ESTIMATES OF POTENTIAL OUTPUT: BENEFITS AND PITFALLS FROM A POLICY PERSPECTIVE

Jean-Philippe Cotis, Jørgen Elmeskov, Annabelle Mourougane
OECD Economics Department

Introduction

1. Notions of potential output and output gaps are crucial in many areas of economic policy. Unfortunately, however, they are not observable. Indeed, important policy mistakes have been made because perceptions about these concepts turned out to be wrong. Potential output and output gaps are also important concepts in economic research aimed at improving our understanding of how economies work. Here again, getting it right is important for the conclusions that are subsequently drawn.

2. Many things have changed over the past two decades in the way economists assess potential output and output gaps. Potential output was originally estimated through mechanistic linear trends or relatively "inert" production functions. This set-up proved ill-suited to coping with large supply side shocks and their ensuing inflationary pressures. Indeed, over-expansionary macroeconomic policies during the 70s and the 80s were probably encouraged by wrong diagnostics based on over-optimistic estimates of potential output.

3. Since then, there has been a proliferation of both conceptualisations and estimation techniques of potential output. It might be thought that this abundance should have helped policy makers. However, economic developments over the past few years do not provide strong evidence of such an effect. In Europe, a number of "policy failures" have emerged in the form of insufficient fiscal consolidation and lack of structural reforms during good times. As a result, both the Stability and Growth Pact and the Lisbon Strategy are now under great strain. Many reasons can be found to explain these difficulties, including failure at the political level to act on existing information. But it was probably also the case that new methods to estimate potential output, for a variety of reasons, helped nurture complacency about the magnitude of potential growth.

4. The blame for this less than satisfactory situation must be shared among those involved. All in all, academics have not provided policy makers with the kind of systematic large-scale assessment of strengths and weaknesses of different methods that was needed. And policy makers have not taken seriously enough the potential for improvement which the new methods could bring to the policy process. Bridging the gap between the policy and the academic spheres is obviously important. This paper attempts to make a very modest contribution in that direction by, first, providing a critical overview of the existing state of affairs from a political economy point of view and, second, suggesting and applying a set of criteria that may help policy makers assess the usefulness of various potential output estimates.

I. Choosing from a wealth of methods: a political economy perspective

Balancing methodological abundance with requirements for policy advice

5. In principle, methodological abundance should be a boon by providing for greater diversity and increased sophistication:

- Diversity should provide greater scope for cross-checking of diagnostics which is helpful both when different methods give broadly the same answers to a given question and when they yield different answers – in the latter case by stimulating frank debate and the selection of the best tools. Diversity may also lead to optimal specialisation according to uses; a simple stochastic filter may prove more cost-effective for historical analysis than other, "heavier" approaches, better suited for forward-looking, long-term structural analysis.

- Greater sophistication should also be a plus. In the case of pure statistical techniques, the stochastic approach has improved our capacity to better capture the inherently unstable nature of potential output. And stochastic techniques can also enrich and improve economic approaches through creative hybrids (structural VARs, multivariate filters, etc.). These techniques impose lighter economic priors than at least some of the alternatives and thereby allow "data to talk more freely".

6. In practice, absorbing and processing information is far from costless. The risk is that the wealth of new instruments is so overwhelming that they are not harnessed correctly, leading to confusion and policy errors. Over and above the costs of dealing with a wealth of information, too much methodological diversity may also be incompatible with the nature of the policy making process. If not for other reasons, this is so because communicating efficiently with public opinion and high level policy makers is not workable in practice without a single set of figures for potential output and output gaps. As a result, the only way for methodological diversity to feed into policy advice is as "ancillary complements" to the results of a preferred method. There is thus a stark contrast between academic life, where intellectual eclecticism and experimentation are key, and the cautious world of policy making.

7. International policy co-ordination further constrains choice. In such a context, transparency speaks in favour of countries jointly choosing the preferred method. The alternative of national diversity might well impair co-ordination. That said, diversity may be justified if country-specific policy needs are really very different and indeed no method is deemed universal enough to address all issues. For international organisations, applying the same methodology across member countries would seem a pre-requisite for consistency in policy advice.

8. The choice of a preferred method for policy involves further constraints. It is not enough that a method is convenient for analytical purposes and for interpretation of past economic experience[1]. Much more important is whether the method will provide reliable "policy feedback". This criterion implies taking into account that policy decisions are taken at sample end-points. Hence, extensive appraisal is needed of a method's behaviour at decision time. This evaluation should not be mechanistic, it is also of utmost importance to endogenise the behaviour of experts and policy-makers who will use the method under a typical set of political and institutional constraints and incentives. In this respect there is clearly a strong suspicion that many methods may pass the analytical test while failing to provide a reliable source of policy feedback (cf. *infra*).

1. In this respect it is nonetheless reassuring that most methods tend to give a reasonably similar account of potential output developments over the past two decades in industrial countries.

Identifying a preferred method for policy advice

9. With a need to identify a preferred method for evaluating potential output, it would be good if this method could be seen as having "absolute advantage" in all policy areas. This, of course, may not necessarily be the case. As a second best, the aim could be to find a method with strong performance in the highest priority policy area, under the constraint that it displays reasonable effectiveness in other policy areas. This method should also meet a number of core requirements which are discussed in section 2 of the paper (transparency, internal consistency, replicability, parsimony in data requirements, etc.).

10. Before choosing a preferred method, policy makers should first agree then on their uppermost policy priority area. We would argue that revealed priority lies in the area of structural long-term issues. For example, at the Lisbon Summit, Europeans have set themselves a very ambitious target for trend growth of 3 per cent per year over the decade. Meeting this objective will imply substantial reforms in labour, product and capital markets. To harness these policies it will be important to rely on estimates of potential output that can serve, at least in a broad sense, as a source of "policy feedback". This also implies that estimates of the impacts of reform in these markets should be able to feed back into estimates of potential output. As a second example of revealed priority, economic reforms are now taking the lion's share within the Broad Economic guidelines of the E.U. Against this background, economic approaches of the production function type, seem to be the unavoidable first choice.

11. Structural long-term challenges extend beyond optimising long-term growth and employment. They also include sustainability issues associated with demographic ageing. Analysis of policies to address the long-term sustainability of public finances requires again an economic approach, using actual demographic projections as a core input for both potential output and public spending projections and considering, for example, how different paths for unemployment and labour force participation might affect outcomes in these areas.

12. Because sustainability issues are becoming more and more pressing there is also an increasing willingness to cast short-run fiscal surveillance, in the Stability and Growth Pact sense, in a unified and more medium-term oriented framework. Hence, fiscal surveillance and sustainability assessment need to rely on the same methodology. For those institutions that still rely on stochastic filters for short-run fiscal assessments it will become increasingly difficult to use two different methods for what amounts to a common purpose.

13. A counter-argument might be that strong automatic stabilisers in Europe reduce the need for discretionary policy feedback and the associated need for information on output gaps so as to calibrate a policy response. However, strong automatic stabilisers do not always eliminate the need for fiscal discretionary policies and the use of indicators such as structural deficits and output gaps. Should a slowdown reflect a negative supply side shock it would be unwise, for example, to let automatic stabilisers play and allow the spontaneous weakening of tax receipts to deteriorate the budget balance. Furthermore, it may also be useful for policy makers to convince themselves that, indeed, automatic stabilisers are effectively impacting the economy and that fiscal impulse do not go beyond what was intented to be normal automatic stabilisation.

14. There are of course other policy issues where potential output and output gap estimates may be important, such as the pursuit of price stability. Indeed output gaps can be seen as summary indicators of the balance between demand and supply conditions in the economy. And, in this regard, they can provide a useful proxy for underlying inflation pressures. Output gaps can thus be considered as leading indicators of inflation to be used, *inter alia*, in the conduct of monetary policy.

15. The methodological requirements for short-run monetary stabilisation policies may be less than for medium-term issues. After all, there are many other conjunctural indicators available besides output gaps. Rates of capacity utilisation and business expectations also provide useful and more timely information about discrepancies between aggregate supply and demand. When they think about short-run stabilisation, monetary policy makers use a plurality of short-run indicators as a source of policy feedback. Output gaps contribute to the policy making through thought provoking normative instruments like the Taylor rule but this is one among many ways to approach monetary policy.

16. Because monetary authorities are often independent and not formally involved in policy co-ordination with the authorities in charge of fiscal and structural issues, there is less of a mechanical need to use similar assessments of potential output. Nevertheless, even in the context of informal policy dialogue there might be a good case for monetary and other public authorities using at least broadly similar methodologies to assess output gaps and potential output. Were for instance central bankers to complain about inadequate structural reforms, it would certainly help them to be able to specify the areas (employment, productivity, capital accumulation) most in need of improvement and their potential contribution to medium term disinflation.

17. Based on this examination of policy makers' needs there seems to be a strong case for choosing the production function as a "flagship" method, likely to cover all areas of policy. Nonetheless, as argued above, this is not sufficient to make it the preferred method. Its "technical" performance against alternative methodologies should not be dramatically weaker across a variety of criteria (internal consistency of the method, transparency, information needed, possibility to estimate uncertainty, reliability of estimates at the end of the sample, etc.). Fortunately, the conclusions of the survey conducted in section 2 of this paper turn out to be reassuring when it comes to the feasibility of applying the production function approach. Even so, it is not without specific drawbacks. It is for instance impossible to evaluate in a formalised manner the degree of uncertainty attached to potential output estimates. Production functions also have to rely, like other methods on judgmental choices. However it is of particular importance for the policymaker to note that a relative point of strength of the production function approach is its reliability at sample end point.

18. Past experience with HP filtering, in the context of European fiscal surveillance, illustrates that not every estimation method has passed this difficult test. With hindsight it has indeed appeared that the sample end point weaknesses characterising HP filtering are beyond practical remedy given the way "official forecasters" generally operate. It is now clear that a combination of HP filtering and "back to average growth" forecasting provided policy makers with an inflated view of potential output growth and structural budget positions during upswings such as in the late '90s and early '00s. Policy mistakes made in that period might have been mitigated had we disposed of more robust warning signals in the area of output gaps and structural fiscal balances.

19. Rather than standing on their own, stochastic filters may be put to best use within a production function framework. Estimating time varying NAIRUs through Kalman filters, filtering total factor productivity and participation rates are good examples of how to design flexible economic approaches to potential output. Filtering such individual components of potential output should not lead, in practise, to the kind of acute end point biases which have been observed when applied directly to "politically charged" GDP series. In those areas, forecasters will indeed feel less pressure to prolong the sample using back to the average growth assumptions.

20. Stochastic filters may also have a role as ancillary methods to directly estimate potential output, thereby helping to explore the margins of uncertainty surrounding the production function approach. The use of alternative, possibly sophisticated methods for "private use" might also be a good way for official experts to keep up with state of the art approaches and to prepare for future methodological improvements.

How far are we today from the optimal framework?

21. Over the recent past, national and international policy makers have taken a much stronger interest in potential output estimates. This renewed interest has been associated with a sense of frustration. The diversity of methods chosen by national and international agencies has been rightly perceived as an impediment to good dialogue and co-ordination. The absence of systematic performance assessment of existing methods has also been regretted. Serious efforts have been made to overcome this lack of collective organisation but it is too soon yet to tell how much progress has been achieved.

22. We have not yet reached the situation where international organisations and national agencies are all sharing reasonably similar instruments based on an economic approach. But there has been substantial convergence over the recent past. The OECD, the EU and to an extent the IMF, are using flexible production functions for country assessments (see Box 1 for more details). Protocols to use these methods within the framework of international organisations have also become much more rigorous. The European System of Central Banks (ESCB) has taken a different course with a specialisation of methods according to policy areas. This polycentric approach may not facilitate its participation in the policy debate, in particular on the fiscal side, where, unlike the OECD and the EU, the chosen method centres around HP filtering. The drive towards better assessment and harmonisation internationally has also had very positive spillovers at the national level.

Box 1. The computation of potential output and output gaps and their use in international organisations

In the context of its bi-annual economic outlook, the OECD estimates potential output and output gaps for member countries. The methodology is based on a Cobb-Douglas production function[1] with Harrod neutral technical progress and is extensively described in Giorno *et al.* (1995). The approach is hybrid in the sense that it relies on economic relationships (to estimate the NAIRU, see Richardson *et al*. (2000)) but also univariate filters (most of the time the HP filter) to calculate trend participation rates, trend hours worked and trend total factor productivity. Moreover, the 'trend' measures that enter in the determination of potential can be qualified by the judgement of country desks[2]. In recent years, certain changes have been introduced. The resulting output gap is generally well correlated with an output gap generated by an HP filter (the correlation being close to 0.9 for all the G7 countries, except Germany for which it is only 0.4). Potential output and output gaps are used to compute cyclically adjusted fiscal variables to assess fiscal policy stances (see Giorno *et al.* (1995)). Estimated NAIRUs and potential outputs are also used to give an ex-post assessment of the impact of past structural reforms.

The reference method used by the EU Commission to assess the stability and convergence programmes has recently changed[3] to an estimation of potential output very close to the OECD's i.e. making use of a Cobb-Douglas production function with exogenous trend (see Denis et al. (2002)). The principal differences with the OECD method are that the EU Commission estimations do not incorporate hours and that a NAWRU rather than a NAIRU is estimated (i.e. using information on wages rather than on prices)[4]. In the context of the Broad Economic Policy Guidelines, estimates of potential output are used to assess the impact of structural reforms in product and labour markets. In addition, potential output is also used to study structural changes in the economy (e.g. the emergence of the New economy, see Mc Morrow and Röger (2001)).

By contrast to the latter two institutions, neither the IMF nor the ECB have an "official" method to compute potential output. In the case of the IMF, each country desk chooses the method that fits the country situation best. This approach should be seen against the background that the IMF with its universal coverage has to monitor very heterogeneous countries with marked differences with regards to the coverage and the quality of economic data. Most of the time, however, the IMF approach is based on a production function method, but with underlying assumptions that vary across countries (see de Masi (1997)). A notable exception is the United States for which a number of selected methods including the split time trend but also the HP filter, the band pass filter and the production function is presented (see IMF (2002)). In addition to their use in the surveillance work, potential output estimates are *inter alia* used by the IMF to give insights on structural trends (e.g. developments of sectors) or on the impact of particular supply shocks (e.g. the terrorist attacks in the United States).

The ECB has so far not published any estimates of potential output for the euro area. In its derivation of the reference value for monetary growth, it looks at several estimates of potential output, including those from international organisations, but does not focus on a single one. However, the ESCB makes use of trend measures of various macro-economic series derived by an HP filter in its calculation of cyclically adjusted budget balances (Bouthevillain et al. (2001)). Regarding the outlook for price developments, measures of potential output and the output gaps, together with other indicators are considered (see ECB (2000))

____

1. The only exception is Japan for which a CES production function is applied as the assumption of a Cobb-Douglas specification was strongly rejected by the data (see Turner et al. (1996)).

2. In practise, trend total factor productivity is rarely modified by Desks' judgement.

3. It was previously based on an HP filter applied to actual GDP.

4. Both estimations use the Kalman filter but the EU Commission imposes a zero sample mean of the unemployment gap.

II. Assessing estimation methods of potential output based on the needs of policymakers

23. As seen in Part 1, potential output and output gaps are extensively used in the economic policy making process. The problem is that potential output can not be observed and has to be estimated using available macroeconomic data. A large number of methods have been developed for this purpose. The aim of this part is not to give an exhaustive list of methods that could be used to estimate potential output and output gaps but rather to review the most common methods and discuss their properties in light of certain criteria that are particularly important for policy makers. Therefore new methods or methods that are scarcely used will not be discussed in this paper.[2]

24. It is worth stressing that there is not a single, "right" ranking of criteria, given that potential output and output gaps can be utilised to analyse different policy issues. The priorities will depend on the objectives of the user and the relevant horizon. A user who is concerned with recent developments and short term forecasts will be interested mostly in the output gap measures at the end of the sample, whilst a user who cares more about long-term trends may be content to have estimates for some sub-samples which do not include the latest observations. A user involved with international comparisons will emphases applicability across countries whereas a user involved with a single country may emphases greater richness in economic content. In what follows a distinction will be made between 'core requirements', i.e. universal properties a method needs to have, and 'user specific requirements', which are criteria that are only relevant when estimates of potential and output gaps are used for specific purposes.

Criteria for assessing different methods

Core requirements

25. One obvious criterion is the *consistency between economic priors and the underlying assumptions of the method*. Priors could be based on economic theory or stylised facts. A useful method should be consistent with economic theory. As well, a method that gives stationary cycles may seem desirable. Priors may also concern the statistical properties of potential output, its volatility or the length of the cycle.

26. Because of concerns for accountability, the policy maker needs to use variables that have been estimated in a *transparent* way to motivate his policy choices. Other users also need to have a transparent method to warrant their forecast or their advice. At the limit, the estimates should be easily replicated by other actors. Transparency also helps ensure equal treatment of countries, which is for instance necessary in a context of international budgetary surveillance. In this regard, a desirable feature for a method is that all the underlying assumptions made during the estimation are clearly identified and justified.

27. It is also important that data updates do not imply very large and unwarranted revisions in estimates to ensure the credibility of the method. Indeed, a method that generates very large revisions will be considered as uncertain. In this regard, a method which produces estimates that are *consistent over time* and not too sensitive to the sample period of estimation will be preferable. In particular, the sensitivity of the estimates to the last observations of the sample, which might be markedly revised when new information becomes available, should not be too high.

28. The ability of a method to provide information on the *precision of the estimates* is important as such information may help to indicate what weight should be put on the estimates relative to other

2. For instance, we will not refer to trend linear methods with unknown breaks, Cochrane (1994) method or methods based on survey data.

information at disposal in the policy making process. Here again it is important to make a distinction between the end point (which as already said is of main importance for the policy maker) and the rest of the sample.

User specific requirements

29. A user who is interested in international comparisons and needs to get regular briefings to monitor the economy will be keen to have a method which is easy to update. In this regard, the *quantity and nature of information that is needed* to implement the method is of importance. Indeed if a method requires a very large set of long time series, it will be more difficult to put into practise. Moreover, a method which relies on information which is available only with a long lag will not be helpful in a number of contexts. This criterion will be important also for users interested in countries which still lack reliable data.

30. Most policy makers have to decide on actions that will affect future economic developments, and are thus more interested in output gap estimates at the end rather than at the middle of the sample. In this regard, a particular attention is addressed to the *end point problem* inherent to a number of techniques i.e. the fact that the estimation of the output gap at the end of the sample is mechanically less reliable. For this reason good end-of-sample performance is a crucial requirement. In this context, it is important that the method allows the policy maker to *detect any permanent structural changes* (e.g. a reduction in the level of potential) so that he can adapt his policy reaction adequately. Policymakers may even be more concerned by expected future values of the output gap which is more demanding than simply estimating the gap at the end of the historical sample.

31. It appears that the extension of the sample with forecasts that has been put forward as a method to overcome this problem is not satisfactory as it basically means substituting a transparent method intented to help the forecaster by the forecaster's intransparent judgement. In European countries, it led to an over-estimation of potential output. Indeed, a need to project estimates into the future call for these to be based on variables and assumptions that can be extended from the past. This may frequently imply a need for economic content in methods if projected estimates are to be seen as credible.

32. Obviously, there is no ideal method which would fulfil all these requirements. The objective of the next sections is to assess how the different classes of methods compare in light of these criteria.

Presentation of the methods

33. Few years ago it was easy to classify the different de-trending methods between statistical methods (which only extract information from the series itself) and the economic methods (which use economic information to make the decomposition between trend and cycle), even though from a practical point of view some approaches could use instruments from both classes. In recent years, however a new type of instruments, multivariate filters (sometimes called the semi-structural methods), has been introduced[3]. These methods explicitly supplement the traditional filters with additional economic information (in most cases, a Phillips curve which provides information on inflation developments, an Okun relationship or information on business indicators). Rather than two distinct classes of methods there appears to be now a continuum of methods with no objective criterion to classify them into groups. For the purpose of simplicity, we will however still use the old typology and the distinction statistical/economic

3. These methods were partially developed in reaction to Quah (1992) who showed that without additional restriction, the univariate techniques are completely uninformative for the relative importance of the underlying permanent and transitory components.

methods, economic meaning any method that is based in some measure on additional economic relationships.

34. In this review the focus will be on the most common methods (see Table 1 for an intuitive description of the different methods and Annex 1 for more details). In particular, international organisations most of the time compute their output gap estimates by using one of these methods (see Box 1).

35. It should be recognised that the classification of methods is questionable. Indeed there are some overlaps as some methods can be seen as particular, restricted version of others. For instance, Harvey (1985) shows that the outcome of the HP filter can be replicated by using a state-space decomposition and estimating it with the Kalman filter. Boone (2000) extends this result to the multivariate HPMV. Morley et al. (2003) formulate the Beveridge-Nelson decomposition in a state-space form and show that under certain conditions the trend resulting from the Beveridge-Nelson decomposition and from the Kalman filter are similar. Moreover, some methods classified as economic may explicitly use inputs based on univariate filters (e.g. production function approach with exogenous trend).

36. Without entering into specific technical details it is important to stress the difference between one and two-sided filters as these terms will be used in the subsequent paragraphs. Two-sided filters (like the HP filter, the HPMV or the Baxter-King filters) use both past and future information whilst one-sided filters use only past information. The Kalman filter can be applied as a one or two-sided filter.

Differences in results from the various methods

37. A relevant question is obviously how quantitatively significant the differences between estimates based on different methods really are. Examining this is not an easy task as no exhaustive comparison has been done in the literature. The existing results are often based on selected methods and most of the time for a single country or area.[4] Differences in the scope of the various analyses imply that contradictory results can sometimes be found. As a result, only very tentative conclusions can be drawn from existing analyses.

38. To start with, it is important to see whether the different measures of output gaps give the same description of the cycles. Despite the lack of comparability across studies, it seems that the correlation between the different measures is in general around 0.7/0.9, though it can be much lower in some cases for some methods or some countries (Chagny and Döpke, 2001; Dupasquier et al., 1997; Scott, 2000). This correlation reflects the fact that estimated output gaps generally move in a synchronised manner, although there can be sizeable discrepancies at particular point in times. Mc Morrow and Röger (2001) come to the same conclusion for a sample of EU countries when comparing output gaps estimated with an HP filter and with a production function. Cerra and Chaman Saxena (2000) also obtain broadly similar pattern for output gaps in Sweden for most of the techniques they use.[5] As expected, methods that are closely related (for instance the HP filter, the HPMV and band-pass filters) give estimates with a higher correlation. Moreover, the use of the concordance statistics (which shows whether two measures consistently agree on whether the output gap is positive or negative) indicates that on average the methods do not contradict each other (Chagny and Döpke, 2001; Claus et al., 2000).[6]

---

4. Exceptions are Mc Morrow and Röger (2001), Bold and Els (1998).

5 A notable exception is for Structural VARS for which implausible results were found.

6. The presence of a structural break may alter these results as some methods (linear time trend, HP filter) do not treat structural breaks in a satisfactory way.

Table 1: Methods to estimate potential output and output gaps

Methods	Intuition
Trend	
Linear trend	The trend component of output is a linear function of time. It thus involves a linear regression of the log of real GDP on a constant and a time trend.
Split trend	Trend output is calculated a linear trend during each cycle, where the cycle is defined as the period between peaks in economic growth
Univariate filters	
Hodrick Prescott (hereafter HP)	This filter extracts a trend component by introducing a trade off between a good fit to the actual series and the degree of smoothness of the trend series.
Baxter-King filter	This is a linear filter which eliminates very slow moving ('trend') components and very high frequency ('irregular') components while retaining intermediate ('business cycle') components.
Beveridge Nelson decomposition	This method imposes restrictions on the trend and the cycle to identify the decomposition trend/cycle
Kalman filter	This technique assumes that macroeconomic time series are composed of trend cycle and erratic components, which are not directly observable. These three components can be recovered by imposing sufficient restrictions on the trend and the cycle process.
Multivariate filters	
Hodrick Prescott (hereafter HPMV)	Potential output minimises a weighted average of deviation of output from potential, changes in the potential rate of growth and errors in the three conditioning structural relationships: a Phillips curve, an Okun Law and a relation between capacity and output gaps.
Beveridge Nelson decomposition	The trend is assumed to be a random walk but the stochastic shock driving this trend is supposed to be a linear combinaison of innovations of GDP and other variables which contains useful information to determine long term GDP.
Kalman filter	Extension of the univariate case by taking into account additional equations e.g. a Phillips curve.
Production function approaches	
Full structural model	All the inputs of the production function are recomputed endogenously using for instance a macro-economic model.
Production function with exogenous trends	All the inputs of the production function are determined exogeneously using uni or multivariate filters.
Structural VAR	Structural autoregressive model that estimates potential output and the output gap based on structural assumptions about the nature of economic disturbances.

39. Although the methods give more or less the same picture in terms of profile the size of the estimated output gap sometimes differ markedly (Canova, 1998). For instance, the trend linear method is found to lead to deeper gaps than the HP filter.[7] Mc Morrow and Röger (2001) find that the HP filter generates larger output gaps than the Kalman filter in the EU countries. By contrast, in the case of New Zealand, Claus et al. (2000) indicate a greater amplitude of output gaps from structural VARs and the Kalman filter than of those from the HP and HPMV. Kuttner (1994) shows that the size of the output gap estimated by the Kalman filter (univariate case) is markedly reduced by including an inflation equation (multivariate case) in the case of the US. In any case, the differences between the amplitude of the output gaps at a point in time have been found to be low compared to the uncertainties surrounding their estimation (Dupasquier et al., 1997).

40. In addition to their own statistical properties, output gap measures are often assessed in light of their relation with other indicators to check whether the different measures give the same picture of the business cycles. Usually the output gaps are compared with capacity utilisation or survey data. In general, available studies suggest that most methods give output gap estimates that are positively correlated with other indicators (Mc Morrow and Röger, 2001; Camba-Mendez and Rodriguez-Palenzuela, 2001). Notable exceptions are the Nelson- Beveridge decomposition and the structural VARs which do not always show a positive correlation with the other business cycle indicators (Mc Morrow and Röger, 2001). Among statistical measures, the HP measure of the gap is found to have a higher correlation with capacity utilisation than methods based on linear trend (Mc Morrow and Röger, 2001). The same authors find that a measure based on the production function approach to be more strongly correlated with other business indicators than a measure generated by a HP filter, though most of the time the difference is not statistically different.

41. Another test is whether estimated output gaps can explain past inflation developments. In general, it is found that output measures have some power in tracking inflation developments (for instance Mc Morrow and Röger, 2001; Chagny and Döpke, 2001), but that a large part of inflation is explained by other factors (see for instance Claus et al., 2000; Slevin, 2001). Mc Morrow and Röger (2001) indicate that the performance of the HP filter and the production function are very similar in this regard. Results reported in Claus et al. (2000) suggest that structural VAR measures of the output gap are less effective at explaining inflation in New Zealand relative to other techniques but Chagny and Döpke (2001) find exactly the contrary for the euro area. Intuitively, methods that make use of an inflation equation (HPMV or multivariate Kalman filter) are likely to track inflation better. Kichian (1999) concludes that multivariate Kalman filter models (which use a price Phillips curve) are useful for estimating output gaps to assess inflationary pressures in Canada[8]. Rünstler (2002) and Proetti *et al.* (forthcoming) also find that multivariate Kalman filter output gaps are significantly related to inflation in the euro area in recent decades.

42. Output gaps appear to be useful to help predicting short-term future inflation (Camba-Mendez and Rodriguez-Palenzuela, 2001; Slevin, 2001; de Brouwer, 1998). De Brouwer (1998) shows, in the case of Australia, that the introduction of the output gap (independently from the method) implies an improvement in out-of-sample prediction performances compared to an equation without gap. The equations using the HPMV or the production function gaps are those for which the prediction errors are the lowest. Kichian (1999) reports that Kalman filter models perform relatively well in predicting inflation. By contrast, Rünstler (2002) indicates that output gaps generated by a multivariate Kalman filter have only limited leading properties for inflation in the euro area. Orphanides and van Norden (2001) have shown

7. The introduction of a time break can diminish the severity of the cycle rending the outcome closer to an HP filter.

8. Kichian (1999) estimates the Gerlach and Smets (1999) model with some changes (introduction of unknown structural breaks).

that for a variety of methodologies, real estimates of the output gap provides little information in terms of out-of-sample inflation forecast.

43. Rennison (2003) takes a different approach and assesses some alternative measures of the output gap on the basis of their ability to accurately estimate the output gap of an economy by using Monte-Carlo simulations to generate data assuming different processes. Restricting the analysis to HP, HPMV filters and structural VARs, he finds that a combination of the HPMV filter and structural VARs generally produces a more efficient estimate of the output gap at the end of the sample and, in most cases, estimates which are the most highly correlated with the true output gap.

The methods in light of the criteria

44. Following on from the review of differences in results of the various estimation methods, this sub-section examines how they compare against the various criteria highlighted above.

Core requirements

Consistency between priors and the underlying assumptions of the methods

45. It is generally agreed that potential output and the corresponding output gap should comply with certain statistical properties, the most obvious being stationarity of the gap. However, some methods, with a linear trend being an extreme example, do not ensure this, as they assume the trend is purely deterministic and do not fully extract the stochastic trend component.

46. Some methods assume a process for potential output which is not in line with economic priors. The linear trend method assumes that the supply side of the economy is deterministic and that changes in demand are the prime factors in economic fluctuations. This assumption is difficult to reconcile with the view that potential output is at least in part driven by technology shocks as pointed out by Lippi and Reichlin (1994). Implicitly, the HP filter assumes that the output gap is white noise. This is inconsistent with historical data which show a high degree of persistence of the gap. By contrast, the multivariate Kalman filter approach can specify a wide range of processes for potential output and the output gap and is thus easy to adapt to economic priors. However, when used as a two-sided filter, this method has difficulty in generating sudden breaks of trend, as the effect of breaks is spread out after and before the break. This implies a bias in the trend that should be corrected, if it is believed that a sudden break has effectively occurred.

47. Most of the methods used to estimate potential output fail to incorporate additional relevant economic information (e.g. information on inflation developments and/or unemployment gaps). As already said, univariate filters do not incorporate any economic information. The production function with inputs estimated by uni- or multivariate filters disregards the mutual dependence of the unemployment and output gaps. By contrast, the HPMV can make use of both an Okun law and a Phillips equation (see Annex 1) but their influence depends on weights which are often assigned on a judgemental basis. More generally, one way to incorporate the relevant information on inflation and gaps is to determine potential output and NAIRUs simultaneously. Such a system can be estimated either by standard econometric methods (Adams and Coe, 1990) or by the use of an unobservable components model and the Kalman filter (Apel and Jansson, 1999).

48. A number of priors exist concerning the expected volatility of potential output. In particular, potential output growth is expected to vary over time (as its components, NAIRUs, trend productivity, hours and participation rates, vary), but its volatility is likely to be lower than that of actual GDP. In this regard, estimates of potential output resulting from a linear trend are implausible as they imply a constant rate of potential output growth. This criticism also holds, though to a lesser extent for split time-trend

methods (as there is no reason to believe that potential output growth varies only between cycles). The Beveridge and Nelson approach implies that much of the variation of the series is attributable to variation in the trend while the cycle component is small and noisy (Morley et al., 2003). Comparing structural VARs with other filtering methods, Mc Morrow and Röger (2001) indicate that the introduction of supply and demand shocks in the determination of the trend and the gap generated by structural VARs can lead to a relatively higher volatility of the trend (compared to the other methods).

49. Regarding the remaining methods, potential output estimates based on the production function method are somewhat sensitive to cyclical factors (as for instance they use series of actual capital stock), but their volatility remains lower than for the majority of the statistical methods (Chagny and Döpke, 2001). However, this depends on how extensive is the use of cyclical factors in the estimation. If for instance a full structural approach is adopted (with a vintage approach to generate capital series) it has been shown that potential output based on a production function approach may be more cyclical than if based on a standard HP filter (See Mc Morrow and Röger, 2001). More generally, the degree of volatility in estimates from statistical filters also depend on the smoothness parameters applied. All in all, most methods (except the linear trend and Beveridge-Nelson decomposition) appear to give acceptable estimates in terms of volatility, and a ranking/choice between methods would clearly have to depend on more specific user priors.

50. The length of cycles implicit in output gap estimates may also conflict with priors though this may have less to do with the chosen method than with how it is actually implemented. In particular, the implied length of the cycles is sensitive to the data and to the smoothness parameter used in the estimation, for instance the λ in a HP filter. This said, it appears that in general statistical measures imply relatively short cycles whilst economic measures tend to allow for persistent output gaps and therefore longer cycles (Chagny and Döpke, 2001; Claus et al., 2000; Scott, 2000[9]). In theory, the Baxter-King filter produces output gaps which satisfy any constraint the user has on the length of the cycle. In fact, as it is only an approximation of an ideal filter, the separation between trend and cycle does not satisfy exactly the given constraint. Regarding the multivariate methods, Saint-Amant and van Norden (1997) report that the HPMV method gives (in the case of Canada) much longer cycles than the usual definition of a typical business cycle (see e.g. Burn and Mitchell, 1946). Comparing the multivariate Beveridge-Nelson decomposition, structural VARs and the Cochrane method[10] on US data, Dupasquier et al. (1997) indicate that only structural VARs generate output gaps with a length that is consistent in most cases, with traditional notions of business cycle length.

Transparency

51. Transparency requires the estimation technique and all the steps of the estimation process to be explainable. In this regard, a useful method should have an intuitive explanation. This criterion is nonetheless not very discriminating as most methods have an intuitive explanation (see Table 1), even those which in fact utilise sophisticated techniques (e.g. the band-pass filter or Kalman filter estimations). A more important condition is that all assumptions made for the estimation can be justified.

52. Although statistical methods are often considered as more transparent than economic methods because they incorporate only a limited set of information, their estimates also depend heavily on technical assumptions that are not always easy to justify. For instance, the choice of the lambda parameter in the

---

9. Scott (2000) reports cycles for New Zealand of higher duration for linear trends and Kalman filter and lower duration for cycles generated by the HP filter, HPMV and band pass filter.

10. Cochrane (1994) method uses a two-variable VAR including GDP and consumption and the cointegration between the two variables to identify the permanent and transitory components of GDP.

Hodrick Prescott filter is still controversial[11] and the estimates obtained are sensitive to this parameter. [12]In the same vein, the weight used in the HPMV filter to introduce economic information is also subject to discussion as well as the smoothness parameter (the so-called signal-to-noise-ratio) in the Kalman filter estimation. In theory, a production function approach is more transparent as it is easier to single out the different components of potential output estimates and to give an economic explanation of their assumed pattern. In practise, however, very often the production function approach is a hybrid method with inputs obtained by univariate filters, very often the HP filter. Moreover, this method relies on assumptions made on specification of the production function, which are still open to discussion.

Size of the revisions between real time and ex-post estimates

53. For credibility purposes, it may be important that revisions between real time and ex-post estimates are not too large. Linear and mechanistic filters appear to be quite sensitive to the estimation period and thus likely to imply large revisions. Using univariate filters, Orphanides and van Norden (1999) show that revisions may be of important amplitude in the case of the US output gap (of the same order as the size of the gap itself). These revisions reflect uncertainties surrounding the model parameters rather than simple data revisions. Cayen and van Norden (2002) found similar results for Canada. Nelson and Nikolov (2001) also report large mismeasurements for the United Kingdom, but reflecting above all data revisions in actual GDP.

54. Regarding multivariate filters, the expected size of the revision depends on the specific techniques. Camba-Mendez and Rodriguez-Palenzuela (2001) find that multivariate Kalman filter models of the output gap as well as multivariate Beveridge- Nelson or structural VAR approaches show small revisions between real-time and final estimates for the euro area and the US. These revisions are likely to be reduced by the introduction of relevant economic information. Planas and Rossi (2000) show that incorporating information on inflation improves the accuracy of real time estimates of potential output. This contrasts with Orphanides and van Norden (2001), Gaiduch and Hunt (2000) and Cayen and van Norden (2002) who report that multivariate methods that incorporate information about inflation are subject to large revisions between real-time and final estimates. Real time estimates can also be improved by using information contained in cyclical co-movements of output with factor inputs and business cycles indicators (Rünstler, 2002; Proetti et al., forthcoming). Focusing on the end of the sample, however, it appears that two-sided filters give a higher weight to the end of sample observations, which are likely to be significantly revised. As a result, their estimates are likely to experience higher revisions than those produced by production function approaches at the last points of the estimation. However, as already mentioned, the latter techniques may also use input based on two-sided filters and thus be subject to revision. Another important feature is that the production function approach gives information on the contributions to potential growth and thereby makes it easier to justify any update.

Estimation of uncertainties

55. In recent years, there has been a growing debate on the influence of output gap uncertainties on economic policy (see Box 2 for a discussion on the implications of such uncertainty for monetary policy).

11. Criteria have been proposed for the choice of λ (e.g. by Kaiser and Maravall (1999)). They propose to choose the value of λ for which the variance of the cyclical component is mostly determined by cycles around the 'critical' length (cycles with length longer than the 'critical' length are allocated to the trend component and cycles with shorter length to the cyclical component). However, they do not document how to select the 'critical' length.

12. In a cross country analysis, there is also the issue of whether the same λ should be applied to all the countries. Assuming all the actual series have the same order of integration and priors on the 'critical' length of cycle in each country do not differ, then a common λ could be warranted.

Against this background, a policy maker is likely to need, in addition to the potential and output gap estimates some information on their precision, in order to qualify these estimates. Amongst the different methods, the use of the Kalman filter approach has been justified by its ability to provide additional information on the uncertainties surrounding the estimates.[13]

User specific requirements

Information needed

56. Statistical methods are obviously more parsimonious than economic methods as they only make use of the GDP series itself. Amongst the economic methods, the multivariate filters (which usually add no more than three equations to the filter) appear to need less data than a full structural model. In practise, however, given that in general most data are easy to access, the quantity of information needed might not be as important a constraint nowadays as it was few decades ago. Nonetheless, it remains important to try to restrict the quantity of data needed to keep the estimation process manageable and transparent, in particular in a context of cross country estimation. It is also a meaningful criterion for some countries which are still lacking reliable economic data (e.g. Eastern European countries as well as most developing countries). In addition to the quantity of information, the nature of information used in the various techniques is also different. For instance, two-sided filters use both past and future observations whilst techniques based on the production function, the Beveridge-Nelson decomposition or structural VARs use past information only. The use of a two-sided filter typically implies a need to predict information in order to provide estimates of potential for the current period, which is more demanding than using historical data.

Ability to detect persistent structural changes at the end of the sample

57. As mentioned, the property of the output gap estimates at the end of the sample is crucial for policy makers who are mostly interested in recent and future economic developments. To this extent a method like the Baxter-King filter which provides no estimate at the end of the sample is of little use. From a policymaker viewpoint, it is moreover important to detect any structural changes, particularly at the end of the sample. In this regard, the imposition of constraints such as the gap being zero on average over the cycle might reduce the ability of the method to detect such persistent breaks. Indeed, implicitly, having a symmetric gap is a way to ensure that actual and trend GDP move closely and that the trend tends to oscillate around actual GDP. By contrast leaving out this constraint allows the gap to be persistent over certain periods (e.g. because for a certain time expectations differ from observed inflation). For instance, if the inflation response is asymmetric, the economy will on average not be at potential if inflation is to be stabilised. Most univariate techniques (in particular the HP filter) but also some multivariate techniques (e.g. HPMV) assume some kind of symmetry of the gap. By contrast, multivariate techniques (in particular the Kalman filter) can be specified so that the output gap is not forced to be on average zero. Similarly, the production function method does not impose this restriction.

---

13. In fact any method that can be re-written in a state-space model, and be estimated by a maximum likelihood estimation, can provide estimated standard errors surrounding the estimation (see Hamilton, 1986). In particular, Harvey (1985) shows that the HP filter can be re-written in a state space format. Boone (2000) extends this result to the multivariate case. As a result, although it is not often done in the literature, both the HP filter and the HPMV can also provide information of the precision of their estimates.

Box 2. Uncertainty and the use of potential in monetary policy formulation

There is by now widespread consensus that monetary policy should focus on ensuring medium-term price stability. Countries differ as to whether output stabilisation has a similar status as an objective (eg. the United States) or is of lower rank in a lexicographic ordering of objectives (eg. the euro area). In either case, to improve the trade-off between variability of output and inflation the objective should be pursued in a forward-looking manner. Since estimates of the current (and projections of the future) output gap provide information on future inflation trends, monetary policy should respond to developments in the output gap independently of whether output stabilisation is a first order objective or not. This role of the output gap has been encapsulated most prominently in the various versions of the Taylor rule for setting policy controlled interest rates and which link these to the equilibrium real interest rate, actual inflation and the output gap.

The usefulness of the output gap for the setting of monetary policy depends on the accuracy with which it is measured. Various papers have shown that the more noise there is in the measurement of the output gap, the smaller its weight should be in the Taylor rule (eg. Smets, 1998; Orphanides et al., 1999). Furthermore, estimates of output gaps have sometimes been systematically biased and reliance on them have arguably led to serious policy mistakes. The most prominent example is perhaps the systematic rise in US inflation from the second half of the 1960s through to 1980 when the output gap estimates by the Council of Economic Advisors tended to show slack in the economy (Orphanides et al., 1999). Although the absence of official output gap estimates makes the argument more tenuous, Nelson and Nikolov (2001) argue based on real-time statistical releases and statements by policy-makers that mis-judgements about the output gap contributed strongly to UK inflation during the 1970s and 1980s. As well, ex post revisions of output gaps measures have been shown to predominantly relate to changes in potential as opposed to actual output (Orphanides and van Norden, 2002).

This said, estimates of output gaps have likely improved as a result of better techniques and better understanding of the economic mechanisms operating. In a study for Australia, Gruen et al.(2002) argue that real-time estimates of output gaps over the past 28 years based on a flexible Phillips-curve framework were un-biased and highly correlated with estimates based on full historical knowledge.

Having phrased the discussion above in terms of the Taylor rule, it should be recognised that, in practice, no central bank would set interest rates just on the basis of this rule. Rather, it may be more of a communications device and/or one among several internal yardsticks for judging the appropriateness of policy interest rates. There is much more information available to central banks about the future path of inflation (and output) than is contained in an estimated output gap, even if the latter may have a predictive value that stretches out relatively far compared to other indicators. Hence, even with considerable uncertainty about the magnitude of the output gap (or indeed its impact on inflation at any particular moment in time) central banks are able to set interest rates in a forward looking mode.

Reliability of the estimates at the end of the sample

58. The end-point problem is a typical feature of two-sided filters which use future data to estimate the current level of potential output. Near to the end of the sample, less information is available regarding the persistence of shocks rending the decomposition trend-cycle less reliable. In contrast, the Beveridge-Nelson decomposition, as well as the structural VARs and the production function approaches, which are purely backward methods, do not suffer from the end-point problem. End-point problems can also occur when the estimation is not run under a number of finite cycles, for instance for the linear trend technique: if the beginning and the end do not reflect similar point in the cycle, then the trend will be pulled upwards or downwards towards the path of actual output for the first few and the last few observations (Giorno et al., 1995).

59. Amongst the two-sided filters, the HP filter has been highly criticised for serious end-point problems. Baxter and King (1995) show that close to the end point (especially the last 3 to 4 observations), the HP filter not only eliminates the low frequency cycles it is supposed to eliminate but also has a tendency to underestimate cycles with higher frequencies. This will affect cyclical components with a period larger than 4 years. Only cycles with shorter periods will be passed fully. This implies that, if no

corrective measures are taken, the HP filter produces a series for the output gap which underestimates the length of the cycle close to the end point.

60. A number of solutions have been put forward to reduce - at least partially - the end point problem. The most frequently used solution has been to extend the dataset with forecasts. However, for techniques that need to be applied on a number of complete cycles, this raises the problem of the the number of periods that need to be added to finish off the current cycle[14]. Moreover, the resulting estimates are going to be influenced by forecasts (especially in two-sided filters), with a risk that the filter re-produces the assumptions behind the forecast data rather than providing additional information. In practise, it appears that generally forecasters tend to predict in the medium run economic growth close to the historical average. As a result, the use of these forecasts can lead to important policy failures (see part 1). Against this background, one-sided filters which do not suffer from the end-point problem are likely to be preferred.

Conclusions from the review

61. The preceding overview does not provide firm conclusions as to which methods are preferable from a policymaker viewpoint, but nonetheless highlights some important points:

- Although it is difficult to give a universal ranking of the methods, the statistical methods (trend and univariate filters) seem to be have more shortcomings than the economic methods (particularly, multivariate filters and production function approaches). This is particularly so on the 'consistency with priors' and the 'difference between real-time and final estimates' criteria. Amongst the multivariate filters, the Kalman filter appears to pass most of the criteria but it is not the most transparent method and, when used as a two-sided filter, is affected by the end point problem. By contrast, the production function approach is somewhat more transparent and has no direct end-point problem, but does not provide information on uncertainties. Moreover, there is an issue as to how inputs to this approach are constructed. As a result, the choice between these methods will depend on the priors and priorities of the user.

- Whatever method is used, it is necessary to make a critical and a non mechanical use of it (in particular, it is important to bear in mind its underlying assumptions and its shortcomings).

- Most methods provide estimates with a similar overall profile of potential output and, to some extent output gaps, but there are large divergences on the assessment of the magnitude of the output gap.

---

14. In the case of the HP filter, Kaiser and Maravall (2001) show that the HP filter can be obtained with a Wiener-Kolmogorov filter, and using the latter only 4 years of forecast are needed to overcome the end point problem.

REFERENCES

Adams C., D.T. Coe (1990), A system approach to estimating the national rate of unemployment and potential output for the United States, IMF Staff Papers, Vol 37, n°2, June

Apel M., P. Janson (1999), System estimates of potential output and the nairu, Empirical Economics 24: 373-388

Baxter M., R.G. King (1995), Measuring business cycles: approximate band-pass filters for economic series, NBER Working paper, n°5022

Beveridge S., C.R. Nelson (1981), A new approach to the decomposition of economic time series into permanent and transient components with particular attention to measurement of the business cycle, Journal of Monetary Economics, Vol. 7

Blanchard O. J., D. Quah (1989), The dynamic effect of aggregate demand and supply disturbances, American Economic Review, n°79

Bolt W., P.J.A. van Els (1998), Output gap and inflation in the EU, de Nederlandsche Bank working paper series 550

Boone L . (2000), Comparing semi-structural methods to estimate unobserved variables : the HPMV and Kalman filter approaches, OECD Economic Department Working Paper, n° 240

Bouthevillain C., P. Cour-Thimann, G. van den Dool, P. Hernandez de Cos, G. Langenus, M. Mohr, S. Momigliano, M. Tujula (2001), Cyclically adjusted budget balance: an alternative approach, ECB Working Paper n°77, September

de Brouwer G. (1998), Estimating output gaps, Research Discussion Paper 9806, Reserve Bank of Australia

Burns A., W.C. Mitchell (1946), Measuring business cycles, National Bureau of Economic Research

Butler L. (1996), A semi-structural method to estimate potential output:combining economic theory with a time-series filter, The Bank of Canada's New Quarterly Model, Part 4, Technical Report n°77

Camba-Mendez G., D. Rodriguez-Palenzuela (2001), Assessment criteria for output gap estimates, ECB Working Paper 54

Canova F. (1998), Detrending and business cycles facts, Journal of Monetary Economics 41(3), 475-512

Cayen J-P., S. van Norden (2002), la fiabilité des estimations de l'écart de production au Canada, Banque du Canada, Document de travail 2002-10

Cerra V., S. Chaman Saxena (2000), Alternative methods of estimating potential output and the output gap : an application to Sweden, IMF WP/00/59

Chagny O., J. Döpke (2001), Measures of the output gap in the euro-zone: an empirical assessment of selected methods, Kiel Working Paper n°1053, June

Claus I. (2000), Is the output gap a useful indicator of inflation, Reserve Bank of New Zealand, Discussion Paper Series n°2000/05

Claus I., P . Conway, A. Scott (2000), The output gap: measurement, comparisons and assessment, Reserve Bank of New Zealand Research Paper, n°44

Cochrane J.H. (1994), Permanent and transitory components of GNP and stock prices, Quarterly Journal of Economics 109(1), pp241-65

Cogley T., J.M. Nason (1995), Effect of the Hodrick-Prescott filter on trend and difference stationary time series: implication for business cycle research, Journal of Economic Dynamics and Control 19, 253-278

Cooley T.J., Dwyer (1998), Business cycle analysis without much theory: a look at structural vars, Journal of Econometrics, 83, 57-88

De Masi P. (1997), IMF estimates of potential output : theory and practice, Staff Studies for the World Economic Outlook, December

Denis C., K. Mc Morrow and W. Röger (2002), Production function approach to calculating potential growth and output gaps – estimates for the EU Member States and the US, ECFIN Economic paper, n°176

Dupasquier C., A. Guay, P. St-Amant (1997), A survey of alternative methodologies for estimating potential output and the output gap, Journal of Macroeconomics 21:577-595

European Central Bank (2000), Potential output growth and output gaps; concepts, uses and estimates, Monthly Bulletin, June

Elmeskov J. (1993), High and Persistent unemployment: assessment of the problem ans its causes, OECD Economic Department Working Paper, n°132

Fagan G., J. Henry, R. Mestre (2001), An area-wide model for the euro area, ECB working paper 42

Forni M., L. Reichlin (1998), Cyclical adjustment of government budget balances: evaluation of alternative trend estimation methods and of the cyclical sensitivity of budgetary components, Internal study for the DG ECFIN, July

Gaiduch V., B. Hunt (2000), Inflation targeting under potential output uncertainty, Reserve Bank of New Zealand, DP2000/08

Gerlach S., F. Smets (1999), Output gaps and monetary policy in the EMU area, European Economic Review 43

Giorno C., P . Richardson, D. Roseveare, P. van den Noord (1995), Estimating potential output, output gaps and structural budget balances, OECD Economic Department Working Paper, n°157

Hamilton J. D. (1986), A standard error for the estimated state vector of a state-space model, Journal of Econometrics, 33, 387-97

Harvey A.C. (1985), Trend and cycles in macroeconmic time series, Journal of Business and Economic Statistics 3: 216-27

Harvey A.C., A. Jaeger (1993), Detrending, stylised facts and the business cycle, Journal of Applied Econometrics 8, pp. 231-47

Hodrick R.J., E.C. Prescott (1997), Postwar US business cycles: an empirical investigation, Journal of Money, Credit and Banking 29:1-16

IMF (2002), United States Selected Issues, Chapter 1

Kaiser R., A. Maravall (1999), Estimation of the business cycle : a modified Hodrick-Prescott, Banco de España, Servicio de Estudios, Documento de Trabajo, n°9912

Kaiser R., A. Maravall (2001), Measuring business cycles in economic time series, Lecture Notes in Statistics 154, New York; Springer-Verlag

Kichian M. (1999), Measuring potential output with a state space framework, Working Paper 99-9, Bank of Canada

King R. G., S.T. Rebelo (1993), Low frequency filtering and real business cycles, Journal of Economic Dynamics and Control

Kuttner K.N. (1994), Estimating potential output as a latent variable, Journal of Business and Economic Statistics 12, pp. 361-68

Laxton D., R. Tetlow (1992), A simple multivariate filter for the measurement of potential output, Technical Report n°59, Bank of Canada

Lippi M., L. Reichlin (1994), Diffusion of Technical change and the decomposition of output into trend and cycle, Review of Economic Studies 61(1), pp. 19-30

Mc Morrow K., W. Röger (2001), Potential output: measurement methods, “new” economy influences and scenarios for 2001-2010, ECFIN Economic Paper n°150

Mise E., K. Tae-Hwan, P. Newbold, On suboptimality of the Hodrick-Prescott filter at time series endpoints, mimeo, University of Edinburg, University of Nottingham

Morley J.C., C.R. Nelson, E. Zivot (2003), Why are Beveridge-Nelson and Unobserved components decompositions of GDP so different?, Review of Economics and Statistics, Vol. 85, no 2, May

Murray C.J. (2002), Cyclical properties of Baxter-King filtered time series, University of Houston

Nelson E., K. Nikolov (2001), UK inflation in the 1970s and the 1980s : the role of output gap mismeasurement, Bank of England working paper series 148

Orphanides A., S. van Norden (1999), The reliability of output gap estimates in real time, Board of Governor of the Federal Reserve System, Finance and Economics Discussion Series, 99/38

Orphanides A., S. van Norden (2001), The unreliability of output gap estimates in real time, WP 2001s-57, Cirano, Montreal

Quah D. (1992), The relative importance of permanent and transitory components: identification and some theoretical bounds, Econometrica 60:107-118

Planas C., A. Rossi (2000), Univariate versus bivariate decomposition and reliability of real-time output gap estimates, Institute for systems, informatics and safety, TP361, Joint Research Centre of European Commission

Proetti T., A. Musso, T. Westermann (forthcoming), Estimating potential output and the output gap for the euro area: a model-based production function approach, ECB working paper

Razzak W., R. Dennis (1995), The output gap using the Hodrick-Prescott filter with a non-constant smoothing parameter: an application to New Zealand. Reserve Bank of New Zealand Discussion Paper n° 95/8

Rennison A. (2003), Comparing alternative output gap estimator: a monte carlo approach, Bank of Canada, Working Paper 2003-8

Richardson P., L. Boone, C. Giorno, M. Meacci, D. Rae, 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 Paper n°250, June

Rünstler G. (2002), The information content of real-time output gap estimates: an application to the euro area, ECB Working Paper n°182

Saint-Amant P., S. van Norden (1997), Measure of the output gap: a discussion of recent research at the bank of Canada, Technical Report n°79

Scott A. (2000), Stylised facts from output gap measures, Reserve Bank of New Zealand, Discussion Paper Series

Slevin G. (2001), Potential output and the output gap in Ireland, Central Bank of Ireland technical paper, 5/RT/01, September

Turner D., P. Richardson and S. Rauffet (1996), Modelling the supply side of the seven major OECD economies, OECD Economic Department Working Paper n°167

ANNEX 1 : OVERVIEW OF THE MAIN METHODS

This annex aims to review the principal methods that are discussed in the paper. These descriptions rely heavily on Mc Morrow and Röger (2001), de Brouwer (1998), Chagny and Döpke (2001).

Statistical methods

Trending methods

Linear trend

This is the simplest method of estimating potential output. It assumes that the trend component of output is a linear function of time. It thus involves a linear regression of the log of real GDP on a constant and a time trend.

$$\ln Y_t = \alpha + \beta t + e_t$$

This method builds on the assumption that GDP can be decomposed into a deterministic trend component and a cyclical component. Potential output in this equation is given by the trend component ($\alpha + \beta t$). This method cannot allow for any supply shocks to the system and implies a constant potential output growth rate (estimated slope). Moreover, the resulting gaps might not always been stationary (as the stochastic trend is not fully eliminated). As a result, the linear trend can bias the output gap by partially allocating trend components into the cyclical component.

Split time trend

Trend output is calculated during each cycle, where the cycle is defined as the period between peaks in economic growth:

$$\ln Y_t = \alpha + \sum_{j=1}^{n} \beta_i t_i + e_t$$

This specification allows estimated trend growth to change between cycles but not within each cycle. This method is straightforward but in practise it may be complicated to determine the peaks. Output gaps are symmetric over each complete cycle.

Univariate filters

Hodrick Prescott filter

This filter extracts a trend component by introducing a trade off between a good fit to the actual series and the degree of smoothness of the trend series. Formally, it minimises following objective function:

$$Min \sum_{t=1}^{T} \left(\ln Y_t - \ln Y_t^*\right)^2 + \lambda \sum_{t=1}^{T-1} \left[\left(\ln Y_{t+1}^* - \ln Y_t^*\right) - \left(\ln Y_t^* - \ln Y_{t-1}^*\right)\right]^2$$

where Y_t is actual output, Y_t^* trend output and λ Lagrange multiplier

λ controls the smoothness of the resulting trend and its appropriate value depends on the relative size of the variances of the shocks to permanent and transitory components to output (Hodrick and Prescott (1997)). Initially the value of 1600 was chosen for the US GNP on the basis of Hodrick and Prescott assessment of the relative size of shocks of that series. A low value of λ will produce a trend that follows actual output very closely, whilst a high value of λ reduces the sensitivity of the trend to short-term fluctuations in actual output. For λ very large, the filter will converge to the linear time trend method, with a linear time trend close to the mean growth rate of real GDP over the sample. λ determines the resulting length of the cycles. A λ of 100 (for annual data) gives cycles up to 15-16 years, whilst a λ of 10 corresponds to cycles of no more than 8 years.

The HP filter estimate of trend output may not be an ideal estimate of potential output. Mise et al. (unpublished) show that the HP filter is suboptimal at end-point. Cogley and Nason (1995) observe that the HP filter when applied to persistent time series can generate business cycle dynamics even when they are not present in the original data. Bouthevillain et al. (2001) report however that for small value of λ (λ=30 for annual data), this effect is unlikely to occur. There is also evidence that the accuracy of the decomposition varies over different data-generating processes and different datasets (King and Rebelo (1993), Harvey and Jaeger (1993)).

This filter fits trend through all the observations of real GDP, regardless of any structural breaks that may have occurred.

Some strategies have been developed to improve the HP filter ability to identify potential output and the output gap. One approach is to choose the parameter λ in line with prior beliefs about the ratio of demand and supply shocks (Razzak and Dennis (1995)). Another approach is to supplement the filter with a number of macroeconomic relationships (Laxton and Tetlow (1992)). Lastly, Butler (1996) conditions filter estimates of potential output at the end of the sample using a long run growth rate restriction on potential output to overcome the end-point problem. The LRX filter, developed by Laxton, Rose and Xie is a way to implement these constraints. The most common way to overcome the end point problem has been to extent the dataset with forecast. Mise et al (unpublished) confirm the finding of Kaiser Maravall (1999) that the use of forecast-augmented series can reduce the revision errors at the end of the sample. The size of reduction of the revision errors depend on the value of the smoothness parameter. This solution is only warranted when the forecasts used are reliable enough. There is also the question of the number of additional observations to be added.

Baxter-King filter

This is a band pass filter developed by Baxter and King (1995). The idea is that business cycles can be defined as fluctuations of a certain frequency. This is a linear filter which eliminate very slow moving ('trend') components and very high frequency ('irregular') components while retaining intermediate ('business cycle') components. When applying the filter, the critical frequency band to be allocated to the cycle has to be exogenously determined.

This filter, like the Hodrick-Prescott filter represents finite order, two-sided and symmetric moving averages able to eliminate stochastic trends.

Advantages: Flexible, easy to change the filter construction when the frequency changes.

Shortcomings: the filter is calculated by a moving average and thus has no value for the recent quarters.

Murray (2002) show that the Baxter King filter - and in fact all band-pass filters does not isolate the cycle but rather passes the first difference of the trend through the filtered series. Moreover the Baxter-King filter can potentially overstate the importance of transitory dynamics at business cycle frequencies.

Beveridge Nelson decomposition

As there are a lot of possibility to extract trend and cycle from a given time series, Beveridge and Nelson (1981) suggest to make the two following identifying assumptions: the trend is modelled as a random walk and shocks of trend and cycle are perfectly negatively correlated. To compute the output gap, the given series (here real GDP) is transform in a stationary series. Then an ARMA model is estimated. This model is used to forecast the series over an horizon s. Then for each period of the sample t, the output gap is defined by

$c_t = E_t(\Delta y_{t+s} + \Delta y_{t+s-1} + ... + \Delta y_{t+1}) - s\hat{\alpha}$ where $\hat{\alpha}$ is the constant of the estimated ARMA model.

This decomposition imposes a very specific functional form on the trend component and it is also assumed that trend and cycle are driven by the same shock.

This is a backward filter so it has no end point problem.

However, it can generate very noisy cycles and there could be some negative correlation between the cycle and actual GDP growth.

Kalman filter

The Unobserved Components model approach - estimated with the Kalman filter technique - assumes that macroeconomic time series are composed of trend cycle and, in some cases erratic components, which are not directly observable. These three components can be recovered by imposing sufficient restrictions on the trend and the cycle process.

For example, in the Watson model reviewed by Cayen and van Norden (2002) the log of real GDP is assumed to be:

$y_t = y_t^p + z_t$

where y_t^p is the permanent component, z_t the cyclical component, the two components are un-correlated with each other.

The permanent component can be seen as an estimate of potential GDP whereas the transitory component is an estimate of the output gap. Permanent output is often specified as

$y_t^p = y_{t-1}^p + \mu_{t-1} + \eta_t$ and $\mu_t = \mu_{t-1} + \zeta_t$, where η_t and ζ_t are orthogonal white noise

In the literature, the cycle is specified as an AR(2) to introduce persistance

$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \gamma_t$ with γ_t a white noise.

In that case, output gap will be on average zero.

This model is then re-written on a state-space representation and estimated with the Kalman filter and maximum likelihood. Alternative models can be found *inter alia* in Clark (1987) and Harvey Jaeger (1993).

Economic techniques

Multivariate techniques

Multivariate Hodrick Prescott (HPMV):

The multivariate filter proposed by Laxton and Teltow (1992) sets potential output to minimise a weighted average of deviation of output from potential, changes in the potential rate of growth and errors in the three conditioning structural relationship. By conditioning the HP filter estimate of output gap on additional relevant information, a more precise estimate of potential output and hence the output gap should be obtained.

Potential output is thus defined as the series which minimises the following loss function:

$$Min\sum_{t=1}^{T}\left(\ln Y_t - \ln Y_t^*\right)^2 + \lambda\sum_{t=1}^{T-1}\left[\left(\ln Y_{t+1}^* - \ln Y_t^*\right) - \left(\ln Y_t^* - \ln Y_{t-1}^*\right)\right]^2 + \sum_{t=1}^{T}\beta_t\varepsilon_{\pi,t}^2 + \sum_{t=1}^{T}\mu_t\varepsilon_{U,t}^2 + \sum_{t=1}^{T}\psi_t\varepsilon_{CU,t}^2$$

where $\varepsilon_{\pi,t}^2, \varepsilon_{U,t}^2, \varepsilon_{CU,t}^2$ are respectively the residual of a price Phillips curve, of an Okun relationship and from a relation between capacity utilisation and output gap.

The issue of the weighting of the various of the various components in the loss function needs to be resolved. Although the weights could be estimated, they are in generally assumed to be known in the literature.

Multivariate Beveridge Nelson

This method was developed by Forni and Reichlin (1998). The trend is still assumed to be a random walk but the stochastic shock driving this trend is supposed to be a linear combinaison of innovations of GDP and other variables which contains useful information to determine long term GDP. For example a change in output correlated with a change in employment would indicate a supply side shock and therefore a change in potential. In contrast, if the change in output is correlated with the change in consumption, a demand shock is more likely. The multivariate Beveridge Nelson decomposition defines potential output as the level of output that is reached after all transitory dynamics have worked themselves out.

This method gives a transitory component whose importance increases with the number of series used. Moreover, the properties of the trend are highly sensitive to the additional variables used.

Multivariate Kalman Filter

This extension of the univariate case incorporate additional equations based on economic relationships (see Kuttner (1994)). For instance, Gerlach and Smets (1999) add the following relations: potential output is assumed to follow a random walk, inflation is linked to past inflation and the lagged output gap and there is a reduced form aggregate demand equation which relates the output gap to its own lags and the real interest rate). Multivariate Kalman filter models can be specified in various forms (for instance Kichian (1999) modifies the Gerlach and Smet model to estimate potential output in Canada. Apel and Jansson (1999) specify a system approach to estimate output gap and the nairu at the same time). The advantages of this approach rely on a correct specification of the additional equations.

Production function :

The production function method usually postulates a Cobb-Douglas technology[15] with employment and capital as input. Assuming in addition that technical progress is Harrod neutral the production function has the following form:

$$Y_t = (TFP_t \; N_t)^{\alpha} K_t^{(1-\alpha)}$$

with Y_t actual output, TFP_t total factor productivity, N employment and K capital, α the labour share

As TFP is un-observable it is usually computed as the Solow residual i.e. by substracting contribution of capital and labour to actual GDP. Thus in log form :

$$tf\dot{p}_t = 1/\alpha\left(y_t - (1-\alpha)k_t\right) - n_t$$

Potential output series is then calculated by substituting trend variables in the production function, along with actual capital :

$$Y_t^* = (TFP_t^* N_t^*)^{\alpha} K_t^{\ (1-\alpha)}$$

with $N^*_{t,}$ defined as $N^*_{t,} = hrs^*_{t,}\ pop_{t,}\ pr^*_{t,}(1 - u^*_{t,})$

where *hrs** is an index of trend hours,
pop is the working age population,
*pr** is the trend participation rate,
*u** is the trend unemployment rate.

The capital stock series is not detrended because the maximum potential contribution of capital is given by the full utilisation of the existing capital in the economy.

In a full structural model, all the star-variables are recomputed endogenously using for instance a macro-economic model (see for instance Fagan et al. (2001) or Mc Morrow and Röger (2001) who attempt to use a vintage approach with the age of capital to compute trend tfp).

Most of the time, however the function production approach is an hybrid method which uses both economic relationships (e.g. the Elmeskov method (1994) or a Kalman filter model to generate a u*) and exogenous trend using univariate filters (most of the time, the HP filter) for variable on which very few information is available.

The most important shortcomings of this method are the following. The choice of the specification of the production function is open to discussion. The poor quality of capital stock data is often put forward. Different assumptions on the trend components, particularly tfp* et u*, are likely to lead to very different estimates of the level of potential output.

15. In most estimations, the Cobb Douglas specification has been used. An exception is Bolt and van Els (1998) who use a CES function.

Structural vars (SVARs)

Developed by Blanchard and Quah (1989), this technique has been is based on a structural autoregressive model that estimates potential output and the output gap using structural assumptions about the nature of economic disturbances.

The SVAR model uses information from the labour market (employment) and capacity utilisation to aid in the decomposition of actual output into a permanent trend component (supply) and a temporary cyclical component (demand). The trend is interpreted as a measure of potential output and the cycle as a measure of the output gap.

In the first stage of the estimation process the relationships are captured by regressing each of the three variables in the system on their own lags and the lags of the other variables. The reduced form shocks are composites of the supply and demand shocks that drive potential output and the output gap respectively. To recover these structural shocks a minimal set of identifying restrictions is imposed on the system. The key identifying restrictions are that demand shocks do not affect output in the long run whereas supply shocks do. The restrictions are imposed on the long-run dynamics of the variables. Once the structural shocks have been recovered the variables of the system can be expressed as the sum of current and past realisation of these shocks.

Cooley and Dwyer (1998) observe that results from SVARs are sensitive to the identifying assumptions which can not be tested.