by most of the other procedures. Subsequently, the approaches of Pesaran, PS, MP and BN are sketched out. The advantage of the test procedures from DHT and Pesaran is that they can be applied to unbalanced panels, while the tests of PS, MP and BN require balanced panels. In that case, balancing the panel reduces the cross section dimension to $N = 24$ and fixes the time dimension to $T = 56$.[12]

[11] Cf. Gengenbach et al. 2004 for the relation between the MP and BN tests.

[12] Balancing the panel drops the observations for the Czech Republic, Hungary, Mexico, Poland, Slovak Republic and the Republic of Korea.

Demetrescu et al. (2006, DHT)

The DHT test directly builds on the test statistics of the outcomes of the individual ADF tests from section 3.3. The test statistic is constructed as a linear combination of individual specific probits t_i corresponding to the p-values p_i resulting from the individual unit root tests. The probits are quantiles from the standard normal distribution of the respective p-values. This proceeding corresponds to the inverse normal method and DHT propose a modified version for panel unit root testing to account for dependencies in the probits. These dependencies in turn stem from the dependencies in the underlying test statistics and reflect cross section dependence.

The recommended (unweighted) test statistic by DHT due to Hartung (1999) to test for a unit roots in the panel against the heterogeneous alternative is[13]

$$t(\widehat{\rho}^*, \kappa) = \frac{\sum_{i=1}^{N} t_i}{\sqrt{N + N(N-1)\left[\widehat{\rho}^* + \kappa\sqrt{\frac{2}{N+1}}(1-\widehat{\rho}^*)\right]}} \tag{3.5}$$

where $\widehat{\rho}^* = max\left(-\frac{1}{N-1}, \widehat{\rho}\right)$, $\widehat{\rho} = 1 - \frac{1}{N-1}\sum_{i=1}^{N}(t_i - \bar{t})^2$, $\bar{t} = \frac{1}{N}\sum_{i=1}^{N} t_i$ as the arithmetic mean of the probits t_i, which are calculated from the inverse of the standard normal distribution Φ^{-1}.

The contribution of DHT is to show under which conditions the statistic of (3.5) follows a standard normal distribution. In addition, it is demonstrated that the test is robust if the correlation of the probits varies to a certain degree. Furthermore, on experimental grounds, DHT provide evidence that the modified inverse normal method is reasonably reliable when applied to ADF tests in correlated panels. This holds also when $N = 25$ and $T = 50$ but it is shown that the modified inverse normal method results in an undersized test in the presence of weak correlation.

The test statistic $t(\widehat{\rho}^*, \kappa)$ is readily computed with the p-values from table 3.1. The value of $\widehat{\rho}^*$ amounts to 0.16 in the intercept case and to 0.05 in the intercept and trend case. DHT and Hartung (1999) propose to use $\kappa = \kappa_1 = 0.2$ or $\kappa = \kappa_2 = 0.1(1 + \frac{1}{N-1} - \widehat{\rho}^*)$. This parameter is intended to regulate the actual significance level in small samples.[14] In the simulation studies of DHT, the experimental size of the test is not sensitive to the choice of κ.[15] The test statistic here is slightly influenced by the choice of κ in the intercept and trend case. However, the test decision is not affected by this option. Table 3.4 shows test results.

The low level of significance for both the intercept only and intercept and trend specification clearly suggests that the unit root hypothesis should not be rejected.

[13]Cf. DHT, p. 5.
[14]Cf. Hartung (1999) for details.
[15]However, DHT assume stronger correlation in their simulation study than it is indicated for hours worked in the OECD panel according to table 3.2.

Table 3.4: Results of the DHT test

	Intercept	Intercept and trend
$t(\widehat{\rho^*, \kappa_1})$	0.72	3.63
P-value	0.76	1.00
$t(\widehat{\rho^*, \kappa_2})$	0.76	4.02
P-value	0.78	1.00

Notes: Test statistics are based on MacKinnon (1996) p-values of individual ADF tests. $N = 30$ and $T_i \in [17, 56]$.

Pesaran (2005, Pesaran)

It is highly conceivable that cross section dependence in international data on hours worked can occur because of common global factors like a global trend or cyclical element. The country figures suggest that there is co-movement between hours worked (see figures 3.4 to 3.6 in the appendix). The Pesaran test and also the subsequent methods do account for cross section dependence through the assumption that one or more common unobserved factors are driving the dependence structure.

Pesaran builds on the assumption that the error terms u_{it} of equation (3.3) follow a single common factor structure

$$u_{it} = \lambda_i f_t + \epsilon_{it} \tag{3.6}$$

The common unobserved factor f_t is always assumed to be stationary and impacts the cross section times series with a fraction determined by the individual specific factor loading λ_i. For the idiosyncratic errors ϵ_{it}, the same assumptions as under the panel unit root tests of the first generation hold, i.e. they are are i.i.d. across i and t with $E(\epsilon_{it}) = 0$, $E(\epsilon_{it}^2) = \sigma_i^2$ and $E(\epsilon_{it}^4) < \infty$. Furthermore, ϵ_{it}, f_t and λ_i are mutually independent distributed for all i.

Thus, cross section dependence arises due to the common factor, which can be approximated by the cross section mean $\overline{h_t} = \frac{1}{N}\sum_{i=1}^{N} h_{it}$.[16] Pesaran proposes the following augmented Dickey-Fuller regression:

[16] If a common time specific effect is the only source of cross section correlation, the correlation can be eliminated by subtracting cross sectional means from the data. Im et al. (1995) propose this proceeding. However, Strauss and Yigit (2003) demonstrate through Monte Carlo simulation that demeaning the data leads to false inference in panel unit root tests if the original data generating process had heterogeneous correlation, i.e. if pair-wise cross-section covariances of the error components differ across units.

$$\Delta h_{it} = a_i + \alpha_i h_{it-1} + \beta_i \overline{h}_{t-1} + \sum_{j=1}^{p} \gamma_{ij} \Delta h_{it-j} + \sum_{j=0}^{p} \theta_{ij} \Delta \overline{h}_{t-j} + d_{it} + \varepsilon_{it} \tag{3.7}$$

The test for the presence of a unit root can now be conducted on the grounds of the t-value of α_i either individually or in a combined fashion. The first statistic is denoted as cross sectionally augmented Dickey-Fuller $CADF_i$ statistic while the latter resembles the familiar IPS statistic of Im et al. (2003) and is constructed as

$$CIPS = \frac{1}{N} \sum_{i=1}^{N} CADF_i \tag{3.8}$$

Pesaran investigates the performance of the $CADF_i$ and $CIPS$ tests by means of Monte Carlo simulations and shows that these tests have satisfactory size and power even for relatively small values of N and T, i.e. even in the case of $N = T = 10$. In the linear trend model, power rises quite rapidly with both N and T if $T > 30$. This small sample property renders the Pesaran test quite appealing for an application to the present OECD cross section.

Due to the presence of the lagged level of the cross sectional average, the limiting distribution of the $CADF_i$ statitics and the $CIPS$ statistic does not follow a standard Dickey-Fuller distribution. However, Peseran provides critical values based on simulations for the CADF and CIPS-distributions for three cases (no intercept and no trend, intercept only, intercept and trend).

Table 3.5 reports the results of the $CIPS$ test for hours worked for the unbalanced OECD panel and different lag length l.

Table 3.5: Results of the $CIPS$ test

l	0	1	2
$CIPS^c$	-1.91	-2.17	-1.77
$CIPS^{c,\tau}$	-2.35	-2.72	-2.25

Notes: Entries are averages of t-values. $CIPS^c$ is based on individual CADF regressions with l lags of differences including an intercept only, while $CIPS^{c,\tau}$ is based on CADF regressions including an intercept and trend. Critical values for $N = 30$ and $T = 50$ are tabulated in Pesaran (2005). They are -2.30/-2.16/-2.08 for the 1%/5%/10% level of significance in the intercept only case, and -2.78/-2.65/-2.58 for the intercept and trend case. $N = 30$ and $T = 50$.

The $CIPS$ statistic is not smaller than any of the critical values corresponding to the 1%, 5% or 10% level of significance for all specifications. The case when $l = 1$ in the trend and intercept model is an exception. Here, the test indicates stationarity at the 5% level of significance. Otherwise, the outcomes are not very sensitive to the choice of number of lagged differences l. Thus, on the basis of the common unobserved factor assumption for the error process, the Pesaran test gives indication of non-stationarity of hours worked.

Phillips and Sul (2003, PS)

PS also assume that cross section dependence arises from a single common factor in u_{it}. The errors u_{it} follow the same data generating process as in equation (3.6). Similar to Pesaran, the idiosyncratic errors ϵ_{it} are i.i.d with variance σ_i^2, the factor loadings are non-stochastic and the common factor f_t is i.i.d. $\mathcal{N}(0, 1)$.

The idea of PS is to remove this common factor effect by pre-multiplying the original data with a projection matrix $\widehat{F}_\lambda$, thereby eliminating cross section dependence. The projection matrix $\widehat{F}_\lambda$ is obtained by an orthogonalization procedure that builds on a moment based method for estimating the factor loadings λ_i and the covariance matrix Σ_ϵ of the idiosyncratic errors.[17] Following the terminology of Jang and Shin (2005), this proceeding will be denoted projection de-factoring.

The transformed data $h_{it}^+ = \widehat{F}_\lambda h_{it}$ is then used to perform individual ADF regressions. Since h_{it}^+ are asymptotically uncorrelated across i, standard panel unit root tests with the de-factored data are feasible.

PS propose combining p-values of the univariate ADF regressions with the de-factored data to construct meta-statistics just as in Choi (2001) or DHT to test for unit roots in the panel.[18] The first test statistic is a Fisher type and given by[19]

$$P = -2 \sum_{i=1}^{N-1} \ln(p_i) \tag{3.9}$$

while the second statistic is an inverse normal test, denoted

$$Z = \frac{1}{\sqrt{N}} \sum_{i=1}^{N-1} \Phi^{-1}(p_i) \tag{3.10}$$

Once again, p_i defines the p-values of the univariate ADF tests with de-factored data and Φ^{-1} is the inverse of the standard normal distribution. For fixed N and

[17] Cf. PS, p. 237, for details on the orthogonalization procedure.

[18] In fact, PS propose two additional statistics to test for a homogeneous unit root in the panel that build directly on the coefficient estimates of the individual autoregressive parameters. PS refer to them as G tests. However, as PS demonstrate by means of simulation experiments that the P and Z test have considerably greater power than the G tests, they are not pursued here.

[19] The sums of the test statistics go over i to $N - 1$ since the transformation due to removal of the cross section dependence in the limit reduces the panel to dimension $N - 1$. Cf. PS, p. 238.

as $T \to \infty$, P converges to a $\chi^2_{2(N-1)}$ distribution and Z to a standard normal distribution.

PS provide guidance to the small sample performance of their proposed tests via Monte Carlo experiments. It is shown that the tests have good size and power properties even in cases were $N = 10$ and $T = 50$.[20] The results for the PS panel test for a homogeneous unit root in average hours worked are reported in table 3.6. Note that the test results rely on a balanced panel.

Table 3.6: Results of the PS test

	Intercept	Intercept and trend
Fisher P test	17.09	33.93
P-value	0.99	0.91
Inverse normal Z test	7.57	2.62
P-value	1.00	1.00

Notes: Computational work was performed in *GAUSS*. An according *GAUSS* code is available from Donggyu Sul. Here, the lag order of the univariate ADF regressions is chosen based on the top-down method. The maximum number of lags was set to 10. $N = 24$ and $T = 56$.

Both the P and the Z statistic strongly imply not to reject the unit root hypothesis for the intercept only as well as the intercept and trend specification.

So far, the test of Pesaran and PS failed to reject the unit root hypothesis for hours worked when allowing a single factor structure in the composite error term. The next section investigates whether there is more than one factor causing the dependence pattern of the data.

Moon and Perron (2004, MP)

The MP test for a homogeneous unit root is similar to the PS test in that it also removes dependency that arises from common factors by projection de-factoring. Yet it differs from the PS proceeding mainly in two ways. First, it allows cross section dependence to originate from more than one common factor. Secondly, the derivation of the projection matrix differs from the PS method. MP estimate the factor loadings λ_i, which are required to obtain the projection matrix, by a principal component estimation scheme.

MP assume that the error term from equation (3.3) follows

[20] Jang and Shin (2005) report an experimental size of 9.7% at the 5% nominal level for the PS panel unit root procedure and a sample with $N = 25$ and $T = 50$. However, their experiment is not strictly comparable to the PS experiment since Jang and Shin (2005) base statistics on simple averages of t-values instead of considering the P and Z statistics.

$$u_{it} = \lambda_i' \mathbf{f}_t + e_{it} \tag{3.11}$$

where in this case $\mathbf{f}_t$ is a $(K \times 1)$ vector of common unobserved factors and λ_i is the corresponding $(K \times 1)$ vector of factor loadings. Similar to the assumptions of Pesaran and PS, the individual specific error components e_{it} and the common factors $\mathbf{f}_t$ follow stationary and invertible $MA(\infty)$ processes that are independent of each other.[21] In the unit root case, $\phi_i = 1$ in equation (3.3) and this implies that the factors and idiosyncratic components integrate to $\sum_{s=1}^{t} \mathbf{f}_s$ and $\sum_{s=1}^{t} e_{is}$, respectively. By assumptions, MP allow the non-stationary factors to cointegrate while cointegration among the integrated idiosyncratic errors is excluded.

MP's testing procedure is summarized as follows. In a first step, under the null hypothesis of a homogeneous unit root in equation (3.3), the pooled OLS estimator $\widehat{\phi}_{pool}$ of the autoregressive coefficient is obtained. This estimator is used to construct an estimate of the composite error terms $\widehat{u}_{it} = h_{it} - \widehat{\phi}_{pool} h_{it-1}$ and by means of principal components analysis, an estimate of the factor loadings $\widehat{\Lambda} = (\widehat{\lambda}_1, ..., \widehat{\lambda}_N)'$ is attained. The $(N \times K)$ matrix $\widehat{\Lambda}$ is then utilized to construct the projection matrix $Q_{\widehat{\Lambda}} = I_N - \widehat{\Lambda}(\widehat{\Lambda}'\widehat{\Lambda})^{-1}\widehat{\Lambda}'$ for removing common factor effects from the original data. However, this procedure requires knowledge of the number of common factors K. In practice, this is not the case and the number of factors needs to be estimated. For these purposes, MP suggest using the information criteria of Bai and Ng (2002) which necessitate the setting of a maximal number K^{max} of factors.

With the projection matrix at hand, MP propose the following modified pooled estimator of the de-factored data:[22]

$$\widehat{\rho}^*_{pool} = \frac{tr(H_{-1} Q_{\widehat{\Lambda}} H') - NT\widehat{\gamma}^N_e}{tr(H_{-1} Q_{\widehat{\Lambda}} H'_{-1})} \tag{3.12}$$

In equation (3.12), $tr(.)$ is the trace operator and $\widehat{\gamma}^N_e$ an estimator of the cross-sectional average of the one-sided long-run variance of the idiosyncratic errors e_{it} in (3.11) and is meant to account for serial correlation in the transformed idiosyncratic errors e_{it}.

MP recommend looking at the following t-statistics for testing the homogeneous unit root hypothesis against the heterogeneous alternative:

$$t^*_a = \frac{\sqrt{N}T(\widehat{\rho}^*_{pool} - 1)}{\sqrt{2\widehat{\varphi}^4_e / \widehat{\omega}^4_e}} \tag{3.13}$$

[21]Cf. MP, pp. 84 for the full set of assumptions.

[22]The vectors $h_i = (h_{i2}, ..., h_{iT})$ and $h_{i,-1} = (h_{i1}, ..., h_{iT-1})$ have been horizontally concatenated to the matrices H and H_{-1}.

$$t_b^* = \sqrt{N}T(\widehat{\rho}_{pool}^* - 1)\sqrt{\frac{1}{NT^2}tr(H_{-1}Q_{\widehat{\Lambda}}H'_{-1})\left(\frac{\widehat{\omega}_e}{\widehat{\varphi}_e^2}\right)} \qquad (3.14)$$

Equation (3.13) and (3.14) involve estimators of the long-run variances $\omega_{e_i}^2$ of e_{it}, where $\widehat{\omega}_e^2$ is an estimator for the cross sectional average of $\widehat{\omega}_{e_i}^2$ and $\widehat{\varphi}_e^4$ a cross sectional average of $\widehat{\omega}_{e_i}^4$. As MP note, averaging the individual specific long-run variances should remove some of the uncertainty inherent in estimation of long-run variances and improve unit root testing over univariate counterparts. However, bias in the estimation of these variances will not be removed through averaging.

MP show that under the null hypothesis, the statistics t_a^* and t_b^* converge to a standard normal distribution as $N \to \infty$ and $T \to \infty$ with $N/T \to 0$.

MP also demonstrate that their tests have no power against local alternatives in the case where heterogeneous deterministic trends exit in the data. Therefore, the tests should not be used if one assumes linear time trends in the deterministic components of the data generating process of (3.3).

The simulation experiments of MP confirm the good power and size results of the t-tests, especially when $T = 300$. They also conclude that the number of factors is estimated imprecisely for a small number of cross sections ($N = 10$). If the number of cross-sections is at least 20, the number of factors can be estimated with high precision.[23] Since MP do not consider samples with less than 100 time series observations in their simulation, the applicability of the MP procedure for the present panel data of hours worked is assessed with the help of the experiments of Gengenbach et al. (2004) and Gutierrez (2006).

From the tables of Gengenbach et al. (2004)[24] it can be seen that both statistics of MP have rejection frequencies lower than the nominal size if $T = 50$ and $N = 10$ or $N = 50$, irrespective of whether the non-stationarity originates from the idiosyncratic components or common factors. For the near unit root case, the power of the MP test is good if $N > 10$.

Although Guiterrez (2005) concludes that the MP tests in general show good size and power for various values of N and T and different model specifications, the results also indicate that for $N = 20$ and $T = 50$ the t_a^* is undersized while t_b^* has in general rejection frequencies higher than the nominal size.[25]

As mentioned above, in applied work the number of common factors needs to be estimated. In conducting the MP test for hours worked, the seven information criteria for estimating the number of factors that are due to Bai and Ng (2002) are considered.[26]

[23] This is due to Bai and Ng (2002).
[24] Cf. Gengenbach et al. (2004), pp. 26. The comments refer to the simulation results assuming a single common factor.
[25] CF. Guiterrez (2005), p. 11, table 1.
[26] Cf. MP, pp. 93, or Bai and Ng (2002), pp. 201, for a detailed description of these information

In the application of the information criteria to the logarithm hours worked in the balanced OCED panel, congruent results are obtained when setting $K^{max} = 6$ and focusing on IC_{p2} and BIC_3 in which case it is recommended to assume one common factor. The latter is the preferred criterion of MP in small samples.[27] However, for robustness check, the case $K = 2$ and $K = 6$ is also considered below. Under the assumption of one common factor, the data generating processes of the MP and PS tests are the same and the only difference lies in the treatment of the common unobserved factor in the estimation strategy.

As in MP, the long-run variances are estimated using the Andrews and Monahan (1992) method. Tabel 3.7 shows results of the MP panel unit root test for hours worked.

Table 3.7: Results of the MP test

K	1	2	6
t_a^*	-0.09	-0.10	-0.10
P-value	0.46	0.46	0.46
t_b^*	-9.71	-9.16	-11.72
P-value	0.00	0.00	0.00

Notes: Computational work was performed in MATLAB. A MATLAB code is available from Benoit Perron. Intercept only case. $N = 24$ and $T = 56$.

The t_a^* statistic implies that the null of a homogenous unit root in the panel for the assumption of one, two or six common factors should not be rejected. In contrast, the t_b^* statistics reject for all specifications. When considering the results of both test statistics, the conclusions to be drawn are highly contradictory. There is some evidence that the t_b^* statistic is oversized in small samples in Gutierrez (2006). Since there is no general guidance as to which t-statistic should be preferred in applied settings, the MP test alone offers no direction in the present analysis. However, the results of the previous tests suggest putting more confidence into the t_a^* in the present estimation and concluding that the MP panel unit root test also fails to reject the null hypothesis.

Bai and Ng (2004, BN)

Instead of treating the common factors as a nuisance, they become a direct object of further investigation in the BN testing framework. BN build on assump-

criteria.

[27] In the absence of a formal criterion, MP and Bai and Ng (2002) set $K^{max} = 8$ in their simulation studies for all values of N and T.

tions very similar to those of MP. BN call their testing procedure Panel Analysis of Non-stationarity in Idiosyncratic and Common components (PANIC). This acronym aptly summarizes the intended aim: While allowing the data to be driven by one or more common factors and idiosyncratic components, the time series properties of these elements are assessed separately without a priori assumptions on whether these elements are stationary or integrated. Since the panel unit root tests give evidence for the non-stationarity of hours worked so far, the BN test is employed to determine the source of non-stationarity. On the grounds of the previous analysis, it is assumed throughout that hours worked are driven by idiosyncratic elements and a single unobserved common factor.[28]

In the presence of a single common factor, the data generating process of BN is

$$h_{it} = d_{it} + \lambda_i F_t + E_{it} \tag{3.15}$$

where the common factor F_t and the idiosyncratic errors E_{it} follow AR(1) models with a polynomial lag structure of i.i.d. shocks. Concerning the deterministic elements d_{it}, an intercept or a trend or both are allowed. If the errors E_{it} are independent across units, i.e. if the cross section dependence can be effectively represented by a common factor structure like in the Pesaran, PS and MP setting, pooled tests for unit roots in the idiosyncratic components are feasible. An appealing feature of the pooled tests is that they can be regarded as a panel test of no cross-member cointegration. The workings of the latter will be demonstrated below.

The strategy of consistently estimating the individual components of (3.15), even if some or all elements of F_t and E_{it} are integrated of order one, can be described as follows. In a first step, the h_{it}'s are differenced if the deterministics include only an intercept or are differenced and demeaned if d_{it} includes and intercept and trend.[29] As in MP, the principal component method is employed with the differenced data and the common factors, factor loadings and residuals are estimated.

In a next step, the estimates of the differenced factors and idiosyncratic error components are re-integrated à la $\widehat{x}_t = \sum_{s=2}^{t} \Delta\widehat{x}_s$ and tested separately for unit roots. Let $\widehat{E}_{it}$ and $\widehat{F}_t$ be the re-integrated estimates of the common factor and idiosyncratic components. Since $\widehat{E}_{it} = h_{it} - \widehat{\lambda}_i \widehat{F}_t$, Jang and Shin (2005) denote such a proceeding as subtraction de-factoring. For unit root testing, BN propose employing the usual t-statistics of ADF regressions in the common factor and idiosyncratic components, respectively. For the model with an intercept only, the t-statistics to test the common factor for a unit root is denoted $ADF^c_{\widehat{F}}$. If the model contains an

[28]If there is more than one common integrated factor, these factors need to be tested for cointegration in order to obtain the number of common stochastic trends. BN explain the testing strategy for this case.

[29]For details to this procedure, cf. BN, pp. 1137.

intercept and linear trend, the statistic is $ADF^{c,\tau}_{\widehat{F}}$. Accordingly, the t-statistics for individual unit root tests of the idiosyncratic components are denoted $ADF^{c}_{\widehat{E}}(i)$ and $ADF^{c,\tau}_{\widehat{E}}(i)$.

BN show that the asymptotic distribution of $ADF^{c}_{\widehat{E}}(i)$ coincides with the usual DF distribution (no intercept), while $ADF^{c}_{\widehat{F}}$ has the same limiting distribution as the DF test for the intercept only case. Furthermore, $ADF^{c,\tau}_{\widehat{F}}$ follows a DF distribution for the case with intercept and trend in the limit. However, the limiting distribution of $ADF^{c,\tau}_{\widehat{E}}(i)$ is proportional to the reciprocal of a Brownian bridge and critical values are not tabulated yet and need to be simulated.[30]

For independent E_{it}, BN propose a pooled test for unit roots in $\widehat{E}_{it}$ due to Choi (2001) that builds on combining p-values $p^{c}_{\widehat{E}}(i)$ of $ADF^{c}_{\widehat{E}}(i)$ and which is similar to the test statistics of PS and DHT.[31] However, in the more general case where some remaining independent structure in the idiosyncratic components is allowed for, it seems more advisable to directly apply the DHT test. Following equation 3.5 and the notational convention of the present paragraph, this pooled test statistic for the idiosyncratic errors will be denoted $t^{c}_{\widehat{E}}(\widehat{\rho}^{*}, \kappa)$.

Such a pooled test can also be regarded as a panel test of no cross-member cointegration since no stationary combination of the individual variables h_{it} can be obtained so that the unit root hypothesis holds for all i. On the other hand, if the common factor is integrated of order one and there are some stationary E_{it}'s, then the common factor and the stationary variables are cointegrated with vector $(1, -\lambda_i)'$. If the $t^{c}_{\widehat{E}}(\widehat{\rho}^{*}, \kappa)$ statistic rejects and all idiosyncratic components can be seen as stationary, the h_{it}'s cointegrate and the matrix $Q_{\widehat{\Lambda}}$ of MP for de-factoring the data serves as a cointegration matrix.[32]

BN enrich their work with simulation studies to investigate the small sample performance of the PANIC procedure. They conclude that the proposed tests have good power even when $N = 40$ and $T = 100$. However, these values are nearly twice as large as the panel dimension of hours worked. Jang and Shin (2005) find that tests based on the BN method have sizes close to the nominal level when $T = 50$ and $N = 25$ and power is slightly better than for the PS and MP procedure. In the case where a unit root is present in the common factor and in all idiosyncratic errors, the simulation experiments of Gengenbach et al. (2004) show that the tests of BN for the intercept only case have rejection frequencies close to the nominal size even when $N = 10$ and $T = 50$. If there is a unit root in the common factor and the idiosyncratic components are near unit root, the $ADF^{c}_{\widehat{F}}$ rejects with a frequency at the nominal level, while the pooled test has more power than $ADF^{c}_{\widehat{E}}(i)$. Although

[30]Since this is beyond the scope of the present chapter, this test will no bet considered below.
[31]Cf. Bai and Ng (2004), p.1140.
[32]Cf. Gengenbach et al. (2004), p. 13, for these points.

the PANIC method is derived for applications with large dimensional panels where a high number of cross sections permits consistent estimation of the common factors, while the large T dimension allows application of central limit theorems, there is some simulation evidence that the BN test also gives reasonable guidance in samples of moderate size.

Table 3.8: BN results for common factor and pooled idiosyncratic components

	$ADF^c_{\widehat{F}}$	$t^c_{\widehat{E}}(\widehat{\rho}^*, \kappa)$
Test statistic	-1.98	0.09
P-value	0.30	0.54

Notes: MacKinnon (1996) p-values. $N = 24$ and $T = 56$, $\kappa = 0.2$, $\widehat{\rho}* = 0.54$.

Results for the $ADF^c_{\widehat{F}}$ and $t^c_{\widehat{E}}(\widehat{\rho}^*, \kappa)$ test are shown in table (3.8). The $ADF^c_{\widehat{F}}$ test does not reject and an integrated common factor can be assumed. In addition, the pooled test of the idiosyncratic errors also fails to reject so that two conclusions can be drawn. First, the non-stationarity of hours worked for which nearly all panel tests of the previous sections found evidence, seems to originate from a common as well as country specific sources. Second, the insignificance of the $t^c_{\widehat{E}}(\widehat{\rho}^*, \kappa)$ statistic implies that the non-stationarity hypothesis can not be rejected jointly. Since testing the idiosyncratic errors jointly for a homogenous unit root amounts to a test of no cointegration, the outcomes here give evidence that there is no cointegration among the individual time series of hours worked.

The results for individual unit root tests of the country specific errors are reported in table 3.9. The insignificance of the $ADF^c_{\widehat{E}}(i)$ for all countries confirms the result of the $t^c_{\widehat{E}}(\widehat{\rho}^*, \kappa)$ test.

Columns 4 and 8 of table 3.9 show the impact of the common factor (Imp) in relation to the idiosyncratic component. This measure is calculated as the standard deviation of the country specific factor effect ($\widehat{\lambda}_i \widehat{F}_t$) divided by the standard deviation of the estimated errors $\widehat{E}_{it}$. This ratio is greater than one for all countries except for Spain and Japan and indicates that most of the variation in the logarithm hours worked arises from the common factor. For Spain and Japan, idiosyncratic elements are more important in driving the evolution of hours worked.

To get a visual impression of the decomposition of the BN procedure, the graphs of the estimated factor component $\widehat{\lambda}_i \widehat{F}_t$ and estimated country specific elements $\widehat{E}_{it}$ for Japan, Germany and Norway are depicted in the figures 3.1 to 3.3. These

Table 3.9: Results of BN's test for unit roots in idiosyncratic errors

	$ADF^c_{\widehat{E}}(i)$	p-val.	Imp		$ADF^c_{\widehat{E}}(i)$	p-val.	Imp
Australia	-0.26	0.59	3.02	Ireland	0.03	0.69	4.62
Austria	-1.10	0.24	3.84	Iceland	-0.09	0.65	3.68
Belgium	-0.15	0.63	3.33	Italy	-0.72	0.40	5.88
Canada	-0.11	0.64	3.17	Japan	-1.09	0.25	0.60
Switzerland	-0.71	0.41	9.18	Luxembourg	0.10	0.71	4.92
Germany	0.52	0.83	4.81	Netherlands	0.06	0.70	4.88
Denmark	0.16	0.73	3.26	Norway	-1.52	0.12	10.72
Spain	-1.92	0.05	0.95	New Zealand	-0.18	0.62	3.67
Finland	0.45	0.81	4.82	Portugal	0.48	0.82	5.38
France	-0.57	0.47	6.01	Sweden	-0.08	0.65	2.20
United Kingdom	-0.88	0.33	9.73	Turkey	-1.62	0.10	5.57
Greece	-1.62	0.10	5.57	USA	-0.38	0.54	3.06

Notes: Computational work was performed in MATLAB. A corresponding MATLAB code is available from Serena Ng. MacKinnon (1996) p-values. Lag length was chosen due to the minimum of the modified Akaike information criterion. $Imp = \frac{St.Dev(\widehat{\lambda}_i \widehat{F}_t)}{St.Dev(\widehat{E}_{it})}$. $N = 24$ and $T = 56$.

three countries were chosen because they show a low, medium and high impact of the common factor in relation to their country specific effects. Also shown in each graph is the sum of the estimated factor component and the country specific element. This sum corresponds to the deviation of hours worked around its mean level, i.e. the lines referred to as "Hours" in the figures 3.1 to 3.3 show $h_{it} - \widehat{d_i} = \widehat{\lambda}_i \widehat{F}_t + \widehat{E}_{it}$, where $\widehat{d_i}$ is the estimated intercept of the relation given by equation 3.15.

For Japan, it can be seen that the time path of hours worked is dominated by country specific influences, while the German evolution of hours worked is marked by the downward trend of the common factor, which is overlaid by the idiosyncratic component. In contrast, Norway shows a development of hours that is mainly in line with the common factor and exhibits country specific influences with a cyclical movement around the factor trend.

It is important to notice that actual hours worked and the common factor do *not* describe some kind of equilibrium as the idiosyncratic errors, defined as the residual from the linear relationship between the country specific factor influence and actual hours, are *non-stationary.* Rather, the decomposition shows that, although a common integrated factor can substantially influence the development of hours (e.g. Norway), the persistent behavior of this variable is not solely due to a common stochastic trend but also characterized by persistent country specific determinants.

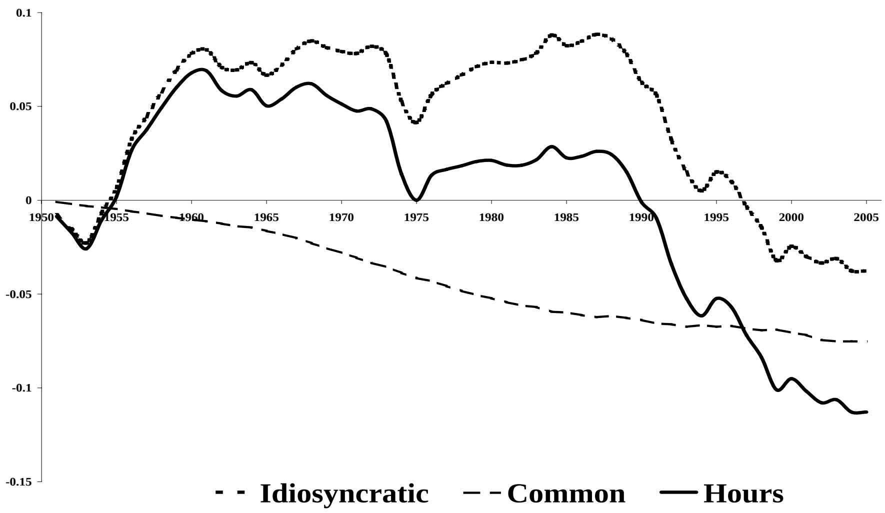


Figure 3.1: PANIC decomposition for Japan

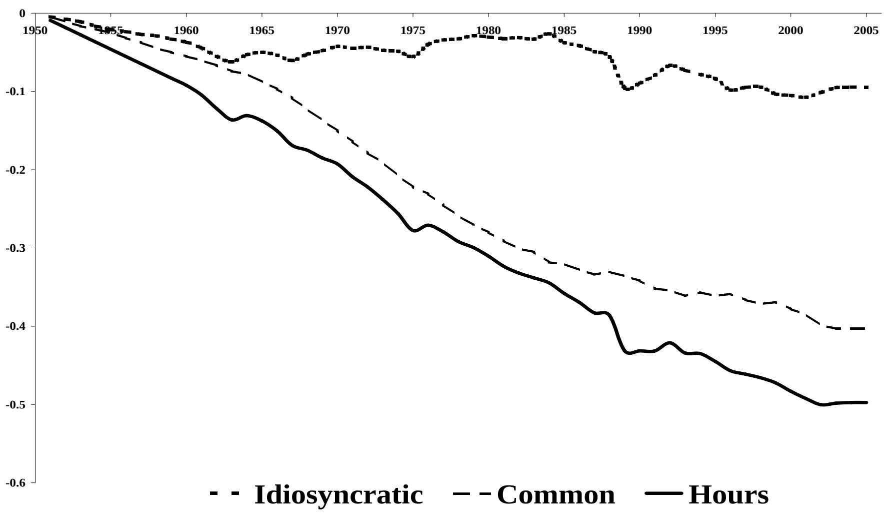


Figure 3.2: PANIC decomposition for Germany

3.5 Summary and conclusion

The results of the present analysis show that evidence in favor of the non-stationarity hypothesis of hours worked per employee in the OECD countries is vast. Simple ADF

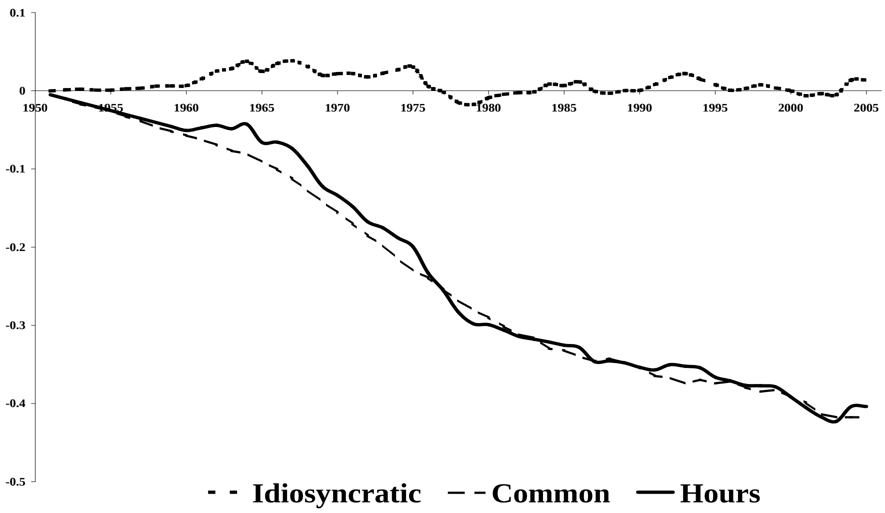


Figure 3.3: PANIC decomposition for Norway

tests for individual countries are not able to reject the unit root hypothesis both in the intercept and intercept and linear trend model.[33] However, univariate unit root tests lack power against local alternatives in finite samples. This is one of the reasons why researchers sometimes doubt the implications of these tests when the alternative under consideration is a stationary but persistent series. Panel unit root tests are able to substantially increase power over univariate tests if the panel data is cross sectionally independent. In the presence of cross section dependence, this needs to be accounted for in order to retain the high power properties of these tests.

In the present analysis, it was first shown that the cross sectional observations for hours worked are characterized by heterogeneous cross section dependence. Second, panel unit root tests of the second generation that account for this dependence were applied. For robustness reasons, five different panel unit root tests were conducted and only under rare circumstances, rejections of the homogeneous unit root hypothesis were observed.

Besides diagnosing the property of non-stationarity as fairly robust, the data revealed further interesting features. When allowing hours worked to be influenced by a common factor and applying the PANIC procedure of Bai and Ng (2004) to decompose the factor structure, the following stands out: Non-stationarity of hours worked originates both from an integrated common factor and an integrated idiosyn-

[33]New Zealand is the solitary exception, where the individual ADF test rejects at the 10% level of significance in the intercept only specification.

cratic component. Since this holds for all countries, it implies that the individual time series are not cointegrated along the cross sectional dimension.

However, the empirical analysis here refers to rather abstract and intangible concepts such as "common unobserved factors" and ' "persistent idiosyncratic components" which help to empirically model the data properties quite well, but give no further insights into economic relations. The cited literature in the introduction to this chapter rudimentarily illustrates that various candidates for persistently influencing the aggregate labor supply and for explaining cross country differences have been proposed and also empirically investigated. Further work in this direction should follow.

Based on the results of the present analysis, it is strongly recommended to transform hours worked to obtain a stationary time series if one employs econometric methods that rely on standard asymptotic theory or to use the analytical tools that have been developed for investigating non-stationary variables if one considers the level of hours worked.

Appendix

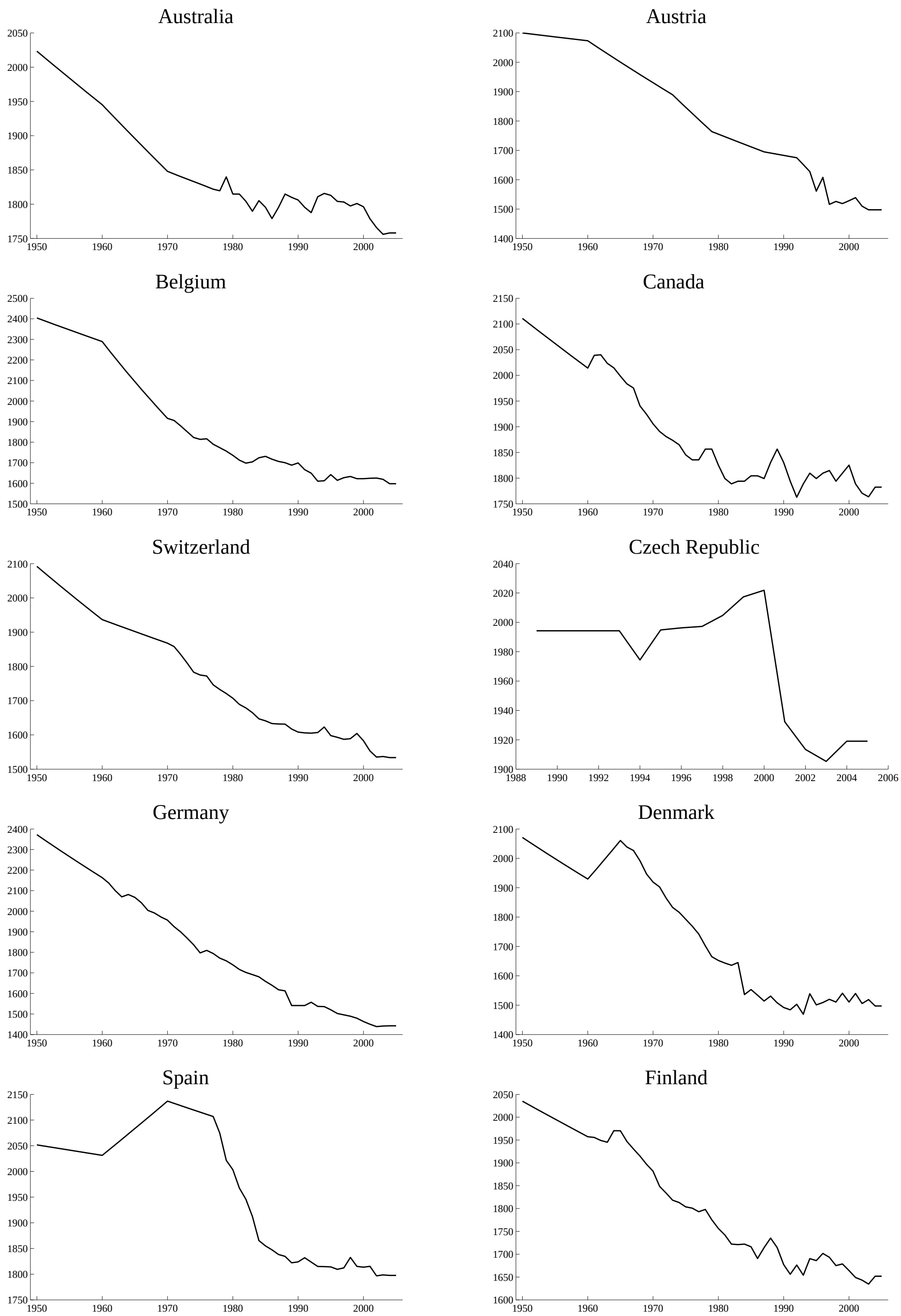


Figure 3.4: Annual hours worked per worker (I)

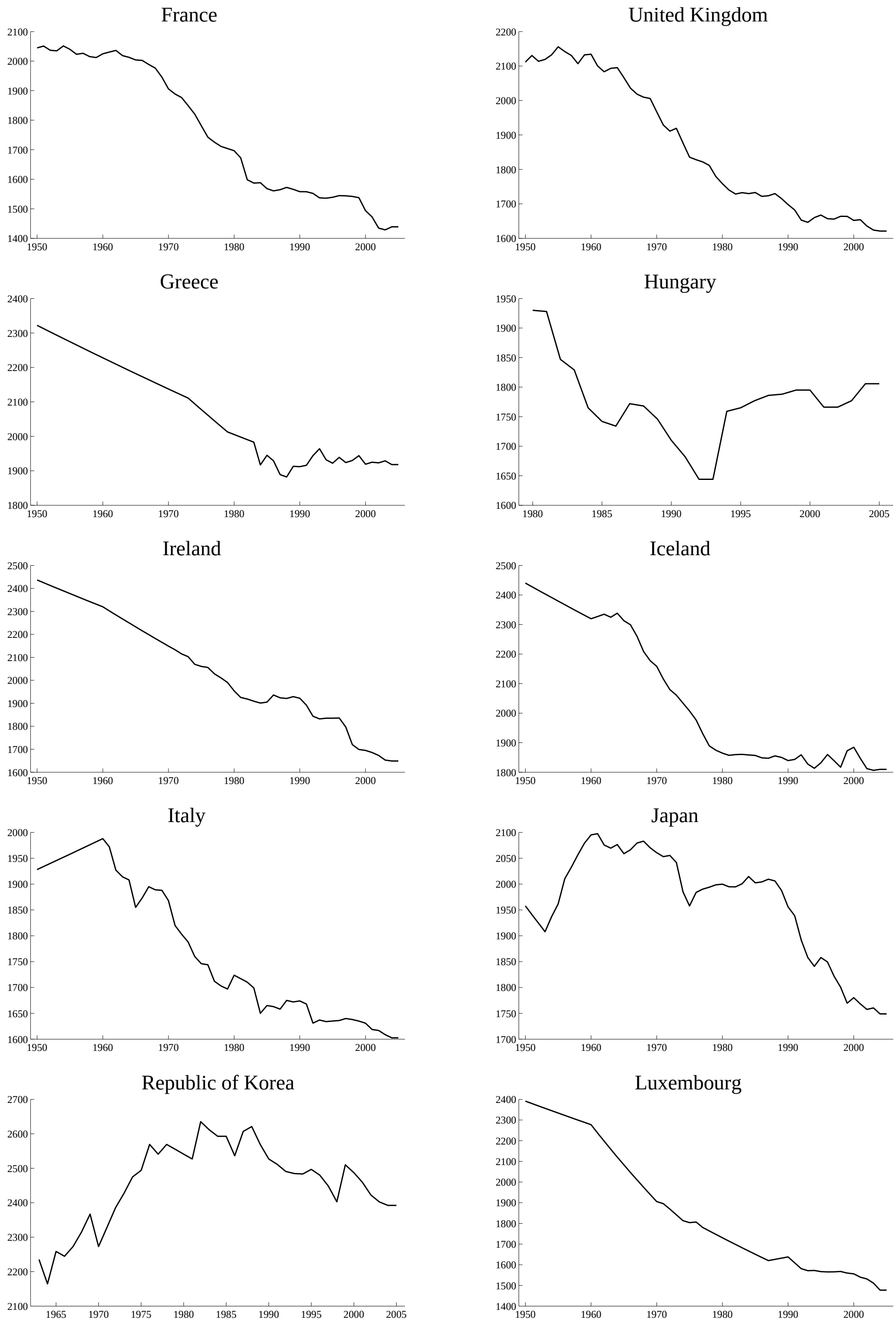


Figure 3.5: Annual hours worked per worker (II)

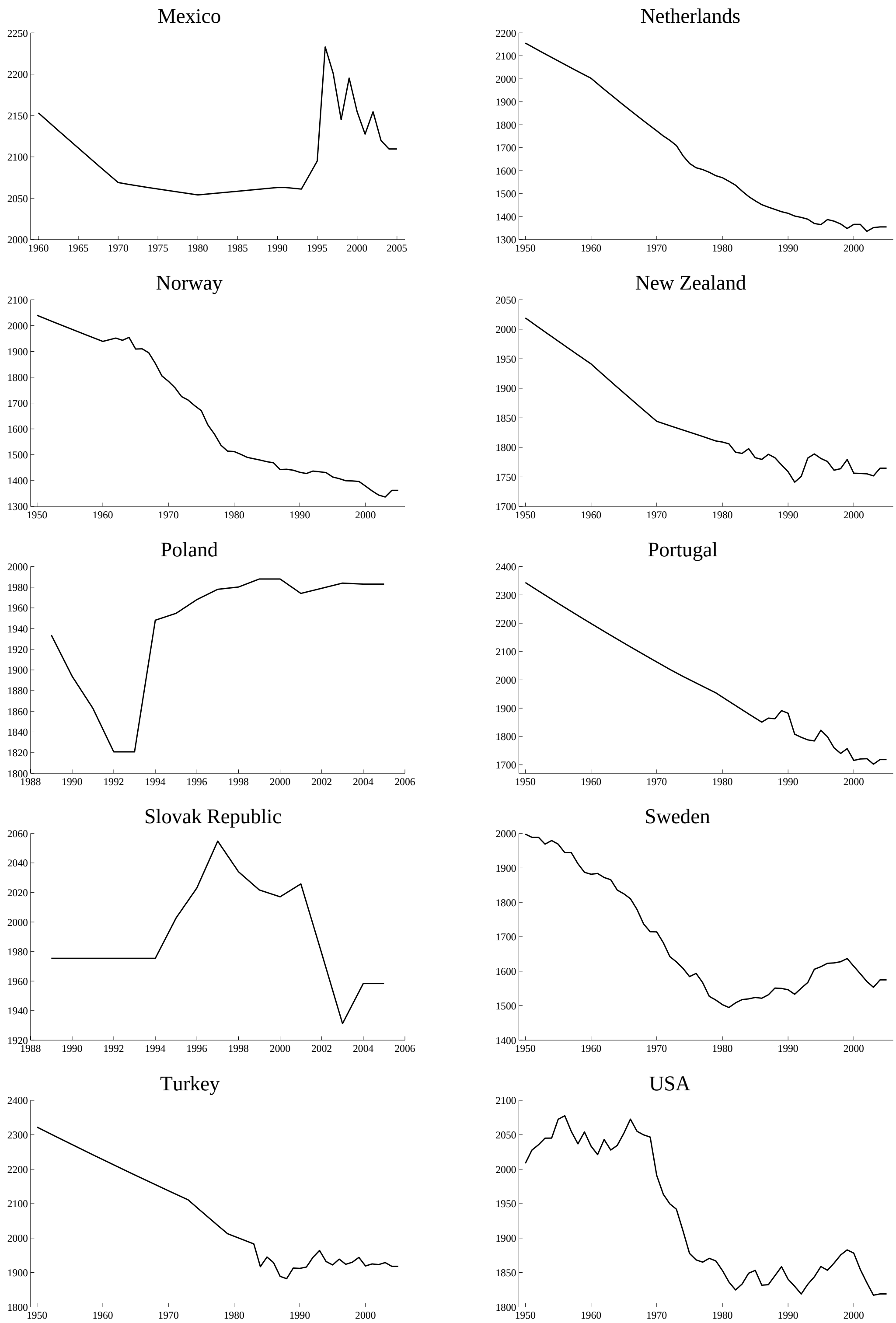


Figure 3.6: Annual hours worked per worker (III)

Bibliography

Acemoglu, D. (2003): "Labor- and Capital-Augmenting Technical Change," *Journal of the European Economic Association*, 1, 1–37.

Agell, J., T. Lindh, and H. Ohlson (1997): "Growth and the Public Sector: A Critical Review Essay," *European Journal of Political Economy*, 13, 33–52.

Aghion, P., and S. N. Durlauf (eds.) (2005): *Handbook of Economic Growth*, vol. 1. Elsevier Science B. V.

Aghion, P., and P. Howitt (1992): "A Model of Growth Through Creative Destruction," *Econometrica*, 60, 232–351.

——— (1998): *Endogeneous Growth Theory*. The MIT Press, Cambridge.

Ahn, S., and P. Hemmings (2000): "Policy Influences on Economic Growth in OECD Countries: An Evaluation of the Evidence," OECD Economics Working Paper No. 246.

Alesina, A., E. Glaeser, and B. Sacerdote (2005): "Work and Leisure in the U.S. and Europe: Why So different?," NBER Working Paper No. 11278.

Allen, F., and D. Gale (2000): *Comparing Financial Systems*. The MIT Press, Cambridge.

Andrews, D. W. K., and J. C. Monahan (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60, 953–966.

Ang, A., and G. G. Bekaert (2006): "Stock Return Predictability: Is it There?," Review of Financial Studies, forthcoming.

Apel, M., and P. Jansson (1999): "System Estimates of Potential Output and the NAIRU," *Empirical Economics*, 24, 373–388.

Ash, J. C. K., D. J. Smyth, and S. M. Heravi (1998): "Are OECD Forecasts Rational and Useful?: A Directional Analysis," *International Journal of Forecasting*, 14, 381–391.

Ashiya, M. (2006): "Testing the Rationality of Forecast Revisions Made by the IMF and the OECD," *Journal of Forecasting*, 25, 25–36.

BAGHLI, M., C. CAHN, AND J. P. VILLETELLE (2006): "Estimating Potential Output with a Production Function for France, Germany and Italy," in *Convergence or Divergence in Europe? Growth and Business Cycles in France, Germany and Italy*, ed. by O. De Bandt, H. Herrmann, and G. Parigi. Springer Berlin, Heidelberg.

BAI, J., AND S. NG (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.

——— (2004): "A PANIC Attack on Unit Roots and Cointegration," *Econometrica*, 72, 1127–1177.

BALDWIN, R. (2000): "Trade and Growth: Still Disagreement About the Relationship," OECD Economics Department Working Papers No. 264.

BANERJEE, A. (1999): "Panel Data Unit Roots and Cointegration: An Overview," *Oxford Bulletin of Economics and Statistics*, 61, 607–629.

BANERJEE, A., M. MARCELLINO, AND C. OSBAT (2005): "Testing for PPP: Should we Use Panel Methods?," *Empirical Economics*, 30, 77–91.

BARRO, R. J. (1991): "Economic Growth in a Cross Section of Countries," *Quarterly Journal of Economics*, 106, 407–443.

——— (1997): *Determinants of Economic Growth: A Cross-Country Empirical Study*. MIT Press, Cambridge.

BARRO, R. J., AND X. SALA-I-MARTIN (2001): *Economic Growth*. The MIT Press, Cambridge, 3 edn.

BASSANETTI, A., J. DÖPKE, R. TORRINI, AND R. ZIZZA (2006): "Capital, Labour and Productivity: What Role Do They Play in the Potential GDP Weakness of France, Germany and Italy?," in *Convergence or Divergence in Europe? Growth and Business Cycles in France, Germany and Italy*, ed. by O. De Bandt, H. Herrmann, and G. Parigi, Berlin, Heidelberg. Springer.

BASSANINI, A., S. SCARPETTA, AND P. HEMMINGS (2001): "Economic Growth: The Role of Policies and Institutions. Panel Data Evidence from OECD Countries," OECD Working Paper No. 283.

BATISTA, C., AND J. ZALUENDO (2004): "Can the IMF's Medium-Term Growth Projections be Improved?," IMF Working Paper No. 04/2003.

BEFFY, P. O., P. OLLIVAUD, P. RICHARDSON, AND F. SÉDILLOT (2006): "New OECD Methods for Supply-Side and Medium-Term Assessments: A Capital Services Approach," OECD Economics Department Working Papers No. 482.

BENES, J., AND P. N'DIAYE (2004): "A Multivariate Filter for Measuring Potential Output and the NAIRU: Application to The Czech Republic," IMF Working Paper No. 04/45.

BERNARD, A. N., AND S. N. DURLAUF (1996): "Interpreting Tests of the Convergence Hypothesis," *Journal of Econometrics*, 71, 161–173.

Binder, M., and S. Bröck (2006): "On the Relation Between Investment and Economic Growth: New Cross-Country Empirical Evidence," Mimeo, Goethe University Frankfurt.

Binder, M., and M. H. Pesaran (1999): "Stochastic Growth Models and Their Econometric Implications," *Journal of Economic Growth*, 4, 139–183.

Bjørnland, H, C., L. Brubakk, and A. S. Jore (2005): "The Output Gap in Norway - A Comparison of Different Methods," *Economic Bulletin (Norges Bank)*, 76, 90–100.

Blanchard, O. J. (2004): "The Economic Future of Europe," MIT Economics Working Paper No. 04-04.

Blanchard, O. J., and D. Quah (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *The American Economic Review*, 79, 655–673.

Bleaney, M., and A. Nishiyama (2002): "Explaining Growth: A Contest Between Models," *Journal of Economic Growth*, 7, 43–56.

Breitung, J. (2000): "The Local Power of Some Unit Root Tests for Panel Data," in *Nonstationary Panels, Panel Cointegration, and Dynamic Panels, Advances in Econometrics, Vol. 15*, ed. by B. Baltagi. JAI, Amsterdam.

Breitung, J., and S. Das (2005): "Panel Unit Root Tests Under Cross Sectional Dependence," *Statistica Neerlandica*, 59, 1–20.

Breitung, J., and M. H. Pesaran (2005): "Unit Roots and Cointegration in Panels," Deutsche Bundesbank Discussion Paper No. 42/2005.

Brown, B. W., and S. Maital (1981): "What Do Economists Know? An Empirical Study of Experts' Expectations," *Econometrica*, 49, 491–504.

Butler, L. (1996): "A Semi-Structural Method to Estimate Potential Output: Combining Economic Theory with a Time-Series Filter," Technical Report No. 77, Bank of Canada, Ottawa.

Carnot, N., V. Koen, and B. Tissot (2005): *Economic Forecasting.* Palgrave MacMillan, New York.

Caselli, F., and J. Feyrer (2007): "The Marginal Product of Capital," *Quarterly Journal of Economics*, 122, 535–568.

Cellini, R. (1997): "Implications of Solow's Growth Model in the Presence of a Stochastic Steady State," *Journal of Macroeconomics*, 19, 135–153.

Cerra, V., and S. C. Saxena (2000): "Alternative Methods of Estimating Potential Output and the Output Gap: An Application to Sweden," IMF Working Paper No. 00/59.

Chagny, O., and J. Döpke (2001): "Measures of the Output Gap in the Euro-Zone: An Empirical Assessment of Selected Methods," Kiel Working Paper No. 1053, Institut für Weltwirtschaft, Kiel.

Choi, I. (2001): "Unit Root Tests for Panel Data," *Journal of International Money and Finance*, 20, 249–272.

——— (2004): "Nonstationary Panels," forthcoming in: Palgrave Handbooks of Econometrics, Vol. 1.

Christiano, L., M. Eichenbaum, and R. Vigfusson (2003): "What Happens After a Technology Shock?," Federal Reserve Board, International Finance Discussion Papers No. 768.

Clark, T. E. (1999): "Finite-Sample Properties of Tests for Equal Forecast Accuracy," *Journal of Forecasting*, 18, 489–504.

Claus, I. (2003): "Estimating Potential Output for New Zealand," *Applied Economics*, 35, 751–760.

Clements, M. P. (2005): *Evaluating Econometric Forecasts of Economic and Financial Variables*. Palgrave MacMillan, New York.

Clements, M. P., F. Joutz, and H. O. Stekler (2007): "An Evaluation of the Forecasts of the Federal Reserve: A Pooled Approach," *Journal of Applied Econoemtrics*, 22, 121–136.

Coe, D. T., and E. Helpman (1995): "International R&D spillovers," *European Economic Review*, 39, 859–887.

Congressional Budget Office (CBO) (2004): "A Summary of Alternative Methods for Estimating Potential GDP," Background Paper, available at www. cbo. gov/ftpdoc. cfm?index=5191&type=1.

Cooper, I., and R. Priestley (2006): "Time-Varying Risk Premia and the Output Gap," Norwegian School of Management (BI) - Department of Financial Economics and Norwegian School of Mana.

Cotis, J. P., J. Elmeskov, and A. Mourougane (2004): "Estimates of Potential Output: Benefits and Pitfalls from a Policy Perspective," in *The Euro Area Business Cycle: Stylized Facts and Measurement Issues*, ed. by L. Reichlin. CEPR, London.

Davies, A., and K. Lahiri (1995): "A New Framework for Analyzing Survey Forecasts Using Three-Dimensional Panel Data," *Journal of Econometrics*, 68, 205–227.

De la Croix, D., T. Lindh, and B. Malmberg (2006): "Growth and Longevity from the Industrial Revolution to the Future of an Aging Society," mimeo.

De La Fuente, A. (2003): "Convergence Equations and Income Dynamics: The Sources of OECD Convergence," *Economica*, 70, 655–671.

De Long, J. B., and L. H. Summers (1991): "Equipment Investment and Economic Growth," *Quarterly Journal of Economics*, 106, 445–502.

De Masi, P. R. (1997): "IMF Estimates of Potential Output: Theory and Practice," IMF Working Paper No. 97/177.

Dees, S., F. Di Mauro, M. H. Pesaran, and L. V. Smith (2005): "Exploring the International Linkages of the Euro Area. A Global VAR Analysis," CESifo Working Paper No. 1425.

Demetrescu, M., U. Hassler, and A. I. Tarcolea (2006): "Combining Significance of Correlated Statistics with Application to Panel Data," *Oxford Bulletin of Economics and Statistics*, 68, 647–663.

Denis, C., K. McMorrow, and W. Röger (2002): "Production Function Approach to Calculating Potential Growth and Output Gaps: Estimates for the EU Member States and the US," European Commission Economic Papers No. 176.

Diebold, F. X., and R. S. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13, 134–144.

Döpke, J., and U. Fritsche (2006): "When Do Forecasters Disagree? An Assessment of German Growth and Inflation Forecast Dispersion," *International Journal of Forecasting*, 22, 125–135.

Dupasquier, C., A. Guay, and P. St-Amant (1997): "A Comparison of Alternative Methodologies for Estimating Potential Output and the Output Gap," Bank of Canada Working Paper No. 97-5.

Durlauf, S. N., P. A. Johnson, and J. R. W. Temple (2005): "Growth Econometrics," in *Handbook of Economic Growth*, ed. by P. Aghion, and S. N. Durlauf, vol. 1, chap. 8, pp. 555–677. Elsevier Science B. V.

Durlauf, S. N., and D. T. Quah (1999): "The New Empirics of Economic Growth," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and M. Woodford, vol. 1, chap. 4, pp. 235–308. Elsevier Science B. V.

Elliott, G., T. J. Rothenberg, and J. H. Stock (1996): "Efficient Tests For An Autoregressive Unit Root," *Econometrica*, 64, 813–836.

Engel, R. F., and C. W. J. Granger (1987): "Co-Integration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 55, 251–276.

Entorf, H. (1997): "Random Walks With Drifts: Nonsense Regression and Spurious Fixed-effects Estimation," *Journal of Econometrics*, 80, 287–296.

European Commission (1997): "Council Regulation No. 1466/97," ECOFIN Council, Brussels.

——— (2005): "Specifications on the Implementation of the Stability and Growth Pact and Guidelines on the Format and Content of Stability and Convergence Programmes," ECOFIN Council, Brussels.

Fabiani, S., and R. Mestre (2004): "A System Approach for Measuring the Euro Area NAIRU," *Empirical Economics*, 29, 311–341.

Feldstein, M. (1983): *Inflation, Tax Rules and Capital Formation*. University of Chicago Press, Chicago.

Fernandez, C., E. Ley, and M. F. J. Steel (2001): "Model Uncertainty in Cross-Country Growth Regressions," *Journal of Applied Econometrics*, 16, 563–576.

Fildes, R., and H. O. Stekler (2002): "The State of Macroeconomic Forecasting," *Journal of Macroeconomics*, 24, 435–468.

Florax, R. J. G. M., H. L. F. De Groot, and R. Heijungs (2002): "The Empirical Economic Growth Literature," Tinbergen Institute Discussion Paper No. 040/3.

Fölster, S., and M. Henrekson (2001): "Growth Effects of Government Expenditure and Taxation in Rich Countries," *European Economic Review*, 45, 1501–1520.

——— (2006): "Growth Effects of Government Expenditure and Taxation in Rich Countries: A Reply," *European Economic Review*, 50, 219–221.

Fritsche, U., and C. Logeay (2002): "Structural Unemployment and the Output Gap in Germany: Evidence from an SVAR Analysis Within a Hysteresis Framework," DIW Berlin Discussion Paper No. 312.

Galí, J. (1999): "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?," *American Economic Review*, 89, 249–271.

Garratt, A., K. Lee, M. H. Pesaran, and Y. Shin (2006): *Global and National Macroeconometric Modelling: A Long-Run Structural Approach*. Oxford University Press, Oxford.

Gengenbach, C., F. Palm, and J. P. Urbain (2004): "Panel Unit Root Tests in the Presence of Cross-Sectional Dependencies: Comparison and Implications for Modelling," Research Memoranda 040, Maastricht Research School of Economics of Technology and Organization.

Gerlach, S., and F. Smets (1999): "Output Gaps And Monetary Policy In The EMU Area," *European Economic Review*, 43, 801–812.

Giorno, C., P. Richardson, D. Roseveare, and P. van den Noord (1995): "Estimating Potential Output, Output Gaps and Structural Budget Balances," OECD Economic Department Working Papers No. 152.

Gosselin, M. A., and R. Lalonde (2006): "An Ecletic Approach to Estimating U. S. Potential GDP," *Empirical Economics*, 31, 951–975.

Granger, C. W. J. (1996): "Can We Improve the Perceived Quality of Economic Forecasts?," *Journal of Applied Econometrics*, 11, 455–473.

Groen, J. J. J., and F. Kleibergen (2003): "Likelihood-Based Cointegration Analysis in Panels of Vector Error-Correction Models," *Journal of Business & Economic Statistics*, 21, 295–318.

GROSSMAN, G. M., AND E. HELPMAN (1991): *Innovation and Growth in the Global Economy*. The MIT Press, Cambridge.

GUTIERREZ, L. (2006): "Panel Unit Roots Tests for Cross-Sectionally Correlated Panels: A Monte Carlo Comparison," *Oxford Bulletin of Economics and Statistics*, 68, 519–540.

HADRI, K. (2000): "Testing for Stationarity in Heterogeneous Panel Data," *Econometrics Journal*, 3, 148–161.

HALDRUP, N. (1998): "A Review of Econometric Analysis of $I(2)$ Variables," *Journal of Economic Surveys*, 12, 595–650.

HAMILTON, J. D. (1994): *Time Series Analysis*. Princeton University Press, Princeton, New Jersey.

HARTUNG, J. (1999): "A Note on Combining Dependent Tests of Significance," *Biometrical Journal*, 41, 849–855.

HARVEY, D. I., S. J. LEYBOURNE, AND P. NEWBOLD (1997): "Testing the Equality of Prediction Mean Squared Errors," *International Journal of Forecasting*, 13, 281–291.

——— (2001): "Analysis of a Panel of UK Macroeconomic Forecasts," *Econometrics Journal*, 4, S37–S55.

HASSLER, U., AND A. I. TARCOLEA (2005): "Combining Multi-Country Evidence on Unit Roots: The Case of Long-term Interest Rates," *Applied Economics Quarterly*, 51, 181–189.

HAUSMAN, J. (1978): "Specification Tests in Econometrics," *Econometrica*, 46, 1251–1271.

HEINEMANN, F. (2006): "Planning or Propaganda? An Evaluation of Germany's Medium-Term Budgetary Planning," *Public Finance Analysis*, 62, 551–578.

HM TREASURY (2002): "Trend Growth: Recent Developments and Prospects," London, available at http://www. hm-treasury. gov. uk.

HODRICK, R. J., AND E. C. PRESCOTT (1997): "Postwar U. S. Business Cycles: An Empirical Investigation," *Journal of Money, Credit, and Banking*, 29, 1–16.

HOLDEN, K., D. PEEL, AND B. SANDHU (1987): "The Accuracy of OECD Forecasts," *Empirical Economics*, 12, 175–186.

HOLDEN, K., AND D. A. PEEL (1990): "On Testing for Unbiasedness and Efficiency of Forecasts," *Manchester School*, 58, 120–127.

IM, K., M. H. PESARAN, AND Y. SHIN (1995): "Testing for Unit Roots in Heterogeneous Panels," DAE Working Papers Amalgamated Series No. 9526, University of Cambridge.

——— (2003): "Testing for Unit Roots in Heterogeneous Panels," *Journal of Econometrics*, 115, 53–74.

Isiklar, G., and K. Lahiri (2007): "How Far Ahead Can We Forecast? Evidence from Cross-Country Surveys," Forthcoming, International Journal of Forecasting.

Islam, N. (1995): "Growth Empirics: A Panel Data Approach," *Quarterly Journal of Economics*, 110, 1127–1170.

Jang, M. J., and D. W. Shin (2005): "Comparison of Panel Unit Root Tests under Cross Sectional Dependence," *Economics Letters*, 89, 12–17.

Jones, C. I. (2005): "The Shape of Production Functions and the Direction of Technical Change," *The Quarterly Journal of Economics*, 120, 517–549.

Kao, C. (1999): "Spurious Regression and Residual-based Tests for Cointegration in Panel Data," *Journal of Econometrics*, 90, 1–44.

Keane, M. P., and D. E. Runkle (1990): "Testing the Rationality of Price Forecasts: New Evidence from Panel Data," *The American Economic Review*, 80, 714–735.

King, R. G., and S. T. Rebelo (1999): "Resuscitating Real Business Cycles," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and M. Woodford, vol. 1, chap. 14, pp. 927–1007. Elsevier Science B. V.

Kirchgässner, G., and U. K. Müller (2006): "Are Forecasters Reluctant to Revise their Predictions? Some German Evidence," *Journal of Forecasting*, 25, 401–413.

Larsson, R., J. Lyhagen, and M. Lothgren (2001): "Likelihood-based Cointegration Tests in Heterogeneous Panels," *Econometrics Journal*, 4, 109–142.

Laxton, D., and R. Tetlow (1992): "A Simple Multivariate Filter for the Measurement of Potential Output," Bank of Canada Technical Report No. 59.

Leamer, E. E. (1983): "Let's Take the Con Out of Econometrics," *American Economic Review*, 73, 31–43.

——— (1985): "Sensitivity Analysis Would Help," *American Economic Review*, 75, 308–313.

Levin, A., and C. F. Lin (1992): "Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties," U. C. San Diego Discussion Paper 92-23.

Levin, A., C. F. Lin, and C. S. J. Chu (2002): "Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties," *Journal of Econometrics*, 108, 1–24.

Levine, R. (1997): "Financial Development and Economic Growth: Views and Agenda," *Journal of Economic Literature*, 35, 688–726.

——— (2005): "Finance and Growth: Theory and Evidence," in *Handbook of Economic Growth*, ed. by P. Aghion, and S. Durlauf, vol. 1, chap. 12, pp. 865–934. Elsevier Science B. V.

Levine, R., and D. Renelt (1992): "A Sensitivity Analysis of Cross-Country Growth Regressions," *The American Economic Review*, 82, 942–963.

Lindh, T. (2004): "Medium-Term Forecasts of Potential GDP and Inflation Using Age Structure Information," *Journal of Forecasting*, 23, 19–49.

Lindh, T., and B. Malmberg (1999): "Age Structure Effects and Growth in the OECD, 1950-1990," *Journal of Population Economics*, 12, 431–449.

Lucas, R. E. J. (1988): "On the Mechanics of Economic Development," *The Journal of Monetary Economics*, 22(1), 3–42.

MacKinnon, J. G. (1996): "Numerical Distribution Functions for Unit Root and Cointegration Tests," *Journal of Applied Econometrics*, 11, 601–618.

Maddala, G. S., and S. Wu (1999): "A Comparative Study of Unit Root Tests with Panel Data and A New Simple Test," *Oxford Bulletin of Economics and Statistics*, 61, 631–652.

Mankiw, N. G., D. Romer, and D. N. Weil (1992): "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics*, 107, 407–37.

McMorrow, K., and W. Roeger (2001): "Potential Output: Measurement Methods, "New" Economy Influences and Scenarios for 2001-2010: A Comparison of the EU15 and the US," European Commission, Economic Papers No. 150.

Milesi-Ferretti, G., and N. Roubini (1998): "Growth Effects of Income and Consumption Taxes," *Journal of Money, Credit and Banking*, 30, 721–744.

Mincer, J., and V. Zarnowitz (1969): "The Evaluation of Economic Forecasts," in *Economic Forecasts and Expectations*, ed. by J. Mincer. National Bureau of Economic Research, New York.

Moon, H. R., and B. Perron (2004): "Testing for a Unit Root in Panels with Dynamic Factors," *Journal of Econometrics*, 122, 81–126.

Musso, A., and T. Westermann (2005): "Assessing Potential Output Growth in the Euro Area: A Growth Accounting Perspective," ECB Occasional Paper Series No. 22.

Newey, W. K., and K. D. West (1987): "A Simple Positive Semi-Definite Heteroskedasticity and Autocorrelation-Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Ng, S., and P. S. Perron (2001): "Lag Length Selection and the Construction of Unit Root Tests with Good Sizeand Power," *Econometrica*, 69, 54–81.

Nickell, S. J., and L. Nunciata (2001): "Labour Market Institutions Database," Centre for Economic Performance, London School of Economics, London.

Ögünç, F., and D. Ece (2004): "Estimating the Output Gap for Turkey: An Unobserved Components Approach," *Applied Economics Letters*, 11, 177–182.

Okun, A. M. (1962): "Potential GNP: Its Measurement and Significance," in *Proceedings of the Business and Economic Statistics Section*, pp. 98–104. American Statistical Association.

Öller, L. E., and B. Barot (2000): "The Accuracy of European Growth and Inflation Forecasts," *International Journal of Forecasting*, 16, 293–315.

Padovano, F., and E. Galli (2002): "Comparing the Growth Effects of Marginal vs. Average Tax Rates and Progressivity," *European Journal of Political Economy*, 18, 529–544.

Pedroni, P. (1997): "Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis: New Results," Working paper, Indiana University.

——— (1999): "Critical Values for Cointegration Tests in Heterogeneous Panels with Multiple Regressors," *Oxford Bulletin of Economics and Statistics, Special Issue*, 61, 653–670.

——— (2000): "Fully-Modified OLS for Heterogeneous Cointegrated Panels," *Advances in Econometrics*, 15, 93–130.

——— (2004): "Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled TimeSeries Tests with an Application to the PPP Hypothesis," *Econometric Theory*, 20, 597–626.

Pesaran, H. M., T. Schuermann, and S. M. Weiner (2004): "Modelling Regional Interdependencies Using a Global Error-Correction Macroeconometric Model," *Journal of Business and Economic Statistics*, 22, 129–162.

Pesaran, M. H. (2004a): "General Diagnostic Tests for Cross Section Dependence in Panels," CESifo Working Paper No. 1229.

——— (2004b): "A Pair-Wise Approach to Testing for Output and Growth Convergence," Cambridge Working Papers in Economics No. 0453, University of Cambridge.

——— (2005): "A Simple Panel Unit Root Test in the Presence of Cross Section Dependence," University of Cambridge Working Paper.

——— (2006): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74, 967–1012.

Pesaran, M. H., Y. Shin, and R. P. Smith (1999): "Pooled Mean Group Estimation of Dynamic Heterogeneous Panels," *Journal of the American Statistical Association*, 94, 621–634.

Phillips, P. C. B., and H. R. Moon (1999): "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67, 1057–1111.

Phillips, P. C. B., and P. Perron (1988): "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, 335–346.

Phillips, P. C. B., and D. Sul (2003): "Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence," *Econometrics Journal*, 6, 217–259.

PONS, J. (2000): "The Accuracy of IMF and OECD Forecasts for G7 Countries," *Journal of Forecasting*, 19, 53–63.

PRESCOTT, E. C. (2004): "Why Do Americans Work So Much More Than Europeans?," *Federal Reserve Bank of Minneapolis Quarterly Review*, 28, 2–13.

RAVN, M., AND H. UHLIG (2002): "On Adjusting the HP-Filter for the Frequency of Observations," *Review of Economics and Statistics*, 84, 371–376.

RENNISON, A. (2003): "Comparing Alternative Output-Gap Estimators: A Monte Carlo Approach," Bank of Canada Working Paper No. 2003-8.

RICHARDSON, P., L. BOONE, C. GIORNO, M. MEACII, D. RAE, AND D. TURNER (2000): "The Concept, Policy Use and Measurement of Structural Unemployment: Estimating a Time Varying NAIRU Across 21 OECD Countries," OECD Economic Department Working Papers No. 250.

RÖGER, W. (2006): "The Production Function Approach to Calculating Potential Growth and Estimates for EU Member States and the US," EU-Commission DG ECFIN Paper, available at eemc.jrc.ec.europa.eu/EEMCArchive/ Publications/2006/Roeger2006%5B3%5D.pdf.

ROMER, P. M. (1986): "Increasing Returns and Long-Run Growth," *The Journal of Political Economy*, 94, 1002–1037.

ROUSSEAU, P. L., AND P. WACHTEL (2005): "Economic Growth and Financial Depth: Is the Relationship Extinct Already?," New York University Department of Economics Working Papers 05-15.

SAIKKONEN, P. (1991): "Asymptotic Efficient Estimation of Cointegration Regressions," *Econometric Theory*, 7, 1–21.

SALA-I-MARTIN, X. (1997): "I Just Ran 2 Million Regressions," *American Economic Review (Paper and Proceedings)*, 87, 178–183.

SALA-I-MARTIN, X., G. DOPPELHOFER, AND R. I. MILLER (2004): "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach," *American Economic Review*, 94, 813–835.

SARNO, L. (1999): "Stochastic Growth: Empirical Evidence from the G7 Countries," *Journal of Macroeconomics*, 21, 691–712.

SCACCIAVILLANI, F., AND P. SWAGEL (2002): "Measures of Potential Output: An Application to Israel," *Applied Economics*, 34, 945–957.

SCHWERT, G. W. (1989): "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business & Economic Statistics*, 7, 147–159.

SIEBERT, H. (2002): *Economic Policy for Aging Societies*. Springer, Berlin, Heidelberg.

SOLOW, R. M. (1956): "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics*, 70, 65–94.

Stadler, G. W. (1990): "Business Cycle Models with Endogeneous Technology," *The American Economic Review*, 80, 763–778.

Stekler, H. O., and G. Petrei (2003): "Diagnostics for Evaluating the Value and Rationality of Economic Forecasts," *International Journal of Forecasting*, 19, 735–742.

Stock, J. H., and M. W. Watson (2003): "Forecasting Output and Inflation: The Role of Asset Prices," *Journal of Economic Literature*, 151, 788–829.

Stockman, A. C. (1981): "Anticipated Inflation and the Capital Stock in a Cash-In-Advance Economy," *Journal of Monetary Economics*, 8, 387–393.

Strauss, J., and T. Yigit (2003): "Shortfalls of Panel Unit Root Testing," *Economics Letters*, 81, 309–313.

Teles, V. K. (2005): "The Role of Human Capital in Economic Growth," *Applied Economics Letters*, 12, 583–587.

Temple, J. (1999): "The New Growth Evidence," *Journal of Economic Literature*, 37, 112–56.

——— (2000): "Inflation and Growth: Stories Short and Tall," *Journal of Economic Surveys*, 14, 395–426.

——— (2006): "Aggregate Production Functions and Growth Economics," *International Review of Applied Economics*, 20, 301–317.

The Conference Board and Groningen Growth and Development Centre (2006): "Total Economy Database," January.

Tobin, J. (1965): "Money and Economic Growth," *Econometrica*, 33, 671–684.

Westerlund, J. (2005): "New Simple Tests for Panel Cointegration," *Econometric Reviews*, 25, 297–316.

Widmalm, F. (2001): "Tax Structure and Growth: Are Some Taxes Better than Others?," *Public Choice*, 107, 199–219.

Wolters, J., and U. Hassler (2006): "Unit Root Testing," *Allgemeines Statistisches Archiv*, 90, 43–58.

Wössmann, L. (2003): "Specifiying Human Capital," *Journal of Economic Surveys*, 17, 239–266.

Curriculum Vitae

Marcus Heribert Kappler
geboren am 5. Oktober 1973 in Stuttgart

Wissenschaftlicher Werdegang

05/1993	Abitur, Theodor-Heuss Gymnasium Mühlacker
1994-1997	Studium an der Eberhard-Karls-Universität Tübingen, Fachbereich Volkswirtschaftslehre
1997-1998	Graduate School of Economics, University of Maryland, USA
1998-2001	Freie Universität Berlin, Fachbereich Volkswirtschafts-lehre
06/2001	Studienabschluss (Diplom-Volkswirt), Betreuer der Diplomarbeit: Prof. Dr. Jürgen Wolters
seit 2002	Wissenschaftlicher Mitarbeiter am Zentrum für Europäische Wirtschaftsforschung (ZEW) GmbH, Mannheim (Forschungsbereiche "Internationale Finanzmärkte, Finanzmanagement und Makroökonomik" sowie "Unternehmensbesteuerung und Öffentliche Finanzwirtschaft")
seit 2003	Goethe-Universität Frankfurt am Main, Promotionsstudium in Volkswirtschaftslehre, Betreuer der Dissertation: Prof. Dr. Uwe Hassler