Discussion Paper No. 07-068

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

Marcus Kappler

ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Discussion Paper No. 07-068

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

Marcus Kappler

Non-Technical Summary

This paper first provides a survey of methods for producing medium-term projections that are used by governmental bodies in the major industrial countries and by international institutions. 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.

In the second part of the study it is assessed whether projections from a widely employed version of the production function approach convey a reliable view about future economic developments. For this purpose an out-of-sample forecast exercise based on quarterly data from National Accounts is carried out and a framework that allows evaluation of forecast errors based on consistent test statistics is developed. Empirical implementation of the proposed test strategy is straightforward and conducted for three to five year cumulative projections of GDP growth for the G7 countries. In addition, available projections from national and international sources are included in the forecast evaluation.

The evaluation of the forecast errors of the out-of-sample experiment for the observation period from 1985 to 2005 highlights the following: The production function approach with its assumption that output gaps close over medium-term forecast horizons yields unbiased projections of real GDP growth for most countries, but misses other important features of actual GDP developments. Root mean squared errors and mean absolute errors are large and there are long periods where the forecasts are not in line with actual GDP growth. However, projections from the production function approach are generally capable of beating naïve forecasts in terms of root mean squared error although the differences in accuracy are mostly not statistically significant. Nevertheless, this is a remarkable result considering the fairly long forecast horizons.

The pseudo forecasts from the out-of-sample exercise serve as a sort of "status quo" or neutral benchmark which incorporates an assessment of the future economic outlook if factor contributions and total factor productivity follow past trends. The comparisons of these pseudo forecasts with projections from official authorities shows that the German government's and the IMF's future assessments of economic developments, in particular, tended to deviate systematically from neutral assumptions in the past and resulted 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.

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

Marcus Kappler*†

November 2007

Abstract

The focus of this paper is the evaluation of a very popular method for potential output estimation and medium-term forecasting— the production function approach—in terms of predictive performance. For this purpose, a forecast evaluation for the three to five years ahead predictions of GDP growth for the individual G7 countries is conducted. To carry out the forecast performance check a particular testing framework is derived that allows the computation of robust test statistics given the specific nature of the generated out-of sample forecasts. In addition, medium-term GDP projections from national and international institutions are examined and it is assessed whether these projections convey a reliable view about future economic developments and whether there is scope for improving their predictive content.

JEL Classification: E 23, E 27, C53
Key Words: Potential output, projections, forecast evaluation

*Very skillful research assistance by Katrin Sommerfeld and Qianying Chen is gratefully acknowledged. Furthermore, I would like to thank the participants of the 2006 DIW macroeconometric workshop in Berlin, especially Jürgen Wolters and Bernd Lucke for helpful comments and suggestions. I am also very grateful to the participants of the IWEF 2007 in Tripolis and the participation of the International Symposium on Forecasting 2007 in New York City, notably Ullrich Heilemann, Albert Caspers and Lars-Erik Öller. I also wish to thank Andreas Schrimpf, Steffen Osterloh, Friedrich Heinemann and Philipp Ager for valuable comments. Of course, all remaining errors are mine.

†Centre for European Economic Research (ZEW), P.O. Box 103443, D-68034 Mannheim, Germany, Phone: +49/621/1235-157, Fax: +49/621/1235-223, E-mail: kappler@zew.de

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, but also central banks and international institutions such as 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

[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).

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 paper first provides a survey of methods of medium-term forecasting that are used by governmental bodies in the major 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 paper, 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 paper 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 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 paper proceeds as follows. Section 2 discusses commonly used approaches for producing medium-term predictions and reviews the relevant literature. Section 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 4 while section 5 summarizes and concludes.

2 Projection approaches

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

[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 paper, 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.

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 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

[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.

[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.

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 paper 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 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.

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

[5]It is not the aim of the present paper 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).

[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).

function or growth accounting approaches are most widely-used. The OECD[8], the IMF[9] and the European Commission[10] 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.[11] 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[12] and the HM Treasury in the United Kingdom[13], for instance, use a growth accounting framework to derive medium-term projections.

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.[14] 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

---

[8]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).

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

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

[11]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.

[12]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.

[13]See HM Treasury (2002).

[14]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.

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.[15]

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 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 3.2.2 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

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

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 paper will focus on analyzing the forecast performance of the PFA based methods.

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.

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:

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

Y_t denotes output , N_t labor input, K_t capital input and E_t the Harrod-neutral labor augmenting Total Factor Productivity (TFP).[16] Labor input comprises several key variables of the labor market and enters the production function on a hours worked basis rather than on number of employed:

$$N_t = H_t L_t \tag{2}$$

$$L_t = PW_t PR_t (1 - U_t) \tag{3}$$

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):

[16] 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.

$$E_t = Y_t^{-\alpha} K_t^{-(1-\alpha)/\alpha} N_t^{-1} \tag{4}$$

In order to obtain a measure of potential output of the economy, several trend variables (indicated with an asterisk) are substituted in equation (1):

$$\begin{aligned} Y_t^* &= (E_t^* N_t^*)^{\alpha} K_t^{(1-\alpha)} && (5) \\ N_t^* &= H_t^* L_t^* && (6) \\ L_t^* &= PW_t PR_t^* (1 - U_t^*) && (7) \end{aligned}$$

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 (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.

3.1.1 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.[17] 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, 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

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

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 shows the averages of the unadjusted and adjusted labor share of the G7 countries computed from National Accounts data.

Table 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 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.[18] 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 is that adjusted labor shares do not vary much across

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

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.[19]

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. 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 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 2). The magnitudes of individual estimates of the labor share do not match the figures computed from National Accounts data and this holds for all countries.[20] In fact, correspondence between econometric estimates and available information on the labor share from National Accounts is 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.

[19] 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.

[20] 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.

Table 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.

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.

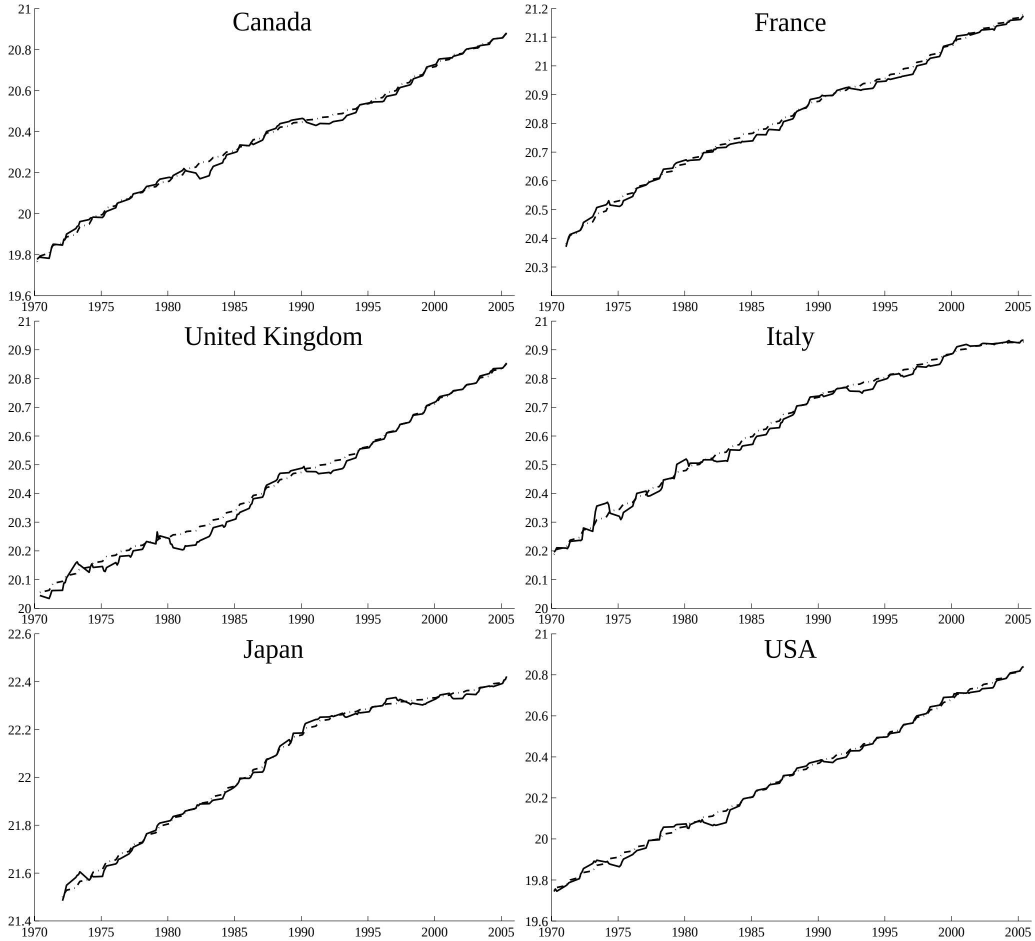
An overall conclusion from the preceding exercise is the following: Econometric support for the usual assumptions of the growth accounting procedures can be provided and econometrically estimated parameters which are broadly in line with the National Accounts data approach can be obtained. However, this 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 and table 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.[21]

[21] 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.

3.2 In-sample estimates of potential output

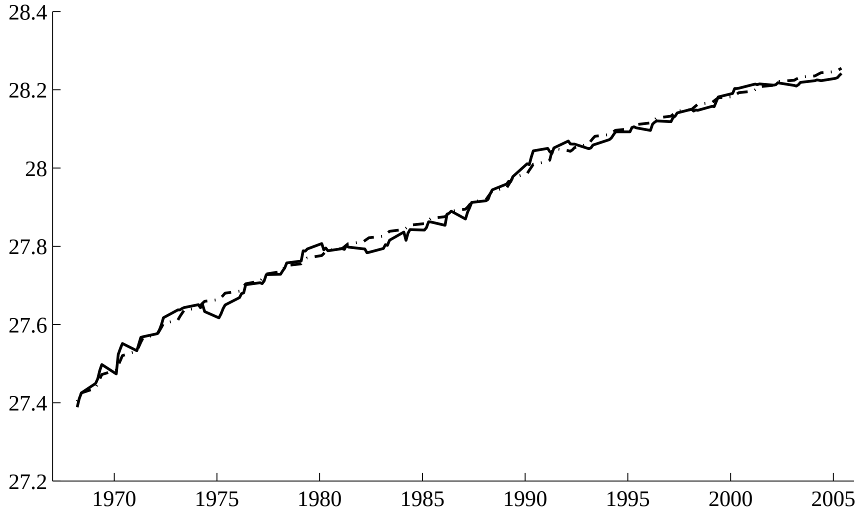
Figure 1 shows the results of the in-sample computation of potential output corresponding to the above outlined production function method in logarithmic form for Canada, France, the United Kingdom, Italy, Japan and the USA. The seasonally adjusted quarterly data is taken from the OECD Economic Outlook database.[22] 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: Potential and actual GDP

[22]Section 5 in the appendix provides more detailed information about the data set.



Notes: See figure 1

Figure 2: Potential and actual GDP for Germany

3.2.1 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.[23]

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

[23]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.

production function version of Giorno et al. (1995) is used to estimate Germany's 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 2 shows the path of potential output for Germany.

3.2.2 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.[24] 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 3 illustrates these points by stylizing the derivation of projections over the period from T_0 to T_1.

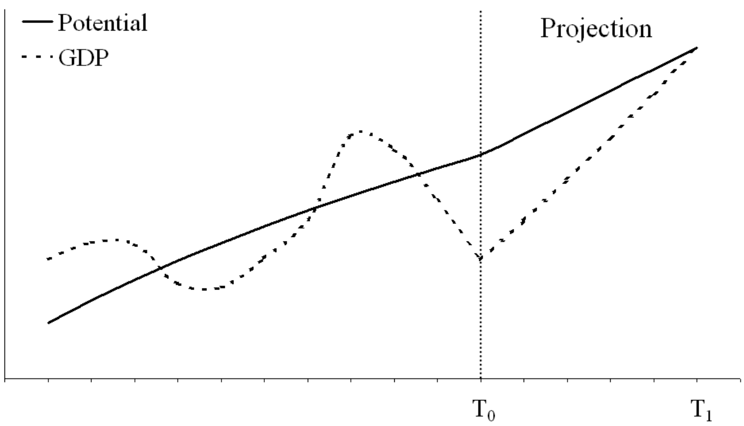


Figure 3: Potential output and the output gap

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

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

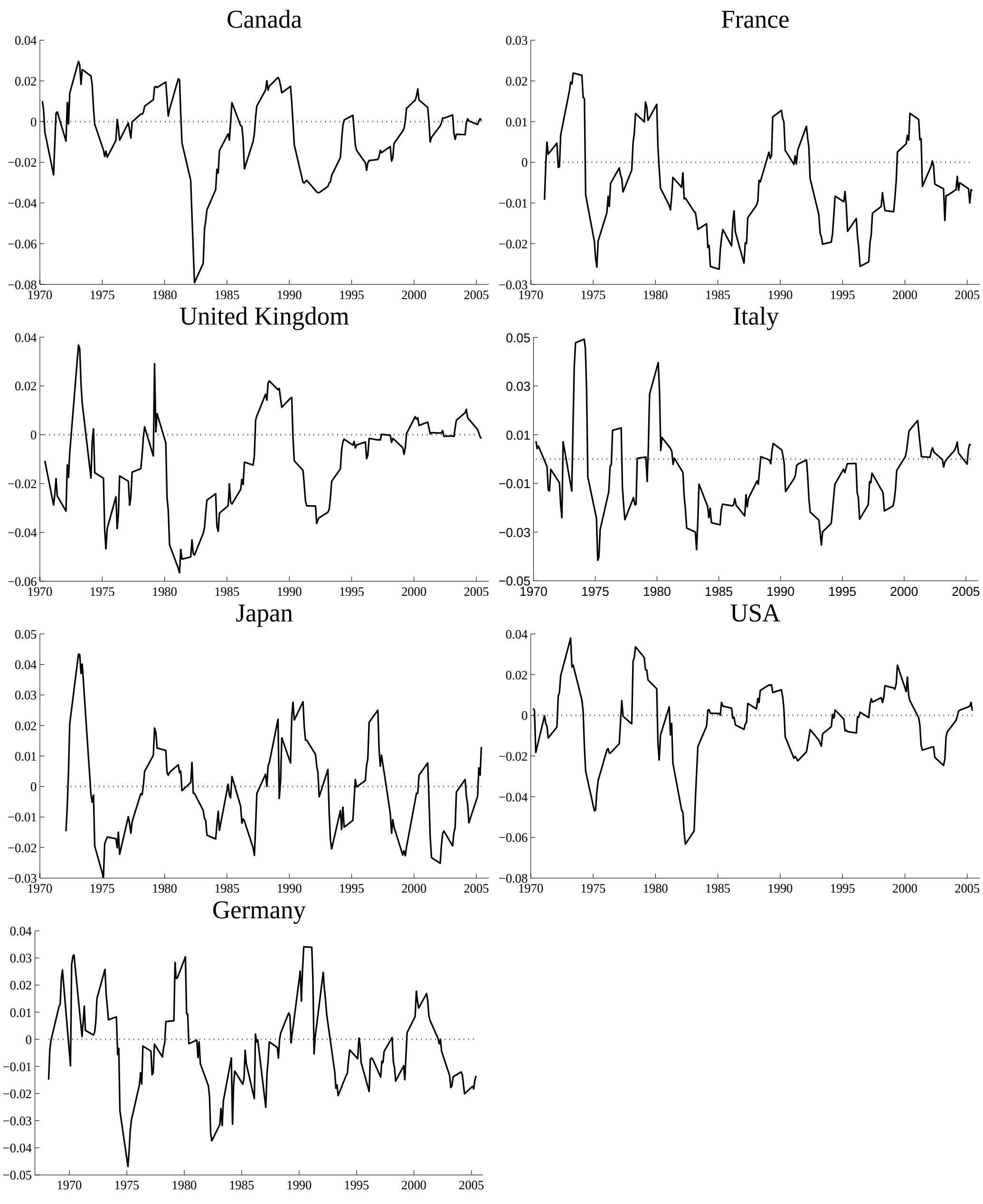


Figure 4: Output gaps from the PFA

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 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 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.[25]

The entries in table 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 grounds.

[25] 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 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.

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 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 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 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.

3.3 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 3.4 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) to (7) from above).[26] 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.

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

- The **Total Factor Productivity** is estimated as the Solow residual corresponding to equation (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.[27]

- 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.[28]

- 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 paper 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.

[27]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.

[28]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.

3.4 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 $F_{t^*}^h = \Delta y_{t^*+1|t^*} + \Delta y_{t^*+2|t^*} + ... + \Delta y_{t^*+h|t^*} = \sum_{i=1}^{h} \Delta y_{t^*+i|t^*}$ 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 $A_{t^*}^h = \sum_{i=1}^{h} \Delta y_{t^*+i}$.
- Forecast errors are computed:

$$e_{t^*}^h = A_{t^*}^h - F_{t^*}^h = \sum_{i=1}^{h} \Delta y_{t^*+i} - \sum_{i=1}^{h} \Delta y_{t^*+i|t^*} \tag{8}$$

- The sample is expanded by one quarter, i.e. the next forecasts are conducted as $F_{t^*+1}^h = \sum_{i=1}^{h} \Delta y_{t^*+1+i|t^*+1}$ and errors are obtained as $e_{t^*+1}^h = A_{t^*+1}^h - F_{t^*+1}^h$.
- 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 (8) have the following structure:[29]

[29]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 paper.

$$e_t^h = A_t^h - F_t^h = \sum_{i=1}^{h} u_{t+i} + \phi = \nu_t^h + \phi, \quad t = t^*, \ldots, T-h \tag{9}$$

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 (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 (9) it follows that

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

Therefore, rather than being diagonal, the variance matrix $E[\nu^h \nu^{h\prime}] = \Sigma_\nu$ takes the following block diagonal form: [30]

$$\underset{(T^* \times T^*)}{\Sigma_\nu} = \sigma_u^2 A \tag{10}$$

with

$$A = \begin{pmatrix}
a^{(0)} & a^{(1)} & \cdots & a^{(h-1)} & 0 & \cdots & & & \cdots & 0 \\
a^{(1)} & a^{(0)} & a^{(1)} & \cdots & a^{(h-1)} & 0 & \cdots & & & \vdots \\
\vdots & a^{(1)} & a^{(0)} & a^{(1)} & \cdots & a^{(h-1)} & 0 & \cdots & & \\
a^{(h-1)} & \cdots & a^{(1)} & a^{(0)} & a^{(1)} & \cdots & a^{(h-1)} & 0 & \cdots & \\
0 & a^{(h-1)} & \cdots & a^{(1)} & a^{(0)} & a^{(1)} & \cdots & a^{(h-1)} & 0 & \\
\vdots & 0 & a^{(h-1)} & \cdots & a^{(1)} & a^{(0)} & a^{(1)} & \cdots & a^{(h-1)} & 0 \\
 & & 0 & a^{(h-1)} & \cdots & a^{(1)} & a^{(0)} & a^{(1)} & \cdots & a^{(h-1)} \\
 & & & 0 & a^{(h-1)} & \cdots & a^{(1)} & a^{(0)} & a^{(1)} & \vdots \\
\vdots & & & & 0 & a^{(h-1)} & \cdots & a^{(1)} & a^{(0)} & a^{(1)} \\
0 & \cdots & & & & 0 & a^{(h-1)} & \cdots & a^{(1)} & a^{(0)}
\end{pmatrix} \tag{11}$$

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

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

From (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.

3.4.1 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 (9), where

$$H_0: \quad \phi = 0 \tag{12}$$

The test statistic of interest is

$$t_\phi = \frac{\widehat{\phi}}{\widehat{\sigma}_\phi} \tag{13}$$

with

$$\widehat{\phi} = \frac{1}{T^*} \sum_{t=t^*}^{T-h} e_t^h \tag{14}$$

and the consistent covariance matrix estimator

$$\widehat{\sigma}_\phi^2 = (X'X)^{-1} X' \widehat{\Sigma}_\nu X (X'X)^{-1} = \frac{1}{T^{*2}} i'_{T^*} \widehat{\Sigma}_\nu i_{T^*} \tag{15}$$

and $X = i_{T^*}$ with i_{T^*} as a vector of ones with dimension T^*.[31] The expressions (14) and (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 (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[32]

$$\widehat{\sigma}_u^2 = \frac{1}{T^*} \widehat{\nu}^{h\prime} A^{-1} \widehat{\nu}^h \tag{16}$$

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

[32] 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 (16).

We refer to the above outlined approach as generalized least squares (GLS) framework although simple averaging (OLS) is used to estimate the bias term ϕ.[33] The focus of interest is rather on the GLS standard errors as given by equation (15).

Table 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 (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.[34] The matrix in (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 (9). This estimator has difficulties in capturing the correct standard errors in finite samples. Table 4 shows results of the size properties of the Newey-West

[33] 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.

[34] 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.

t-statistic in a test under the null hypothesis of unbiased multi-step forecasts provided by a Monte Carlo experiment.[35] The entries of table 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.

3.4.2 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:

$$d = \frac{1}{T^*} \sum_{t=1}^{T} d_t^h, \tag{17}$$

and

$$d_t^h = (\tilde{e}_t^h)^2 - (\overline{e}_t^h)^2 \tag{18}$$

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

[35] 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.

$$\tilde{e}_t^h = \sum_{i=1}^{h} \tilde{u}_{t+i} + \tilde{\phi} = \tilde{\nu}_t^h + \tilde{\phi}, \quad E[\tilde{e}_t^h] = \tilde{\phi}, \quad V[\tilde{e}_t^h] = h\sigma_{\tilde{u}}^2 \tag{19}$$

$$\overline{e}_t^h = \sum_{i=1}^{h} \overline{u}_{t+i} + \overline{\phi} = \overline{\nu}_t^h + \overline{\phi}, \quad E[\overline{e}_t^h] = \overline{\phi}, \quad V[\overline{e}_t^h] = h\sigma_{\overline{u}}^2 \tag{20}$$

The test statistic of interest is given by

$$DM = \frac{d}{\sqrt{\widehat{V}(d)}} \tag{21}$$

$\widehat{V}(d)$ 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:[36]

$$\widehat{V}(d) = \frac{1}{T^*} \sum_{k=-(h-1)}^{(h-1)} \widehat{\gamma}_d(k) \tag{22}$$

and $\widehat{\gamma}_d(k)$ is the estimated autocovariance of d. DM propose to estimate (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 $V(d)$.[37] 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 (19) and (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:[38]

$$\gamma_d(0) = V[(d_t^h)] = V[(\tilde{e}_t^h)^2] + V[(\overline{e}_t^h)^2] = 2h\sigma_{\tilde{u}}^2(h\sigma_{\tilde{u}}^2 + 2\tilde{\phi}^2) + 2h\sigma_{\overline{u}}^2(h\sigma_{\overline{u}}^2 + 2\overline{\phi}^2) \tag{23}$$

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

[37] 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.

[38] 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 $V[a^2] = 2\sigma^4 + 4\sigma^2\mu^2 = 2\sigma^2(\sigma^2 + 2\mu^2)$.

However, the assumption of uncorrelated disturbances resulting from two forecast models that have the same target is not realistic. Dependence arises since the 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 (23). Taking $Cov(\tilde{u}_t, \overline{u}_t) = \sigma_{\tilde{u},\overline{u}}$ into account leads to:

$$Cov\left[(\tilde{e}_t^h)^2, (\overline{e}_t^h)^2\right] = 2h\sigma_{\tilde{u},\overline{u}}(h\sigma_{\tilde{u},\overline{u}} + 2\tilde{\phi}\overline{\phi}) \tag{24}$$

Combining these results and rearranging expressions produces the following formula for the variances and autocovariances of the quadratic loss differential:[39]

$$\begin{aligned}
\gamma_d(k) &= E[d_t^h d_{t-k}^h] - E[d_t^h]E[d_{t-k}^h] = \\
&= 2\check{h}\left(\sigma_{\tilde{u}}^2(\check{h}\sigma_{\tilde{u}}^2 + 2\tilde{\phi}^2) + \sigma_{\overline{u}}^2(\check{h}\sigma_{\overline{u}}^2 + 2\overline{\phi}^2) - 2\sigma_{\tilde{u},\overline{u}}(\check{h}\sigma_{\tilde{u},\overline{u}} + 2\tilde{\phi}\overline{\phi})\right) \\
&\quad \check{h} = h - |k|
\end{aligned} \tag{25}$$

Replacing population moments with sample moments in equation (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 (16) while the covariance is estimated analogously as follows

$$\widehat{\sigma}_{\tilde{u},\overline{u}} = \frac{1}{T^*}\widehat{\tilde{\nu}}^{h\prime} A^{-1}\widehat{\overline{\nu}}^h \tag{26}$$

Estimates of $\tilde{\phi}$ and $\overline{\phi}$ can be obtained by following (14).[40]

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 (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 (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

[39]This result is established more rigorously in appendix 5, page 62.

[40]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.

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 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.[41] Then these forecast errors are cumulated over various horizons and afterwards the modified DM test and the test as described by equations (21),(22) and (25) are performed for sample sizes of $T = 80, 100$ and 120 as well as contemporaneous correlations of $\rho = 0.5$ and 0.9.[42] 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 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

[41]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.

[42]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 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 (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 (19) and (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.

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.

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 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 3.4.1.

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 3.4.2. 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.[43] 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.[44] If a neutral scenario per se produces a biased forecast, these forecasts

[43]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.

[44]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.

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.

4.1 Germany

Table 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 5 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

Table 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.

to the actual growth rates. For the remaining years, neither the pseudo forecasts

Notes: The date always refers to the last year of the projection. See section 5 in the appendix for details. Average error of government projections: -0.486, RMSE of government projections: 1.036

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

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. Indeed , the figure 5 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.

4.2 USA

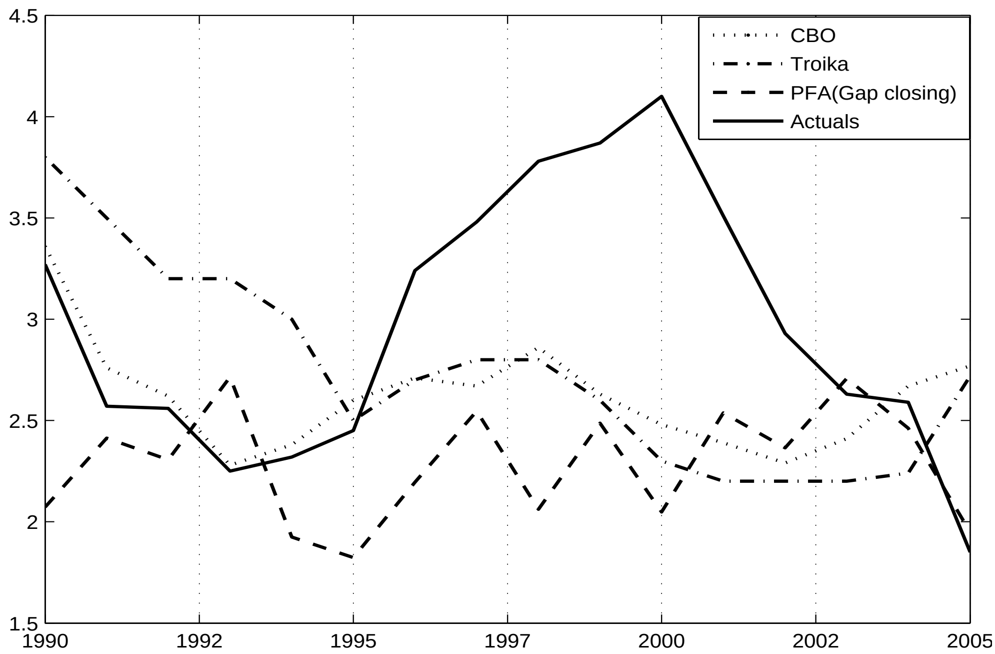
In contrast to the outcomes for Germany, for the USA we find that the random walk model demonstrates better forecast performance than the PFA based forecast.

Table 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.

As depicted in table 7, the forecasts produced with both PFA versions exhibit positive bias for all horizons but at the same time these estimates are not significant



Notes: The date always refers to the last year of the projection. See section 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 6: CBO and Troika projections for 5-year GDP growth in the USA

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 3.4.2, 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.

Figure 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 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 7.

4.3 United Kingdom

The outcomes of the forecast performance tests for the United Kingdom are displayed in table 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.

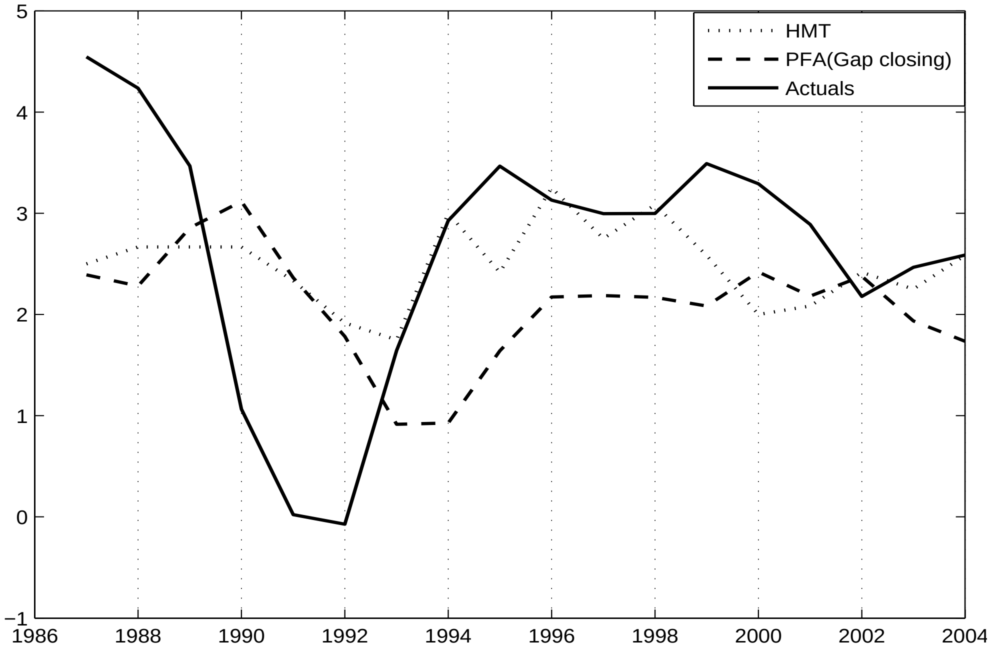
The HAC t-values imply insignificant loss differentials for the PFA (gap closing) predictions at the four and five year horizon, however, we regard the GLS statistics as being more reliable in the light of the experimental outcomes reported in section 3.4 and therefore conclude that the RW generated more accurate forecasts than the other models for horizons beyond three years.

Figure 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.

Table 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 5 in the appendix for details. Average error of HMT projections: 0.134, RMSE of HMT projections: 1.145

Figure 7: HMT projections for 3-year GDP growth in the United Kingdom

4.4 France

Results for France are given in table 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.

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 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

Table 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.

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 5 in the appendix for details. Average error of IMF projections: -0.914, RMSE of IMF projections: 1.462

Figure 8: IMF projections for 3-year GDP growth in the France

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 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.

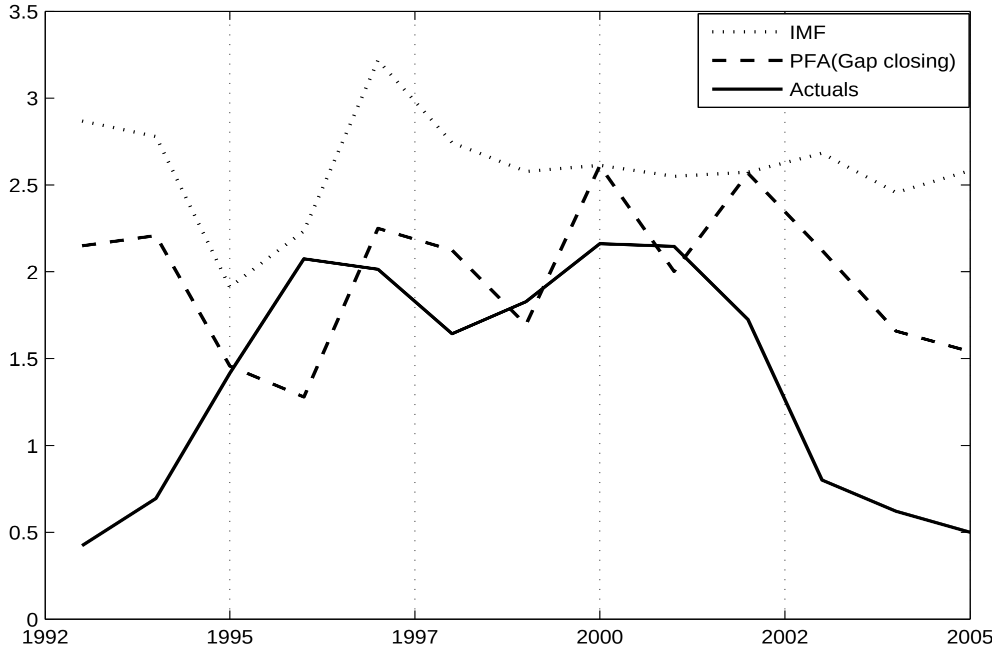
In figure 9, the pseudo forecasts as well as projections from IMF staff are contrasted with the final medium-term growth rates of GDP.

Table 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.

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



Notes: The date always refers to the last year of the projection. See section 5 in the appendix for details. Average error of IMF projections: -1.211, RMSE of IMF projections: 1.419

Figure 9: IMF projections for 3-year GDP growth in Italy

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.

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 11 are based.

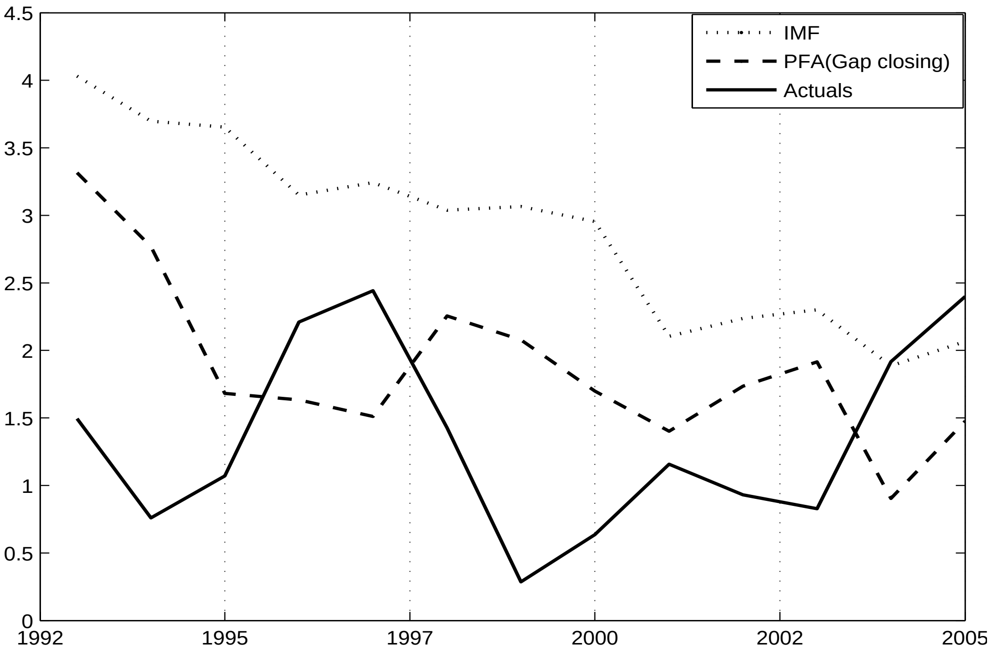
The PFA (direct) and PFA (gap closing) models are able to predict GDP growth rates that approximately match with this development: The mean forecasts are only slightly above two percent over all horizons and bias estimates are not significant even once. The opposite holds for the random walk forecast. 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

Table 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.

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 5 in the appendix for details. Average error of IMF projections: -1.528, RMSE of IMF projections: 1.836

Figure 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 10. 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.

4.7 Canada

The last outcomes to discuss are those for Canada. Table 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 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 5 in the appendix for details. Average error of IMF projections: -0.419, RMSE of IMF projections: 1.013

Figure 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 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.

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 paper 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 paper 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 paper 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 find 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 from the out-of-sample exercise serve as a sort of "status quo" or neutral benchmark which incorporates an assessment of the future economic outlook if factor contributions and total factor productivity follow past trends. The comparisons of these pseudo forecasts with projections from official authorities shows that the German government's and the IMF's future assessments of economic developments, in particular, tended to deviate systematically from neutral assumptions in the past and resulted 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.

References

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

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

Aghion, P., and P. Howitt (1998): *Endogeneous Growth Theory*. The MIT Press, Cambridge.

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.

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.

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.

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

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.

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.

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.

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

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.

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

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.

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

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.

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.

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.

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.

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.

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

Jones, C. I. (2005): "The Shape of Production Functions and the Direction of Technical Change," *The Quarterly Journal of Economics*, 120, 517–549.

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.

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.

MacKinnon, J. G. (1996): "Numerical Distribution Functions for Unit Root and Cointegration Tests," *Journal of Applied Econometrics*, 11, 601–618.

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.

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.

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.

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.

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.

Pons, J. (2000): "The Accuracy of IMF and OECD Forecasts for G7 Countries," *Journal of Forecasting*, 19, 53–63.

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.

Scacciavillani, F., and P. Swagel (2002): "Measures of Potential Output: An Application to Israel," *Applied Economics*, 34, 945–957.

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.

Temple, J. (2006): "Aggregate Production Functions and Growth Economics," *International Review of Applied Economics*, 20, 301–317.

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 (5) to (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.[45]

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.[46]

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 3.2.1, **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.

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

[46] 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.

The **Labor shares** are based on annual data from the OECD Economic Outlook 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.[47]

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.

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

Inconsistency of the kernel-based HAC estimator

Assume that the forecast errors follow the data generating process as given by equation (9), which is repeated for convenience:

$$e_t^h = A_t^h - F_t^h = \sum_{i=1}^{h} u_{t+i} + \phi = \nu_t^h + \phi \tag{27}$$

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 10) and the variance estimator for a test of $\widehat{\phi} = 0$ in an OLS regression is directly given by equation (15). Expanding this expression by applying matrix algebra (and skipping asterisks) results in

$$\widehat{\sigma}_\phi^2 = \frac{1}{T^2} i_T' \widehat{\Sigma}_\nu i_T = \frac{1}{T^2}\left(Th\widehat{\sigma}_u^2 + \sum_{k=1}^{h-1} 2(T-k)(h-k)\widehat{\sigma}_u^2)\right) \tag{28}$$

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 (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 (27) with an intercept as sole regressor becomes:[48]

$$\widehat{\sigma}_{\phi,NW}^2 = (X'X)^{-1}\widehat{\Sigma}_{NW}(X'X)^{-1} = \frac{1}{T^2}\widehat{\Sigma}_{NW} \tag{29}$$

where:

$$\widehat{\Sigma}_{NW} = \sum_{t=1}^{T}(\widehat{\nu}_t^h)^2 + \sum_{k=1}^{h-1} w_k \sum_{t=k+1}^{T} 2(\widehat{\nu}_t^h \widehat{\nu}_{t-k}^h) \tag{30}$$

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

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

covariance is positive definite.[49] Often a Bartlett kernel in the form of $1-\frac{k}{h}$ is used for w_k. Replacing the sample disturbance moments in (30) with the estimates of the corresponding parameterized expressions results in the following expression for the HAC variance estimator for $\widehat{\phi}$:

$$\begin{aligned} \widehat{\sigma}^2_{\phi,NW} &= \frac{1}{T^2}\left(\sum_{t=1}^{T} h\widehat{\sigma}^2_u + \sum_{k=1}^{h-1} w_k \sum_{t=k+1}^{T} 2(h-k)\widehat{\sigma}^2_u\right) \\ &= \frac{1}{T^2}\left(Th\widehat{\sigma}^2_u + \sum_{k=1}^{h-1} w_k 2(T-k)(h-k)\widehat{\sigma}^2_u\right) \end{aligned} \tag{31}$$

On comparing the formula for the consistent GLS estimator of $\widehat{\sigma}^2_\phi$ (see equation 28) with the Newey-West estimator as shown in equation (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 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:

$$E\left[(\sum_{i=1}^{h}\tilde{u}_t)^2(\sum_{i=1}^{h}\tilde{u}_{t-k})^2\right] = \begin{cases} (3h^2-4h|k|+2|k|^2)\sigma^4_{\tilde{u}} & \text{for} \quad h>|k| \\ h^2\sigma^4_{\tilde{u}} & \text{for} \quad h\leq|k| \end{cases} \tag{32}$$

$$\begin{aligned} E\left[(\sum_{i=1}^{h}\tilde{u}_t)^2(\sum_{i=1}^{h}\overline{u}_{t-k})^2\right] &= \\ E\left[(\sum_{i=1}^{h}\overline{u}_t)^2(\sum_{i=1}^{h}\tilde{u}_{t-k})^2\right] &= \begin{cases} h^2\sigma^2_{\tilde{u}}\sigma^2_{\overline{u}} + 2(h-|k|)^2\sigma^2_{\tilde{u},\overline{u}} & \text{for} \quad h>|k| \\ h^2\sigma^2_{\tilde{u}}\sigma^2_{\overline{u}} & \text{for} \quad h\leq|k| \end{cases} \end{aligned} \tag{33}$$

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

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

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

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

Given the forecast error models of equation (19) and (20), the quadratic loss differential becomes

$$d_t^h = (\tilde{e}_t^h)^2 - (\overline{e}_t^h)^2 = (\sum_{i=1}^{h} \tilde{u}_{t+i} + \tilde{\phi})^2 - (\sum_{i=1}^{h} \overline{u}_{t+i} + \overline{\phi})^2 \tag{34}$$

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

$$\begin{aligned} E[d_{t-k}^h] &= E\left[(\sum_{i=1}^{h} \tilde{u}_{t-k+i})^2 + 2\tilde{\phi}\sum_{i=1}^{h} \tilde{u}_{t-k+i} + \tilde{\phi}^2\right] \\ &- E\left[(\sum_{i=1}^{h} \overline{u}_{t-k+i})^2 + 2\overline{\phi}\sum_{i=1}^{h} \overline{u}_{t-k+i} + \overline{\phi}^2\right] & (35) \\ &= h\sigma_{\tilde{u}}^2 - h\sigma_{\overline{u}}^2 + \tilde{\phi}^2 - \overline{\phi}^2 \qquad \forall\, k & (36) \end{aligned}$$

it follows that

$$E[d_{t-k}^h]E[d_t^h] = \left(\tilde{\phi}^2 - \overline{\phi}^2 + h(\sigma_{\tilde{u}}^2 - \sigma_{\overline{u}}^2)\right)^2 \tag{37}$$

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

$$
\begin{aligned}
E[d_t^h d_{t-k}^h] &= E[\tilde{\phi}^4] + E[\overline{\phi}^4] - 2E[\tilde{\phi}^2\overline{\phi}^2] + E\left[\tilde{\phi}^2(\sum_{i=1}^{h}\tilde{u}_{t+i})^2\right] - E\left[\overline{\phi}^2(\sum_{i=1}^{h}\tilde{u}_{t+i})^2\right] \\
&+ 4E\left[\tilde{\phi}^2\sum_{i=1}^{h}\tilde{u}_{t+i}\sum_{i=1}^{h}\tilde{u}_{t-k+i}\right] + E\left[\tilde{\phi}^2(\sum_{i=1}^{h}\tilde{u}_{t-k+i})^2\right] - E\left[\overline{\phi}^2(\sum_{i=1}^{h}\tilde{u}_{t-k+i})^2\right] \\
&+ E\left[(\sum_{i=1}^{h}\tilde{u}_{t+i})^2(\sum_{i=1}^{h}\tilde{u}_{t-k+i})^2\right] - 4E\left[\tilde{\phi}\overline{\phi}\sum_{i=1}^{h}\tilde{u}_{t-k+i}\sum_{i=1}^{h}\overline{u}_{t+i}\right] \\
&- E\left[\tilde{\phi}^2(\sum_{i=1}^{h}\overline{u}_{t+i})^2\right] + E\left[\overline{\phi}^2(\sum_{i=1}^{h}\overline{u}_{t+i})^2\right] - E\left[(\sum_{i=1}^{h}\tilde{u}_{t-k+i})^2(\sum_{i=1}^{h}\overline{u}_{t+i})^2\right] \\
&- 4E\left[\tilde{\phi}\overline{\phi}\sum_{i=1}^{h}\tilde{u}_{t+i}\sum_{i=1}^{h}\overline{u}_{t-k+i}\right] + 4E\left[\overline{\phi}^2\sum_{i=1}^{h}\overline{u}_{t+i}\sum_{i=1}^{h}\overline{u}_{t-k+i}\right] \\
&- E\left[\tilde{\phi}^2(\sum_{i=1}^{h}\overline{u}_{t-k+i})^2\right] + E\left[\overline{\phi}^2(\sum_{i=1}^{h}\overline{u}_{t-k+i})^2\right] - E\left[(\sum_{i=1}^{h}\tilde{u}_{t+i})^2(\sum_{i=1}^{h}\overline{u}_{t-k+i})^2\right] \\
&+ E\left[(\sum_{i=1}^{h}\overline{u}_{t+i})^2(\sum_{i=1}^{h}\overline{u}_{t-k+i})^2\right] \qquad (38)
\end{aligned}
$$

To find the parametric solution of (38), the cases $h > |k|$ and $h \leq |k|$ need to be differentiated.

From corollary (32) and (33), the following results:

For $h > |k|$,

$$
\begin{aligned}
E[d_t^h d_{t-k}^h] &= \tilde{\phi}^4 + \overline{\phi}^4 - 2h\overline{\phi}^2\sigma_{\tilde{u}}^2 - 3h^2\sigma_{\tilde{u}}^2 - 4hk\sigma_{\tilde{u}}^2 + 2k^2\sigma_{\tilde{u}}^2 \\
&- 2\sigma_{\overline{u}}^2(h^2\sigma_{\tilde{u}}^2 + 2k\overline{\phi}^2 - 3h\overline{\phi}^2) + \sigma_{\overline{u}}^2(3h^2 - 4hk + 2k^2) \\
&- 2\tilde{\phi}^2(\overline{\phi}^2 - 3h\sigma_{\tilde{u}}^2 + 2k\sigma_{\tilde{u}}^2 + h\sigma_{\overline{u}}^2) - 4\sigma_{\tilde{u},\overline{u}}^2(h-k)^2 \\
&- 8\tilde{\phi}\overline{\phi}(h-k)\sigma_{\tilde{u},\overline{u}} \qquad (39)
\end{aligned}
$$

and subtracting (37) from (39) gives the non-zero autocovariance formula as shown above and in equation (25) in the text.

For $h \leq |k|$, the term in (38) collapses to

$$
E[d_t^h d_{t-k}^h] = \left(\tilde{\phi}^2 - \overline{\phi}^2 + h(\sigma_{\tilde{u}}^2 - \sigma_{\overline{u}}^2)\right)^2 \qquad (40)
$$

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