Testing output gaps

An Independent Fiscal Institutions' guide

Output gaps continue to play a major role among Independent Fiscal Institutions (IFIs) for assessing the appropriate fiscal stance, underlying or structural deficits and when monitoring fiscal rules. IFIs should therefore expand and improve upon existing models, testing different approaches and incorporating them into the analysis of potential output and the output gap.

This paper sets out a framework to back-test alternative estimates taking account of certain key features needed for IFIs' purposes. While the results suggest there is no one-size-fits-all approach for estimating the output gap, a "suite of models" approach offers advantages over single models. Notwithstanding this, procyclicality remains a major issue with all estimates of potential output. This presents a major challenge that would need to be overcome, possibly allowing a role for judgement, if future reforms of the fiscal rules, including a spending rule, have estimates of potential output at their core.

JEL references: C50, E27, E32, E62. Keywords: output gap, growth cycle, potential output, uncertainty, unobserved component, fiscal authority, independent fiscal institution, production function, trend-cycle decomposition.

This paper was prepared by an ad hoc Working Group of the Network of EU IFIs. The Working Group was coordinated by Eddie Casey (Irish Fiscal Advisory Council), with additional input prepared by Killian Carroll of the Irish Fiscal Advisory Council, staff and former staff of AIReF, including José Miguel Ramos Carrasco, Carlos Cuerpo Caballero, Enrique M. Quilis, Angel Cuevas, and staff of the Portuguese Public Finance Council (CFP), including Nuno Gonçalves and Tiago Martins. The document is not an official position document of the Network. Views, if any, expressed in the paper reflect the views of the technical experts working on the document and do not necessarily represent the views of the individual IFIs of the Network or the collective views of the Network.

Main messages

- Estimates of the cycle and potential output are key tools for assessing fiscal policy. However, estimates are highly uncertain, varying over time and across methods.
- We build a framework to assess standard approaches — univariate filters, multivariate filters, and production functions — based on six criteria important to IFIs. Rather than just focusing on how stable estimates are, we are interested in assessing their plausibility, their performance at key turning points, and how consistent a signal they give to users.
- We find that the multivariate filter outperforms the standard production function approach used for EU fiscal surveillance in important areas aside from stability. On balance, the multivariate filter approach is more desirable for its greater plausibility, its superior performance at turning points, and the consistency of the signal it gives.
- We find promising results for combining models in a "suite of models" approach. As predicted by the forecasting literature, the suite of models approach broadly performs either better than the individual approaches or on a par with the best approach for each criterion.
- This paper therefore advises that practitioners developing output gap estimates for use in assessing the cycle proceed with caution and with an open mind. Putting too much faith in applying one method mechanically would be foolhardy. Ultimately, every cycle will be somewhat different and maintaining a suite of models, and combining these when needed, can help to avail of the best aspects of individual approaches while allowing for changing dynamics in the economy.
- However, we find that all methods produce estimates of potential output growth that are excessively procyclical — rising in good times and falling in bad times. This is a worrying feature, especially if such estimates are to be mechanically incorporated into assessments of "sustainable" net government expenditure growth as is the case with the Expenditure Benchmark in the EU fiscal rules. This problem of procyclicality does not appear to be addressed using ten-year averaging.
- We suggest further work is required to overcome the procyclicality of potential output growth rates. This would be particularly important if estimates of potential output are to serve as the basis for any new set of fiscal rules, which emphasise adherence to an expenditure rule with a debt brake. A role for judgement should be considered to alleviate the procyclicality of potential output estimates that are used for fiscal surveillance.

Table of contents

Introduction

Policymakers and Independent Fiscal Institutions will often base their assessments of the appropriate fiscal stance on the cycle. In addition to this, they are often tasked with assessing deficits adjusted for the cycle, or the sustainable growth rate of government expenditure.[1] Indeed, modern fiscal frameworks — most notably the EU fiscal rules — often tend to place estimates of the output gap or potential output at their core.

The use of output gaps in crucial policy decisions and assessments in the EU points to the need for accurate estimates in real time. However, the output gap is an unobserved variable whose estimate is surrounded by substantial uncertainty (Orphanides and van Norden, 2002). This adds substantial complexity to the EU fiscal framework. Therefore, there have been some proposals to get rid of or adjust fiscal rules that rely on structural balance estimates in the revision of the EU's fiscal framework currently underway.[2]

Given the importance of cyclical conditions for fiscal policy, there is a widespread recognition among IFIs that alternative estimates of potential output and the output gap need to be developed and assessed carefully. Such assessments can allow for a better understanding of cyclical developments, of the underlying budget balance, and of a sustainable medium-term path for net government spending.

This paper builds on the Network's companion paper (EU IFI Network, 2019), which provided an overview of the potential output methodologies used among EU IFIs and in the broader literature. With that paper, the EU IFIs reached some common ground on a definition of potential output relevant for fiscal surveillance. That is, we can define potential output as

> *"…a maximum level of output sustainable in the medium to long run, where "sustainable" implies that output, when at its potential, is not unduly influenced in any particular direction by imbalances in the economy, be they external, internal or financial."*

This definition is broader than simple statistical de-trending methods and inflation-focused definitions often used by Central Banks and others. It reflects a conceptual shift in the literature from a narrow balance concept (internal) to a broader one (internal, external and financial imbalances) driven by the experience from the financial crisis of 2007–2008. This broader view is essential for fiscal authorities to correctly remove all cyclical fluctuations that are beyond their direct control while producing the structural fiscal indicators. Financial cycles (Benetrix and Lane 2011, Borio, Disyatat and Juselius 2017), absorption cycles (Lendvai, Moulin and Turrini 2011, Darvas and Simon 2015) or commodity price cycles (Bornhorst, et al. 2011)

[1] In the current framework, EU countries have to reach a medium-term objective (MTO) in terms of the structural balance, and if the structural deficit is greater than that objective, countries must implement fiscal consolidation. Uncertain estimates of structural balances complicate the assessment of the attainment of the MTO, or the adjustment path towards it, and might even lead to some policy mistakes.

[2] See Darvas (2019) or the "Campaign against 'nonsense' output gaps" (Tooze, 2019).

are among the potential sources of imbalances that could be considered as important for influencing the fiscal position.

In practise, the approach followed in many IFIs is similar to that adopted by other institutions in terms of the methods used to assess the unobserved potential output and the output gap. IFIs are typically faced with many trade-offs when developing meaningful output gaps for the purposes of fiscal surveillance. There are a variety of methodological approaches that can be adopted. Often these will give disparate results. IFIs use lots of methods in practice and rely on national data sources, with some preferring bespoke adjustments to account for distortions in national data (e.g., Ireland for the Multinational sector; UK for the North Sea Oil). Moreover, estimates can be highly volatile, revision-prone, and—in many instances—downright implausible. This means that there can potentially be only marginal gains made by investing substantial efforts in terms of developing and maintaining alternative estimates of the cycle. At the same time, IFIs frequently have limited time and staff. Their focus may therefore often shift to more immediate challenges such as unravelling the latest fiscal developments. This means that IFIs tend to use standard methods in practice: univariate and multivariate filters and production function approaches.

There are also questions about the desirable features that a set of output gap estimates should have. For instance, how much do we value stable real-time estimates (small revisions) as compared to more plausible estimates according to experts' judgement. As noted in Cuerpo, Cuevas and Quilis (2018), revisions are sometimes the price paid for having the most reliable and up-to-date output gap estimates. However, for policymakers, uncertainty about the output gap and structural balances in real time complicates decision making and the monitoring of fiscal rules.

Solving these challenges while seeking a clear narrative is problematic when every cycle is different. With the challenges faced by IFIs in mind, this paper sets out some tools for examining alternative output gap estimates across multiple dimensions. We consider output gap estimates derived using various methods, data vintages, and across different countries. While the assessment in this paper is limited to certain large countries, the approach is one that can easily be replicated for other countries and the models and data used are provided as an accompaniment to the paper.

In terms of our findings, we find that, on balance, the multivariate filter approach has more benefits over the next best performer, the production function approach. This is evident in most of the important areas that would concern an IFI aside from the method's performance in terms of the stability of estimates. Both the multivariate filter and the production function produce reasonably plausible results. However, the multivariate filter has more desirable features in terms of having reliable estimates in real time, at turning points, and in terms of providing clear and consistent policy signals.

A concern with our findings is that all approaches assessed for estimating potential output growth appear excessively procyclical. Sharp increases in potential growth rates are visible during booms, with pronounced falls evident during recessions. This remains the case even when estimates are averaged over ten years. This finding suggests that using mechanical estimates based on such methods to inform "sustainable" growth in spending — as is done for the Expenditure Benchmark — could be problematic.

Our results are informative but there are three important caveats. First, we would like to be able to extend our analysis to more countries beyond the five countries that we consider in this paper. Second, there are a variety of multivariate filter approaches that can be used and we do not take a view on what the best method is. This is ultimately something for IFIs to assess at a national level. Third, it is important to note that the production function estimates do not allow for a strictly like-for-like assessment. That is, estimates from the production function approach (a) do not just incorporate real-time data but also real-time forecasts, which due to data availability, could not be applied to the other approaches; (b) they involve a methodology that has changed over time; and (c) they involve a certain degree of pre-smoothing whereby inputs are filtered before entering the model.

The results — more than anything — stress the need for IFIs to consider multiple approaches to estimate the output gap. That is, a "one-size-fits-all" approach for every country and every time period is unlikely to be a viable solution to the challenges faced in terms of developing a well-formed picture of how the cycle is evolving. Indeed, the experience of IFIs to date suggests that getting closer to "true" estimates of the cycle is helped by avoiding an over-reliance on single models.

In conclusion, we argue that maintaining a suite of tools and combining these estimates when needed is the best solution to help deal with the uncertainty and pitfalls associated with measuring the cycle in real time, while a role for judgement is perhaps unavoidable. As we show, the suite of models approach has a superior performance that retains the best characteristics of other individual methods. Moreover, the suite of models approach can also help to incorporate the broader scope of what it is IFIs care about. That is, a recognition of how imbalances from various sources can potentially drive fiscal developments and the cycle. Maintaining a number of models also gives adequate attention to the problem of changing economic drivers or paradigms for how the economy should be assessed.

A good avenue for future reform would be to continue efforts to reduce the dependence placed on output gap and potential output estimates in the EU's fiscal framework. It would also be useful to augment the role of IFIs in providing alternative estimates to those of the European Commission, while allowing a role for judgement if the rules are to continue to have potential output at their core.

Methodologies and Data

This section sets out the methods and data used to produce output gap estimates as well as the framework for assessing their respective performance. For analysis of the various approaches, we focus on five countries, Spain, Italy, the United Kingdom, Portugal and Ireland.[3]

Producing output gap estimates

Theoretically, there is an infinite number of ways to break down an economic series into a trend and a cyclical component. However, neither economic theory nor econometrics provides clear guidance on how best to produce such a breakdown. This has led to a proliferation of techniques for measuring business cycles and potential output ranging from data-driven univariate filters to more complex structural general equilibrium models.[4]

In this paper, we assess three common approaches to estimate output gaps. We consider univariate filters, multivariate filters, and production function approaches.[5]

Univariate Filters: Univariate filters estimate trend output on the basis only of the information content of actual output data. The approach is purely statistical in nature as these filters entail statistical assumptions that determine the amplitude of the cycle and the dynamics of trend output. While they are generally recognised as being a crude option for the identification of the cycle, their use has nonetheless proven resilient to newer more complex techniques. They are generally simple procedures that do not require judgmental assumptions about the structure of the economy, so they can be applied to many countries in a homogeneous way. The main critique to these methods is the lack of economic theory criteria in their application, and the fact that they do not incorporate potentially useful information on some other variables. Another drawback is the substantial amount of the end-of-sample uncertainty that leads to procyclical (biased towards trend) and unstable assessment of the output gaps and so undermines the use of the methods in real-time applications, especially for the small and open economies with many structural breaks as they can be smoothed to an unreasonable degree (Ódor and Jurašekova Kucserová 2014).

The most popular non-parametric univariate techniques among IFIs include the HP filter, the Kalman filter (KF), Baxter-King's band-pass filter and its generalization, the Christiano-Fitzgerald filter.

The real-time data used in this paper are pulled from national sources, with inputs provided by members of the working group for their respective countries. The

[3] This selection of countries was largely determined by the availability of real time data.

[4] See, for example, Alvarez and Gomez-Loscos (2018) for a review of the main advantages and drawbacks of the different methods.

[5] For a detailed description and literature review of all approaches to estimating the output gap see EU IFI Network (2019).

univariate filter used here is the HP filter on quarterly GDP with a smoothing parameter, λ = 1600.

Multivariate Filters: The next set of estimates produced fall under the broad category of multivariate filters. These filters allow for the consideration of additional economic relationships (Okun´s Law, Phillips Curve, etc.), while at the same time they impose less economic priors than fully structural models.

Although the literature provides a variety of approaches, in this paper we use these state-space models based primarily on the selection criteria set out in Cuerpo, Cuevas and Quilis (2018).[6] In particular, these authors present an approach for the selection of an output gap estimate that pivots around a multivariate unobserved components (MUC) Kalman filter estimation. Different specifications of the model are tested by combining GDP with potential candidate variables sharing relevant information about the domestic business cycle (capacity utilization, unemployment), external imbalances (current account, exchange rate), the financial situation (credit to non-financial corporations) and price pressures (GDP deflator, CPI, house prices). This approach allows for country-specific cycle definitions, generalizing the work in Borio et al. (2017) and Alberola et al. (2013).

Six criteria relating to the output gap estimates' statistical and economic features are used when determining which multivariate model will be chosen:

1) Statistical significance of coefficients — this considers the importance of the observable variables used to identify the cycle. Variables are selected only if statistically significant.
2) Average Relative Revision (ARR) — this is defined as the difference between the initial output gap estimates and the final output gap estimates produced. It is obtained as the average distance between the filtered (one-sided or first estimate) and the smoothed (two-sided) estimates of the output gap, normalized by the maximum range of the filtered estimation. A lower revision is preferred in the selection process.
3) Average Relative Uncertainty (ARU) — this measure captures the relative uncertainty around estimates. It is defined as the average standard error, again scaled against the maximum amplitude of the output gap estimates produced.
4) Economic soundness – meaning that some textbook macro relationships can be captured in the selected models (Okun's law, PC, etc.).
5) Amplitude and profile alignment – relative to consensus figures (range given by a panel of official institutions) and in agreement with commonly accepted business cycle chronology (e.g. ECRI dating). The quantification of the profile alignment can be made by means of the cross-correlation function and different measures of conformity, e.g. Harding and Pagan (2006).

[6] Due to a lack of convergence when using estimates produced for Ireland under the "Beauty contest" approach, we use a similarly derived state-space model for the multivariate filter estimates for Ireland that take the modified current account balance as the chosen cyclical indicator.

6) **Stability** – Stability of the one-sided cycle estimate, as this would mimic the practitioner’s need for updated estimates as new data is added in real time.

This approach puts the focus on the specification of the model rather than on a prior selection of the methodology itself. It also allows for considerable flexibility in terms of how country-specific cycles are derived.

Again, pseudo-real time data used are pulled from national sources, with inputs provided by members of the working group for their respective countries. In the case of Spain, only three candidates made it all the way down to the fourth criteria: (i) the unemployment rate; (ii) the current account balance over GDP; and (iii) investment in construction over GDP (see Table 1 for Spain); For the United Kingdom the model includes the unemployment rate, new building permits and the exchange rate. For Italy, the variables selected are the unemployment rate, employment, investment in construction and investment in equipment. For Portugal, the unemployment rate and net exports are chosen. For Ireland, we use estimates based on the current account balance.[7]

The differences in variable selection reflects the significance of those variables in explaining the cycle.

Table 1: Multivariate model selection

Variables selected	Spain	Italy	UK	PT	IE
Unemployment rate	✓	✓	✓	✓	
Total employment		✓			
Investment, construction	✓	✓			
Investment, equipment		✓			
Current Account Balance	✓				✓
Net exports				✓	
Exchange Rate			✓		
Housing completions			✓		

The Spanish economy presents an interesting case study to test the methodology. According to traditional visions of the cycle, such as the Phillips curve, the run up to the 2008 financial crisis was not perceived as an overheating period. Unemployment developments since the trough in 1994 to the peak in 2007 (from 24 per cent to 8 per cent) were not mirrored by rising inflationary pressures. These gains were thus interpreted as structural and real-time estimates of the non-

[7] A modified current account balance produced by the national Central Statistics Office is used, which (1) subtracts net factor income of re-domiciled PLCs, depreciation of R&D imports, traded intellectual property, and leased aircraft; and (2) adds back the cost of imported investment in net aircraft related to leasing, R&D related intellectual property, and the imports of R&D services.

accelerating inflation rate of unemployment (NAIRU) moved in line with observed data.

With hindsight, this vision was clearly misguided. By the early 2000s, Spain was already accumulating large imbalances and overheating pressures were present, although not visible, in headline inflation figures. For example, the current account was building up large deficits. Extending the concept of structural unemployment from the NAIRU to include a balanced external sector already reveals a downward bias in the former as it did not take into account all of the dimensions relevant to the cycle. Other variables might have also been relevant in defining and identifying the Spanish cycle, such as investment in construction, which was soaring, together with prices in non-financial assets (mainly in dwellings). By letting the 'beauty contest' between the different candidate variables take place, the methodology developed provides an efficient algorithm for variable selection. Table 1 in Annex A gives an example of the multivariate estimation results for Spain.

Table 2: Data set for Spain

Variable	Unit	Source
GDP	Volume index (base 2010=100)	INE
Internal demand		
Investment, Construction	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	INE
Investment, Equipment	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	INE
Productive Capacity Utilization	%	MINETUR
External sector		
Real Effective Exchange Rate	Index 1999 I=100	Bank of Spain
Current Account Balance	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	Bank of Spain
Gross National Savings	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	INE
Prices		
CPI, General	(i) Price index (base 2011=100); (ii) growth rate, % change	INE
GDP Deflator	(i) Price index (base 2010=100); (ii) growth rate, % change	INE
Compensation per employee	Euros per employee	INE
Housing prices	Euros per square meter	MFOM
Labour market		
Unemployment Rate	%	
Employment, full-time equivalent	Thousands	INE
Hours worked per employee	Units	INE
Compensation of employees	(i) Volume index (base 2010=100); (ii) M€;	INE
Financial and Monetary sector		
Credit to Non-Financial Corporations	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	Bank of Spain
Credit to Households	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	Bank of Spain
Broad Money (M3 aggregate)	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	Bank of Spain
Narrow Money (M1 aggregate)	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	Bank of Spain
Fiscal Variables		
Public Debt, Excessive Deficit Procedure	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	Bank of Spain
Net Lending (+), Net Borrowing (-): General Government	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	INE
Taxes on Production and Imports	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	INE
Taxes on Income and Wealth	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	INE
Social Contributions	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	INE
Unemployment Benefits	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	MEYSS

* Total number of variables included: 52

Sources: INE: National Statistics Institute; BDE: Bank of Spain; MFOM: Ministry of Public Works; MINETUR: Ministry of Industry, Energy and Tourism; MEYSS: Ministry of Employment and Social Security.

All the variables are corrected for seasonal and calendar effects. In the case of the series from the Quarterly National Accounts, they are already published corrected of such effects. For the remaining time series, Tramo-Seats is used (Gómez and Maravall, 1996, Caporello and Maravall, 2004). In addition, there are three main issues to set before estimation: (a) the cyclical behaviour of the selected variables accompanying GDP; (b) their order of integration; and (c) unit specification.

The approach puts the focus on the specification of the model amongst candidate variables, rather than on a prior selection of the methodology itself (model "horse race"). It also allows to internalise some of the criteria "ideal" properties of output gap estimates in terms of economic soundness and statistical goodness. It also takes into account the consistency with estimates of the output gap made by official institutions. Nevertheless, the results are likely to be driven by these specific requirements or assumptions that can fail to be tested. In comparison to the production function approach, multivariate techniques allow for an integrated estimation of uncertainty surrounding the estimates. They tend to be also more parsimonious than fully-fledged economic models and thus easily replicable.

Production Function: The third approach is widely used by international organisations, including the European Commission, (see Havik, et al., 2014) and the OECD (see Chalaux and Guillemette, 2019). This approach involves the use of the neoclassical Solow-Swan growth model. This approach is relatively more structural and comprehensive relative to other approaches.

The method explicitly models output in terms of underlying factor inputs (and not just labour, as in an Okun's law approach), and involves specifying and estimating production functions that link output to capital, labour and total factor productivity.

Potential output is calculated as the level of output that results when the rates of capacity utilisation are normal, when labour input is consistent with the natural rate of unemployment, and when total factor productivity is at its trend level. This method allows for easier identification of the source of the changes in the output gap.

However, as noted in Paper 1 from the working group (EU IFI Network, 2019), the application of the Production Function approach in practice also has some drawbacks. For example, the approach requires some assumptions on the structure of the economy that may not fully correspond to reality (perfect competition, constant returns to scale, for example). Consequently, cross-country comparisons have to be made with care due to differences in the economic structure of different countries. Moreover, estimating the output gap with a Production Function approach entails using measures of the trend of the inputs which are not straightforward to obtain. Furthermore, as noted by Fernald (2014), production function measures of potential output are inherently cyclical because investment is cyclical.

For analysis in the following section, we use the biannual vintages of the output gap and potential output from the European Commission (see Havik, et al., 2014 for details). Data for these output gap estimates are available from the AMECO database.

While efforts have be made to compare the different approaches on as comparable basis as possible, this is not fully possible using the European Commission's estimates of the production function. First, key parameters used in the production function approach have changed over time. Second, the estimates of the output gap and potential output from the production function approach are derived using real-time forecasts by the European Commission. This reduces the end-point bias in the estimates of the output gap from the production function approach, but may introduce bias/procyclicality if the forecasts are biased/procyclical. Unfortunately, given that the univariate and multivariate approaches do not incorporate real-time forecasts of variables, it is not possible to assess these approaches on an entirely comparable basis. As a result, while the subsequent results should be treated with caution, we still believe that they are informative.

Assessing the different methods

Once we have narrowed our selection of multivariate models considered to a single model, the next step is to compare the performance of each of the three types of estimates (univariate, multivariate and production function).

We assess the models across six criteria. The criteria chosen are based on desirable attributes for output gaps and potential output estimates that would best avoid large policy mistakes.

The six criteria we look for in the estimates produced under each method are that they:

1) **Avoid procyclicality** – we estimate the ability of 10-year averages of potential output growth rates to avoid procyclicality based on the range of peak-to-trough estimates. The use of ten-year averaging, although applied backward-looking from the latest outturn (i.e., averaged over year t to year t-9), broadly mirrors the approach adopted in the EU fiscal rules for the Expenditure Benchmark. Larger ranges would imply that potential output gets revised substantially over a given period. We focus on the period 2001–2018, with the peak taken as the maximum for the period 2001–2009 and the trough taken as the minimum for the period 2010–2018. The method with the largest peak-to-trough range or "most procyclical" method is scored as 100-100 = 0, with others scored relative to that: $100 - (range/range_{largest}) \times 100$.

2) **Perform well at turning points** – we assess this in terms of (1) the initial vs final estimate of the output gap, and (2) the extent to which the output gap are revised at key turning points. We take two common turning points across these countries, 2006 and 2012. The score for each approach based on an equal weighting of 1) the correct sign of the initial estimate, and 2) the size of the revision. The sign criteria is scored out of 10 (2 periods*5 countries). The size of the revision is scored based on the sum of the absolute revisions for each method, with the largest sum of absolute revisions scored as 0, and the others scored relative to that: $100-(revision/revision_{largest}) \times 100$.

3) **Give consistent signals** – we assess the consistency of output gaps produced under each method, first, by identifying instances where there are large positive or large negative output gaps based on initial estimates. Second, we assess the consistency with this signal over subsequent vintages of estimates for the same period. If subsequent vintages continue to show a large output gap of the same direction, then we deem it consistent. Formally, consistency is measured as: where estimates for initial estimates >= 1 in either direction, consistency = % subsequent vintages it stays as such (100% = a score of 100).

4) **Have a plausible narrative** – We assess plausibility qualitatively. Our assessment is based on the idea of a "smell test" — a qualitative "expert" judgement of how well the output gap estimates align with the historical economic narrative for each

country. While this test is ultimately subjective, it is nonetheless a vital dimension and essential for ruling out poor estimates of the output gap. It can also be helpfully informed by business cycle dating committees such as the CEPR-EABCN Euro Area Business Cycle Dating Committee.[8] Drawing on backward-looking assessments of each economy's performance and past performance of the cycle, we grade the output gap estimates from: 0 = implausible; 25 = not very plausible; 50 somewhat plausible; 75 = very plausible; 100 = fully plausible.

5) **Have less variable potential growth** – We assess the variability of potential output growth from one year to the next using the most recent vintage of potential output growth.[9] The approach with the highest standard deviation of potential output growth is given the highest score (score = 100), while the scores of the approaches are measure relative to the highest scoring approach: 100 – $(SD/SD_{smallest} -1)\times100$.

6) **Are stable over vintages** – we test the stability of the output gaps and potential output growth using metrics such as, the mean absolute revision; the maximum absolute revision; and the number of sign changes (output gap only). Revisions are assessed relative to the most recent available vintage, with equal weight (50%) applied to the assessment of the output gap and potential output growth using these metrics. Each individual metric is scored using the following scoring approach: the worst method is scored as 100-100 = 0, with others scored relative to that: 100 – $(metric/metric_{largest})\times100$.

We do not put any specific weights on these respective criteria.

[8] For instance, see the chronology of Euro Area business cycles available at: https://eabcn.org/dc/chronology-euro-area-business-cycles

[9] A five-year moving average of potential output growth for both the multivariate and univariate approach are used in the assessment here (See Figure 3). In estimating the production function approach detrending is applied, which is not the case for the multivariate and univariate approach.

Results

This section assesses the estimates of the output gap and potential output from each of our three approaches using our assessment framework. We briefly summarise the main results based on the 6 criteria outlined in section 2, before providing more background detail on these results.

The results can be summarised as follows, with a graphical summary in Figure 1:

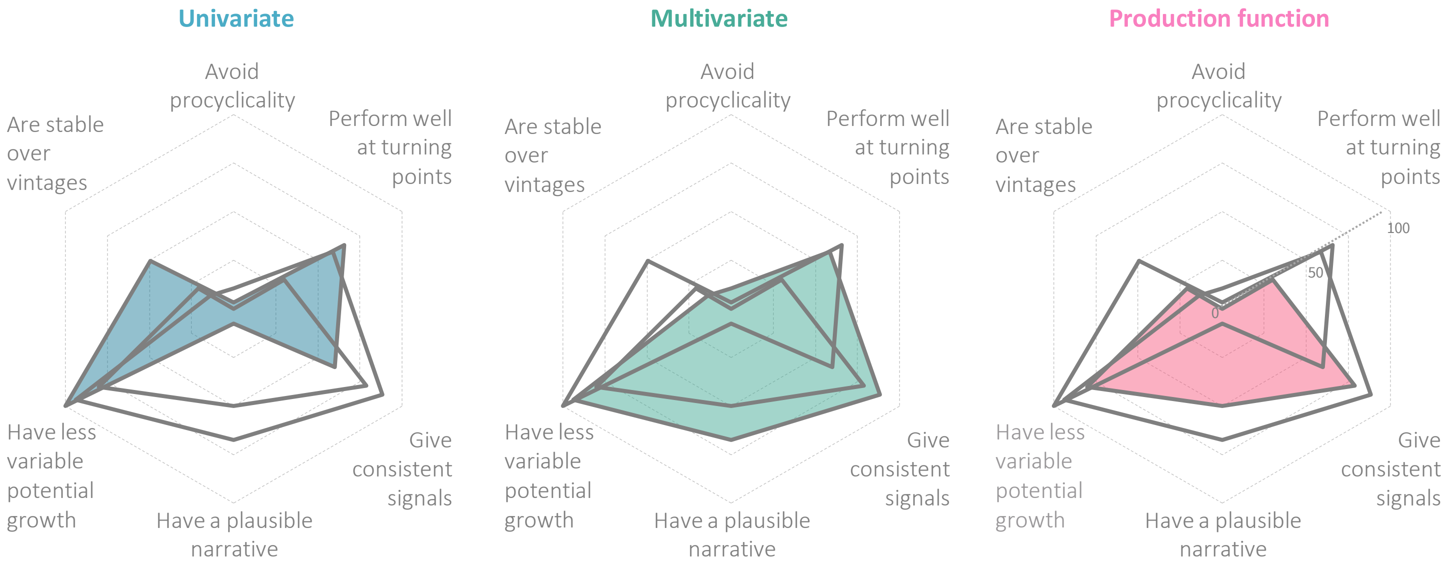
1) Univariate — The univariate approach is the most stable over vintages, has less variable potential growth (when smoothed) and performs well at turning points. However, the stability over various vintages and the performance at turning points reflects the relatively small amplitude of the estimated cycles, with the output gap being close to zero for most of the sample. Moreover, the estimates do not have a plausible narrative and seldom give consistent signals. Therefore, the method does not provide a clear signal as to the extent of overheating pressures or spare capacity.

2) Multivariate — The multivariate approach aligns well the economic narrative and ECRI business cycle dates, has less variable potential output growth, performs relatively well at turning points and gives consistent signals at turning points. However, it can be procyclical—although less so than the other approaches—and, on average is less stable over vintages than the other approaches (much less than the univariate filter; slightly less than the production function approach).

3) Production Function — The production function approach performs slightly better than multivariate filter estimates on stability. However, it is particularly bad around turning points and has the most procyclical estimates. The production function approach does manage to give reasonably consistent signals for policy, although less so than the multivariate approach. Potential growth is more variable than for other approaches and the overall narrative appears less plausible than with the multivariate filter.

Figure 1: Methods compared

Higher figures represent better performance with figures standardised from 0 to 100



For brevity, in the discussion below, we focus on the three largest countries, Spain, Italy and the UK. However, the overall results, including those shown above, are based on all five countries.

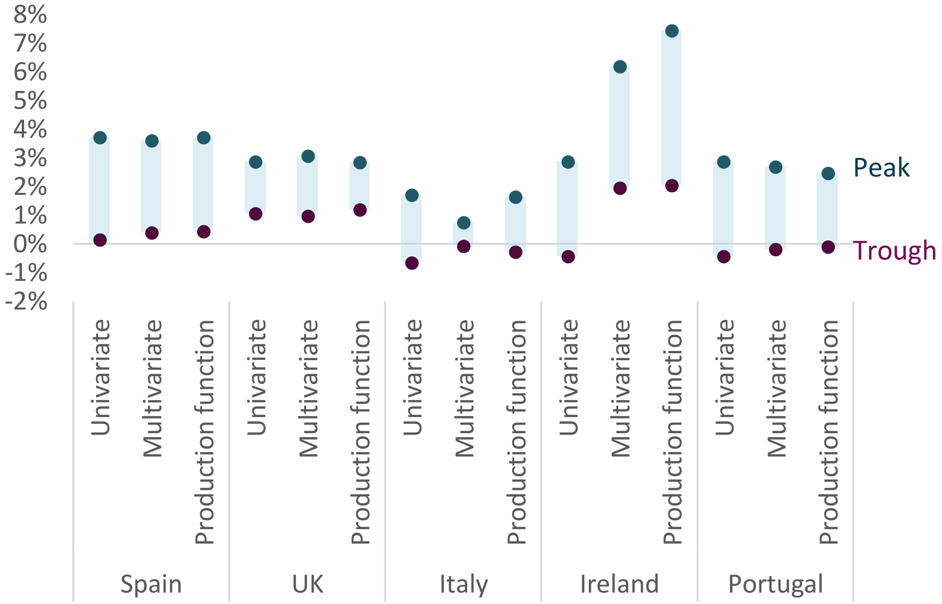
Avoiding procyclicality

None of the methods appear to fare that well in terms of avoiding procyclicality. That is—even when we use the smoothed 10-year averages of potential output growth, all of the estimates produced show a large drop from their cyclical peaks to their cyclical troughs. On average, the falls from peak to trough for potential output growth rates for 2010–2018 as compared to 2001–2009 are slightly smaller for the multivariate filters than for the other methods assessed. However, this is driven by the results for Italy. Results for Spain and Italy are not that different depending on the method used (Figure 2).

This is a poor result. It suggests that methods commonly used to calculate potential output may not serve as a useful basis for assessing things like sustainable growth in net government expenditure. Of course, the results could reflect trend declines as well as cyclical falls in growth rates.

Figure 2: Peak to trough estimates of potential output growth

10-year average potential growth rate estimates (peak 2001-2009; trough 2010-2018)



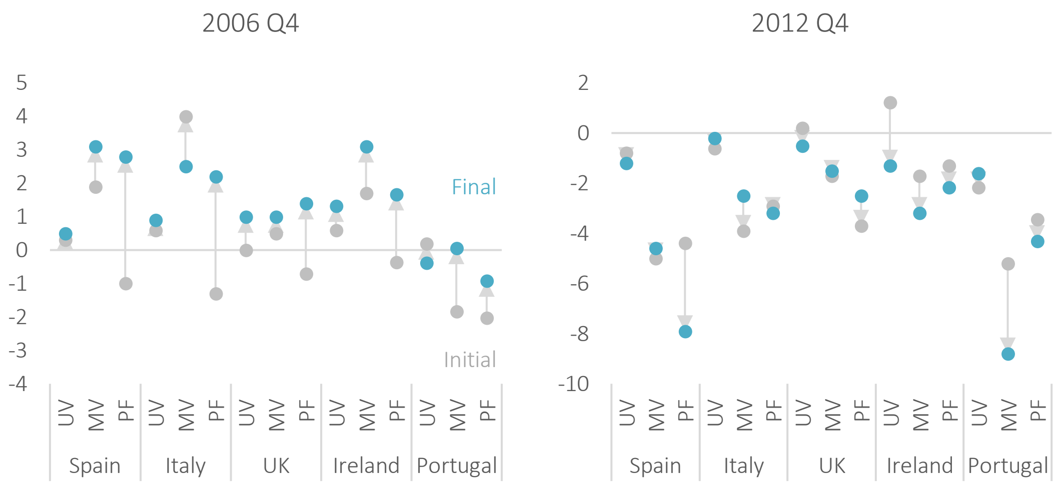
Performance at turning points

An important test of the plausibility of output gap estimates is their **performance at turning points.** We assess this in terms of (1) the initial vs final estimate, and (2) the extent to which they are revised at key turning points. In this manner, we consider two time periods that mark, or are close to, common turning points across these countries. Identifying turning points in real time is important for fiscal policy. Arguably, fiscal policy should be most active at the peaks and troughs of the cycle — leaning against the wind in booms and supporting the economy in recessions.

We take the first vintage of full year estimates for 2006, and 2012 for each approach, and consider these in terms of what is indicated by their initial and most recent vintage. The production function approach appears to be the poorest method in terms of both criteria. First, the initial estimates produced for the final quarter of 2006 using the production function approach indicate a reasonably large magnitude of spare capacity. This does not align with the now well-documented narrative for the pre-financial crisis period of substantial imbalances in economic and financial activity, nor does it align with business cycle dating assessments, nor does it conform with later vintages, which point to more positive output gaps. Second, the scale of revisions associated with the production function is wide — exceeding more than 2 percentage points in most cases — and larger than revisions seen under other methods (Figure 3). The multivariate approach performs the best on this metric for the UK, while for Spain and Italy, the univariate approach performs the best on this metric.

Figure 3: Output gap revisions at key turning points largest for production function

Initial and final output gap estimate for period shown, % potential output



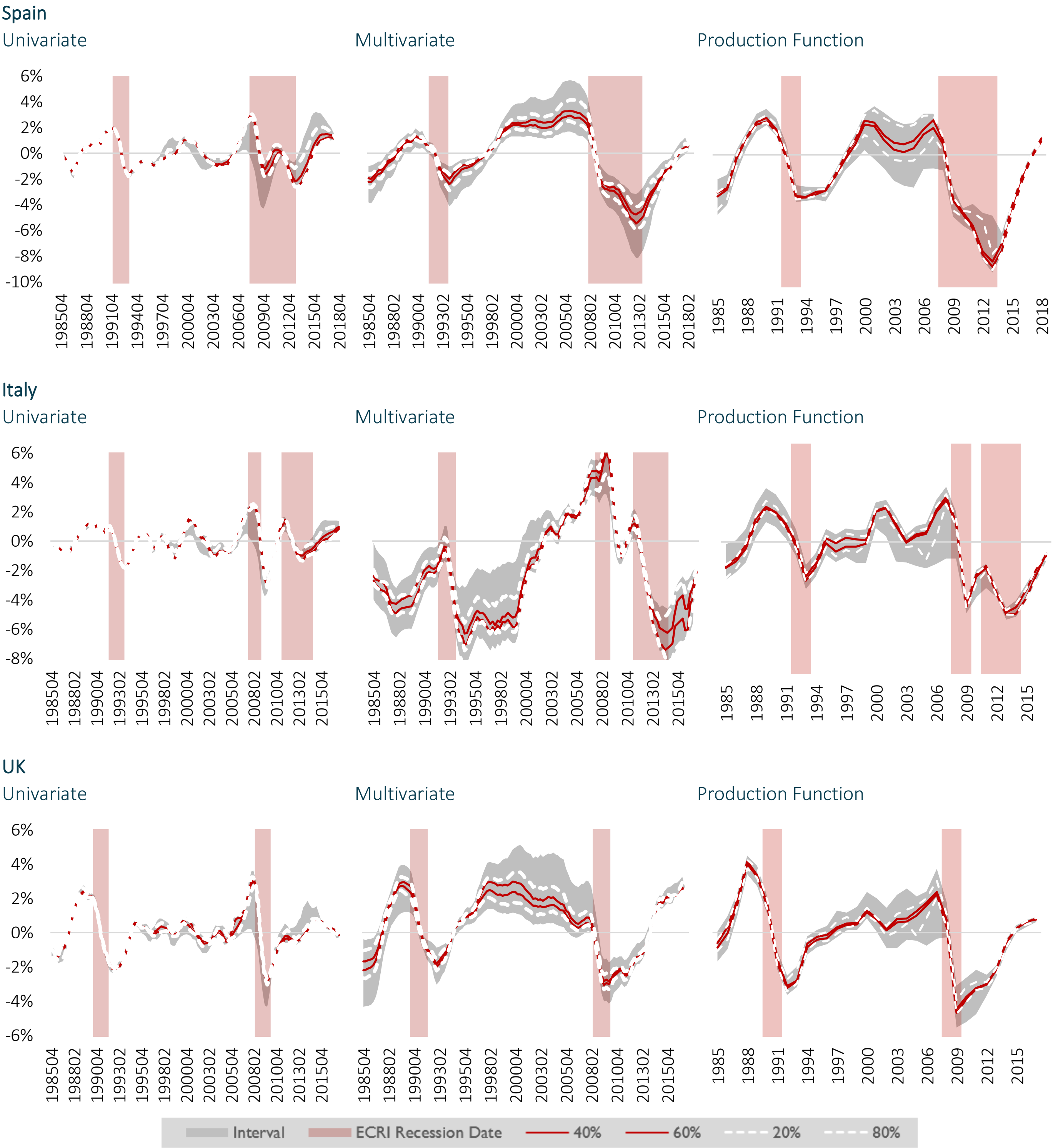
Note: Figures show the revision from the first full year vintage for each approach for the corresponding time period, relative to the most recent vintage.

Consistent signals

Another aspect of plausibility to consider is the **consistency of signals** that are presented by the output gap. In some instances, the size of the output gap is so small or so close to zero as to not really offer any strong signal in either direction of either overheating or spare capacity. This property partly relates to how convincing estimates will ultimately be in terms of their signal to users. Taking the performance of the methods from 2000 to 2008, we can see that only the multivariate filter approach provided clear and consistent signals of an emerging positive output gap in the case of Spain and Italy (Figure 4). By comparison, the other methods were regularly close to zero or, indeed crossed the zero mark with subsequent vintages, hence changing sign and their signal in terms of what was happening the cycle. For the UK, the production function offered relatively clearer signals.

More formally, we examine consistency for each method in terms of the percentage of subsequent vintages that accord with initial estimates. We examine initial estimates of large output gaps in either direction (positive or negative and greater than 1) and find that the multivariate filter accords with initial estimates 89% of the time. The production function does so 78% of the time, while the univariate filter only does so only about half the time (for 56% of subsequent vintages).

Figure 4: Vintages of output gap estimates



Note: Data covers 1985-2017 (2018 for Spain). Figures show the interval of estimates of various vintages of the output gap in grey. The 20-80 percentile of vintages of the output gap estimates for each year are shown in white dashed lines, while the 40-80 percentile are shown in red. Shaded pink region shows the ECRI recession dates for each country. Due to data availability issues, the multivariate output gap estimation for Italy only has a maximum of 16 vintages (the UK estimate has 76, and the Spanish estimate has 80). Both the univariate and multivariate figures show data based on quarterly estimation. The production function figures show data based on biannual estimation.

Plausible narrative

We next assess the plausibility of estimates.

The first aspect of the estimates to consider is the degree to which they conform with **wider economic narratives**. In this sense, the univariate estimates of the output gap do not appear to be entirely plausible. What is quite clear from the results is that the univariate estimates have a much shorter length of cycle and smaller amplitude. By comparison, both the production function and multivariate approaches having persistent positive/negative output gaps. The sharp build up and unwinding of imbalances that the univariate approach implies, does not seem to fit well with (1) typical assessments of the length of expansions, and (2) the economic narrative surrounding economic activity in these countries over the time periods considered.

The multivariate and the production function approaches both offer relatively more plausible results. The peaks and troughs of each cycle generally coincide with the estimates produced under both approaches for each country. In addition, the length of the cycle in each case appears similar between the production function approach and the multivariate approach.

Turning to potential output growth, we ask whether these estimates of potential output growth align with the wider economic narrative? All estimates of potential output growth appear to imply a secular decline in potential growth rates for each country. This seems to fit the consensus around the decline in sustainable growth rates for each country. However, each approach to estimating potential output growth appears to exhibit overly procyclical tendencies. For instance, there are large increases in potential output growth rates estimated in the run up to the Great Recession, and large declines in growth rates estimated during the recession. These procyclical tendencies are evident across all of the approaches to estimating potential output growth rates.

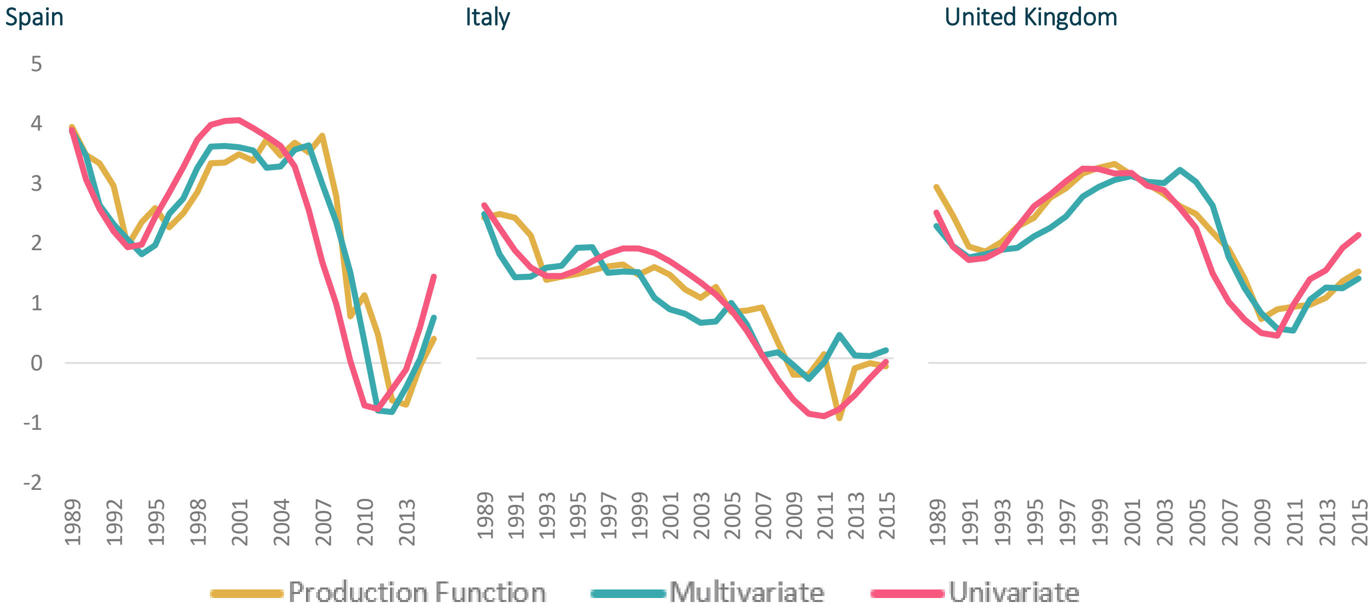
Variable potential growth

To be a reliable indicator to inform fiscal policy and to provide a theoretically plausible result, potential output growth should be reasonably smooth and should not be overly sensitive to the economic cycle—that is they should not be procyclical.

What is noticeable across the various approaches is that the potential output growth from year-to-year is relatively smoother under the production function approach (Figure 4). This reflects the fact that inputs to the production function approach are pre-smoothed, in the sense that (1) TFP is a filtered estimate of the Solow residual; and (2) the labour input relies on a filtered estimate of the natural unemployment rate.

Figure 5: Potential output growth rates compared

% year-on-year growth rates



Note: The figures shows a five-year moving average (+/- 2.5 years from each date) of potential output growth for the multivariate and univariate approach, alongside the actual figures for the production function approach.

By comparison, the multivariate approach shows a significant degree of variability in the annual estimates of potential output. Similar to the production function approach, for Italy, the univariate approach has relatively stable annual estimates of potential output growth, but for Spain and the UK the univariate approach has a large degree of variability in the year-to-year estimates of potential growth.

However, if a moving average of the multivariate and the univariate approach are used to smooth the year-on-year estimates, the figures align relatively closely with those of the production function (which is pre-smoothed). Shown in Figure 5 is a symmetric five-year moving average of the multivariate and univariate approaches alongside the production function approach.[10] For all countries the three estimates are closely aligned, and relatively smooth from year-to-year.

Stability

Output gap

The stability of estimates is one feature considered desirable. We assess this primarily in terms of how large revisions are from an initial vintage of output gap estimates to the latest set of available estimates.

On average, we can see that, for Spain the univariate and multivariate are more stable than production function estimates, with smaller revisions and fewer sign changes (Table B1). For Italy, univariate filters are the most stable by far, but

[10] Symmetric here means that the figure for each date is taken as an average of the previous 2.5 years and the following 2.5 years.

multivariate estimates are quite unstable. For the UK, all approaches produce reasonably stable estimates.

However, we can see that stability varies over time. For instance, from Figure 4, we can see that in the run up to the financial crisis, multivariate estimates were relatively stable and consistent in signalling overheating. By contrast, the univariate and production function approaches showed more tepid signs of overheating over this period, with estimates close to zero and sign changes frequently occurring.

Potential output growth

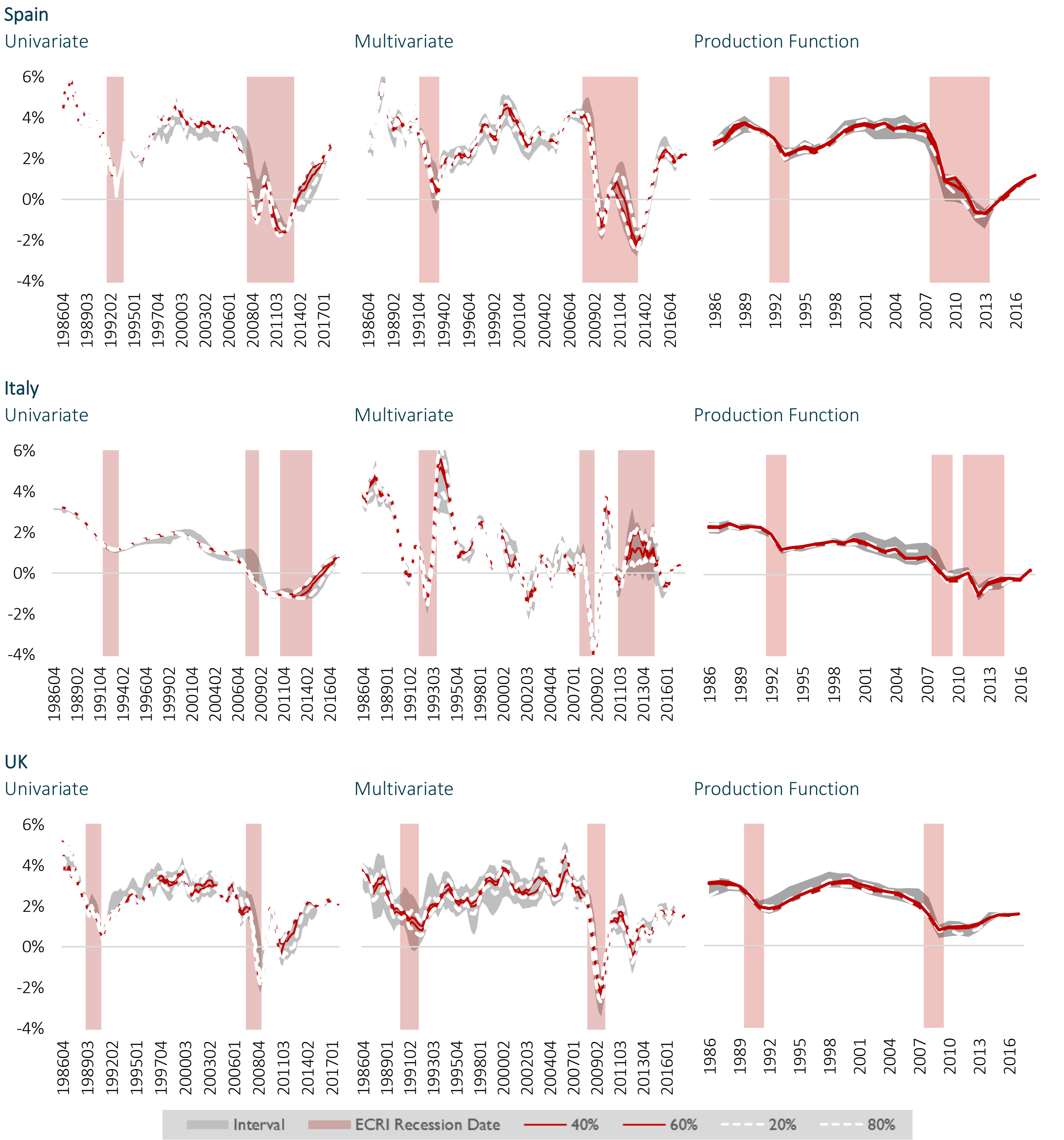
In terms of the stability of vintages of potential output growth estimates, there appears to be a clear winner, with the production function approach having the best stability characteristics for potential output growth across all countries (Figure 6 and Annex B Table B2). The multivariate approach appears to have the worst stability of potential output growth estimates.

For Spain, the production function approach has the smallest maximum absolute revision of all approaches (1.1 percentage points) and has the same standard deviation of revisions (0.3 percentage points) as the other two approaches. The production function approach is marginally worse than the univariate approach in terms of the mean absolute revision (0.2 percentage points vs 0.1 percentage points). The production function approach also has the smallest mean interval width (0.7 percentage points) and the smallest standard deviation of interval width (0.2 percentage points).

Again, the production function approach has the best stability characteristics of potential output growth for Italy. The production function approach has the smallest maximum absolute average revision (0.9 percentage points) and the smallest mean absolute revision (0.1 percentage points), whereas the multivariate approach has the largest maximum absolute average revision (3.2 percentage points). The production function approach and the univariate approach have the same mean interval width of vintages (0.5 percentage points), with the multivariate approach having a much larger width (1.3 percentage points).

Similarly, for the UK, the production function approach has the smallest mean interval width of vintages (0.6 percentage points), the smallest maximum absolute revision (0.9 percentage points), the smallest mean absolute revision (0.2 percentage points) and the smallest standard deviation of revisions (0.2 percentage points). On the other hand, the multivariate approach has the highest values across all these three approaches, indicating that it has the worst stability characteristics of potential output growth for the UK.

Figure 6: Vintages of potential output growth estimates



Note: Data covers 1986-2017 (2018 for Spain). Figures show the interval of estimates of various vintages of potential output growth in grey. The 20-80 percentile of vintages of the output gap estimates for each year are shown in white dashed lines, while the 40-80 percentile are shown in red. Shaded pink region shows the ECRI recession dates for each country. Due to data availability issues, the multivariate output gap estimation for Italy only has a maximum of 16 vintages (the UK estimate has 76, and the Spanish estimate has 80). Both the univariate and multivariate figures show data based on quarterly estimation. The production function figures show data based on biannual estimation.

Best practice

The output gap and potential output growth are key inputs for assessing appropriate fiscal policy. However, the output gap and potential output growth are unobservable, and therefore need to be estimated. To support fiscal policy in a reliable way these estimates need to be both plausible and stable.

Nominating the winner

The results from section 3 indicate that the multivariate filter has several advantages over the production function approach, while neither are perfect. Results produced by univariate filters appear to have little value by comparison.

Both the multivariate filter and the production function produce reasonably plausible results. However, the multivariate filter has more desirable features in real time, at turning points, and in terms of policy signals but can be less stable than the production function in some cases. Both the production function and the multivariate filter have varying degrees of success depending on the country assessed in terms of stability and plausibility and depending on the period considered. It is therefore more difficult to assess what the superior approach of the two might be.

Across the countries we assessed, the favourability of the multivariate approach over the production function approach varies a little. The multivariate filter fares better for Spain. For Italy and the UK, both the multivariate filter and production function methods fare reasonably well. The multivariate approach can be less stable than the production function approach. However, this is not always the case. For instance, the method is remarkably stable in signalling overheating in Italy in the lead up to the financial crisis. On balance, the multivariate approach has more favourable properties given that it produces more reasonable results at turning points, with smaller subsequent revisions, and it also provides more concrete signals of potential imbalances in the economy compared to other approaches. That is, it offers clear (not close to zero) and more consistent indications of large output gaps at key periods.

That there is no uniformly "best method" is perhaps unsurprising. This is evidenced in the literature on output gaps elsewhere and is discussed in the companion literature review (Network of EU IFIs, 2019).

The univariate approach, though more stable, offers limited value in terms of plausibility. This result is reasonably consistent across the countries considered. That is not to say that the univariate approach is never useful, but that it can tend to produce implausible estimates in terms of the size of output gaps and the duration of cycles, with no clear policy signal resulting.

For potential output growth, all of the methods appear to exhibit a high degree of procyclicality. This raises doubts about their usefulness for applying fiscal rules — particularly spending limits that are meant to get at the concept of a sustainable growth rate for government expenditure over the medium term. There are clear tendencies towards large increases in the growth rates of potential output in upturns and reductions in growth rates in downturns. In some cases, most notably for Italy with the production function, negative potential growth rates are estimated. Some falls in the growth rate may be plausible theoretically. For example, due to emigration resulting from a downturn. Yet the magnitude of declines shown across the methods does not seem reasonable. Further work may be needed to overcome this excessive procyclicality of potential output growth and an approach that incorporates judgement rather than purely mechanically derived estimates could be considered.

Possible best practice: a suite of models approach

The results—more than anything—stress the need to consider multiple approaches rather than over-relying on a single, standardised approach. That is, a "one-size-fits-all" approach is unlikely to be a viable solution to the challenges faced in terms of developing a well-formed picture of how the cycle is evolving. Every cycle is different, with different dynamics and drivers. What may have been a relative cyclical indicator in last 20 years may not be so in next 20 years.

Indeed, the experience of IFIs to date suggests that getting closer to "true" estimates of the cycle is helped by avoiding an over-reliance on single models. This approach can also help to incorporate the broader scope of what it is IFIs care about (with imbalances from various sources potentially driving fiscal developments). And it gives adequate attention to the problem of changing economic drivers or paradigms.

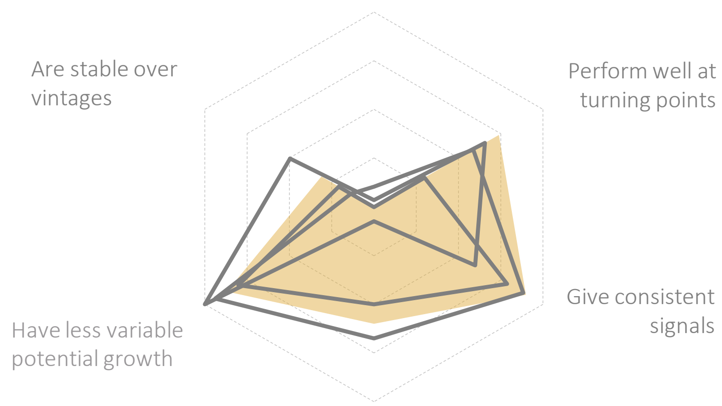
The beauty contest approach of Cuevas and Quilis (2018) applied for the multivariate approach assumes "one true model" or one "best model". While this practice has its merits and is shown to produce good alternative estimates of the output gap, it might not be the best approach to use in isolation.

There is an extensive literature assessing ways to improve predictions, which focuses on the benefits of aggregation. That is, combining multiple estimates to produce better judgment. Kahneman *et al.* (2021) offer a useful and recent review of the literature.

To illustrate the benefits of a suite of models approach, we consider combining the three approaches to estimating the output gap. We combine the three approaches using the mid-point of the output gaps for each year. We then assess this estimate based on the same six criteria outlined in Section 2.

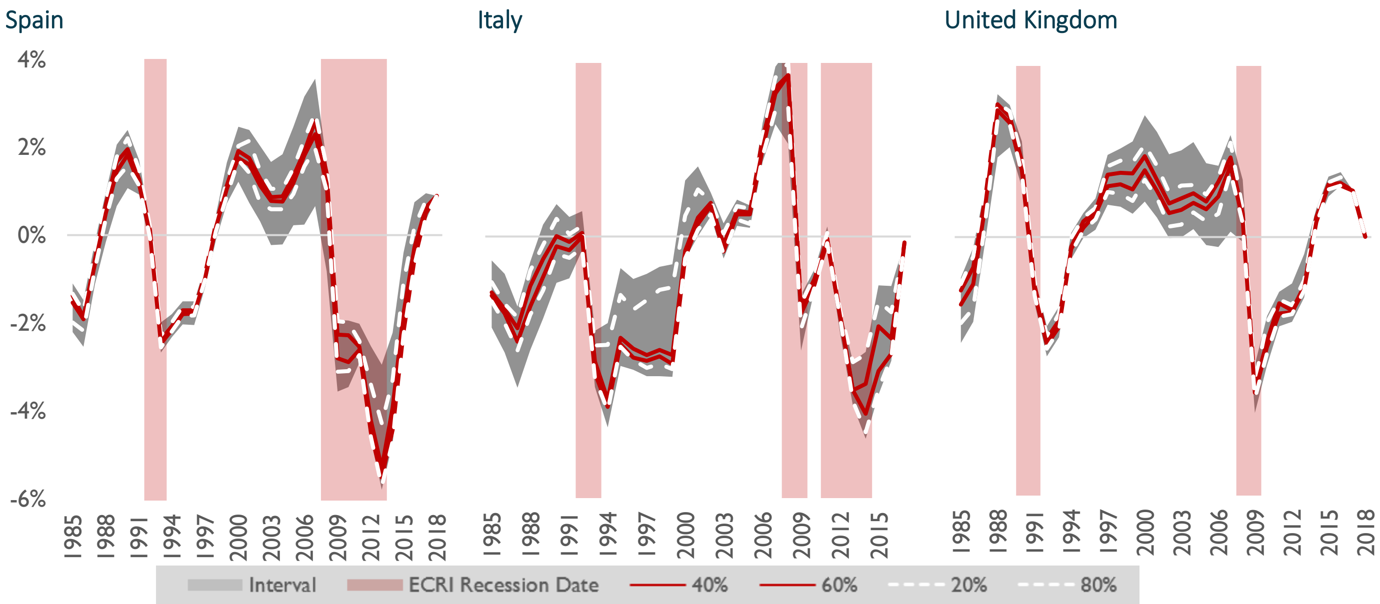
Figure 7 shows the graphical summary of the results. The estimates produced using this suite of models approach are more stable across vintages than both the production function and multivariate filter approaches are, have a plausible narrative (see Figure 8), perform better at turning points, give consistent signals and yield less variable potential output growth estimates (once smoothed). While the estimates of the output gap are still procyclical, they are about as procyclical as other approaches. Broadly speaking, the approach seems to perform better than individual approaches or it manages to perform relatively on a par with the best performing approach for each criteria. The only exception is for stability, where the univariate approach performs best. However, that is not a terribly useful benchmark to assess it against as the univariate approach yields conservative estimates that are broadly close to zero in most circumstances.

Figure 7: Suite of models – comparison with individual models



As a result, the best practice would seem to be to maintain a suite of several models. These could be chosen so as to possess necessary characteristics based on form of the "beauty contest approach" advocated in Cuevas and Quilis (2018). Maintaining models in this fashion would allow for the construction of a transparent and plausible uncertainty range within which the output gap may lie. Further work would be beneficial to consider the effectiveness of this approach for different countries though its success for the output gap and in forecasting, more generally, is well documented.

Figure 8: Vintages of the mid-point of the output gaps



Note: The figure shows the output gap vintages of the mid-point of the estimates of the output gap from the 3 approaches.

How best to combine the estimates of several models into a single estimate of the output gap, for the purposes of a numerical fiscal rule, is an open question. The "Estimate Combination Approach" is proposed by Ódor and Kucserová (2014). Casey (2019) proposed using the mid-point of all the models used. The median output gap estimate in a given year and the mid-point of the interquartile range of estimates are other possibilities as are Bayesian approaches. Regardless of the sophistication of the aggregation method, the results would appear to be superior in the broader literature, as we show.

We combine estimates in this paper as, sometimes, it is necessary to have point estimates of the output gap. This often proves to be the case for fiscal rules, for example, and for giving policy advice that can be clearly communicated.

While combining estimates has its benefits, notably in terms of superior point estimates, it is also useful to maintain multiple models, without necessarily combining them. It may be the case that one model performs best at turning points, so that this could prove useful if users are concerned about a potential correction in the economy arising in the near term. Similarly, other models might be good at giving consistent signals, such that users may opt to use these when worried about the possibility of subsequent revisions.

Conclusions

The output gap and potential output growth are key indicators against which the stance of fiscal policy is assessed. This paper assesses the performance of three standard approaches to estimating the output gap: the univariate approach, the multivariate approach, and the production function approach.

First, the results show that output gap estimates are surrounded by considerable uncertainty, not just in real-time but also with hindsight. The sizes of output gap estimates vary considerably across approaches. As discussed in a previous paper (EU IFI Network, 2019), this uncertainty stems from three main sources: model uncertainty both within a particular method and across different approaches; data uncertainty associated with the data revisions and methodological changes in statistical data definitions; and end-of-sample uncertainty that reflects the differences between one-sided (ex-ante, concurrent) and two-sided (ex-post, historic) estimates.

Second, assessing the various approaches on six key criteria, we find that the multivariate filter approach outperforms the production function approach in important areas aside from stability. On balance, the multivariate filter approach has desirable features in terms of its plausibility, its performance at turning points, and the consistency of the signal it gives. These features would tend to make it a more favourable approach for IFIs when assessing the output gap.

Third, estimates of potential output growth appear excessively procyclical across all approaches. We find that this is a common shortcoming and one which is a worrying feature if the estimates are to be mechanically incorporated into assessments of "sustainable" government expenditure growth as is the case with the Expenditure Benchmark in the EU fiscal rules. This problem of procyclicality does not appear to be addressed using ten-year averaging — a shortcoming also explored in Barnes and Casey (2019).

We would suggest that further work is required to overcome the procyclicality of potential output growth rates. This would be particularly important if estimates of potential output are to serve as the basis for any new set of fiscal rules, which emphasise adherence to an expenditure rule with a debt brake. A role for judgement should be considered to alleviate the procyclicality of potential output estimates used for fiscal surveillance.

It is clear from the results, that no single method for producing output gaps will work for all countries at all times. This chimes with the initial Network paper on the output gap (Network of EU IFIs, 2019; p.56), which concluded that:

> *"The view of different EU IFIs seems to converge on the view that it is best to use many alternative methodologies jointly for comparison and assessment. This leads one to propose the use of a suite of models approach."*

Some models may be more appropriate for larger, more self-contained, economies, while some models may be more appropriate for smaller more open economies. Some models may fare poorly than others for a time, for instance, due to asset price booms, but may at a later date prove useful as different economic drivers captured by these models return to prominence. As a result, best practice may involve maintaining several models of the output gap, assessing these with care in real time, and considering all models collectively as an input to what is happening the cycle.

A "suite of models" approach can capture the best aspects of individual methods. Combining models in a "suite of models" approach has been shown to be superior than a single model in terms of stability and plausibility of output gap estimates (Casey, 2019; Ódor and Kucserová, 2014). There is also wider literature in support of aggregation when difficult judgments are vulnerable to a high degree of noise (Kahneman *et al.,* 2021). We show that—in terms of our six key criteria—a suite of models approach broadly performs either better than the individual approaches or on a par with the best approach for each criteria. Maintaining a number of models, without necessarily combining them, can also help to ensure that changing dynamics and drivers in the economy are more adequately monitored.

This paper therefore advises that practitioners developing output gap estimates for use in assessing the cycle proceed with caution and with an open mind. Putting too much faith in applying one method mechanically would be foolhardy. Ultimately, every cycle will be somewhat different. As Blagrave *et al.* (2015) note in relation to estimation challenges: "designing a "least bad" solution among a host of mediocre choices might be the only realistic goal". Maintaining a suite of models and complementing this with a richer analysis of potential imbalances in the economy offers a way to break through the fog.

References

- Alberola E, Estrada A, Santabárbara D (2013). Growth beyond imbalances. Sustainable growth rates and output gap reassessment. Bank of Spain, working paper no. 1313
- Álvarez, L. J., and Gómez-Loscos, A., (2018). A Menu On Output Gap Estimation Methods. Banco de España, Documentos de Trabajo N.º 1720.
- Barnes, S. and E. Casey (2019). Overcoming Procyclicality in the EU Spending Rule. Paper in European Fiscal Board (2019), Independent Fiscal Institutions in the EU Fiscal Framework.
- Benetrix, A., and P.R. Lane. Financial Cycles and Fiscal Cycles. EUI-IMF conference on Fiscal Policy, Stabilization and Sustainability, 2011.
- Blagrave, P., R. Garcia-Saltos, D. Laxton and F. Zhang, 2015. A Simple Multivariate Filter for Estimating Potential Output. IMF Working Paper WP/15/79.
- Borio, C., P. Disyatat, and M. Juselius. Rethinking potential output: Embedding information about the financial cycle. Oxford Economic Papers vol. 69(3), 2017.
- Bornhorst, F., G. Dobrescu, A. Fedelino, J. Gottschalk, and T. Nakata. When and How to Adjust Beyond the Business Cycle? A Guide to Structural Fiscal Balances. IMF Technical Notes and Manuals, 2011.
- Casey, E. (2019). Inside the "Upside Down": Estimating Ireland's Output Gap. The Economic and Social Review, 50 (1, Spring), pp.5-34., (2019).
- Chalaux, T., and Guillemette, Y. (2019). The OECD potential output estimation methodology. Economics Department Working Papers no. 1563. https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ECO/WKP(2019)32&docLanguage=En
- Cuerpo, C., Á. Cuevas, and E.M. Quilis., (2018). Estimating Output Gap: a Beauty Contest Approach. AIReF. Working Paper Number 18/02, 2018.
- Darvas, Z., and A. Simon. Filling the gap: open economy considerations for more reliable potential output estimates. Bruegel Working Paper No. 2015/11, 2015.
- Darvas, Z., (2019). Bruegel Blog post, June 17, 2019, Uncertainty over output gap and structural-balance estimates remains elevated. Available at: https://www.bruegel.org/2019/06/uncertainty-over-output-gap-and-structural-balance-estimates-remains-elevated
- EU IFI Network (2019). A Practitioner's Guide to Potential Output and the Output Gap. Working Paper prepared by the EU IFI Network's Output Gap Working Group. Available at: https://www.euifis.eu/download/ogwg_paper.pdf

- Fernald, J. G. (2014). Productivity and Potential Output before, during, and after the Great Recession. NBER Macroeconomics Annual 29(1):1-51.
- Harding, D. and Pagan, A., (2006). Synchronization of cycles. Journal of econometrics, 132(1), pp.59-79.
- Havik, K., Mc Morrow, K., Orlandi, F., Planas, C., Raciborski, R., Roeger, W., Rossi, A., Thum-Thysen, A., & Vandermeulen, V. The production function methodology for calculating potential growth rates & output gaps. Economic Papers 535, Directorate-General for Economic and Financial Affairs, European Commission, 2014.
- Kahneman,D., O. Sibony, and C.R. Sunstein (2021). Noise: A flaw in human judgment.
- Lendvai, J., L. Moulin, and A. Turrini. From CAB to CAAB? Correcting Indicators of Structural Fiscal Positions for Current Account Imbalances. European Economy, Economic Papers, 2011: 442.
- Network of EU IFIs (2019). A Practitioner's Guide to Potential Output and the Output Gap. Network of EU Independent Fiscal Institutions, working paper, 2019.
- Orphanides, A. and van Norden, S. (2002). The Unreliability of Output-Gap Estimates in Real Time. The Review of Economics and Statistics 2002; 84 (4): 569–583. doi: https://doi.org/10.1162/003465302760556422
- Ódor, L., and J. Jurašekova Kucserová (2014). Finding Yeti: More Robust Estimates of Output Gap in Slovakia. Council for Budget Responsibility Working Paper No. 2, 2014.
- Tooze, A (2019). Output gap nonsense. Social Europe, 30 April.

Annex A: Example of multivariate filter selection —the case of Spain

This annex shows an example of the results of the multivariate filter selection process for Spain. The selection process is set out in detail in Cuerpo et al. (2018).

The selection of potential candidate variables follows an encompassing approach, aiming at capturing the build-up of potential imbalances across all relevant dimensions: (i) domestic economy; (ii) external sector; (iii) prices; (iv) labour market, and (v) financial and monetary conditions.

The selection of the relevant variables follows a reductionist approach according to the six criteria specified above. In this context, reductionist means that the complete list of potential variables is pruned through a specification process to derive a shorter list that will form the basis for the final econometric model. Every variable is modelled in a bivariate framework together with real GDP.

In the first place, the candidates not passing the significance test are removed. Two sets of variables are left out in this first round, most labour market series and, somewhat surprisingly, financial variables.

The average revision indicator provides the second screening for the remaining variables. This indicator reflects the average gap between the filtered (one-sided) and smoothed (two-sided) estimates of the output gap, normalized by the maximum range of the filtered estimation. Variables experimenting large revisions relative to their volatility are thus penalized (e.g. public debt, housing prices). The defining threshold is set at 0.25, to include two thirds of the remaining sample. Third, goodness of fit is assessed in relative terms as the ratio between the average standard error and the maximum range of the filtered estimate. Again, the threshold is set to keep two thirds of the competing variables (at 0.4). Prices and monetary variables are discarded at this stage as can be seen in Table A.

Once the necessary conditions are checked out, the 4th criterion looks at the amplitude and profile of the output gap estimates. Small cycles, defined by a small amplitude (lower than 4 pp.) are first left out. These include productive investment and most of the remaining fiscal variables (net income, social security contributions, direct and indirect taxes). A closer look at the specific profiles and ECRI dating allows for a further screening by removing unemployment benefits (as it does not properly identify the beginning of the last cycle) and capacity utilization (as it advances the recovery after the last cycle and points to positive output gap figures already in 2016).

Only three candidates made it all the way to the 4th criteria: (i) the unemployment rate; (ii) the current account balance; and (iii) construction investment.

When turning from the bivariate to the full model set-up, which includes GDP altogether with the three selected variables, the transition is far from smooth. Collinearity amongst the cyclical components can potentially generate imprecise point estimates that, combined with a flat likelihood function, may cause "jumps" in the estimations, rendering output gap estimates unstable.

Table A: The multivariate filter estimation results

Variable	Transformation	Criteria 1	Criteria 2	Criteria 3	Criteria 4	Criteria 5
GDP		t-statistic	ARR	ARU	Profile	Stability
Internal demand						
Investment, Construction	Volume index (base 2010=100)	5.32	~~0.28~~			
	% GDP	8.88	0.10	0.38	YES	YES
Investment, Equipment	Volume index (base 2010=100)	6.30	0.24	~~0.41~~		
	% GDP	4.87	0.11	0.18	NO	
Productive Capacity Utilization	%	3.22	0.02	0.11	NO	
External sector						
Real Effective Exchange Rate	Index 1999 I=100	~~-0.66~~				
Current Account Balance	Volume index (base 2010=100)	~~0.00~~				
	% GDP	-6.98	0.13	0.34	YES	YES
Gross National Savings	Volume index (base 2010=100)	~~0.97~~				
	% GDP	~~-1.64~~				
Prices						
CPI, General	Price index (base 2011=100)	25.22	0.25	~~0.87~~		
	Growth rate, % change	~~0.03~~				
GDP Deflator	Price index (base 2011=100)	~~1.34~~				
	Growth rate, % change	~~0.02~~				
Housing prices	Euros per square meter	2.26	~~0.29~~			
Labour market						
Unemployment Rate	%	-7.59	0.06	0.23	YES	YES
Employment, full-time equivalent	Thousands	3.17	~~0.28~~			
Hours worked per employee	Units	~~-0.27~~				
Compensation per employee	Euros per employee	~~1.59~~				
Compensation of employees	Volume index (base 2010=100)	~~1.84~~				
	M€	~~1.71~~				
Financial and Monetary sector						
Credit to Non-Financial Corporations	Volume index (base 2010=100)	~~-0.14~~				
	M€	~~0.23~~				
	% GDP	~~-1.44~~				
Credit to Households	Volume index (base 2010=100)	~~-0.19~~				
	M€	~~0.43~~				
	% GDP	~~-1.58~~				
Broad Money (M3 aggregate)	M€	~~0.12~~				
	% GDP	5.43	0.13	~~1.54~~		
Narrow Money (M1 aggregate)	M€	~~-0.22~~				
	% GDP	~~-1.55~~				
Fiscal Variables						
Public Debt, Excessive Deficit Procedure	Volume index (base 2010=100)	-2.57	~~0.31~~			
	M€	-6.93	~~0.29~~			
	% GDP	-8.23	0.25	0.36	NO	
Net Lending (+), Net Borrowing (-): General Government	Volume index (base 2010=100)	~~0.00~~				
	M€	~~-0.01~~				
	% GDP	~~1.11~~				
Taxes on Production and Imports	Volume index (base 2010=100)	~~0.47~~				
	M€	1.92	0.20	~~0.92~~		
	% GDP	-3.95	0.10	0.07	NO	
Taxes on Income and Wealth	Volume index (base 2010=100)					
	M€	~~0.06~~				
	% GDP	2.14	0.11	0.12	NO	
Social Contributions	Volume index (base 2010=100)					
	M€	1.90	~~0.26~~			
	% GDP	-5.43	0.10	0.20	NO	
Unemployment Benefits	Volume index (base 2010=100)					
	M€	~~-0.75~~				
	% GDP	-8.84	0.04	0.18	NO	
Net Income	Volume index (base 2010=100)					
	M€	6.41	~~0.26~~			
	% GDP	-5.12	0.16	0.16	NO	

Annex B: Summary statistics for estimates

Table B1: Summary statistics of output gap estimates across vintages

Country & method	Variables used	Sample				Intervals				Revisions			
		Period	Obs	First sample	Vintages	Mean Interval width	SD	40%-60%	20%-80%	Max abs revision	Avg abs revision	SD	No. of sign changes
Spain													
UV	GDP	1985Q4-2018Q4	34	1985Q4-1999Q4	40	1.1%	0.4%	0.1%	0.7%	4.1%	0.2%	0.5%	12
MV	GDP, UR, CA, Cons_emp	1985Q4 - 2018Q4	34	1985Q4-1998Q4	40	2.2%	0.6%	0.2%	0.8%	3.3%	0.4%	0.6%	25
PF	GDP (CAM)*	1985-2018	34	1985-2002	33	2.3%	0.7%	0.3%	1.2%	4.7%	0.7%	1.0%	40
Italy													
UV	GDP	1985Q4-2018Q4	34	1985Q4-1998Q4	40	0.7%	0.2%	0.1%	0.5%	3.0%	0.1%	0.3%	12
MV	Emp, UR, M&E, Cons_inv	1985Q1 - 2017Q4	33	1985Q4-2005Q4	10	3.0%	1.1%	0.5%	1.7%	4.8%	1.4%	1.7%	N/A
PF	GDP (CAM)*	1985-2018	34	1981-2002	32	1.8%	0.5%	0.2%	0.7%	3.8%	0.5%	0.7%	37
UK													
UV	GDP	1985Q1-2018Q4	34	1985Q4-1999Q4	40	0.8%	0.3%	0.1%	0.6%	2.9%	0.3%	0.4%	37
MV	Comps, UR, FX	1985Q1 - 2017Q4	33	1985Q4-1998Q1	38	2.1%	0.6%	0.3%	0.8%	2.3%	0.6%	0.7%	26
PF	GDP (CAM)*	1985-2018	34	1985-2002	36	1.6%	0.4%	0.1%	0.6%	2.3%	0.5%	0.6%	31
PT													
UV	GDP	1990Q4-2018Q4	29	1990Q4-1999Q4	38	1.2%	0.3%	0.1%	0.7%	2.8%	0.2%	0.4%	11
MV	GDP, UR, NX	1990Q4-2018Q4	29	1990Q4-1998Q4	40	2.8%	0.8%	0.3%	1.1%	4.4%	0.9%	1.0%	44
PF	GDP (CAM)*	1990-2018	29	1990-2003	33	2.2%	0.6%	0.3%	1.0%	3.6%	0.6%	0.9%	15
IE													
UV	Domestic GVA	1995Q4-2018Q4	24	1995Q4-1998Q4	40	3.8%	1.2%	0.1%	1.4%	7.7%	0.9%	1.5%	17
MV	Domestic GVA, Modified current account	1995-2017	23	1995-1999	19	2.5%	0.9%	0.5%	1.3%	3.1%	1.0%	1.6%	17
PF	GDP (CAM)*	1995-2018	24	1995-2003	33	3.3%	1.0%	0.4%	1.4%	4.2%	1.0%	1.4%	27

Note: Figures relate to annual estimates of the output gap. Intervals refers to the range of estimates for a given year across the vintages of the output gap. While the univariate and multivariate approaches are estimated on a quarterly basis, for comparability purposes with the production function approach, the figures shown here are for biannual vintages. Due to data availability issues, there are only 10 vintages of the multivariate output gap for Italy. Sign changes refer to a simple count of the number of changes of the sign on the output gap estimate for a given year from one vintage to the next.

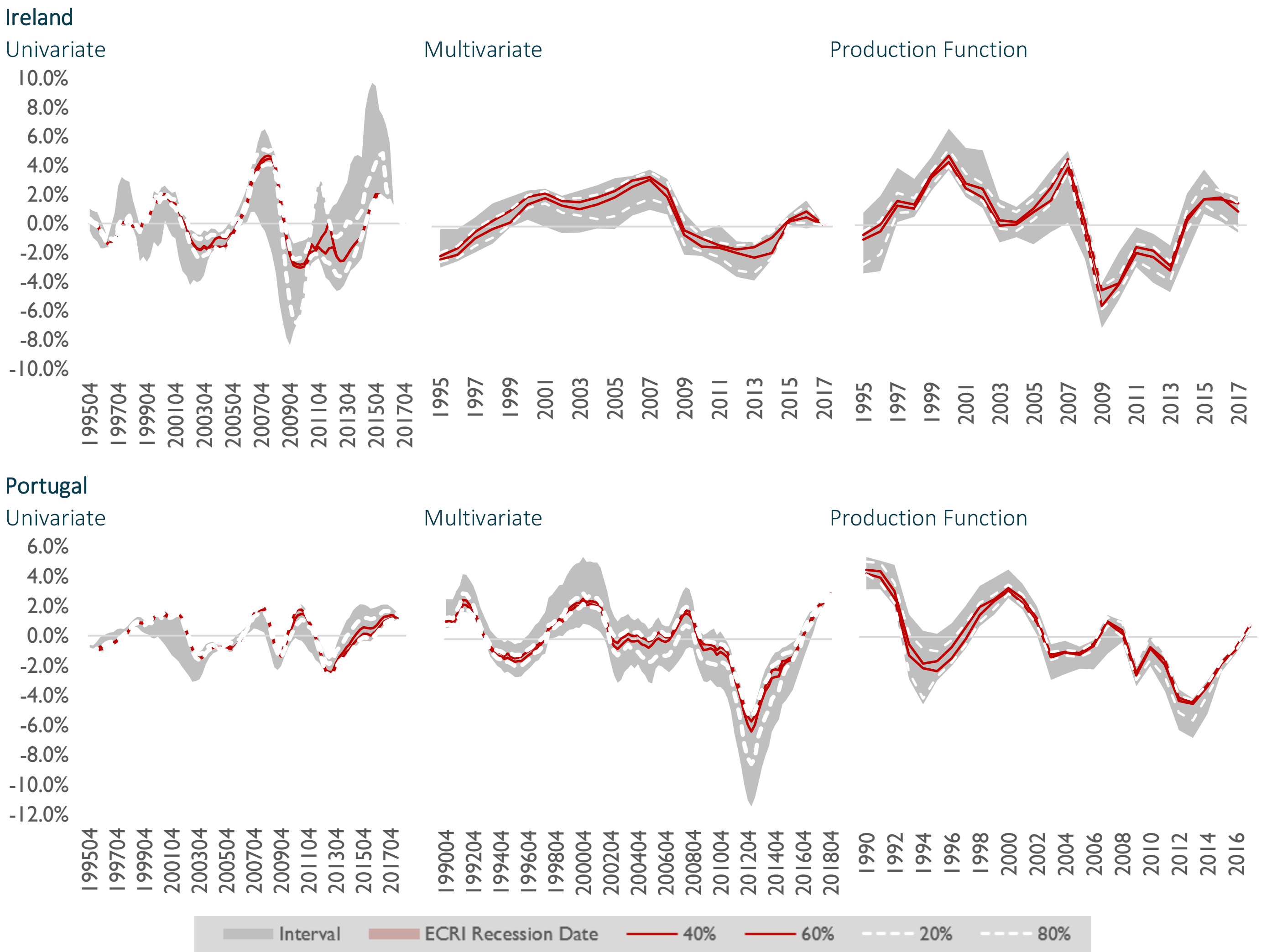
Table B2: Summary statistics of potential output growth estimates across vintages

	Variables	Sample				Intervals				Revision (relative to most recent vintage)		
		Total period	Obs	First sample	Vintages	Mean Interval width	SD	40%-60%	20%-80%	Max abs rev	Mean abs rev	SD
Spain												
UV	GDP	1986Q4-2018Q4	33	1986Q4-1998Q4	40	0.8%	0.3%	0.1%	0.3%	2.4%	0.1%	0.3%
MV	GDP, UR, CA, Cons_emp	1986Q1 - 2018Q4	33	1986Q4 - 1998Q4	40	1.0%	0.3%	0.1%	0.4%	2.4%	0.2%	0.3%
PF	GDP (CAM)*	1965-2018	33	1981-2003	32	0.7%	0.2%	0.1%	0.3%	1.1%	0.2%	0.3%
Italy												
UV	GDP	1986Q4-2018Q4	33	1986Q4-1998Q4	40	0.5%	0.2%	0.1%	0.3%	1.6%	0.1%	0.2%
MV	Emp, UR, M&E, Cons_inv	1986Q1 - 2017Q4	32	1986Q4 - 2005Q3	10	1.3%	0.5%	0.1%	0.7%	3.2%	0.6%	0.8%
PF	GDP (CAM)*	1985-2018	33	1985-2003	32	0.5%	0.1%	0.0%	0.2%	0.9%	0.1%	0.2%
UK												
UV	GDP	1986Q4-2018Q4	33	1986Q4-1998Q4	40	1.0%	0.3%	0.1%	0.2%	2.6%	0.3%	0.4%
MV	Comps, UR, FX	1986Q1 - 2017Q4	32	1985Q4 - 1999Q1	38	1.5%	0.4%	0.2%	0.6%	1.8%	0.5%	0.6%
PF	GDP (CAM)*	1985-2018	33	1985-2003	33	0.6%	0.2%	0.1%	0.3%	0.9%	0.2%	0.2%
PT												
UV	GDP	1991Q4-2018Q4	28	1991Q4-2000Q4	36	0.70%	0.21%	0.10%	0.30%	1.60%	0.10%	0.20%
MV	GDP, UR, NX	1991Q4-2018Q4	28	1991Q4-1990Q4	38	0.50%	0.15%	0.10%	0.20%	1.20%	0.10%	0.20%
PF	GDP (CAM)*	1991-2018	27	1991-2002	32	0.60%	0.20%	0.10%	0.30%	1.20%	0.20%	0.30%
IE												
UV	Domestic GVA	1996Q4-2018Q4	23	1996Q4-2000Q4	36	0.80%	0.25%	0.10%	0.40%	1.60%	0.10%	0.20%
MV	Domestic GVA, Modified current account	1996-2017	22	1996-1999	19	2.42%	0.67%	0.21%	0.59%	5.00%	1.20%	1.10%
PF	GDP (CAM)*	1996-2018	22	1996-2002	32	2.50%	0.70%	0.20%	0.80%	20.70%	0.40%	1.10%

Note: Figures relate to annual estimates of potential output growth. Intervals refers to the range of estimates for a given year across the vintages of potential output growth. While the univariate and multivariate approaches are estimated on a quarterly basis, for comparability purposes with the production function approach, the figures shown here are for biannual vintages. Due to data availability issues, there are only 10 vintages of the multivariate output gap for Italy. Sign changes refer to a simple count of the number of changes of the sign on the output gap estimate for a given year from one vintage to the next.

Annex C: Charts for Ireland and Portugal

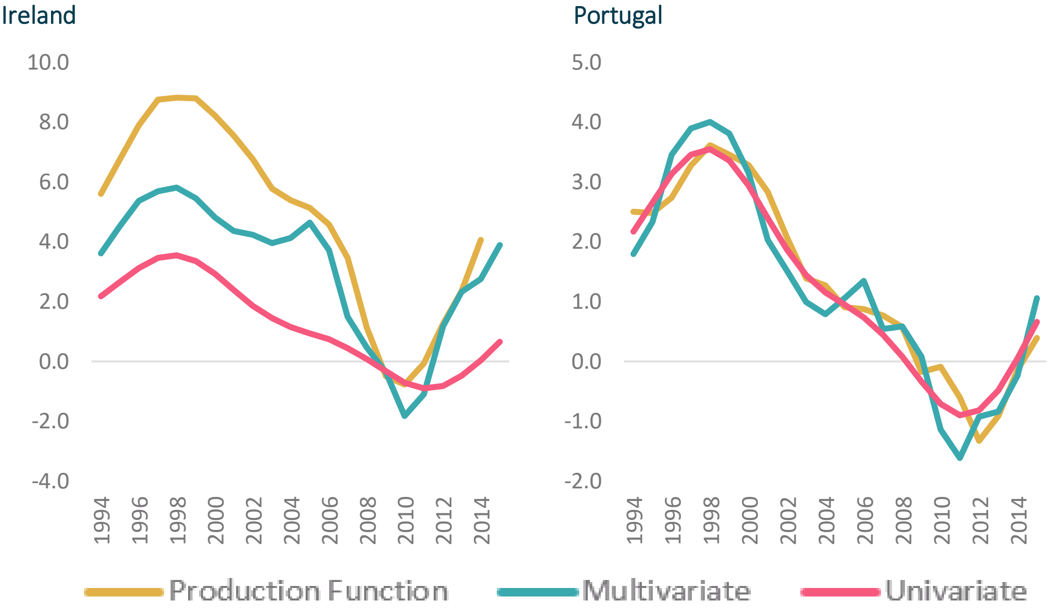
Figure C1: Vintages of output gap estimates



Note: Data covers 1990-2018 for Portugal, and 1995-2017 for Ireland. Figures show the interval of estimates of various vintages of the output gap in grey. The 20-80 percentile of vintages of the output gap estimates for each year are shown in white dashed lines, while the 40-80 percentile are shown in red. Shaded pink region shows the ECRI recession dates for each country. Due to data availability issues, the multivariate output gap estimation for Italy only has a maximum of 16 vintages (the UK estimate has 76, and the Spanish estimate has 80). Both the univariate and multivariate figures show data based on quarterly estimation. The production function figures show data based on biannual estimation.

Figure C2: Potential output growth rates compared

% year-on-year growth rates



Note: The figures shows a five-year moving average (+/- 2.5 years from each date) of potential output growth for the multivariate and univariate approach, alongside the actual figures for the production function approach.