Potential output growth and output gaps: concept, uses and estimates

Measures of potential output play a prominent role in a wide range of economic models as they are useful for distinguishing between medium-term trends and shorter-term cyclical movements in the economy. This article examines the concepts and uses of potential output and discusses different methods of estimation. Typically, these methods can be divided into those which rely heavily on statistical techniques to break output down into trend and cycle components and structural approaches which are more firmly rooted in economic theory. While different methods tend to lead to fairly similar results regarding the rates of growth of potential output, estimates of the level of the output gap are more diverse and are surrounded by a greater degree of uncertainty. For this reason, particular caution is required when drawing conclusions with regard to policy which are based on estimates of the level of the output gap. Furthermore, as estimates of potential output are based on historical data, they are unlikely to capture adequately the effects of structural change until well after such change has occurred.

1 Introduction

The main interest of the ECB in the concept and estimation of potential output arises in the context of its stability-oriented monetary policy strategy, the primary objective of which is the maintenance of price stability in the euro area. As has been explained in detail in previous issues of the ECB Monthly Bulletin, this strategy has two pillars. First, there is a prominent role for money signalled by the announcement of a reference value for broad money growth, including a detailed analysis of monetary developments. Second, there is an analysis of (predominantly) non-monetary indicators in order to form a broadly based assessment of the outlook for future price developments (see, in particular, the January, February and April 1999 issues of the ECB Monthly Bulletin).

Measuring potential output and its growth rate is an important issue under both pillars. Under the first pillar, when coupled with the ECB's quantitative definition of price stability and an estimate of the trend decline in the velocity of circulation of money, a measure of trend growth helps derive the reference value for growth in the broad monetary aggregate M3. Under the second pillar, in terms of the outlook for future price developments, potential output growth and the relation of actual output to potential output (i.e. the output gap) may be useful indicators for assessing the potential for inflationary pressures in the short to medium term. In this regard, potential output and its growth rate represent two indicators among many which are used to assess the outlook for future price developments. However, measures of potential output are unobservable and, therefore, need to be estimated. In the light of these considerations, the present article examines the concept of potential output and assesses the usefulness and reliability of different methods of estimation.

2 The concept and uses of potential output growth and output gap measures

It is possible to view the concept of potential output from different angles. From a purely statistical point of view, it can be seen as the trend or smooth component of the actual output series, without taking a view on the underlying reasons for divergence between actual and trend output. Alternatively, in order to provide some economic rationale, the notion of potential output is often seen as characterising the sustainable aggregate supply capabilities of an economy, as determined by the structure of production, the state of technology and the available inputs. Such a concept of potential output is

different from one where potential output is the level of output which is attainable, in an engineering sense, with the maximum utilisation of the factors of production. In practice, it is not feasible to make full use of all the factors of production, as marginal costs rise steeply (and exceed the marginal benefits) at high degrees of factor utilisation. From an economic point of view, meaningful measures of potential output thus reflect prevailing economic constraints in factor markets.

Various explanations have been given in economic literature as to why actual and potential output often diverge, i.e. the reasons for the emergence of an output gap. One theory is that actual output can differ from potential output, because rigidities in the economy mean that it takes time for prices and wages to adjust. In this case, the output gap is an important measure of the balance of overall demand and supply conditions in the economy and may provide useful information on price pressures. Another theory is that the economy is best characterised by what are known as real business cycle models, where actual output differs from trend output according to random productivity shocks. In this case, the output gap reflects temporary disturbances caused by the adjustment of the production process to technological changes and unexpected developments on the supply side.

As discussed in the introduction, measures of potential output and its growth rate are of relevance to the stability-oriented monetary policy strategy of the ECB. Another important application of these measures arises in the cyclical adjustment of macroeconomic series. For instance, output gaps are used, along with measures of the sensitivity of government revenue and expenditure to cyclical movements in output, to produce an estimate of a cyclically adjusted government budget balance. Such adjustments may be useful in deriving an indicator of the underlying stance of fiscal policy. Also, in terms of monitoring developments in international competitiveness, output gaps are used to calculate real exchange rates based on cyclically adjusted unit labour costs.

It is important for economic policy-makers to be able to detect at an early stage whether the rate of growth of potential output has changed. Given that developments in the capital stock and the labour force tend to be rather smooth, potential output growth is expected to have a rather low variability. Nevertheless, pronounced changes in the rate of growth of potential output can arise as a result of structural change, for instance stemming from institutional reform in the labour market. However, as is discussed in the following section, conventional measures of potential output growth predominantly reflect the institutions and structures existing at the time.

3 Methods for estimating potential output growth

There are a variety of methods available for estimating potential output and they can be grouped into two broad categories: the "production function" and "statistical" approaches. The former attempt to create an explicit model of the supply side of the economy using economic theory. The latter attempt to break the real GDP series down directly into a trend and a cyclical component.

3.1 The production function approach

The relationship between potential output and its structural determinants can be encapsulated in a production function. This framework is useful for explaining the key economic forces underlying developments in output and growth in the medium term. In this regard, there can be important effects from changes in labour supply (stemming from demographics, migration, shifts in labour participation or changes in structural

unemployment) or rates of capital accumulation. A simple growth accounting exercise for the euro area, with capital and labour as inputs, reveals that for the period from 1991 to 1997 the contribution of capital growth to total output growth was 67%, the contribution of employment growth -13% and the residual contribution, reflecting a number of productivity-related factors including technological change, was 45%. Over the longer term, technological change is likely to be the dominant factor systematically leading to growth in per capita output.

The production function approach relates output to the level of technology and factor inputs, usually labour and capital. There are a number of possible functional forms of production functions, including the widely used "Cobb-Douglas" production function. The following formula gives an example of a Cobb-Douglas function with capital and labour as inputs:

$$Y_t = A_t K_t^{1-\alpha} N_t^{\alpha},$$

where Y_t is potential output, A_t is the trend component of total factor productivity, K_t is the capital stock and N_t is the trend component of the effective labour supply (i.e. the labour force adjusted for equilibrium unemployment). The parameter α is the elasticity of labour with regard to output: if labour is increased by 1% and capital is held constant, output increases by α%. Under the conditions of perfect competition, parameter α should coincide with the labour share. In the context of growth accounting, this restriction is used to obtain the "Solow residuals", that is to say the part of output growth not accounted for by labour and capital accumulation. However, direct estimation of the production function can yield values for α which differ markedly from the labour share, as measured by the national accounts, possibly reflecting the fact that an assumption of perfect competition does not hold true at the economy-wide level.

The production function approach is widely used, for example by international organisations such as the OECD and the IMF, to derive estimates of potential output. It is often seen as providing a comprehensive economic framework for estimating potential output, which makes a clear link between output and its long-term fundamental determinants. It can thus be used to assess the impact of structural changes and policies on potential output, although this requires a quantitative assessment of the impact of such changes on the key determinants, such as structural unemployment and total factor productivity. One advantage of the production function approach is that it is possible to forecast potential output, as forecasts of important components (employment, unemployment, investment and therefore capital stock) are usually available.

Nevertheless, there are also certain disadvantages associated with this approach:

- It is subject to important data problems, most notably it is usually the case that measures of capital stock are not very reliable and data on hours of work are often not available.

- The production function approach relies on deriving measures of the trend components of the inputs. However, large fluctuations in productivity levels and labour supply make it very difficult to disentangle the trends of both total factor productivity and the labour force. For example, there is a wide range of views on how technological progress should be modelled. Similarly, alternative views are held with regard to the trend levels of the effective labour supply which, to some extent, depend on the presence of rigidities in labour markets. Different assumptions of these trend components will lead to very different estimates of potential output.

3.2 Statistical methods

These methods of estimating potential output are essentially based on the idea of extracting

the trend from the output series using statistical techniques. They can be divided into two categories. The first comprises "univariate approaches", which include methods which extract the trend from the information contained in the output series in isolation, without using the information contained in other variables. The second comprises methods which attempt to extract the trend using the information in the output series in conjunction with information contained in other variables, most notably inflation. In doing so, these techniques typically attempt to take into account empirical relationships, such as the short-run Phillips curve, which postulates an inverse relationship between unemployment (or changes in unemployment) and inflation.

Univariate approaches

Univariate approaches derive an estimate of potential output by essentially fitting a trend through the series. A key feature of these methods is that they are relatively simple to use. There are a number of univariate approaches available and although it is not possible to provide an exhaustive description of all the techniques, two methods give an indication of the approaches that can be followed. These are the "Deterministic Trend" and the "Hodrick-Prescott filter" approaches. As the discussion of these methods below indicates, a key issue is the degree of smoothness in the estimates of trend output.

The Deterministic Trend method is based on the assumption that the trend component of output is a linear function of time. The output gap is simply the deviation from this linear trend. An obvious advantage of this method is its simplicity, as it essentially involves drawing a straight line through the actual GDP series, amounting to a constant rate of trend growth. However, it requires the very strong assumption that trend output is merely a function of time, irrespective of developments taking place on the supply side of the economy. As an alternative, the method can be adapted to allow for breaks in the trend. Examples of breaks in the trend are instances of major structural change, such as those associated with the oil shocks at the beginning and end of the 1970s. However, it is usually necessary to apply this method to full economic cycles in order to derive a reliable estimate of the trend output growth, as otherwise this estimate could be heavily influenced by cyclical developments. This places a considerable limit on the ability of this method to generate estimates of the current level or growth rate of potential output.

A very widely used approach in the estimation of potential output is the Hodrick-Prescott filter (e.g. this method is used by the European Commission in preparing estimates of the potential output gap which are used for the cyclical adjustment of government budget balances). This filter extracts a trend component by trying to balance a good fit to the actual series with a certain degree of smoothness. A key parameter of the filter, usually referred to as "lambda", determines the respective weight given to each of the two characteristics. If "lambda" is infinite, then all the weight will be on a high degree of smoothness leading to a linear trend. If "lambda" is zero, then all the weight will be on goodness of fit to the original series and, hence, the estimated trend will always be the same as actual output. A key advantage of the Hodrick-Prescott filter is that it is also relatively straightforward to implement. However, the filter has a number of major drawbacks:

- The filter can mistakenly identify a cycle in a series which has none.

- The choice of the smoothing parameter lambda is critical, but ultimately arbitrary. Usually, a fixed value of lambda is used across countries, but this is not particularly satisfactory, as the speed of adjustment to shocks is likely to differ.

- The filter suffers from the poor reliability of the end of sample estimates, which limits

its usefulness for estimating the current value of potential output. One common attempt to deal with this problem is to extend the GDP series by incorporating forecasts for GDP. However, in this way the current estimate of potential output becomes sensitive to the quality of the macroeconomic forecasts used to construct it.

- The method, along with other univariate methods, takes no account of either information contained in other series which may help to identify the trend-cycle breakdown or economic theory.

- The filter will smooth structural breaks, even if these imply clear shifts in the level or the rate of growth of output and, therefore, it generates misleading estimates of potential output over these periods.

Multivariate approaches

In order to overcome some of the drawbacks inherent in the univariate approaches, a number of multivariate statistical techniques have been developed and have become increasingly popular in academic literature in recent years. These methods derive an estimate of potential output using information from more than one series and they can also make use of well-known economic relationships. One widely used multivariate approach is the unobserved component (UC) modelling technique. This typically builds upon a Phillips curve and adopts the strategy of defining a common cyclical component for output and inflation. Economic theory is incorporated in this approach insofar as there is a theoretical short-run link between inflation and a measure of economic activity arising from the existence of price and wage rigidities.

Multivariate methods have a number of advantages and limitations:

- While being less straightforward to implement than the univariate methods, these methods are easier to adopt than the production function approach as they require a simpler modelling approach and reduced data input.

- In common with the production function approach, they also take account of the information contained in economic series other than real GDP and incorporate some elements of economic theory.

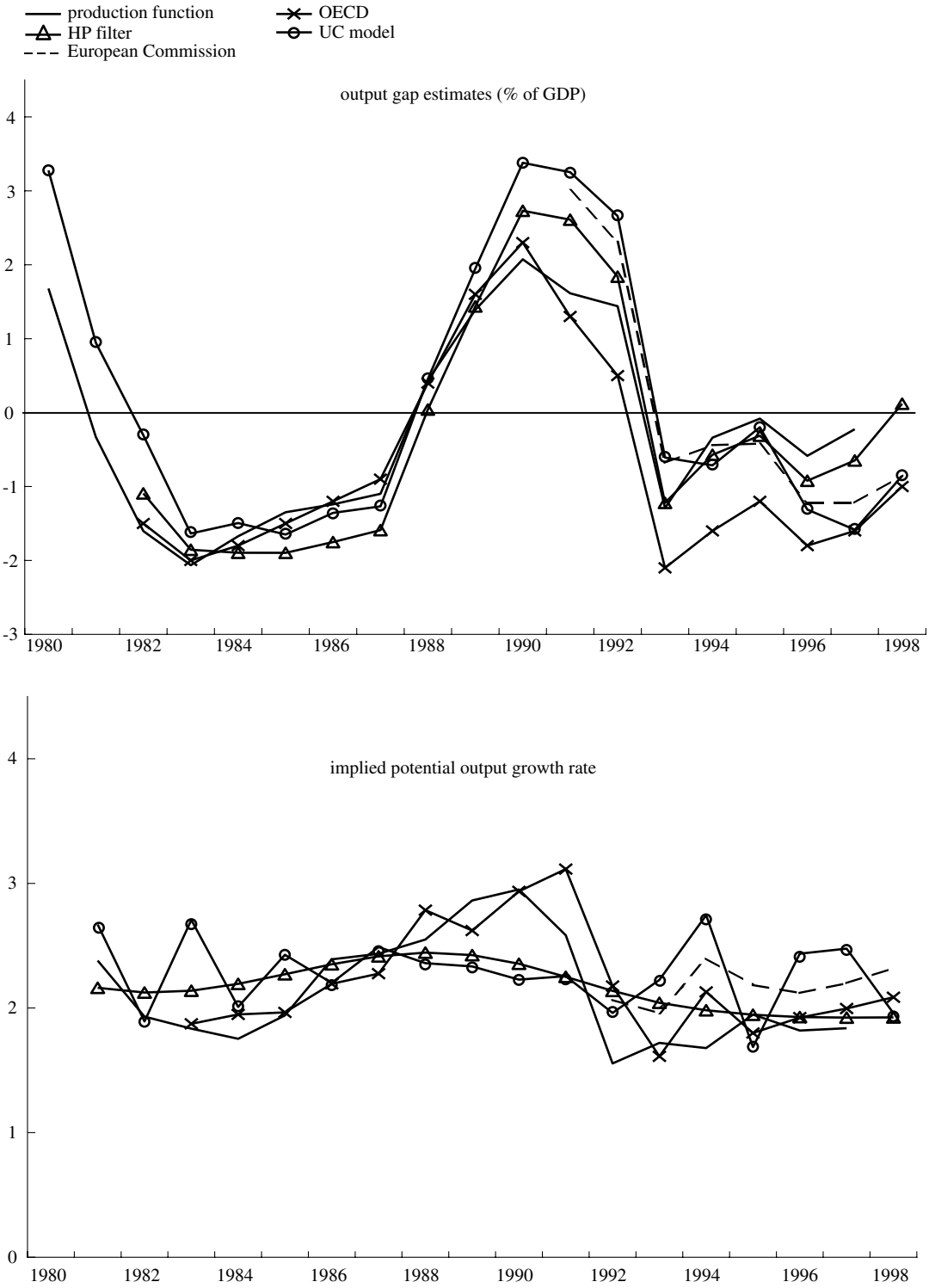
- However, important technical details relating to the implementation of these methods can have very significant effects on the magnitude of the estimates of potential output. For example, with the UC model an assumption has to be made with regard to the underlying trend component of a series (e.g. whether it is a deterministic, smooth or volatile trend).

4 Estimates of the output gap and potential output estimates and sources of uncertainty

In order to illustrate the methods discussed above, Chart 1 gives details of five different estimates of the output gap and the growth rate of potential output for the euro area for the period from 1980 to 1998 (not all the estimates are available for the entire time period). These are estimates from the OECD (based on a production function approach) and an estimate from the European Commission (based on the Hodrick-Prescott filter), together with some illustrative estimates that we have prepared using a production function approach, a Hodrick-Prescott filter and an unobserved components model. As the chart shows, for some years there has been great variation in the magnitude of the output gap estimates generated by these methods. For example, the range of estimates for much of the 1990s is close to 1½ percentage points. Moreover, the estimates generated by these methods do not always match in terms of identifying

Chart 1

Output gap estimates as a percentage of potential output and estimates of the potential output growth rate in the euro area between 1980 and 1998



Source: ECB calculations, OECD and European Commission.

Notes: The Hodrick-Prescott filter output gap estimate reported (HP filter in the chart) has been computed for a parameter lambda equal to 100. This is the standard value when using annual data. The average length of the cycle implied by this parameter lies between five and eight years. For the sake of simplicity, no forecasts of future values of output have been used to extend the sample. The value for the trend components of total factor productivity and employment used in the production function estimate has been computed by means of the Hodrick-Prescott filter (again with a lambda parameter of 100).

whether output is above or below potential (for instance in 1998). Nevertheless, the different estimates do coincide in terms of the timing of cyclical turning-points.

In respect of the growth rate of potential output, the figures show that although there is some variation between the different estimates in any one year, this variation is generally less, as a percentage of output, than in the case of the output gap series. Furthermore, there is a high degree of consistency between the estimates of the growth rate of potential output over the longer term. For the period from 1982 to 1998, the average of the estimates of potential output growth available are all close to 2.2%.

These figures highlight the uncertainty surrounding estimates of potential output and the output gap. In this respect, it is useful to distinguish between three sources of uncertainty: model, parameter and data uncertainty, all of which are considered in more detail below.

Model uncertainty

As discussed above, there is no common view on the appropriate specification of a model to identify potential output. Instead, the academic literature develops a number of approaches which build on different statistical methodologies and macroeconomic theories. The lack of unanimity on the appropriate formal framework to be used affects statistical approaches as well as those based on explicit models of the economy.

The sensitivity of estimates of potential output to the specification chosen is particularly relevant in the case of univariate models compared with multivariate models and the production function approach. It is generally difficult to impose restrictions from economic theory on univariate methods. By contrast, multivariate methods and the production function approach are guided by views on the relationships between relevant macroeconomic variables, which make them less arbitrary and potentially more robust (to the extent that the macroeconomic relationships are supported by the data).

Parameter uncertainty

Even if there were common agreement on the appropriate modelling strategy to measure potential output, estimates would be subject to parameter uncertainty. This is because there is no unique way of determining the parameters of a given model, as the true value of these parameters is not known and has to be either assumed or estimated using econometric techniques. Parameter uncertainty becomes severe when the true parameters of the model change over time as a result of structural changes in the economy. The effects of such changes can only be identified some time after they occur. For example, the reduction in aggregate productivity growth in the United States after 1965 would clearly have affected the rate of growth of potential output to an extent that would not have been fully apparent at the time. In Europe, there has been a significant amount of structural change over the past decade which is still ongoing. This structural change stems from developments such as German unification, the introduction of the Single Market and Economic and Monetary Union. The failure to identify a shock to potential output implies that the estimated magnitude of the output gap may be incorrect and may, therefore, be a misleading indicator.

The discussion above illustrates the point that as more information becomes available it is possible to become more confident about the true parameters of a model. Therefore, an estimate of the output gap for 1990 made at the time is likely to be less reliable than one made today, given all the information that has become available in the intervening period. In order to illustrate the magnitude of this parameter uncertainty, Chart 2 plots the sequential estimates of the output gaps generated using a Hodrick-Prescott filter starting with a sample from 1971 to 1980 and

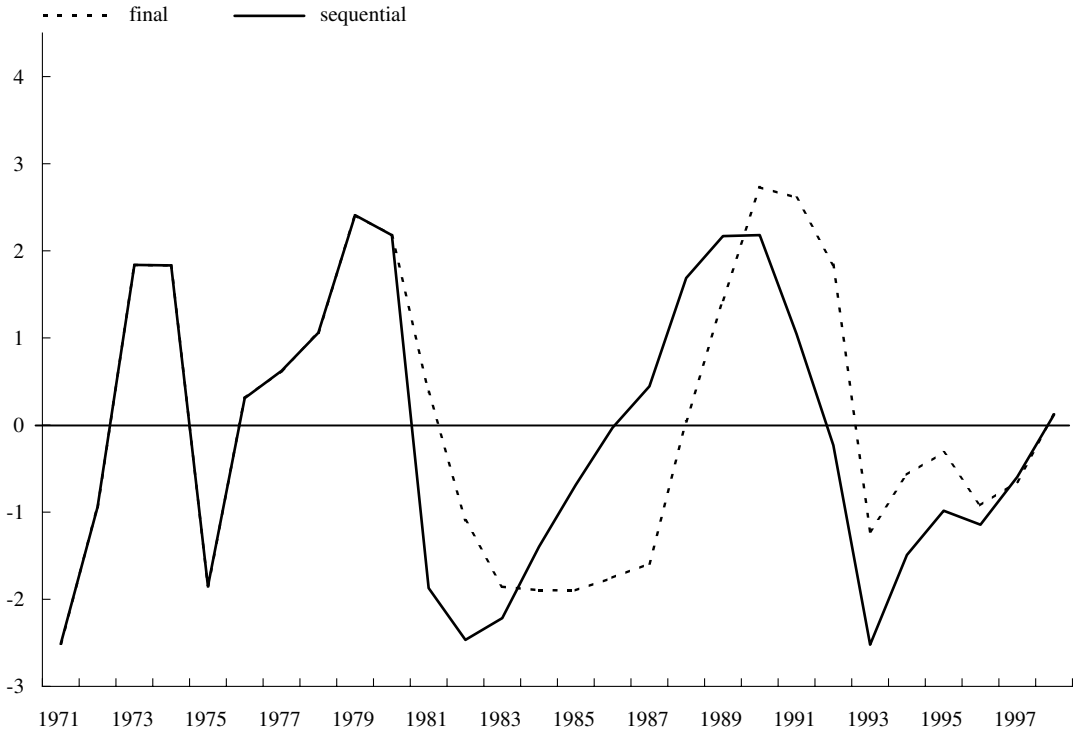
gradually prolonging the estimation period by three months at a time. For the purpose of comparison, the corresponding values of the output gaps using the whole data set up to 1998 are also included and this reveals that the effects of parameter uncertainty can be quite substantial. For instance, the discrepancy in sign between the two estimates for 1988 implies that what is estimated to be a positive output gap using data up to 1988, is in fact revised in the light of subsequent observations so that the output gap for 1988 is negative in the final estimation.

Data uncertainty

The exact value of many economic series is not known immediately and there can be quite some delay before reliable estimates become available. This is particularly the case for estimates of output, where initial estimates are typically available with a few months' delay and can then be subject to significant revisions in subsequent months or even years. This data uncertainty about recent output figures will, to some extent, have an impact on the estimates of the potential output growth rate, particularly when univariate methods are used. Furthermore, other economic series which may be used in the construction of estimates of potential output in the multivariate and production function approaches – such as information on prices, labour supply and capital stock – are also subject to data uncertainty. However, the impact of uncertainty with regard to recent data on the estimates of potential output growth is unlikely to be large if long time series have been used. More importantly, as the most recent value of actual output is used directly in the calculation of the latest output gap, data uncertainty will have a much greater impact here. This matters because it is the estimate of the output gap in the more recent period that is a potentially useful cyclical indicator for policy-makers.

Chart 2

Final and sequential output gap estimate using the HP filter in the euro area between 1971 and 1998



Source: ECB calculations.

The combined effects of parameter and data uncertainty can be assessed through the concept of the "reliability" of output gap estimates. An output gap estimate can be considered to be reliable if there are no sizeable discrepancies between the real-time and ex post estimates. Recent research using US and euro area data shows that output gap estimates are not, in general, highly reliable, particularly when univariate methods are used. Indeed, initial estimates are typically subject to sizeable revisions as new observations become available. Results using US data also indicate that uncertainty arising from parameters and the model specification is relatively more important than data uncertainty.

Estimates of future potential output growth and structural change

While the figures presented in Chart 1 suggest that it is not possible to be very confident about the magnitude of the output gap in any particular year, they do indicate that it is possible to be far more confident that, in recent decades, potential output growth has been around 2-2½% per annum. However, this degree of confidence in the historical estimates of potential output growth does not apply to the estimation of the current or future growth rates of potential output, which are clearly of relevance to a forward-looking monetary policy. For example, if there were a significant change in the pace of technological progress, or in the structural environment, this could alter the trend growth rate.

In assessing the likely growth rate of potential output in the coming years it is important to take account of structural developments in the economy that may have a bearing on the supply side. As discussed in the box, there is, for instance, the potential for New Economy-type developments to stimulate growth through structural and technological changes. More generally, it is necessary to consider other supply side factors, such as developments in the capital stock and the labour market. For example, as discussed in the third section of this article, labour supply plays an important role in determining potential output growth. In many European countries, population growth is slowing and the average age of the population is increasing. However, this may, to some extent, be offset by an ongoing rise in female labour force participation and there may also be scope for a rise in the participation rate of older workers. Furthermore, given the deep-seated problems of the euro area labour markets, which are predominantly structural in nature, reforms which actually lead to a fall in structural unemployment, or which raise labour force participation, should contribute to an increase in potential output (see, for instance, the May 2000 issue of the ECB Monthly Bulletin for a discussion of structural factors affecting labour markets and European labour market initiatives).

The potential for New Economy-type developments and more general structural changes highlights the limitations of estimates of potential output growth which are based on historical information on the economy. However, while it is relatively easy to identify possible reasons why the growth rate of potential output may have changed, it is far more difficult to assess whether it has actually changed, let alone to quantify the magnitude of any change. Therefore, when it is suspected that such change may be occurring, it is particularly important to take account of all the other relevant sources of information in order to gauge the extent to which they support such a conjecture.

Box

The notion and main elements of the New Economy

Based on the experience of the United States, the hypothesis underlying the notion of a New Economy is that structural and technological changes have raised the rate of growth of potential output. This notion has steadily gained support as US economic growth in the current cyclical upturn has proven to be stronger and much more durable than in previous upturns. At the same time, the pattern of wage developments and price inflation also appears to have changed. While the rate of unemployment has fallen to historically low levels, wage developments have been more moderate than would have been expected. Instead of rising, as was generally the experience in earlier periods of strong economic expansion, inflation actually fell in the 1990s.

This remarkable performance has given rise to the idea of a new economic era in the United States. Pronounced changes in information and communication technology (ICT) are seen as a major driving force behind this performance, having an impact through a number of channels.

Conceptual issues in the debate on the New Economy

At this juncture, while there appears to be no generally accepted definition of the New Economy, there seems to be a broad consensus that it encapsulates two main elements. First, higher potential output growth, reflecting mainly an increase in the trend growth of total factor productivity. Second, a permanent reduction in the rate of structural unemployment. The first of these elements could entail two distinct interpretations. One is that the long-run *growth rate* of output and productivity has increased compared with the corresponding rate in previous periods. Alternatively, it could be that the *level* of output and productivity has increased, but that some time is required to reach this new level and, therefore, this gives rise to temporarily higher rates of growth. Indeed, to the extent that the ICT revolution is seen as characterising a major technological innovation, giving rise to secular growth phenomena, the time that is needed to reach this new level could be quite protracted and could be in the region of decades. However, while there is broad consensus about the main defining features, the issue of how enduring these two elements would have to be for an economy to qualify as a New Economy remains the subject of ongoing discussion.

There are considerable difficulties in making inferences about the existence of a New Economy based on data relating to these elements. While there is growing microeconomic evidence that ICT is having substantial beneficial effects on the productive and organisational efficiency of businesses, the macroeconomic evidence to date is limited. There are, for instance, considerable measurement problems, such as those related to the application of appropriate deflators to derive constant price data. The experience gained in the United States has shown that the measured price increases in national accounts are typically being overstated, with the result that the growth of real variables is being understated.

While the possible existence of a New Economy in the United States is still subject to debate, it has already led to speculation as to whether the experience of higher, non-inflationary growth could be repeated in Europe. With regard to the euro area, some progress has been made, but there are still significant differences compared with the United States, in particular in respect of structural rigidities, which diminish the benefits of the new technological possibilities. Strengthening the process of structural reform increases the chances that strong output growth in an environment of price stability, as is currently being observed in the euro area, can be sustained in the years to come.

Driving forces behind the New Economy

There are a number of mutually reinforcing driving forces behind the New Economy. First, the pronounced technological progress in the fields of data processing and telecommunication is believed to have led to higher

productivity. Facilitated by ICT advances, productivity may also be benefiting from improved organisational techniques. Second, there is the ongoing process of globalisation and increased competition, which, in the goods and services markets, has stimulated efficiency and reduced pricing power. A direct favourable effect on inflation is also resulting from significant output price reductions in the ICT industries themselves. Third, a highly competitive and unregulated market structure is facilitating the emergence of new lines of business and the diffusion of the new technologies to promote efficiency. This is aided by a deepening of financial markets – again facilitated by ICT advances – which enables entrepreneurs in high-tech sectors to start up new lines of activity. Finally, one major ingredient is a favourable policy environment, comprising a commitment to sound public finances and price stability at the macro level, and a business-friendly low-tax and low-regulation approach at the micro level. This reduces uncertainty, encourages entrepreneurial activity and facilitates the expansion of investment.

All of these benign influences have been observed individually and, in the case of the US economy, a number of them appear to have come together over the past decade, producing the remarkable performance seen in terms of high growth at comparatively low rates of inflation. With regard to the euro area, progress has been made in a number of areas, but there remain significant differences compared with the United States. In particular, realising any potential for a New Economy in Europe will depend, in a fundamental way, on the adoption of a comprehensive process of structural reform.

5 Concluding remarks

Measures of potential output are of interest to the ECB as a gauge of the sustainability of economic developments. This is relevant in the context of its stability-oriented monetary policy strategy for two reasons. First, as a measure of trend output growth in the derivation of the reference value for monetary growth and, second, in the broadly based assessment of the outlook for future price developments.

In choosing between the various approaches for estimating potential output described in this article, there is inevitably some trade-off between the degree of simplicity of the individual approaches and their ability to take into account the insights of economic theory. Different methods tend to yield broadly comparable estimates of the rate of growth of potential output, particularly over longer time horizons. As a consequence, the different methods also give similar estimates of the change in the output gap. However, estimates of the level of the output gap at any particular point in time tend to be surrounded by a greater degree of uncertainty. Therefore, while the level of the output gap is a potentially useful indicator of short-term price pressures, particular caution is required in drawing policy conclusions based on such estimates.

A general difficulty with estimates of potential output based on historical data is their inability to adequately capture structural change until well after the change has occurred. Therefore, if important structural changes are believed to be under way, then this would point to a correspondingly greater degree of uncertainty surrounding estimates of current potential output. In these circumstances, a forward-looking monetary policy aimed at the maintenance of price stability would clearly wish to take account of the possible effects of such developments. However, in doing so there would be a need for clear and compelling evidence that structural changes were actually leading to a change in the growth rate of potential output. This can only be achieved by means of a general assessment of all other relevant sources of information on the economy.