EU **INDEPENDENT**
FISCAL INSTITUTIONS

A Practitioner’s Guide to Potential Output and the Output Gap

Definition · Estimation · Validation

www.euifis.eu

Abstract

This paper provides a comprehensive literature review of potential output and output gap estimates from the perspective of a fiscal authority and, by extension, an independent fiscal institution tasked with assessing cyclically adjusted fiscal indicators. Considering the mandate of these institutions, the focus is often broader in the sense that more sources of imbalances and longer horizons are considered. Yet it is similar to other institutions in terms of the methods used to assess the unobserved potential output and the output gap. The paper reviews univariate and multivariate trend-cycle decomposition methods that are actually used within the Network of European Union's independent fiscal institutions. It summarizes their salient features and provides a critical review of commonly used methods. This literature review preludes the back-testing exercise assessing the quality of output gap estimates and the discussion of their real-time applied issues in the context of cyclically adjusted fiscal indicators.

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.

The paper has been prepared by the Network's Output Gap Working Group coordinated by Dmitrij Celov from the Lithuanian IFI. Special thanks to Eddie Casey of the Irish IFI for providing a thorough review of the document in addition to contributing to it.

Table of Contents

List of Abbreviations

Abbreviation	Full Form
A	Autumn forecast
AMECO	Annual macro-economic database of the European Commission
ARIMA	Autoregressive Integrated Moving Average model
BK	Baxter-King filter
BN	Beveridge-Nelson decomposition/filter
CAB	Current account balance
CAM	Commonly agreed methodology
CD	Cobb-Douglas
CEE	Central and Eastern Europe countries
CES	Constant elasticity of substitution
CF	Christiano-Fitzgerald filter
CIRCABC	Communication and Information Resource Centre for Administrations, Businesses and Citizens
CPI	Consumer price index
CUBS	Capacity utilization and business surveys indicator
D	Downturn point
DSGE	Dynamic stochastic general equilibrium
EC	European Commission
ECB	European central bank
ESA 2010	European System of National and Regional Accounts 2010
EU	European Union
GDP	Gross domestic product
GG	General government
GNI	Gross national income
GVA	Gross value-added
HP	Hodrick-Prescott filter
IBP	Ideal band-pass filter
IFI	Independent fiscal institution
IMF	International Monetary Fund
KF	Kalman filter
L	Latest known forecast
M1	Narrow money
M3	Broad money
MAR	Mean Absolute Revision
MIMIC	Multiple indicators and multiple causes model
MLE	Maximum likelihood estimation
MR	Maximal Revision
MS	Microsoft
MUC	Multivariate unobserved components model
MUF	Multivariate filter
NAIRU	Non-accelerating inflation rate of unemployment
NAWRU	Non-accelerating wage rate of unemployment

N_i	Necessary condition *i*
NKPC	New Keynesian Philips Curve
OECD	Organisation for Economic Co-operation and Development
OG	Output gap
P	Peak point
PC	Prior consistent filter
PCA	Principal components analysis
PF	Production function
PIM	Perpetual inventory method
S	Spring forecast
SEATS	Signal Extraction in ARIMA Time Series
S_i	Sufficient condition *i*
STS	Structural time series model
T	Trough point
TC	Trend-cycle decomposition/filter
TFP	Total factor productivity
TRAMO	Time series Regression with ARIMA noise, Missing values and Outliers
U	Upturn point
UC	Unobserved components model
UF	Univariate filter
UK	United Kingdom
VBA	Visual Basic
VCV	Variance-covariance matrix

Introduction

Actual General Government (GG) budget balances are imperfect indicators for assessing public finances and the fiscal policy stance as they are affected by a number of temporary and cyclical factors that are beyond the direct control of fiscal authorities. In implementing their functions, Independent Fiscal Institutions (IFIs) strive to understand the dynamics of the output gap and reveal its specific location as it decisively determines the current budgetary stance and its sustainability (Cuerpo, Cuevas and Quilis 2018, Casey 2018). It is also important to have a medium-term view as to economic growth projections in the context of expenditure planning and GG debt sustainability assessments – something that estimates of potential output can help to assess (Casey 2018). Therefore, plausible real-time potential output and output gap estimates become an essential prerequisite to timely implement required countercyclical fiscal policy measures.

The source of substantial uncertainty surrounding the cyclical position of the economy primarily stems from the *unobserved* nature of the potential output used to derive the output gap. Such estimates are a huge problem especially for a small open economy[1] with a short history of data, transitionary nature of economic development and many structural breaks (Ódor and Jurašekova Kucserová 2014). Lacking an objective benchmark, practitioners from different institutions may disagree on the conceptual issues, the theory that defines the potential output and the methods of its assessment. The recent economic crisis appears to have made the agreement even more complex (Frale and De Nardis 2017).

The practitioners often acknowledge that an estimation of the unobserved potential output and the output gap is more an art than a science. Any commonly agreed approach eventually requires addressing country-specific issues when applied to the diverse subset of countries such as the EU (Ódor and Jurašekova Kucserová 2014). The authors noted that revisiting calculations of the output gap in small and open economies is an essential task not only from a fiscal policy perspective since policymakers often use this concept also in relation to monetary or structural policies. Very volatile estimates of output gap with weak information content can quickly undermine the credibility of a fiscal framework and IFIs, which aim to assess the performance against cyclically adjusted indicators, or monetary authorities, who target to control the inflationary pressures and/or ensure financial stability.

Although there is an extensive literature on the topic, this paper aims to guide practitioners who wish to assess potential output and the output gap from the perspective of a fiscal authority and IFI. This is also relevant in terms of implementing the Fiscal Compact's watchdog function. Considering the broad mandate of IFIs the focus often goes beyond inflation in the sense that more imbalances and longer horizons are taken into account. Yet it is similar in terms of the methods used to assess the unobserved potential output and the output gap. Having this objective in mind, the first section of the paper provides a literature review of the conceptual issues surrounding the appropriate definition of the output gap for the purpose of assessing fiscal sustainability. The second section reviews the

[1] A small and open economy is an economy that is small enough compared to world markets that its export volumes and trade policies do not alter world prices or incomes. The country is, thus, a price taker in world markets. In that case, the export dynamics is determined by the supply-side constraints similar to Holly and Wade (1991): mark-up over marginal costs, productive capacity that could be approximated by the potential output of the open sector or the whole economy, and the relative export and domestic prices. The concept of small and open economy could also include the financial markets, where nominal interest rates (apart from a risk premium) are exogenously driven. The financial balance then may have a relevant role in a broad definition of stability.

univariate and multivariate trend-cycle decomposition methods that are actually used within the Network of EU IFIs summarizing their salient features. The last section concludes and also provides a roadmap for further research.

Conceptual issues

This chapter concerns the practical and theoretical aspects of potential output and output gap definitions. It considers the relevant time horizon, required data, sources of uncertainty, computation, and validation issues from the perspective of the fiscal authority and, by extension, of an independent fiscal institution. It is particularly relevant for those tasked with assessing performance against cyclically adjusted fiscal indicators (Murray 2014).

Minding the gap

The concept of the output gap is widely used in academic and economic policy debate, though it is not uniquely defined and estimation methods can vary widely (Ódor and Jurašekova Kucserová 2014). The idea of recurring booms and recessions in overall economic activity is not new and was put forward already at the end of the 19th century. A prominent early work in this field is Juglar (1862), which considers the link between the credit cycle—variation in productive fixed investments—and the economic cycle. In particular, it focuses on the role played by the speculative behaviour of agents (contagion effects). The 7 to 11 year cycles observed by Juglar are, perhaps, the closest in terms of definition to the modern meaning of the output gap.

Today almost every economist agrees on the existence of the cyclical fluctuations, but questions remain as regards the theory that best explains their causes and nature. A useful classification (Mazzi and Ozylidririm 2017) distinguishes cycles in terms of those that are:

1. self-correcting – mainly the classical theories;
2. non-self-correcting – (post-)Keynesian approaches allowing for active countercyclical policy interventions;
3. due to policy ineffectiveness – based on neo-classical theories; and
4. driven by wage and price rigidities – based on new-Keynesian theories.

Keynesian theory directly considers the active role of fiscal policy and is a central part of modern macroeconomics. The classical self-correcting mechanisms fail to adjust both: in aggregate demand due to the conventional monetary policy or deflation inability (the liquidity trap); and in aggregate supply as a result of rigid wages leading to persistent underemployment equilibrium. Hence, there is room for unconventional monetary policy, which can encourage private sector activities even under extremely low entrepreneurs' profit expectations; and countercyclical fiscal policy by direct government expenditure, mainly in infrastructure, and taxes (Mazzi and Ozylidririm 2017). The Keynesian theory also includes psychological factors ("animal spirits") as a potential irrational source of economic instability, especially concerning investment decisions (Keynes 1936).

This Guide adopts a specific output gap definition:

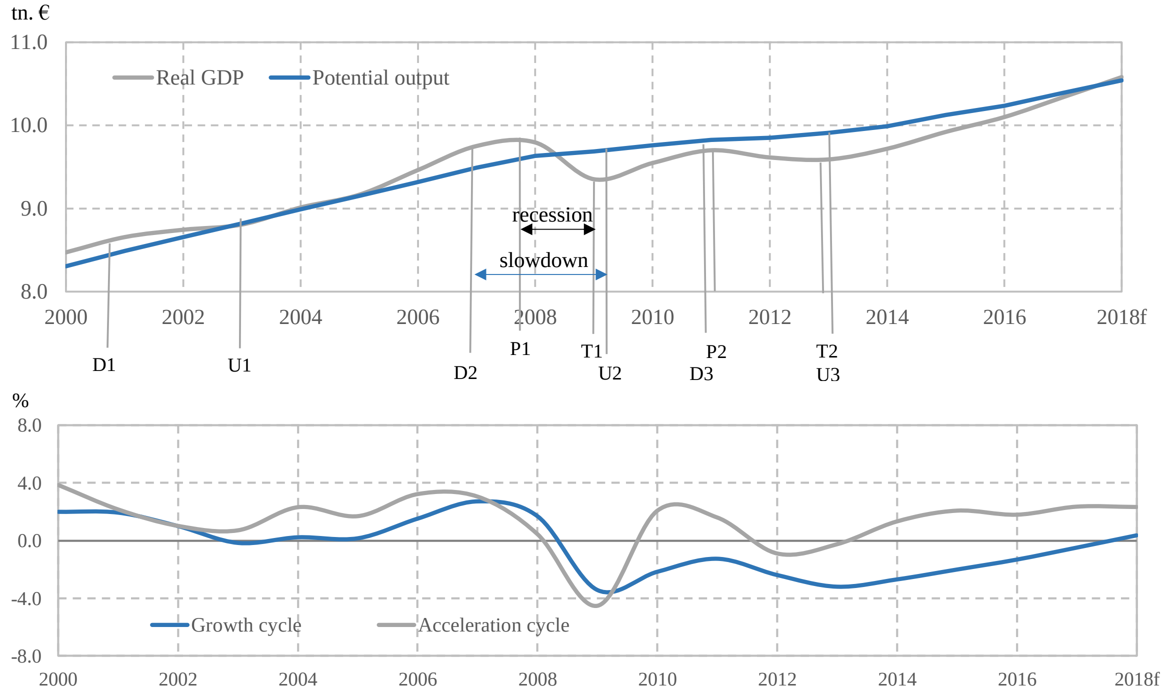
The output gap is a growth cycle – the difference between actual output and potential output expressed as a per cent of potential output.

It is important to distinguish between the "growth" measure of the cycle (Mintz 1969) or the output gap, and the "classical" measure or the business cycle (Burns and Mitchel 1946). The concepts are often confused in the literature **(Fig. 1)**. The growth cycle is dependent on the application of a particular trend-cycle decomposition method, while the classical cycle is not. Following Burns and Mitchel, the classical cycle is defined as aperiodic[2], recurrent fluctuations in aggregate economic activities of market economies. It relates to a sequence of expansions and contractions in the *levels* of a large set of aggregate macroeconomic variables, including the actual output and employment. However, such a definition may not adequately account for economies that experience a fast and stable growth path or for small and open (converging) economies. In such cases, it is unlikely that the classical definition will detect many cyclical fluctuations. This may simply be because the cyclical component will be dominated by the trend. In such cases, de-trending methods can help in making fluctuations more visible (Mazzi, Ozylidririm and Mitchell 2017).

The timing of downturns and upturns can differ in growth cycles and classical cycles. Downturns in the growth cycle will usually lead the peak in the classical cycle and vice versa for the upturns and troughs, implying that the slowdown phases are typically longer in output gaps than classical cycles, hence, the former is more symmetric **(Fig. 1)**. The recent estimates for the euro area show that the growth cycle may happen, while the classical cycle is still in the expansion phase, or the classical cycle may reach its new peak (W-type P1-T1-P2-T2 of cyclical recovery of the euro area), while the output gap is still negative **(Fig. 1)**. Therefore, classical cycles may be relatively less common than growth cycles because accelerations and decelerations in growth might occur without a decline in the level of economic activity.

[2] In duration varying from more than one year to ten or twelve years, but not divisible into shorter cycles.

Figure 1 – Business, growth and acceleration cycles in the euro area in 2000–2018



Notes: P and T are peaks and troughs of a business cycle, D and U denote downturns and upturns in the growth of a growth cycle. Acceleration or growth rate cycle is determined by the growth rate of the actual output (e.g. real GDP). A peak point of the classical cycle then is defined in the place where the acceleration cycle changes its sign from positive to negative, while a trough – from negative to positive. A growth cycle is defined as the difference between actual output and potential output expressed as a per cent of potential output.

Source: European Commission's 2018 spring forecast

Practitioners often acknowledge that an estimation of the unobserved potential output and the output gap is more of an art than a science. Facing such unobserved components is like "finding Yeti" (Ódor and Jurašekova Kucserová 2014), questioning "which gap" (Frale and De Nardis 2017), or "who holds the potential" (Constantinescu and Nguyen 2017) and often leads to a "beauty contest" or "horse race"-like choices (Cuerpo, Cuevas and Quilis 2018). Using indirect evidence at best, the researchers are seeking to assess the uncertain cyclical position of the economy.So how does one define unobserved potential output and the output gap?

The concept of potential output may be viewed from different perspectives (Anderton, et al. 2014).

- First, it could be associated with particular methods and the amount of theory being put into its development. In this sense, there are "*statistical*" univariate filters; *semi-structural* multivariate filters (Melolinna and Tóth 2016); and *structural* models (e.g. production function approach). The first category rests on very limited economic theory, if any at all, and assumes that it is possible to filter out cyclical fluctuations from the data (Ódor and Jurašekova Kucserová 2014).
- The second category can be associated with multivariate unobserved components models. These partly rest on either Phillips curve concepts of short-run zero inflationary pressures or "beyond-inflation" long-run sustainability concepts.

- The third view on the potential output originates from the structural model of the supply side or the economy's capacity to produce, where potential output in the economy is determined by a production function (e.g., Cobb-Douglas or CES)[3]. Murray (2014) defines potential output as being dependent on how many people are available to work and how many hours they are willing to put in (labour); the number of buildings, machines and computers that are available to work with (productive capital); and the efficiency with which they can be combined (total factor productivity).

Following Ódor and Kucserová (2014) and Casey (2018), in this paper, we define the potential output in a pragmatic way 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 view is broader than simple statistical de-trending methods or inflation-based models. For example inDSGE[4] models potential output conventionally follows from Phillips curve and simple Okun's law concepts referring to the maximum feasible level of output without inflationary pressures (Okun 1964). The practice has shifted from a narrow balance concept (internal) to a broader one (internal, external and financial) driven by the experience from 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 corresponding structural fiscal indicators. Therefore, practitioners consider other variables beyond inflation that help to better filter out underlying trends highlighting imbalances in other markets. 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) are among the most recently discussed candidates.

Prospecting the future

The fiscal authorities setting their spending plans are more concerned about sustainability, focusing on the medium to the long-term concept of fiscal space and imbalances that might influence budgetary position. Havik, et al. (2014) provide a comprehensive generalisation of the potential output concept regarding how far the authority plans its corresponding policy actions:

- In *the short run,* the physical productive capacity of the economy may be considered quasi-fixed. Its comparison against actual or technologically effective output reveals by how much aggregate demand can rise in the short period without inducing supply constraints that lead to inflationary pressures, in line with Walrasian view on how markets move towards equilibrium by adjusting the prices, keeping quantity supplied (real output) almost unchanged.
- In *the medium run*, the dynamics of aggregate demand when supported by corresponding changes in the number of productive investments and favourable demographic development may endogenously induce the changes in capacity needed for its own support. The latter is

[3] In many actual applications the production function approach involves the use of statistical univariate or multivariate filters, hence could be also considered as a semi-structural one.
[4] Actually in DSGE models there exist two notions of the output gap related to New Keynesian Philips Curve (NKPC). The first is the "natural" level of output that would prevail under flexible prices under imperfectly competitive markets. Another is the "efficient" level of output that would prevail if both goods and labour markets were perfectly competitive, implying that the output gap considers both imperfect competition and nominal rigidities (Vetlov, et al. 2011).

likely linked to the expected profitability and adequate medium run wage dynamics with respect to labour productivity. This view is in line with the Marshallian interpretation of market driving forces – changing the number of primary factor inputs (labour and capital) implies output adjustments, whilst the technological progress (total factor productivity) remains almost unchanged.

- In *the long run*, the potential output is driven by a neoclassical Solow-Swan type exogenous growth model that is dependent on the demographic trends, especially of the working age population, and the dynamics of technological progress (or total factor productivity).

The medium and long run views correspond to the “beyond-inflation” concept of the potential output, where the output gap is defined as the difference between the actual output and a steady state (sustainable) growth path, without any explicit reference to inflation (Melolinna and Tóth 2016). The difference of this broader view from (dis)inflationary concept is seen in Murray’s (2014) example, where the central bank, in-line with duration theories, takes into consideration that long-term unemployed may exert less downward pressure on wages over the recessions (Blanchard and Summers 1987). Hence, constructing the growth cycle estimates is consistent with time-varying medium-term equilibrium rate (like NAIRU or NAWRU) and may improve the accuracy of inflation projections.

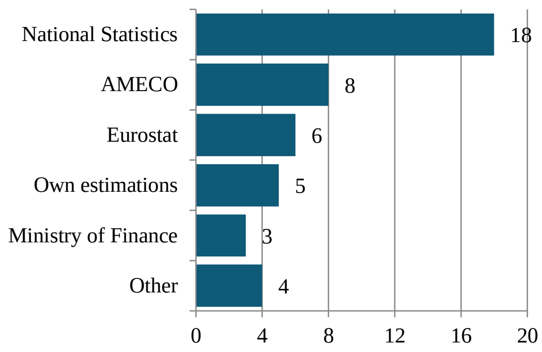
Over a longer time horizon, however, the long-term unemployed may well return to employment. Unless unemployment hysteresis prevails (Blanchard and Summers 1987), from membership theories perspective, then one could assume that automatic stabilizers will improve the corresponding fiscal indicators. In such cases, a government might be expected to conduct proactive discrete labour policy measures. Murray (2014) points out that, when long-term unemployment is still high, inflationary pressures may subside before the public finance fully recover. Therefore, a relevant cyclical correction of government finances has to be consistent with the long-term structural rate of unemployment. This consistency could be achieved, for example, by carefully choosing the long-run anchors, while applying the production function approach, or by judgements surrounding smoothing parameters of univariate and multivariate filters.

Requiring data

The data used for output gap estimation comes from several sources. The data used mostly originates from National statistical agencies **(Fig. 2)**. Some of the data series are own IFI's calculations or are taken from European Commission's database AMECO, which includes Eurostat data. AMECO is a useful primary source of data vintages for IFIs that have not stored the data series or produced their own estimates of the output gap[5].

The available vintages of output gap estimates within the Network resemble the actual lifespan of EU IFIs. A few older members having vintages dating back to the pre-crisis period, while vintages of the new IFIs start mostly from the post-crisis period with hikes in 2010 and 2015. The data sample may start from 1970 as in Denmark, Ireland or the UK, 1980 in Spain, but is limited to begin in 1995 or 1999 for the most of the EU countries. This can mean that the data samples are sometimes relatively short (covering one to two classical cycles) for the purposes of determining long-term trends, and ssmpled may include many structural breaks (Ódor and Jurašekova Kucserová 2014).

Figure 2 – Data sources used for estimation



Source: the Network of EU IFIs

The data frequency is either annual or quarterly. Seeking to exclude seasonal impacts, in the quarterly case, the data is adjusted for seasonally and calendar effects either by official statistical agencies or by own IFI calculations applying, for instance, TRAMO-SEATS method (Gómez and Maravall 1996). Quarterly estimates are aggregated to annual frequencies using the average for level variables and the summation for flows. Mostdata, aside from some rates and ratios, are included in a logarithmically transformed form. We will conventionally denote transformed variables as $x_t = \log X_t$, while X_t will correspond to actual seasonally and calendar adjusted series.

The quality of the potential output estimates is largely determined by an appropriate macroeconomic aggregate for perceiving the cyclical position of the economy (Casey 2018). Bearing in mind country-specific features, the IFIs often scrutinize an appropriate measure of actual output. The default option is real GDP data. However, alternative real gross national income (GNI) or gross value-added (GVA) measures might be usedand can be expressed in per capita terms and/or net of certain components. In some cases, this can help to better match the salient features of a particular economy (Murray 2014). Such alternatives tend to consider sectors of the economy that are not influenced by the country's domestic resources. Examples of sectors removed for the estimation of potential outputmin practice include oil production in the UK (Murray 2014) or the value-added produced by large foreign-owned multinational enterprises, which are not very integrated with the domestic economy, in Ireland (Casey 2018). The latter exclusion is equivalent to an assumption that the omitted sectors always operate at a technologically efficient full capacity so thatthe output gap is assumed to be driven primarily by domestic developments.

[5] An additional source of vintages for actual output, inflation and current account balances are published in IMF World Economic Outlooks.

Aside from output aggregates, IFIs may also explore the inclusion of a number of additional observable variables in a multivariate setting and while applying a production function approach. The selection of potential candidates could follow an encompassing approach such as in Cuerpo, Cuevas and Quilis (2018). This classification covers—in line with a broad potential output concept sense—all relevant sources of imbalances: (i) domestic demand; (ii) external sector; (iii) prices; (iv) labour market; (v) financial sector and asset prices; (vi) fiscal conditions. The relative ease of replicability of the data set for different countries allows a practitioner to reflect different sources of imbalances that may underpin growth cycles. This motivated the selection of similar groups (with the only difference that (i) and (ii) were joined into *GDP and output category*) of variables and vintages of data for other members of the Network of EU IFIs to be used in the back-testing exercise **(Appendix A)**.

The GDP and output category of variables include coincident and leading indicators that complement the corresponding aggregate measures of economic activity. Measures considered include the current account balance; output surveys; gross national savings; housing completions; real investments in equipment/machinery and investments in construction; net capital stock (probably excluding dwellings); new car registrations; and a measure of capacity utilisation.

The inflationary channel is defined either by the GDP deflator, CPI inflation or more domestically aligned CPI subcomponents. The latter may, for instance, include: core inflation – CPI inflation net of food and energy prices (Jarociński and Lenza 2016); or services inflation that might more closely represent the non-traded element of domestic inflation (Casey 2018). This block of variables could also include explicit data for inflation expectations as a component of NKPC corresponding models.

The non-inflationary concepts of potential output tightly relate the price block variables with the third block on labour market developments, in particular with an unemployment rate and the corresponding unemployment gap. For the labour input in production function approach a decomposition into the total population, participation rate and average hours worked is required, whilst nominal earnings growth is a key determinant of unit labour costs (cost-push inflationary factor). This block of variables might also include data on recruitment difficulties and labour force mobility captured by net migration.

Following Borio, et al. (2017) a financial channel and monetary policy impacts are represented by a battery of financial market and assets indicators, although the conclusions regarding their relevance are rather mixed. In these applications practitioners consider various credit and commodity price cycles associated variables: credit growth (could be separated into credit to households and to non-financial corporations); real interest rates; real effective exchange rates; house prices; asset prices; oil prices; and money supply in the narrow (M1 aggregate) and broad (M3 aggregate) sense.

The final block of variables considers the (past) fiscal policy impacts on the cyclical position of the economy and covers: net lending or borrowing of the general government; public sector debt; and general government receipts and spending or their corresponding subcomponents (taxes and unemployment benefits), which are a subject of cyclical adjustment.

Accommodating uncertainty

The output gap—as the growth cycle—is an unobserved object. It is constructed by analysts to better understand the cyclical fluctuations in the economy and to help to inform an adequate counter-cyclical policy in a timely manner. Murray (2014) argues that it is not the type of data which could be retrieved with certainty, even with the benefit of hindsight. Hence, frequent and substantial revisions of potential output estimates are often the rule rather than the exception (Ódor and Jurašekova Kucserová 2014).

The output gap, by definition, is the difference between two variables: *actual output* and *potential output*. The actual output is officially measured by the statistical agencies and, hence, is observed. Being a statistical estimate itself, however, the actual value of output is uncertain and is a subject of further revisions. Therefore, the output gap is surrounded by considerable uncertainty originating from both the observed data revisions and the unobserved potential output estimates. This uncertainty can be characterized by three main sources (Murray 2014, Blagrave, et al. 2015, Mazzi, Ozylidririm and Mitchell 2017):

- *Model uncertainty*. The notion of potential output is tightly related to the variety of methods of its assessment and the uncertainty regarding estimated or calibrated parameters, including hyper-parameters for filters. Many alternative econometric methods used to conduct trend-cycle decomposition can produce a range of estimates. This mightprovide a rationale for model-averaging or interval predictions as discussed below. This should be done taking into account (1) "within" model uncertainty – a density surrounding the point estimates of the output gap for a selected method, and (2) "between" model uncertainty – the range of point estimates from different models for a given period of time. It is crucial to recognise that, even for a single method, ("within") uncertainty in real time could be huge. A limited IFI practice[6] is to consider just "between" model uncertainty constructing either arithmetic average (Murray 2014, Ódor and Jurašekova Kucserová 2014); weighted average, where the choice of the weights is based on expert judgement; or mid-ranges (Casey 2018) of the point estimates obtained from different models. Drawing on the forecasting literature, a useful way to conceive of the differences is in terms of model uncertainty (i.e., within uncertainty), and model disagreement (i.e., between uncertainty).
- *Data uncertainty* arises because (1) the statistical information available is not the final vintage of the data and (2) the data definition can change substantially (e.g., high uncertainty to estimate capital input). The problem is more prominent for small and open catching-up economies dealing with short-time series to estimate long-term trends with many structural breaks. Some methods will be less sensitive to this source of uncertainty than others, depending on which revisions attribute more to changes in potential output or the output gap. In general, a richer structural model is expected to be more sensitive to data revisions than a simpler method.
- *End-of-sample uncertainty* is crucial for policymakers as the policy decisions require cyclical estimates *in real-time*. The real-time estimates are, however, conditional on incomplete information regarding future data revisions as well as on the uncertain projections of certain economic indicators. The latter is a more relevant issue for univariate, purely statistical filters than for semi-parametric models estimated by Kalman filter[7] (Kalman 1960). Most univariate filters imply a smooth two-sided moving average in the middle of the sample, while at the ends of the sample the filter becomes, first, asymmetric and then one-sided. Since the

[6] Only 3 of 20 respondents of the survey of the Network of EU IFIs conducted in October 2017, mentioned the use of model averaging.

[7] A state-space representation of univariate filters could be also solved by Kalman approach. It is the absence of additional information including the explicit model for cycle makes the real-time estimates by statistical filters biased and the errors huge.

revisions of one-sided filters are huge, seeking to apply the two-sided filter at the end-of-sample the future values of variables are projected. Any revisions in these projections hence will translate into end-of-sample uncertainty. For the unobserved components models, the end-of-sample uncertainty becomes relevant if the model is misspecified.

All these types of uncertainty will be at the heart of back-testing exercise. The effect of data revisions and end-of-sample uncertainty are assessed by comparing the real-time and ex-post estimates for a variety of the methods; while model uncertainty is reflected in the range of output gap estimates produced by different methods (Murray 2014, Casey 2018, Cuerpo, Cuevas and Quilis 2018).

Assessing performance

Since the potential output of the economy is an unobserved ingredient of the output gap, a practitioner is not able to measure the accuracy of the trend-cycle decomposition, even *ex-post* (Murray 2014). While economic theory or more structural models are believed to highlight the "true" data generation processes for trend-cycle decomposition, in fact, there is also a significant uncertainty on what the "right" economic theory should be applied (Mazzi, Ozylidririm and Mitchell 2017). The absence of observable targets or benchmarks with which to compare any estimates produced complicates the assessment of different output gap estimation methods. Therefore, in practice, the fiscal authorities and IFIs have to agree on a set of desired attributes that a preferred method is expected to possess.

There are a number of desiable attributes of output gaps. Murray (2014), Ódor and Kucserová (2014) among others emphasize two features: the *stability* of the real-time estimates in the sense of small ex-post revisions at the end-of-sample; and the *plausibility* of the estimates when new data arrives. To maintain the credibility of the output gap estimates in the fiscal framework context, it is recommended to avoid hikes in the revisions of the output gap estimates, especially in the short-term horizon[8]. This is why the end-point stability has to be a primary concern for the EU countries (Darvas and Simon 2015). On the other hand, it is not reasonable to rank the output gap estimation methods simply on their tendency to be revised. An extreme illustration of this is the fact that athe method that always sets potential and actual output equal at all points in time will imply an always zero-valued output gap that will never be revised (Murray 2014). Such a method, however, is clearly inconsistent with economic theory. Among other things, it would implicitly assume that unemployed never return to work. Moreover, there is no point in favouring an unchanged output gap value when it is at odds with empirical evidence, and when hindsight underscores the country-specific trade-off between the stability and plausibility features.

To test the stability of output gap estimates, Casey (2018) and Turner, et al. (2016) suggest examining three measures: (i) the Mean Absolute Revision (MAR); (ii) the Maximal Revision (MR); and (iii) the number of sign changes observed. The revisions could be calculated in different ways and compared using different vintages of output gap estimates. According to the survey of the EU IFIs, the output gaps often are estimated twice per year correspondingly in spring (S) and autumn (A) updates. Therefore, a practitioner can test the revision within the same year, comparing A and S estimates; period-to-period updates on a semi-annual basis; year-to-year updates comparing either SS or AA revisions; and to compare any value with the latest (L) known estimate. The bulk of the analysis of

[8] Ódor and Kucserová (2014) explain this in terms of deviation from the medium-term objective (MTO), where a significant deviation in one year is defined as 0.5 per cent in one year or 0.25 per cent on average for two years. Therefore, the preferred method should be relatively stable at least for one year.

revisions is paid to the end-of-sample revisions, where the projected data is often used. Actual data and model revisions are the key drivers of the output gap revisions in the middle-of-sample and illustrate realized in-sample uncertainty of the selected method.

In general, if t denotes time (a year or spring/autumn updates of the year), and i is the corresponding update of the output gap estimate $x_{t,i}$ for a given year t, then a revision is defined as: $R_t = x_{t,i} - x_{t,i-k}$, where k is an integer higher than 1 and depends on the chosen comparison. A distribution of the revisions R_t depicted, for instance, as a box-plot, a histogram or a density plot would illustrate then the descriptive statistics of the revisions for different methods without significant loss of information. Another option is to analyse the absolute value of the revision $AR_t = |R_t|$ that shows the amplitude of revisions ignoring their direction or the final absolute revision $FAR_t = \left|x_{t,n(t)} - x_{t,1}\right|$ that highlights the differences between the first and the last values of the output gap estimates. Any descriptive statistic – an often the choice is mean and max values – of the absolute revisions is then used for the comparison of methods (Casey 2018). Revisions could be also defined in relative terms as the average distance between one-sided and two-sided estimates (or the mean final absolute revision if the latter are not available), relative to the maximum amplitude of the estimate (Cuerpo, Cuevas and Quilis 2018).

Since the fiscal rules are often dependent on the sign of the output gap, a supplementary characteristic of the stability is the number of sign changes from positive to negative and vice versa, comparing different vintages of the output gap estimates (Casey 2018). In a similar way, one could explore the ability of different methods to detect turning-points (Mohr 2005).

Another conventional evaluation criterion is to choose the reference economic theory based on the objectives of the policy applied either by the monetary or fiscal authority and the ability of the method to predict the theory based outcomes (Murray 2014). For the central banks and some IFIs (Frale and De Nardis 2017, Casey 2018), the estimates of the output gap then could be assessed on their ability to predict inflation, explaining the cyclical element that arises from the demand pressures. For fiscal authorities and IFIs as fiscal watchdogs, a reasonable metric might be the ability of the output gap estimates to explain the cyclical variations in the public finances (or the imbalances in the broader sense). However, the bottleneck in such reasoning is the circularity[9] of the trend-cycle decomposition – how a practitioner could define the cyclical parts of the inflation or the fiscal balance without the estimates of the growth cycle? As a result, for instance, the inflation targeting output gap estimates could place the methods based on the Phillips curve (undeservedly) high (Murray 2014). Hence, it is reasonable to judge the methods by considering the *general plausibility* of the output gap estimates – an intuitive "smell" test or country-specific narrative approach (Casey 2018, Cuerpo, Cuevas and Quilis 2018), keeping an eye on the descriptive statistics of their revisions.

Mazzi, Ozylidririm and Mitchell (2017) stress that a comparative analysis among different detrending techniques only makes sense for the same family of methods, otherwise, the underlying definition of the "cycle" is very different. Besides, the methods which aim at extracting cycles (high-pass, band-pass filters) will typically introduce smoother cycles than those aiming to extract smoother trends

[9] Martti (2015) argued that the potential GDP estimates are dependent on projections, which themselves typically depend on the cyclical phase, i.e. the potential GDP (and the output gap) estimates are pro-cyclical. As the cyclically adjusted budget balance (CAB) depends on the elasticity of the output gap, one can show that, with given nominal budget balance change and given GDP growth, the change of cyclically adjusted budget balance depends on the change of the potential GDP. Thus, whenever estimates of the potential GDP projections depend on the cycle the interpretation on the fiscal policy stance (as a change of CAB) will depend on the cycle too.

(low-pass filters), which is a typical outcome for the methods without explicitly defined irregular component (1). Indeed, the latter is allocated either to trend or cycle making the focus component smoother, while the residual more volatile.

Bearing all these aspects in mind, a comprehensive selection approach on how a "beauty contest" among different models could be organized is provided in Cuerpo, Cuevas and Quilis (2018). First of all, the contest is organized for the models within the same class of unobserved components methods, which are cast in the state-space form. Notice, that widespread filters, such as Hodrick-Prescott filter or Beveridge-Nelson decomposition, belong to the same class of methods and admit a state-space representation. Then the authors split the selection criteria into two categories: the statistical-based ones define the *necessary* (N) conditions, and the economically and policy-oriented ones underline the *sufficient* (S) conditions. The criteria listed in **Table 1** form the core of the back-testing exercise.

Table 1 – Necessary and sufficient selection criteria.

Criterion	Description
N_1	Statistical significance of the coefficients, focusing on the loadings of the observables on the cycle.
N_2	Average relative revision, defined as the average distance between one-sided and two-sided estimates, relative to the maximum amplitude of the output gap estimate.
N_3	Average relative uncertainty surrounding the cycle estimates, as the average standard error relative to the maximum amplitude.
S_1	Economic soundness, meaning that some key macroeconomic relationships could be captured by variables if included in the model (e.g. Okun's Law, Phillips Curve, etc.).
S_2	The amplitude and profile alignment with 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.
S_3	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. Stability can be measured using the revisions of the one-sided estimates.

Source: Cuerpo, Cuevas and Quilis (2018).

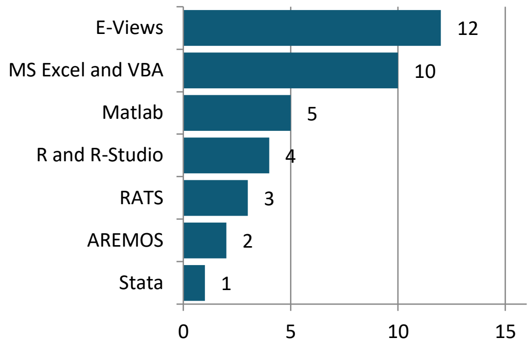
The necessary criterion N_1 could be also extended by the requirement of stability of parameter estimates, measured in terms of the number of observations necessary to achieve stable parameters (Frale and De Nardis 2017).

Computing without pain

Dissemination of the codes and sharing the best practices within the Network of EU IFIs require the use of the similar software. Roughly speaking, any statistical package or programming language capable to work with the vectors and matrices could be used.

The commonest software used by the EU IFIs for the assessment of the potential output and the output gap is E-Views and MS Excel **(Fig. 3)**, where for the latter the required statistical algorithms are either coded directly with a help of visual basic (VBA) or the spreadsheets are integrated with Matlab or R. Matlab and R are often considered as the second best choice. Besides, 9 of 20 respondents noted that they consider the European Commission CAM Toolbox (Havik, et al. 2014) to replicate the output gap estimation.

Figure 3 – Software used for estimation



Source: the Network of EU IFIs

In order to integrate the whole estimation process and to be able to consider the different variable combinations, Cuerpo, Cuevas and Quilis (2018) suggested designing an Excel platform that integrates the database, the estimation functions in Matlab and a stability analysis (back-test). In similar applications by other IFIs, an Excel platform is integrated with E-Views, R or the corresponding algorithms are coded directly with VBA. Noteworthy, the graphical and computational capabilities of Matlab and R are similar, not to mention that it is straightforward to adopt the code written for Matlab to R and vice versa. Yet R is an *open source* software, which is easy to distribute and share with all interested parties increasing the transparency of the methods been used for the output gap assessment.

Assessing practical attractiveness of different estimation approaches Casey (2018) assesses the *computational complexity* of the estimation processes involved. Complexity could be a useful predictor of the likelihood that some computation errors may happen, not to mention a growing probability of system errors occurring as more complex and structural systems of equations are developed with many unknown parameters (a statistical definition of a model complexity). In the field of computer science, one useful approach to testing the complexity of an algorithm involves examining the number of statistical operations involved. This is relatively straightforward for a practitioner to investigate, given that models are often coded in the same software package (Matlab, R or E-Views) so that the operations employed are comparable. One can count all of the statistical operations (operational commands) used (e.g., sample selection, arithmetic, comparisons, accessing array's elements, assignment (Casey 2018)). For instance, Casey (2018) showed that the production function approach based on the CAM toolbox and replicated in E-Views required more than 160 such operations, while the number of operations required for univariate/multivariate filters and unobserved components models varied from 10 to 34. Though costly from a computational perspective, due to the number of procedures involved, mid-ranges or averages of a suite of model estimates can reduce the likelihood of output gap estimates being misspecified due to the properties of model averaging.

Estimating the Output Gap

This section considers the different potential output and output gap estimation methods focusing on the main challenges in their application. Each of the methods entails some arbitrary decisions and comes with its own advantages and drawbacks. These depend on the required information set, filtering technique, economic rationale, statistical properties, and objectives of the analysis (Bouthevillain, et al. 2001). By definition, a detrending method is the decomposition of a time series y_t (log of actual output) into a trend part y_t^*, a cyclical part c_t and an irregular shock ε_t that can be of varying magnitude and direction:

$$y_t = y_t^* + c_t + \varepsilon_t, \ \forall t = 1, \dots, T, \tag{1}$$

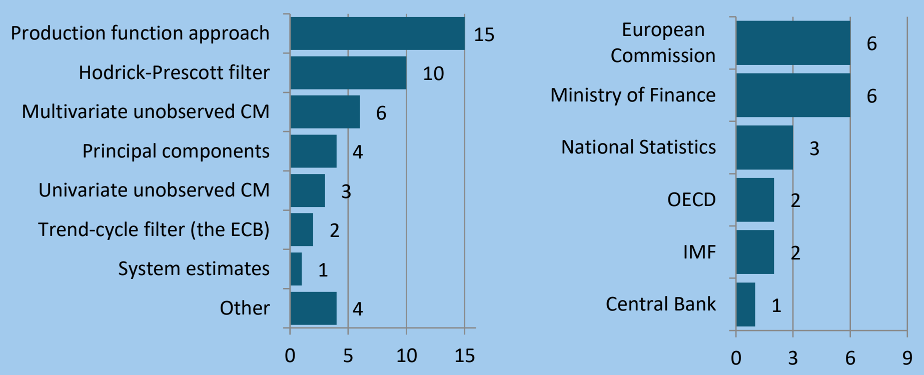
where t denotes time, the trend component y_t^* is interpreted as the potential output, while the cyclical component c_t is identified as the output gap. The decomposition (1) corresponds to the notion of a growth cycle, where the actual level of output should hover around the level of potential output, with deviations being contained and limited in time.

Trend-cycle decomposition methods could be grouped into several overlapping categories highlighting their particular properties: (a) parametric, semi-parametric and non-parametric; (b) univariate and multivariate; (c) statistical, semi-structural and structural; (d) linear and non-linear.

Box 1. An overview of the potential output methodologies used within the Network of EU IFIs

Most members of the Network of EU IFIs (15 of 20 respondents) produce their own independent estimates of potential output and output gaps, typically twice per year. There are two conventional methods to estimate the potential output: Hodrick-Prescott (HP) filter and a production function approach **(Fig. 4)**. Some advanced applications also cover univariate or multivariate unobserved components models and the principal components analysis (PCA). Currently, only 3 EU IFIs that participated in the survey mentioned the application of the combinations of output gap estimates produced by different approaches or sources.

Figure 4 – Potential output estimation methods within the Network of EU IFIs (left), other sources of output gap estimates (right)



Source: the Network of EU IFIs

Other sources of output gap estimates are either the European Commission or the Ministry of Finance, where the latter often replicates the Commission's commonly agreed methodology (CAM, Havik, et al. (2014)) based on their own economic development scenarios and assumptions. In some cases, output gap estimates used are sourced from the OECD and the IMF **(Fig. 4)**. All these alternative sources mostly apply the same two prevalent methods: the production function approach and/or HP filter.

In what follows we will focus on a short description of the methods actually used at least by one member of the Network of EU IFIs **(Box 1)** focusing on the economic soundness, the statistical goodness of fit and the transparency of computation (an optimality triplet suggested in Cuerpo, et al. (2018)). The first subsection considers purely statistical univariate methods – those that utilise just the output series. The next subsection covers semi-structural multivariate refinements, which make use of more than one variable: multivariate unobserved components models and principal component analysis. The third subsection will discuss the production function approach as applied in the European Commission's CAM and the best practice of application within the Network of EU IFIs. Then, we will turn to a few cases of the model averaging application that often lead to desired properties of the output gap estimates. The final subsection compares the methods and concludes the literature review.

Going alone

The least data-greedy detrending approach is to exploit only the information content of actual output data. This approach is purely *statistical* in nature and is categorised as a *univariate* since it does not involve other determinants than actual output. The simplicity comes at the cost of limited economic content if any at all – the univariate methods do not incorporate potentially useful information derived from 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 that undermines the use of the methods in real-time applications, especially for the small and open economies or the economies with many structural breaks, where structural breaks may be spuriously smoothed to an unreasonable degree (Ódor and Jurašekova Kucserová 2014, Hamilton 2017).

Ódor and Kucserová (2014) consider the most popular *non-parametric* univariate techniques. Among these, they list the HP filter, the Kalman filter (KF), Baxter-King's band-pass filter and its generalization Christiano-Fitzgerald filter; and the *parametric* Beveridge-Nelson decomposition. Mazzi, et al. (2017) and Murray (2014) also mention a *parametric* linear detrending probably with structural breaks. However, the practical relevance of deterministic trends in macroeconomic data omitting a stochastic part, not to mention the real-time assessment of structural breaks, is low. Hence, the latter method if not accomplished with the stochastic counterpart is often disregarded from the suite of models as implausible. Other univariate filters in use within the Network of EU IFIs include the prior consistent (PC) filter proposed by Laxton, et al. (1998), and the trend-cycle (TC) decomposition pioneered within the Eurosystem by Mohr (2005).

Hodrick-Prescott filter

The most prevalent univariate method among EU IFIs **(Fig. 4)** is the Hodrick-Prescott filter (Hodrick and Prescott 1997). The practical attractiveness of the filter is its computational simplicity and transparency. At the same time, the method is frequently criticized in the academic literature (see Hamilton (2017) among others). Hamilton criticises what are seen as implausible assumptions on the prior data generating process; severe end-of-sample distortions that challenge the use of HP filter for real-time applications; and, last but not least, the arbitrary choice of the smoothing penalty parameter that is vastly at odds with the statistical formulation of the problem. On top of that, Bouthevillain, et al. (2001) points to the structural breaks problem. Namely, that HP filters (as with any other two-sided moving average) are not able to detect sudden breaks in trends in real-time, e.g., in the course of the recent recession in 2009. A practitioner must be aware of the possibility of spurious dynamics in the output gap being implied as a result of all these distortions.

HP filtering problem rests on two assumptions (Murray 2014):

- an actual output does not deviate too far from its trend level (the output gap is not too big);
- the growth of the potential output is relatively smooth (the potential output is not too volatile).

The trade-off between these assumptions is formulated as the minimisation problem[10] in the squared loss function of the output gap with respect to y_t^* and subject to a constraint that penalizes (de-)accelerations in the growth rate of the potential output (Pedersen 2002):

$$\min_{\{y_t^*\}_{t=1}^T} \underbrace{\sum_{t=1}^{T}(y_t - y_t^*)^2}_{\text{goodness-of-fit}} + \lambda \underbrace{\sum_{t=2}^{T-1}[(y_{t+1}^* - y_t^*) - (y_t^* - y_{t-1}^*)]^2}_{\text{degrees-of-smoothness}}, \lambda = \frac{\sigma_1^2}{\sigma_2^2} > 0, \qquad (2)$$

where y_t^* is the unobserved potential output, y_t is the observed actual output, σ_1^2 is the variance of the output gap, σ_2^2 is the variance of the trend growth dynamics, and λ is a positive *penalty* parameter (Lagrange multiplier) that places a relative weight on the output gap “goodness-of-fit” and potential output “degrees-of-smoothness”. When the smoothness penalty approaches zero ($\lambda \to 0$), potential output estimates are forced to adhere to the actual output, i.e., the loss function’s value is exactly zero and the output gap vanishes. Whereas when it tends to infinity ($\lambda \to \infty$) the HP filter approaches a regression on a linear time trend, for which the second difference is exactly zero, implying all above-mentioned practical implausibility of the parametric linear de-trending. Therefore, the crucial question for the practitioner is what is a reasonable (though still arbitrary) choice of penalty λ to apply, which will directly influence the outcome of the trend-cycle decomposition **(Box 2)**.

Box 2. The Value of Penalty λ

The value of the penalty parameter λ is the only explicit choice associated with the application of the HP filter. The penalty parameter is a source of significant debate in the literature about its proper value, however, and there is no definitive way to choose or calibrate the optimal value of λ.

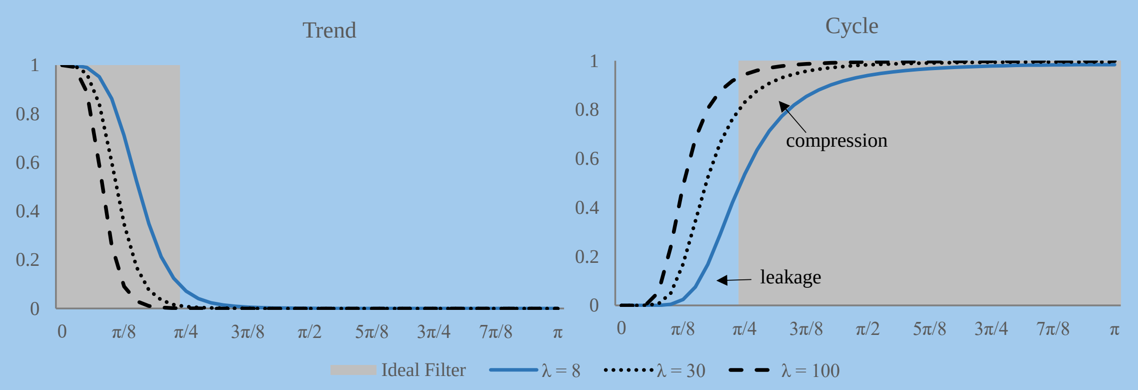
Ideally, the choice of the penalty parameter λ might reflect prior knowledge of the cycle length (Mohr 2005). However, the output gap, when defined as aperiodic recurrent fluctuations, does not provide any particular guidance as to what might be an empirically appropriate length of a cycle for a particular dataset and for a specific period of time. Besides, the choice of λ affects both the size of the cycle and the volatility of the trend – a direct consequence of the cycle in the HP filter framework not explicitly following any particular model. The problem becomes more visible at the end of a sample, when a new data point has to be allocated either to trend or cycle, even if it is an

[10] For simplicity of notations it is convenient to express objective function of (2) in the matrix form (Laxton and Tetlow 1992, Butler 1996, Mohr 2005): $(Y - Y^*)'W_c(Y - Y^*) + \lambda Y^{*\prime}\nabla^{2\prime}\nabla^2 Y^*$, where Y and Y^* are $T \times 1$ vectors of actual data and trend and ∇^2 denotes the second difference matrix. The form is convenient to include additional assumptions on the relative importance of particular periods of time via matrix W_c (default assumption is a unity matrix I) and to add additional restrictions. For instance the sought solution to the HP minimization problem will be $Y^* = \left(I + \lambda\nabla^{2\prime}\nabla^2\right)^{-1} Y$ and therefore is easy to program. For more details on lag and difference operators in matrix form see Appendix A in Mohr (2005).

outlier.

The choice of penalty parameter often reflects a general consensus. For example, common practice is to use a λ value of 1,600 for quarterly data and 100 for annual data as initially suggested by Hodrick and Prescott (1997). The authors argued that a 5 per cent deviation from trend (the output gap) is as moderate as the eighth percentage acceleration per quarter in the trend component. If the cycle and the acceleration in trend are mutually uncorrelated and normally distributed white noise processes, the solution to the minimisation problem (2) will be *an optimal filter* if and only if the parameter λ is equal to the inverse signal-to-noise variance ratio, i.e. $\lambda = \sigma_1^2/\sigma_2^2 = 5^2/(1/8)^2 = 1600$. These prior assumptions, however, are often violated in practice. Therefore, the HP filter might be interpreted not as an optimal filter, but rather as an approximation to an ideal infinite moving average (Pedersen 2002). Going from quarterly reasoning to annual, Ravn and Uhlig (2002) asserted that the default quarterly value is inconsistent with a 2nd order frequency conversion factor that results in value of 100 for annual data and provided frequency-domain arguments for 4th order frequency conversion leading to 6.25–8.25 recommendations[11]. Pedersen (2002), based on an optimal approximation of an ideal filter, argued for a value 1000 for quarterly data in the euro area and 3–5 for annual data, while in Bouthevillain, et al. (2001) the filter is applied with λ = 30 to annual data. From the IFIs point of view higher values of the penalty are preferable as they are more consistent with longer cycle horizons relevant to fiscal authorities.

Figure 5 – power transfer functions of the trend and the cyclical components of the HP filter for different values for λ and the ideal filer cut-off frequency $2\pi/8$



Source: the Network of EU IFIs calculations based on Mohr (2005)

The impact of different values of λ can be best illustrated in the frequency domain byanalysing the power transfer functions of the trend and the cyclical component for different values of the penalty parameter **(Fig. 5)**. The frequency response (gain) function for the cyclical component of the HP filter is defined as:

$$G^{HP}(\lambda,\omega) = \frac{4\lambda(1-\cos(\omega))^2}{4\lambda(1-\cos(\omega))^2+1}, \qquad (3)$$

where ω is the corresponding frequency to be filtered and the low-pass gain is $1 - G^{HP}(\lambda,\omega)$. The squared gain functions are the corresponding power transfer functions depicted in **Figure 5**. Lower frequencies would be ideally allocated to the trend and higher frequencies to the cycle. Higher

[11] A relevant analysis on the choice of λ across sampling frequencies is found in Maravall and del Rio (2007).

values of λ shift the gain function of trend closer to zero, hence the latter becomes smoother and approaches linear trend in the limit. On the contrary, with lower values of λ, the trend becomes more volatile as it will contain more of high-frequency spectrum approaching the original data when the penalty value drops to zero. A cyclical component as a residual of the trend-cycle decomposition will clearly have the opposite effects.

Following Pedersen (2002), the HP filter can be viewed as an approximation to the ideal filter. Suppose that the objective is to filter out a cycle of 8 (or fewer) years. This implies an ideal rectangular gain function in the frequency domain with the critical cut off frequency of 2π/8. By adjusting the value of the penalty parameter λ, we can approximate an ideal filter to some extent. But this approximation will introduce two types of distortions: (1) compression when a part of high frequency goes to the trend, and (2) leakage when a part of low-frequency data goes to the cycle. This frequency domain fact explains the trade-off between two objectives – by changing the value of penalty either the trend becomes too volatile or the cycle will contain too much of the trend. The compromise solution hence requires an appropriate modification of the HP filter to lower the impact of the trade-off.

The approximation result will depend on the length of the cycle assumption and the overall distortionary effect of the filter that depends on the properties of approximate power transfer function as compared with the ideal filter (Pedersen 2002). Therefore an optimal choice of penalty will minimize the deviations from the ideal filter in the sense of a specific loss function. Pedersen suggested to use a symmetric view on both types of distortions and considered a squared loss function, while in some applications an asymmetric weight is put on leakage and compression types of distortions, e.g. Bouthevillain, et al. (2001) placed a higher weight on the compression seeking to minimize the deviations from the latter by more.

Conditional on the chosen smoothness parameter λ the solution to minimisation problem (2) yields a non-parametric moving average representation of the cyclical component, i.e. the filtering problem may be viewed as a high-pass problem (King and Rebelo 1993):

$$a(L) = \frac{\lambda(1-L)^2(1-L^{-1})^2}{\lambda(1-L)^2(1-L^{-1})^2+1} = \frac{\lambda L^{-2}(1-L)^4}{\lambda L^{-2}(1-L)^4+1} \Rightarrow c_t = a(L)y_t, y_t^* = (1-a(L))y_t, \quad (4)$$

where L is a backshift (or lag) operator: $Ly_t = y_{t-1}$, while its inverse L^{-1} shifts time series forward (a lead operator). King and Rebelo (1993) showed that given the four first difference terms in the numerator of (4) the HP filter can render stationarity in c_t for any integrated process (possibly with deterministic polynomial trends) up to the fourth degree. Besides, equation (4) implies that the cycle is proportional to the forth difference of the trend, which is shifted forwards by exactly two periods: $c_t = \lambda L^{-2}(1-L)^4 y_t^*,\ 2 < t < T-2$, but since the filter in the middle of the sample is symmetric and depends on the past and future values of the actual data the phase shift in practical applications will be close to zero (St-Amant and Van Norden 1997).

For a more general stochastic interpretation of the HP filter, it is convenient to represent the minimisation problem (2) in the state space form (Murray 2014):

Signal: $$y_t - y_t^* - c_t = 0, \quad (5)$$

State: $$y_{t+1}^* = 2y_t^* - y_{t-1}^* + \varepsilon_{2,t}, \varepsilon_{2,t} \sim NID(0, \sigma_1^2/\lambda), \quad (6)$$

State: $$c_t = \varepsilon_{1,t}, \varepsilon_{1,t} \sim NID(0, \sigma_1^2). \quad (7)$$

The state-space representation of the HP filter can then be used for real-time and ex-post validation of the method applying the corresponding one-sided (Kalman filter) or two-sided (Kalman smoother) filters (Kalman 1960) as described in **Appendix B**. The same trick is operational for any other type of filters. The state-space representation is used to highlight a number of specific implicit assumptions and judgements other than the earlier discussed value of penalty λ (Mohr 2005, Murray 2014):

- First, equation (6) means that the prior belief for a data generating process of the trend is the second order random walk (a random walk with *a stochastic drift*), which is implausible for many macroeconomic series that are oft-assumed to be an integrated process of the first order with a deterministic drift. Besides, the best next period prediction of potential output growth is current potential output growth, which could be inconsistent, for instance, with the catching-up and convergence assumptions that are often relevant for the CEE countries, among others.
- Second, the signal equation (5) implies that potential output and output gap estimates sum up to original data, i.e. there is no irregular shock ε_t as in (1) to capture non-cyclical behaviour. According to the time-domain representation of HP filter, the cycle is not explicitly modelled, but is defined as a residual process. If a higher frequency data (e.g., a quarterly real GDP) is not seasonally adjusted, the cyclical part will also contain the seasonal component. It is thus recommended to seasonally adjust the data before the application of HP filter, yet the quality of adjustment matters too.
- Third, if $\varepsilon_{1,t}$ is defined as a zero-mean white noise process the best prediction of the output gap is its mean, i.e. zero. In practice, $\varepsilon_{1,t}$ could be explicitly defined as an autoregressive process as, for instance, in trend-cycle decomposition by Mohr (2005) or suggested in Hamilton (2017). On the other hand, the cycle's persistence is introduced through the distortional autocorrelation effects, the properties of which depend on the value of penalty λ.
- Finally, the assumption that shocks to demand $\varepsilon_{1,t}$ and to supply $\varepsilon_{2,t}$ are assumed to be uncorrelated is implausible as it rules out the possibility of Keynesian hysteresis – a large negative output gap that can have very persistent effects on potential output via underemployment equilibrium in the labour market or drop in the level of productive investments.

Bouthevillain, et al. (2001) points to another problem, which is not directly linked to the state-space representation, that the HP filter is unable to detect and reflect a sudden structural break in trend. This problem is, however, less severe the lower the chosen value for penalty λ.

Finally, the HP filter may be viewed as a member of the Butterworth family of filters (Gómez 2001) the low-pass gain function of the two-sided version of which when based on sine function is given by:

$$G^B(n,\omega) = \frac{1}{1+\lambda'\{\sin(\omega/2)\}^{2n}}, \tag{8}$$

where $\omega \in [0,\pi]$ is a frequency, $\omega_c \in (0,\pi)$ is a cut-off frequency when $G^B(n,\omega_c) = 0.5$, λ' is a signal-to-noise ratio parameter, n is a smoothness parameter that has to be not less than the order of the stochastic trend. When n equals 2 the low-pass gain (8) is the low-pass gain for the HP filter:

$$1 - G^{HP}(\lambda,\omega) = \frac{1}{4\lambda(1-\cos(\omega))^2+1} = \frac{1}{\lambda(2\sin(\omega/2))^4+1} = G^B(2,\omega). \tag{9}$$

In general, there could exist a better approximation of an ideal low-pass filter by a more general member of a Butterworth family of filters with $n > 2$, since higher values of the n result in a steeper approximation of the ideal filter around the cut-off frequency at the expense of worse approximation at the end-of-sample. For the most of macroeconomic time series, stochastic trends of order higher than two do not make much sense, therefore such filters will have less economic meaning (Mohr 2005). The Butterworth filter interpretation is useful as it links the penalty with the cut-off frequency by $\lambda = (2\sin(\omega_c/2))^{-4}$. It turns out that for annual data Ravn and Uhlig (2002) 6.25 recommendation is consistent with 10 years cut-off, while 100 – with 19.8 years (Casey 2018).

The most often discussed weakness of the HP filter (as any other two-sided filter) is the so-called "end-of-sample problem". A moving average representation (4) at the end of the sample first becomes asymmetric and then one-sided. Thus, the value of potential output y_t^* towards the end of the period is driven to a large extent by the value of actual output. It follows that potential output growth is biased down for negative output gaps, and biased up for positive output gaps by much more than when applying a two-sided HP filter (Murray 2014). This can be particularly problematic when assessing the output gap conditions in real time, especially in the case when the data is significantly revised. The problem is partly mitigated by extending the actual data with projections, yet a certain level of bias remains (Bouthevillain, et al. 2001, Ódor and Jurašekova Kucserová 2014). Another interesting solution, which was introduced in Butler (1996) at the Bank of Canada, is to put more restrictions on the original problem. However, this solution comes at the cost of higher phase shifts and higher prior trend dominance on the cyclical component **(Box 3)**.

Box 3. Will more restrictions end the end-of-sample problem?

A major problem with the HP filter is that trend or potential output estimates may be driven to a large extent by the value of actual output at the end of a sample. So how do we prevent the tail wagging the dog?

One option could be to place more restrictions on the original problem (2). Ignoring the structural information for expositional simplicity the modified filter as in Butler (1996) solves the following problem:

$$\min_{\{y_t^*\}_{t=1}^T} \underbrace{\sum_{t=1}^{T}(y_t - y_t^*)^2}_{\text{goodness-of-fit}} + \underbrace{\lambda\sum_{t=2}^{T-1}[\Delta^2 y_{t+1}^*]^2}_{\text{degrees-of-smoothness}} + \underbrace{\lambda_{pr}\sum_{t=2}^{T-1}[\Delta y_t^*]^2}_{\text{prior predictions}} + \underbrace{\lambda_{ss}\sum_{t=T-j}^{T}(\Delta y_t^* - \mu_{ss})^2}_{\text{end-of-sample anchor}}, \quad (10)$$

where $\Delta^d = (1-L)^d$ is the d-th difference operator; μ_{ss} is a steady-state growth rate (or any other deterministic drift) of the potential output in the last $j+1$ periods of the sample – an anchor that implies the reversion of the potential output growth towards a constant (or a drift) exactly at the end-of-sample, while a stochastic drift in (2) implies none. For quarterly data Butler (1996) suggested to restrict the final 15 quarters ($j = 14$) of the sample and to set the penalty parameter λ_{ss} to 64. Another novelty in (10) is a recursive weight with which new data updates the potential output growth, where higher values of penalty λ_{pr} would mitigate the impacts of new observations by more. Butler (1996) suggests setting this parameter to 1. Both modifications result in a moving average representation of the modified one-sided filter associated with (10) to behave close to the weights defined by two-sided HP-filter in the middle of the sample. This makes estimates of the output gap behave in a more regular way at the end of sample.

Such restrictions are not costless. St-Amant and van Norden (1997) showed that this regularisation comes at the cost of much higher leakage effect and the extracted cycle becomes dominated by low-frequency movements not normally associated with the output gap by much more than a simple HP filter, but wins in terms of much lower compression. Therefore the net distortionary effect of the filter is rather neutral.

Another, perhaps more important, impact of (10) is the greater phase shift of the modified one-sided filter as compared with its HP counterpart. An empirical investigation in St-Amant and van Norden (1997) shows that the output gap extracted with the modified filter for Canada's data may come with a lag of 2–4 quarters, while the lag is roughly zero for one-sided HP filter. Hence, more restrictions and regularity do not necessarily imply a better net effect on the accuracy of the estimated output gap.

All of the issues discussed above may lead to spurious (artificial) distortions in the cyclical estimates. These may therefore have no basis in the underlying data-generation process (Hamilton 2017). Of course, other filters (for instance, as is shown in Murray (2014) for the UK) do not necessarily lead to a better balance of statistical properties. Pollock (1999) and Pedersen (2002) noticed that the spurious cycle claim is, in general, inaccurate, since spurious cycles, in the sense of distorted auto-correlation functions, may be induced by the ideal filters too. Therefore, one should pay attention to the general distortion as compared with the power transfer function of the ideal filter and seek to minimize the integral sum of compression and leakage effects. At the same time, Mohr (2005) used HP filter as a starting example to generalize the stochastic representation of the model to a trend-cycle (TC) filter and to resolve the most of the abovementioned problems also including the class of Butterworth filters as a particular case.

Prior-Consistent filter

Before putting more structural content into the estimation methods, it is worth questioning the prior belief for the data generating process of the trend directly. "A close cousin" to the HP filter is a simple Prior-Consistent (PC) filter[12] introduced in Laxton, et al. (1998). The PC filter shares similarities with the HP filter assumptions. The first assumption is exactly the same as for the HP filter, but the second differs (Murray 2014):

- actual output does not deviate too far from its trend level (the output gap is not too big);
- potential output growth is consistent with the recent past behaviour or drift (the growth of the potential output reverts to a drift).

Then a PC minimisation problem is formulated as:

$$\min_{\{y_t^*\}_{t=1}^{T}} \underbrace{\sum_{t=1}^{T}(y_t - y_t^*)^2}_{\text{goodness-of-fit}} + \underbrace{k\sum_{t=2}^{T-1}[(y_{t+1}^* - y_t^*) - drift_t]^2}_{\text{degrees-of-smoothness}},\ drift_t = \bar{y}_t - \bar{y}_{t-1},\ k = \frac{\sigma_1^2}{\sigma_2^2} > 0, \qquad (11)$$

where y_t^* is the unobserved potential output; y_t is the observed actual output; $\bar{y}_t$ is the prior belief regarding the potential output, the change of which is *a deterministic drift*; σ_1^2 is the variance of the output gap; σ_2^2 is the variance of the trend growth deviations from its pre-defined historical rate of drift and k is a positive *penalty* parameter with a similar meaning to a corresponding parameter of a HP filter. If the dynamics of the drift term are defined as stochastic (the first difference in the potential

[12] In Murray (2014) the method is called Prior-Constrained.

output), then the PC problem (11) becomes equivalent to (2). The PC filter has a clear advantage over the HP filter in that it allows practitioners to include any prior beliefs regarding the future development of potential output . It also adequately represents the empirically observed data generating process – the first order integrated process with a deterministic drift – for the most of the macroeconomic variables. These beliefs could reflect, for instance, the catch-up nature of economic growth or any anticipated structural breaks. In another extreme, if none of the information exists, the drift could be represented by a constant steady state rate of growth in potential output that serves as a long-term anchor for the potential growth dynamics.

The arbitrary choice of the PC penalty k is similar to the HP problem. Murray (2014) argues that if a 5 per cent output gap is as moderate as the fifth percentage change in the level of potential output, then $k = \sigma_1^2/\sigma_2^2 = 5^2/(1/5)^2 = 625$ is a good choice for quarterly data. In annual terms, this choice is equivalent to the output gap shocks to be about 6.25 times as large as those to the level of potential output[13] or the parameter $k \approx 39$. Laxton, et al. (1998) suggested that the PC penalty defined as the ratio between squares of the prior for the large gap and the prior for the large change in the potential output should be around 5, hence the annual PC penalty equals 25, while the corresponding quarterly value 400. Since prior beliefs are formed on annual basis it is reasonable also to consider Ravn and Uhlig (2002) type frequency conversion factors suggested for HP filters and leading to a higher value of 6400 for quarterly data. From the long-term standpoint of fiscal authorities and IFIs, higher values are preferable, yet the prior beliefs on what is a large output gap and what is a large deviation from the prior rate of potential growth remain arbitrary and country-specific.

For a more general stochastic interpretation of the PC filter and ex-ante and ex-post validation through Kalman algorithm (Kalman 1960), it is convenient to represent the minimisation problem (11) in the state-space form (Murray 2014):

Signal: $$y_t - y_t^* - c_t = 0, \tag{12}$$

State: $$y_{t+1}^* = y_t^* + drift_t + \varepsilon_{2,t},\ \varepsilon_{2,t} \sim NID(0, \sigma_1^2/k), \tag{13}$$

State: $$drift_t = drift_{t-1} \tag{14}$$

State: $$c_t = \varepsilon_{1,t}, \varepsilon_{1,t} \sim NID(0, \sigma_1^2). \tag{15}$$

Murray (2014) shows that the implicit assumptions would be similar in nature to those of HP filter, with the key difference being that the HP filter is consistent with a stochastic drift, when shocks are transmitted to both the potential output and its growth rate; while the PC filter is consistent with constant drift, when shocks will affect only the level of potential output (shifts) but not the growth rate. As we said before, any prior belief on the deterministic drift dynamics is possible. With a similar moving average behaviour in the middle of the sample, the differences become more visible by the end of the sample, because, for one-sided filters, it is more costly to close the output gap for a PC filter than by adjusting also the estimates of potential output growth. Therefore, the output gap bias at the end-of-sample becomes smaller and the overheating or recession problem becomes more visible.

At the same time, the PC filter has a higher leakage effect similar to that discussed in **Box 3**. For instance, for small and open economies the cycle becomes more dominated by a fast past converging growth pace of the trend component and is slower to respond to any structural changes. The latter,

[13] Murray (2014) claims that this ratio is about 5 times as large, but then it would be a parameter 400 as in Laxton, et al. (1998) case.

however, could be an explicit part of the deterministic drift if the structural change could be identified and anticipated in real-time.

Band-pass filters: ideal and approximations

Both classical and growth cycles seek to isolate the patterns in duration varying from more than one year to ten or twelve years, but not divisible into shorter cycles (Burns and Mitchel 1946, Mintz 1969). Hence, an ideal band-pass (IBP) filter aims to directly isolate the components of time series that belong to a given range of frequencies $[\omega_1, \omega_2]$, where the period of the cycle (in quarters) is given by
$p_f = 2\pi/\omega_f,\ f \in \{1,2\}$. The IBP filter provides an ideal split into a trend, a cycle and an irregular components as in (1). The common prior beliefs are similar to that of the HP filter considering anything with a frequency over 8 years as a trend, between 2 and 8 years is a cycle and below 2 years is a noise. Band-pass filters can therefore be applied to seasonally adjusted or unadjusted data, given that outliers[14] are removed (Mazzi, Ozylidririm and Mitchell 2017). Murray (2014) points out that the convention on the choice of cut-offs is not supported by a strong empirical evidence, so this choice is a judgement.

Koopmans (1974) showed that it is possible to construct an ideal low-pass filter in the time domain as an infinite dimensional symmetric linear time-invariant filter:

$$y_t^f = \sum_{i=-\infty}^{\infty} h_i^f\, y_{t-i}, \quad \sum_{i=-\infty}^{\infty} \left|h_i^f\right| < \infty,\ h_0^f = \frac{\omega_f}{\pi},\ h_i^f = \sin(i\omega_f)/i\pi,\ f \in \{1,2\}, \tag{16}$$

the difference of two low-pass filters at different cut-off frequencies then defines the ideal band-pass filter with the coefficients $h_i^{bp} = h_i^1 - h_i^2$.

Since it is impossible to estimate (16) within a finite sample, in practice, an approximation is required. Band-pass filters will then differ in the way they specify the weights for the finite sample approximation.

Baxter and King (1999) proposed a symmetric Baxter-King (BK) approximation to the ideal filter by solving the following minimization problem in the frequency domain:

$$\hat{y}_t^{BK} = \sum_{i=-n}^{n} \hat{h}_i^{BK}\, y_{t-i}, \quad \min_{\{\hat{h}_j^{BK}\}_{j=-n}^{n}} \frac{1}{2\pi}\int_{-\pi}^{\pi} \left(h(\omega) - \hat{h}^{BK}(\omega)\right)^2 d\omega + \lambda\left[\sum_{i=-n}^{n} \hat{h}_i^{BK} - \phi\right], \tag{17}$$

where $h(\omega) = \sum_{i=-\infty}^{\infty} h_i^{bp}\, e^{-i\omega j}$ and $\hat{h}^{BK}(\omega) = \sum_{i=-n}^{n} \hat{h}_i^{BK}\, e^{-i\omega j}$ are the Fourier transforms of the corresponding linear filters (16) and (17), the zero frequency restriction ϕ means that at a zero frequency the filter can deviate from zero. From the first order conditions and (16) than follows that: $\hat{h}_i^{BK} = h_0^{bp} - (n+1)^{-1}\sum_{i=-n}^{n} h_i^{bp}$. A symmetry insures that the approximation does not imply phase shifts, but could distort amplitudes in the same sense as any other approximation of an ideal filter. From the corresponding gain function of the BK filter given by:

$$G^{BK}(n,\omega) = \hat{h}_0^{BK} + 2\sum_{i=1}^{n} \hat{h}_i^{BK} \cos(i\omega), \tag{18}$$

[14] The outliers could be removed applying automatic detection procedures similar to TRAMO-SEATS (Gómez and Maravall 1996)

follows that the distortions are smaller with higher values of n. This, however, requires exclusion of n observations both at the beginning and at the end of the sample (an oft-used cut-off is about 3 years or 12 quarters) and hence BF filter is impractical in real-time applications. The problem could be partly mitigated by extending the time series with the forecasts of at least cut-off length. In that case, the end-of-sample problem becomes a forecasting problem.

Christiano and Fitzgerald (1999) explain that an optimal approximation to the ideal band-pass filter requires the knowledge of the true data generating process of y_t and should account for it. In a more general setting, the authors suggested an asymmetric time-varying Christiano-Fitzgerald (CF) filter (also known as a random walk filter) which solves the following optimisation problem:

$$\hat{y}_t^{CF} = \sum_{i=-n_{t,1}}^{n_{t,2}} \hat{h}_{t,i}^{CF}\, y_{t-i}, \quad \min_{\{\hat{h}_j^{CF}\}_{j=-n_{t,1}}^{n_{t,2}}} \frac{1}{2\pi}\int_{-\pi}^{\pi} \left(h(\omega) - \hat{h}^{CF}(\omega)\right)^2 f_y(\omega)d\omega, \tag{19}$$

where $f_y(\omega)$ is the spectral density of the actual data y_t and the approximation window depends on the choice of time-varying parameters $n_{t,1}$ and $n_{t,2}$. Both parameters could be fixed, inducing stationarity and equal resulting symmetry. In the latter case the corresponding gain function is equivalent to (18), where the estimated parameters are changed to $\hat{h}_i^{CF} = -0.5h_0^{bp} - \sum_{i=1}^{n-1} h_i^{bp}$, and the optimal cut-off periods will depend on the assumptions regarding the spectral density $f_y(\omega)$. Christiano and Fitzgerald (1999) conventionally assumed that the data generating process is an integrated of the first order with probably deterministic drift. If the assumption is violated, then the extracted cycle will be distorted. At the end of the sample, the CF filter will also induce phase shifts in the estimated cyclical components. Another drawback is that the time-varying in parameters problem (19) is not conveniently defined in the state-space form, therefore ex-ante and ex-post validation via the Kalman algorithm is hard to achieve.

In general, band-pass filters are not currently widely used within the Network of EU IFIs, some other interesting extensions like Butterworth filter or wavelet-based methods could be considered (Álvarez and Gómez-Loscos 2017). The Butterworth filters are members of a general family of TC filters pioneered within the Eurosystem by Mohr (2005) and will be discussed at the end of this section.

Beveridge-Nelson decomposition

The Beveridge-Nelson (BN) decomposition, as suggested by Beveridge and Nelson (1981), is one of the most prominent *parametric* procedures for finding the cyclical component of the nonstationary time series. The BN decomposition rests on a number of assumptions (Murray 2014):

- The actual output growth is a stationary process, i.e. the data generating process of actual output indeed follows an autoregressive integrated moving average ARIMA(*p*, 1, *q*) process:

$$a(L)\Delta y_t = \mu + \Delta f_t + b(L)\varepsilon_t,\ \varepsilon_t \sim NID(0, \sigma^2), \tag{20}$$

  where μ is a constant long term growth rate of output (convergence or steady-state anchor), f_t is any deterministic continuous convergence function $f_t \xrightarrow[t\to\infty]{} 0$ defined e.g. as in Celov (2015), $a(L) = \sum_{j=0}^{p} a_j L^j$ and $b(L) = \sum_{j=0}^{q} b_j L^j$ are two lag polynomials of orders *p* and *q* respectively. In practical applications, the *q* parameter is often restricted to 0, while *p* is up to 3 lags for annual data or 12 lags for quarterly (Murray 2014, Kamber, Morley and Wong 2017).

- The trend is equal to the limiting forecast of the series adjusted for its mean growth rate (Mazzi, Ozylidririm and Mitchell 2017):

$$y_t^* = \lim_{h\to\infty}[E(y_{t+h}|\Omega_t) - h\tau], \tag{21}$$

  where h is a forecasting horizon, Ω_t is all information available when the forecast is made, $\tau = E(\Delta y_t)$ is the mean growth of the first difference of the logarithmically transformed actual output y_t (the long-run drift). The cyclical component is defined as the residual $c_t = y_t - y_t^*$.

- Both trend and cycle are affected by a common shock, thus in the corresponding convenient for estimation and validation state-space form (Morley, Nelson and Zivot 2003) the shocks to cycle and trend have to be correlated. Indeed integration of (20) with respect to BN decomposition $b(L)/a(L) =: c(L) = c^*(L)(1-L) + c(1)$ yields:

$$y_t = \underbrace{y_0 + \mu \cdot t + f_t}_{\text{deterministic trend}} + \underbrace{c(1)\sum_{j=1}^{t}\varepsilon_j}_{\text{stochastic trend}} + \underbrace{c^*(L)\varepsilon_t}_{\text{cycle}},\ \varepsilon_t \sim NID(0,\sigma^2), \tag{22}$$

  where stochastic trend and cycle are driven by the same innovations ε_t and the potential output is the sum of a (non-linear) deterministic trend (that could contain also structural breaks) and a stochastic trend (a random walk). The most general definition of BN trend (21) and decomposition (22) implies that in the long-run all possible trend components will converge to the BN trend as the forecasting horizon *h* tends to infinity. This is exactly due to the fact that the value of the long-run projection of any cycle is by definition equal to zero. Hence the BN trend is the limiting trend (Mazzi, Ozylidririm and Mitchell 2017).

- The ARIMA specification is correct. The decomposition (22) is very sensitive to the actual specification of the ARIMA model that may result in very unstable output gap estimates.

The third assumption highlights a key difference from the typical assumptions in the state-space representations of the unobserved components models or HP/PC filters. That is, shocks to potential output are assumed to be negatively correlated with cyclical shocks. In other words, a positive shock pushes potential output up, whilst it pushes aggregated demand down. This restrictive view is, however, consistent with a productivity shock interpretation that may occur due to automatization, robotization and similar 4th industrial revolution outcomes. On the other hand, the direct consequence of the common shock dependence is that the extracted trend tends to be highly irregular due to a stochastic trend, i.e., potential output could be even more volatile than an actual output (Murray 2014), if not efficiently restricted solving the model in the state-space form as recently suggested in BN filter by Kamber, Morley and Wong (2017). The authors provided an intuitive explanation for the case of unrestricted estimates of an AR(*p*) model when the signal-to-noise ratio in terms of trend shocks as a fraction of the overall forecast error variance:

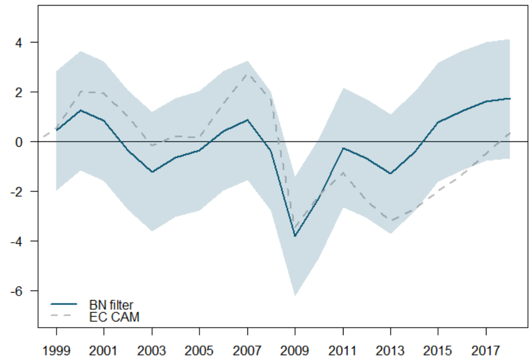
$$\delta = \sigma_{y^*}^2/\sigma^2 = c(1)^2, \tag{23}$$

where $c(1)$ is the long-run multiplier that captures the permanent effect of a forecast error on the long-run conditional expectation of y_t in (21). Kamber, Morley and Wong (2017) showed that for the US quarterly real GDP growth the estimated signal-to-noise ratio is $\hat{\delta} = 2.22$ and implies that the trend volatility will be more than twice volatile than the forecast errors in the log of real GDP, leading to the output gap as the residual to be with smaller amplitude and counterintuitive sign. Notably, for any unrestricted stationary AR(*p*) model $\hat{\delta} = 1/a(1) > 1$. Therefore, all we need is to restrict the penalty parameter δ rather than allow the model to freely estimate it. For instance, under dogmatic prior similar to Hodrick and Prescott (1997) $\hat{\delta} = 0.05$ will be consistent with the strict belief that only 5 per cent of the quarterly forecast error variance for output growth is due to trend shocks. Kamber, Morley and Wong (2017) proposed an algorithm for the automatic selection of $\hat{\delta}$ by maximising the implied amplitude-to-noise ratio, i.e. making the output gap more visible.

The suggested procedure is promising as it results in much smaller revisions of real-time estimates as compared with ex-post estimates, i.e. the amount of end-of-sample problem is substantially reduced, not to mention the possibility to account for likely structural breaks. An application of the BN filter to the logarithmically transformed annual real GDP data for the euro area in **Figure 6** showed that the automatically chosen penalty $\hat{\delta} \approx 0.21$ when an AR(3) model is fitted. The quality of the parametric BN filtering method is expected to increase with longer time series and/or higher frequency data.

Morley, Nelson and Zivot (2003) showed that the state-space representation of BN filter, which is useful for ex-ante and ex-post validation of the method by Kalman filter (Kalman 1960), is equivalent to the univariate unobserved components model discussed below.

Figure 6 – BN filter's application for the euro area's output gap with a penalty parameter assessed within 10 years window, $p = 3$



Source: European Commission's 2018 spring forecast, the Network of EU IFIs' calculations

Univariate unobserved components model and trend-cycle filter

A common *parametric* way to conduct trend-cycle decompositions is the *structural time series* approach. The approach frames the problem in terms of unobserved components (UC), which have a direct economic interpretation. According to Mazzi, Ozylidririm and Mitchell (2017), such models could be viewed as a flexible data-driven approach to detrending. Flexibility is achieved through accommodation of both deterministic and/or stochastic non-stationarity.

A generic UC representation takes the following state-space form (Harvey 1985):

Signal: $$y_t - y_t^* - c_t = \varepsilon_t,\ \varepsilon_t \sim NID(0, \sigma^2), \quad (24)$$

State: $$y_t^* = y_{t-1}^* + \mu + \varepsilon_{2,t},\ \varepsilon_{2,t} \sim NID(0, \sigma_2^2), \quad (25)$$

State: $$c_t \text{ is stationary and ergodic, } \ \varepsilon_{1,t} \sim NID(0, \sigma_1^2), \quad (26)$$

where y_t^* is the unobserved potential output, y_t is the observed actual output, σ_1^2 is the variance of the output gap, σ_2^2 is the variance of the trend growth deviations from constant drift, and ε_t is an irregular component[15] with variance σ^2. All error terms are assumed to be mutually uncorrelated. One may

[15] In the original formulation of the decomposition by Harvey (1985) this irregular term was omitted. In general, ε_t can follow any stationary ARMA process fulfilling requirement to be uncorrelated with other model residuals.

view the state-space form as a further generalisation of a PC filter problem. The generalisation explicitly introduces the irregular component to absorb outliers, especially at the end-of-sample, and introduces the explicit parametric model for the cycle. Since the macroeconomic variables may follow a process different from a random walk with constant drift, a generalized stochastic trend model in the implicit form as in Mohr (2005) is given by:

State: $$\Delta^{d-1}(\Delta y_t^* - \mu) = \varepsilon_{2,t},\ \varepsilon_{2,t} \sim NID(0, \sigma_2^2), \quad (27)$$

where $\Delta^d = (1-L)^d$ is the d-th difference operator; and d denotes the order of the stochastic trend. Mohr (2005) points out that this generalization is well known in the literature. The case $d = 1$ with $\mu = 0$ corresponds to exponential smoothing, $d = 1$ with non-zero drift as in (25) will correspond to extended exponential smoothing. If $d = 2$ with $\mu = 0$ we will get the HP filter problem, while the stochastic trend of order $d > 2$ will correspond to other members of the Butterworth class of filters. The author recommends paying more attention to the cases $d \in \{1, 2\}$. This is because the stochastic trends of a higher order than 2 make little sense when applied to typical macroeconomic data, as real GDP or inflation. Another possible extension is a local linear trend model, which introduces two states (Celov 2015, Jarociński and Lenza 2016, Constantinescu and Nguyen 2017, Cuerpo, Cuevas and Quilis 2018):

State: $$y_t^* = y_{t-1}^* + \mu_{t-1} + \varepsilon_{2,t},\ \varepsilon_{2,t} \sim NID(0, \sigma_2^2), \quad (28)$$

State: $$\mu_t = \rho_\mu \mu_{t-1} + \varepsilon_{\mu,t},\ \varepsilon_{\mu,t} \sim NID(0, \sigma_\mu^2),\ \rho_\mu \in [0, 1], \quad (29)$$

where the local linear trend (29) allows capturing U or inverted-U shaped patterns in the trend components.

A key extension of the UC framework is an explicit model for the cycle. In HP (7) and PC (15) cases, the cyclical model was assumed to follow a normally distributed white noise process, where the visibility of the cycle was implicitly insured by corresponding penalty parameters λ and k chosen in an arbitrary way. Following the stochastic cycle model of order *c* (Harvey 1985) one can derive an explicit stochastic ARMA(2*c*, *c*) representation of the cycle (Mohr 2005):

State: $$(1 - 2\rho\cos(\omega_c)L + \rho^2 L^2)^c c_t = (1 - \rho\cos(\omega_c)L)^c \varepsilon_{1,t},\ \varepsilon_{1,t} \sim NID(0, \sigma_1^2), \quad (30)$$

where $\rho \in (0, 1)$ determines a damping factor and is typically assigned a value close to unity (e.g. 0.975), or could be estimated[16]; ω_c is a frequency in radians $\omega_c = 2\pi/T_c$ corresponding to a cycle of length T_c, which also could be either assumed (e.g. 8 years or 32 quarters) or estimated. In the most of applications, the first order stochastic cycle ($c = 1$) is assumed. The HP and PC cases hence correspond to the stochastic cycle of the zero order ($c = 0$). It is straightforward to extend (30) by allowing additional cyclical determinants such as CUBS from European Commission's CAM (Havik, et al. 2014), capacity utilisation or principal components that represent the cyclical position of the country. In most] practical applications, however, a simpler approach is used allowing the cycle c_t to follow an unrestricted AR(*p*) process, where *p* is often chosen to be equal 1 or 2 (Frale and De Nardis 2017, Cuerpo, Cuevas and Quilis 2018):

State: $$(1 - \phi_1 L - \phi_2 L^2)c_t = \varepsilon_{1,t},\ \varepsilon_{1,t} \sim NID(0, \sigma_1^2). \quad (31)$$

---

[16] To estimate a restricted parameter a useful trick is to define $\hat{\rho}(\alpha) = (tol + (1 - 2 \cdot tol)/(1 + e^\alpha))$, where tol is a tolerance margin close to zero (e.g. 0.01) and α is the parameter to be estimated by either maximum likelihood. In Bayesian models the restricted estimation is achieved by choosing Beta distribution as a prior belief.

To understand the shock interdependence of the state-space form it is convenient to represent it in the reduced-form MA. For instance, assuming $\sigma^2 = 0$ and $\rho_\mu = 1$ Cuerpo, Cuevas and Quilis (2018) get:

$$y_t = y_t^* + c_t = \frac{1}{1-\phi_1 L - \phi_2 L^2}\varepsilon_{1,t} + \frac{1}{(1-L)^2}\varepsilon_{2,t} + \frac{1}{1-L}\varepsilon_{\mu,t}. \tag{32}$$

The state-space representation, hence, is flexible enough to accommodate alternative specifications for the trend and cycle components of GDP. In this way, the model provides a flexible and parsimonious way to represent different non-stationary dynamics[17].

The univariate TC filter applied by Lithuania's and Ireland's IFIs is defined as the state-space model represented by (24), (27) and (30). Seeking to highlight the differences from the HP and PC problems the corresponding minimization problem is (Mohr 2005):

$$\begin{aligned}\min_{\{y_t^*\}_{t=1}^T,\{c_t\}_{t=1}^T,\mu} \underbrace{\sum_{t=1}^{T}(y_t - c_t - y_t^*)^2}_{\text{goodness-of-fit}} + \lambda_2 \underbrace{\sum_{t=d}^{T-d+1}[\Delta^{d-1}(\Delta y_t^* - \mu)]^2}_{\text{stochastic trend}} + \\ + \lambda_1 \underbrace{\sum_{t=2c}^{T}[(1-2\rho\cos(\omega_c)L + \rho^2L^2)^c(1-\rho\cos(\omega_c)L)^{-c}c_t]^2}_{\text{stochastic cycle}},\end{aligned} \tag{33}$$

where trend and cycle are extracted *simultaneously* and $c \geq 1$ is the order of the stochastic cycle. Similar to HP and PC problems, under additional assumptions that all three error terms are normally distributed, the penalty weights could be set to equal to the respective inverse signal-to-noise variance ratios: $\lambda_1 = \sigma^2/\sigma_1^2$ and $\lambda_2 = \sigma^2/\sigma_2^2$. For the simplicity of calculations Mohr (2005) restricts both ratios to 1, noting that the impact of HP prior assumptions are fully reflected by the estimated or calibrated values of ρ and ω_c. This allowed the author to obtain a convenient solution in the matrix form, which is relatively easy to program and integrate with spreadsheets. However, the second restriction in UC models often deviates from unity, therefore the suggested restriction in general does not imply an optimal filter.

The more computationally-intensive way to approach the state-space model (24), (27) and (30) is to estimate unknown parameters either by maximum likelihood (Cuerpo, Cuevas and Quilis 2018), using the prediction error decomposition, or by the Bayesian methods (Melolinna and Tóth 2016, Constantinescu and Nguyen 2017). Given the estimated parameters and applying the corresponding one-sided (Kalman filter) or two-sided (Kalman smoother) filters (Kalman 1960) we then obtain ex-ante and ex-post estimates of the corresponding cycle c_t and the trend y_t^* **(Appendix B)**. Notice, however, that at the end-of-sample both estimates are identical and the proper validation is obtained only in the middle of the sample.

Finally, since the UC model implies a restricted ARMA, the UC trend could be associated with the corresponding BN trend (Morley, Nelson and Zivot 2003). Indeed, the UC conditional on information Ω_t will equal the BN trend when the UC model implies the same reduced-form ARMA model as the BN trend model is based (Mazzi, Ozylidririm and Mitchell 2017). Differences arise because in the UC model all error terms are explicitly assumed to be uncorrelated, in particular, $E(\varepsilon_{1,t}\varepsilon_{2,t}) = 0$. When the latter zero restriction is relaxed the solution of the UC and BN decompositions will be identical.

[17] For instance, In the Spanish case, GDP can be modelled following an I(1) structure plus a highly persistent Markov-switching drift, as shown in Cuevas and Quilis (2017). This specific structure can be linearly approximated by a random walk plus an evolving AR(1) drift.

Since the correlation between disturbances in BN decomposition are explicit there is a corresponding representation of BN filter in a state space form in terms of UC model.

More complex detrending methods could be achieved by relaxing other assumptions of the UC state-space representation. One prominent generalisation is a multivariate UC model that includes other signal equations that bring a semi-structural interpretation to the extracted trends and cycles. Such techniques essentially redistribute some additional stationary element between the BN trend and cyclical components (Mazzi, Ozylidririm and Mitchell 2017).

Seeking refinement

A critical issue with the univariate trend-cycle decompositions of (1), shown above, is that the methods rely heavily on prior judgements over the amplitude of the cycle and the dynamics of the potential output. It is crucial therefore to refine these priors using more information. This view is widely accepted by the Network of EU IFIs (Murray 2014, Ódor and Jurašekova Kucserová 2014, Frale and De Nardis 2017, Casey 2018, Cuerpo, Cuevas and Quilis 2018), central banks (Borio, Disyatat and Juselius 2017, Constantinescu and Nguyen 2017) and international institutions and has led to a growing body of empirical studies that seek to extract relevant information from more variables than just the observed actual output of the economy.

Multivariate detrending methods are essentially more *structural* and combine their univariate counterpart by including additional macroeconomic variables whose stationary component is related to the output gap (Cuerpo, Cuevas and Quilis 2018). These macroeconomic variables may contain crucial information about the aggregated demand and supply side of the economy, internal and external imbalances that stems from many sources. Ódor and Kucserová (2014) reviewed the most common sources of cyclical variation "beyond-inflation" including financial cycles (Benetrix and Lane 2011, Borio, Disyatat and Juselius 2017), absorption cycles (Lendvai, Moulin and Turrini 2011) and commodity price cycles (Bornhorst, et al. 2011).

The quality of the output gap extraction is higher with a better predictability of the unobserved potential output deviating from simple assumptions that the trend follows a random walk (with drift). This comes at the cost of increased complexity, both in terms of the estimation methods involved and in terms of the uncertainty regarding which variables to include. More structural models require using the judgement implying that the net outcome could be more uncertain than in univariate models discussed above. To bring more transparency to the judgement process, Cuerpo, Cuevas and Quilis (2018) summarized various aspects of the economy and proposed a "beauty contest" approach. Essentially, the approach lets the dataset speak for itself within the multivariate unobserved components framework.

The section will cover all semi-structural methods used by the Network of EU IFIs. This mainly involves Multivariate filters (HP and PC), Multivariate Unobserved Components (MUC) models and Principal Components Analysis (PCA). We will, first of all, analyse the bivariate case, which extends the univariate models by an additional signal equation. We will then present the general case, i.e. with many additional variables. This is done by means of the associated state-space form of the general multivariate filter. We will finally turn to an unsupervised learning approach represented by PCA.

Bivariate filters

As we saw for univariate methods, it is straightforward to augment the minimisation problem (2), (10), (11) or (33) including additional restrictions into the Lagrangian function. Laxton et al. (1992) at the Bank of Canada suggested augmenting the objective function of the HP filter (2) with the sum of squared residuals from any relevant *signal* relationship (as specified below). Targeting the inflation the authors included a variant of a Phillips curve. Extending their approach, Murray (2014) used the Prior-consistent (PC) filter (11) in an analogous way minimizing:

$$\min_{\{y_t^*\}_{t=1}^T} \underbrace{\sum_{t=1}^{T}(y_t - y_t^*)^2}_{\text{goodness-of-fit}} + \underbrace{\lambda_1 \sum_{t=2}^{T-1}[(y_{t+1}^* - y_t^*) - drift_t]^2}_{\text{degrees-of-smoothness}} + \underbrace{\lambda_2 \sum_{t=1}^{T}(\varepsilon_{3,t})^2}_{\text{economic restriction}}, \tag{34}$$

where y_t^* is the unobserved potential output; y_t is the observed actual output; $\lambda_1$is a positive number that weights the relative importance of the 'smoothness' criterion of the filter (i.e. the growth rate of potential output does not differ too much from its historical average or rate of drift); λ_2 is a positive number that weights the relative importance of maximizing the empirical fit of the extra signal equation; drift is the prior value for the growth rate in the unobserved variable (potential output); and $\varepsilon_{3,t}$ denotes the errors of the additional signal equation. The latter corresponds to the third additional belief (besides of the first two beliefs of univariate PC filter (11)) that other indicators, for instance, inflation or unemployment, are informative about the cyclical position of the country. In the same way, we could extend the most general specification of the stochastic trend and an explicit model of the stochastic cycle in TC filter (33) pioneered by Mohr (2005).

Evidently, the smaller the signal errors $\sum_{t=1}^{T}(\varepsilon_t)^2$ the better the empirical fit of the additional signal equation to empirical data. The relative importance of this 'empirical fit' criterion on the filter's minimization program is measured by an extra penalty parameter λ_2. Similar to HP and PC problems, under additional assumptions that all three error terms are normally distributed, the penalty weights could be set to equal to the respective inverse signal-to-noise variance ratios: $\lambda_1 = \sigma_1^2/\sigma_2^2$ and $\lambda_2 = \sigma_1^2/\sigma_3^2$. Hence, the uncertainty regarding an optimal choice of penalty parameters becomes higher and the choice is still arbitrary. Murray (2014) suggests to set $\lambda_1 = \lambda_2 = k$, i.e. the ratio between the variance of the stochastic trend and the economic relationship is restricted to 1. Another popular solution is to restrict all variances to be equal by setting both parameter restrictions to 1. This choice is appealing if there is an explicit model for the stochastic cycle as in UC models (30).

The state-space representations of the corresponding optimisation problems (for example, (12)–(15)) used for ex-ante and ex-post validation via Kalman algorithm (Kalman 1960) are then extended by adding a particular economic relationship as a signal equation and a set of UC trend-cycle decomposition state-space equations ((24) and (25) or another stochastic model of trend) to extract the supplementary cyclical component c_t^e. Roughly speaking, a practitioner aims to add an explicit model of the cycle for a particular economic variable in the form of the linear regression:

Signal: $$c_t^e = \alpha_1 c_t + \sum_{i=1}^{k} \beta_i x_{i,t} + \varepsilon_{3,t},\ \varepsilon_{3,t} \sim NID(0, \sigma_3^2), \tag{35}$$

where c_t is the output gap (the treatment variable), c_t^e the cyclical component of the additional economic variable (the response variable), $x_{i,t}$ is a set of k additional demand and supply side variables (controls) expressed as the corresponding cyclical components[18].

Murray (2014) considered three alternative signal equations (35) that link inflation, unemployment and capacity utilization with the output gap of the economy through individual relations stemming from economic theory.

(a) *The (New) Keynesian Phillips Curve* (Gali and Monacelli 2005):

$$c_t^\pi = \alpha_1 c_{t-1} + \beta_1 E c_{t+1}^\pi + (1-\beta_1) c_{t-1}^\pi + \sum_{i=2}^{k} \beta_i x_{i,t} + \varepsilon_{3,t}, \tag{36}$$

where c_t^π stands for the deviation of inflation π_t from its time-varying trend determined by a random walk with zero drift (25); $E c_{t+1}^\pi$ is the expected deviation one period ahead; the inertia effect is caught by the lagged deviation c_{t-1}^π; while c_{t-1} denotes the lag of the output gap. Besides, the domestic inflation could be boosted by a foreign demand shock (trade partner's output gap), the gaps in commodity prices or deviations in the terms of trade among many other variables caught by a $k-1$ set of controls $x_{i,t}$. The strength of the link between the cycles is measured by the positive coefficient α_1. The neutrality of money is preserved by restricting the sum of the coefficients for lagged and future deviations of inflation from its trend to unity. Anchoring the expectations to, in general, the time-varying trend of inflation the expected deviation becomes zero, i.e. $E c_{t+1}^\pi = 0$. This restriction implies "accelerationist" specification that is consistent with the forward looking expectations explicitly linked to backward looking heuristic behaviour and inertia. Then, omitting the impacts of other controls, the oft-estimated form of the Philips curve is (Murray 2014, Melolinna and Tóth 2016, Frale and De Nardis 2017):

$$c_t^\pi = \alpha_1 c_{t-1} + (1-\beta_1) c_{t-1}^\pi + \sum_{i=2}^{k} \beta_i x_{i,t} + \varepsilon_{3,t}. \tag{37}$$

Inflation is defined either by the GDP deflator, CPI inflation, or by more domestically aligned CPI subcomponents. For instance, core inflation (CPI inflation net of food and energy prices (Jarociński and Lenza 2016, Murray 2014)) may be considered, or services inflation measures that more closely represent the non-traded element of domestic inflation (Casey 2018). The model parameters could be calibrated (Murray 2014) or estimated by maximum likelihood (Cuerpo, Cuevas and Quilis 2018) and Bayesian methods (Constantinescu and Nguyen 2017).

Murray (2014) points out that the quality of the output gap extraction incorporating NKPC information (37) will depend on the number of assumptions and implicit judgements, including among others:

- the correct specification and estimation/calibration of (37);
- the choice of the inflation measure;
- time-invariance and stability of NKPC over time;
- the quality of trend-cycle decomposition of the inflation;
- the subjective selection of penalty parameters λ_1 and λ_2.

(b) The Okun's law (Murray 2014, Cuerpo, Cuevas and Quilis 2018):

[18] In some empirical applications the more distortionary first difference filter $c_t^e = \Delta x_t$ is used.

$$c_t^u = -\alpha_1 c_t + \varepsilon_{3,t}, \qquad (38)$$

where c_t^u stands for the deviation of unemployment from its natural rate as set out in Bailey and Okun (1965), and the correlation with the output gap c_t is expected to be negative. The idea of Okun's law follows from the narrow notion of potential output as the full-employment level of output. To be consistent with the production function approach Murray (2014) suggested to apply the time-varying non-accelerating wage rate of unemployment (the NAWRU) as produced using a European Commission's CAM (Havik, et al. 2014). Another possible candidate could be the NAIRU or the time-varying trend extracted by a set of UC trend-cycle decomposition state-space equations (Cuerpo, Cuevas and Quilis 2018).

Similar to the NKPC case (37) the simplest form of equation (38) could be augmented with the inertia and other controls (Constantinescu and Nguyen 2017). Besides, a practitioner has to remember that the quality of the output gap augmentation depends very much on a set of highly uncertain assumptions (Murray 2014):

- that the hypothesised Okun's relation is correctly defined;
- that the estimated coefficients are correct and not time-varying;
- that the filter estimate of the NAWRU (the NAIRU) is accurate.

(c) The Capacity utilization (Murray 2014, Cuerpo, Cuevas and Quilis 2018):

$$c_t^{cu} = \alpha_1 c_t + \varepsilon_{3,t}, \qquad (39)$$

where c_t^{cu} stands for the deviation of capacity utilisation from its long-term mean or assumed technologically efficient level and the correlation with the output gap c_t is expected to be positive. The long-term mean could be estimated as an intercept term in the regression (39) or to follow the state-space representation as in Cuerpo, Cuevas and Quilis (2018).

The incomplete definition of capacity utilisation data is the main drawback of this approach. The full sample time series on capacity utilisation is typically available for industry only, whilst services and construction surveyed only starting from 2011. The European Commission uses a synthetic indicator – combined capacity utilisation and business surveys (CUBS) indicator, which is used as a proxy for the unobserved true level of capacity utilization in the economies of the EU Member States (Havik, et al. 2014). CUBS combines capacity utilisation in the industry with economic sentiment indicators in the services and construction sectors. However, for a majority of EU member states (especially for the EU15), the CUBS series is taken to be equivalent to capacity utilisation.

Since there is no explicit link of (39) to theoretical foundations, the capacity utilisation in industry serves as a liable indicator of cyclical fluctuations that commoves with the output gap. This observation motivated Cuerpo, Cuevas and Quilis (2018) to perform a "beauty-contest" with other similar cyclical indicators that are instruments for the imbalances in different aspects of economic activity.

To sum up, implementation of the bivariate filter rests on the estimation of the α_1 coefficients in the above specifications. Noteworthy, there is no unique optimal specification for the above relationships as their empirical fit largely depends on the country and/or the time period empirically investigated (Blagrave, et al. 2015, Casey 2018). A different number of lags or additional explanatory variables can, therefore, be included in a case-by-case basis seeking to improve the empirical performance of the filter in identifying the cyclical and trend components of the economy's output. In addition, in order to combine and exploit the merits of each individual bivariate model, the above relations can be jointly estimated as a system of equations (Murray 2014). The general case, i.e., the case with any number of variables and equations in the system, is presented next.

Multivariate unobserved components models

In this subsection, we present the multivariate filters by means of the combination of an explicit structural multivariate time series model and the Kalman filter as in Cuerpo, Cuevas and Quilis (2018). The structural decomposition provides an efficient way to estimate the output gap or, more generally, to decompose an observed time series as the sum of an arbitrary number of unobserved elements including supplementary variables whose stationary components are related to the output gap.

The decomposition is based on the well-known (univariate) *structural time series* (STS) representation of a time series vector specified as a generic UC state-space model (24)–(26), see Clark (1987), Harvey (1985) and Durbin and Koopman (2012), among others. This method is rather general and flexible albeit keeping the number of parameters tightly controlled, in contrast with other econometric approaches (e.g., Vector Autoregressive model, VAR).

As is common in the literature (Casey 2018, Cuerpo, Cuevas and Quilis 2018), the system of equations describing the laws of motion of the observed variables (GDP, inflation rate, etc.) and unobserved variables (potential output, natural rate of unemployment, etc.) could be compactly represented in *state-space form* using matrix notations. In this paper we use the general multivariate Gaussian form based on Durbin and Koopman (2012):

Signal: $$y_t = H(\alpha)S_t + v_t, v_t \sim NID(0, \Sigma_v), \quad (40)$$

State: $$S_t = F(\phi)S_{t-1} + u_t, u_t \sim NID(0, \Sigma_u), S_1 \sim N(a_1, P_1), \quad (41)$$

where $t = 1, 2, \dots, T$ denotes time, y_t is a $p \times 1$ vector of p *observed* endogenous variables and is called the *observation* (measurement or signal) vector; S_t is a $m \times 1$ vector of m *unobserved* states and is called the *state* (transition) vector. Error terms v_t and u_t are assumed to be normally distributed white noise processes that are serially independent at all time points. S_1 is the *initial* state vector, which depends on the calibrated choice of initial states mean vector a_1, variance matrix P_1 and is assumed to be serially independent with v_t and u_t. $H(\alpha)$ denotes a matrix of coefficients, regulating the relations between the observed and the unobserved variables in the system. $F(\phi)$ denotes a matrix of coefficients regulating the autoregressive law of motion of the state variables in the system. α and ϕ collect the unknown parameters that a practitioners needs to estimate. Variance-covariance matrices of the extended state-space model (40) and (41) are often assumed to be diagonal with some zero restrictions for the identities. The assumption of orthogonality, however, could be relaxed at the cost of making shock identification more difficult (Clark 1987). For example, to represent hysteresis in the trend dynamics the shocks that determine the long-term trend ($\varepsilon_{2,t}$) would need to be correlated with those that drive its short-term rate of growth ($\varepsilon_{\mu,t}$), while the dependence between demand and supply shocks (e.g., as in BN filter) requires $\varepsilon_{2,t}$ and $\varepsilon_{1,t}$ to be (negatively) correlated. All the parameters in the system may be also time-varying and include vectors of exogenous variables similar to (35).

The matrix of coefficients can be estimated by Maximum Likelihood conditional on distributional assumptions about the disturbance terms' vectors. The distributional assumptions in our case are that they are mutually independent and follow the multivariate normal distribution with zero mean and constant variance-covariance matrix (Cuerpo, Cuevas and Quilis 2018). Another prevalent solution is to estimate the parameters by Bayesian methods (Melolinna and Tóth 2016, Constantinescu and

Nguyen 2017). Given estimated or calibrated values for the coefficients of the model and initial conditions, filtered estimates for the unobserved state variables are obtained through a recursive algorithm known as the Kalman filter (Kalman 1960) the details of which are presented in **Appendix B**.

The general specification of the state-space model (40) and (41) means that structural links of observable variables with cyclical and trend components are a straightforward expansion of corresponding vectors and matrices. For example, Casey (2018) includes additional observable variables aiming to expand the baseline univariate model with additional economic information about the cycle. The variables under consideration cover measures of private sector credit growth (Borio, Disyatat and Juselius 2017), residential property prices, real interest rates, real effective exchange rates and current account balance (Darvas and Simon 2015), all subject to alternative measurement adjustments and modifications. In addition, Cuerpo, Cuevas and Quilis (2018) considered fiscal variables, such as public debt, general government primary balance and taxes, among others. The inclusion of fiscal variables is especially relevant to the scope of IFI practitioners.

To conclude, the main advantage of the multivariate approach is that it explicitly draws on well-established relationships between the output gap and other macroeconomic variables (Pybus 2011). Nevertheless, estimates produced using this MUC approach are inevitably subject to the set of conditioning links and implicit assumptions that need to be imposed in advance.

Principal Components Analysis

Principal Component Analysis (PCA) is another conventional empirical tool that helps identify the latent source of cyclical variation in the economy's output. PCA is a statistical unsupervised learning approach that identifies common determinants of a number of variables using linear combinations (components) of the k cyclical variables $c_{i,t}, i = 1,2, \ldots, k$ in order to model their variance-covariance structure. The *unsupervised* nature of the approach means that the output gap c_t does not guide the signal extraction from $c_{i,t}$ as a response in the form of equation (35). However, the extracted first principal component could later be used as a *latent factor* exactly in the same way we did before. A set of linearly uncorrelated and orthogonal linear combinations that are ordered in terms of variance are generated and are called principal components. In terms of ordering, the first element contributes the most to the sample variability..

Various studies apply the PCA method to draw a common signal from a range of cyclical indicators within the context of output gap estimation. Studies include Pybus (2011), Ódor and Kucserová (2014), Murray (2014) and Casey (2018). The studies consider a range of cyclical indicators, including survey measures of capacity utilization, recruitment difficulties, inflation or wage earnings growth. According to Pybus (2011), the first principal component, i.e., the linear combination with the greatest variance, can be interpreted as a proxy for the output gap. This assumes that the output gap is the most important common determinant of the cyclical indicators.

The first principal component is a vector of *loadings* that maximizes the quadratic form:

$$\max_{\lambda_1} \lambda_1' C' C \lambda_1, \text{ subject to } \lambda_1' \lambda_1 = 1, \tag{42}$$

where C is the $k \times T$ matrix of cyclical components. The solution of (42) is a standard matrix algebra – the value of the objective function is an eigenvalue and λ_1 is the corresponding eigenvector.

An important aspect of the process concerns how the cyclical indicator variables used are standardised and demeaned. This should be done in such a way as to ensure that each indicator has a comparable contribution to the total sample variability. Appropriate scaling of the estimated principal component(s) is also crucial for a correct output gap inference. On the other hand, the standardization procedure produces an output gap with zero mean and unit variance. Therefore, to obtain output gap estimates that are comparable with other methods in real terms, the output gap series obtained from PCA requires rescaling by the statistics, for instance, of the HP-filtered output gap (Ódor and Jurašekova Kucserová 2014). Alternatively, the cyclical indicators could be decomposed by any univariate trend-cycle decomposition method discussed above and then standardized by a standard deviation. Finally, PCA can be complemented with the 'Aggregate Composite' method (Pybus 2011), where instead of estimating the cyclical indicators' weights on the principal component(s), they can be explicitly set by the user.

Accounting for production

The most widespread method of estimating potential output practised by EU IFIs is the Production fFunction (PF) approach **(Fig. 4)**. It is also widely applied by central banks and international organizations, including the European Commission: using the CAM (Havik, et al. 2014). The advantage of the PF method is its direct link to economic theory. The supply-side driven neoclassical Solow-Swan growth model is invoked. This means that the method is more structural and comprehensive compared to other approaches. Hence, the PF approach allows for a more direct link to sources of structural information and for an easier interpretation of the source of changes in the output gap or potential output.

While the PF approach is relatively more structural and comprehensive compared to other approaches, it is not free from some of the limitations present in other methods. The bottom-up feature of the approach does not filter actual output directly, but it does decompose output into a number of components that are themselves individually filtered (Butler, 1996). The method can therefore be considered a semi-structural multivariate filter. This can mean that problems associated with filtering output are in some senses shifted to the trend estimates of inputs (Cerra and Saxena 2000).

The general idea of a production function decomposition is to split the potential output into three components: a labour input, a capital input and a total factor productivity (TFP), each of which could be further decomposed into even smaller subcomponents to reflect demographic changes, capital or TFP composition. There are various functional types of the production function. The majority of practitioners prefer the simplest Cobb-Douglas (CD) type and a few applications consider a more general constant elasticity of substitution (CES) type, where CD is a particular case. Following the majority choice, let us consider an aggregate Cobb-Douglas constant-returns-to-scale production function[19] as determined in the CAM (Havik, et al. 2014):

$$Y_t^* = \left(U_{L_t} L_t E_{L_t}\right)^{\alpha} \left(U_{K_t} K_t E_{K_t}\right)^{1-\alpha} = L_t^{\alpha} K_t^{1-\alpha} TFP_t, \tag{43}$$

where Y_t^* is the potential output, TFP_t is the level of total factor productivity (Solow residual), L_t is the labour input, K_t is the physical (fixed assets) capital stock, which we will refer to as capital, and α is the labour share of income, (U_{L_t}, U_{K_t}) are the corresponding degrees of excess capacity, and $(E_{L_t},$

[19] Since the production function depends only on the primary inputs ignoring the explicit use of intermediate inputs (materials, energy, etc.), the correct name for the PF would be a *value-added function* (OECD 2001).

E_{K_t}) the levels of efficiency. TFP which summarises both the degree of utilisation of factor inputs as well as their technological level is set equal to:

$$TFP_t = \left(E_{L_t}^{\alpha} E_{K_t}^{1-\alpha}\right)\left(U_{L_t}^{\alpha} U_{K_t}^{1-\alpha}\right). \tag{44}$$

The parameter α is often calibrated to the country-specific average in-sample ratio of wage costs and value added in the economy. To ensure the comparability of the countries it could be also fixed to a particular value for all countries (e.g., in the CAM it equals 0.63 – the average wage share for the EU15 over the period 1960–2003).

Total factor productivity is the portion of output not explained by primary inputs. It is computed as a Solow residual in an agnostic way, i.e., often there is no explicit narrative behind the estimates although there could be. Murray (2014) notes that (43) is consistent with a number of assumptions:

- constant returns to scale;
- the marginal productivity of each primary factor is proportional to its average productivity;
- technology experiences Hicks, Harrod, and Solow neutrality at the same time;
- the unit elasticity of substitution between the capital and labour inputs (a particular CES case);
- under the additional assumption of perfect competition in the product market, factors are paid their marginal products, implying the stable steady-state labour share in factor inputs income allowing to calibrate the parameter α to the observed data.

The assumption of perfect competition could be relaxed though, implying that the part of value-added that goes to factor inputs income is deduced by the average mark-up of the imperfectly competitive market. Assuming that in the long-run mark-up is constant, implies that in the case of CD technology the difference will be absorbed by the constant part of Solow residual.

Capital inputs are often a challenge from a conceptual and data perspective even in advanced economies **(Box 4).** The capital input time series tend to be based on a wide range of assumptions and calibrated parameters, including the rate and form of depreciation profiles and initial values. Approaches such as the CAM see growth in the level of the actual net capital stock as driving the capital contribution to potential output (Casey 2018). The actual levels of the capital input are implicitly considered as sustainable (unsmoothed as in Havik et al. (2014)). However, they may not be sustainable. For example, overinvestment in housing can lead to boom-bust dynamics in net capital stock series, hence inflating potential output and resulting in unsustainable (and cyclical) development of the potential output in the long-run. This can be especially evident in small open economies, whereas a neoclassical Solow-Swan growth model for closed economies ignores the international trade flows and the mobility of primary inputs.

Bearing in mind the difficulties with the determination of the relevant capital inputs **(Box 4),** Butler (1996) proposes focusing on the definition of the marginal product of labour from which the actual and potential outputs after logarithmic transformation could be decomposed into the sum of labour and the marginal product of labour inputs:

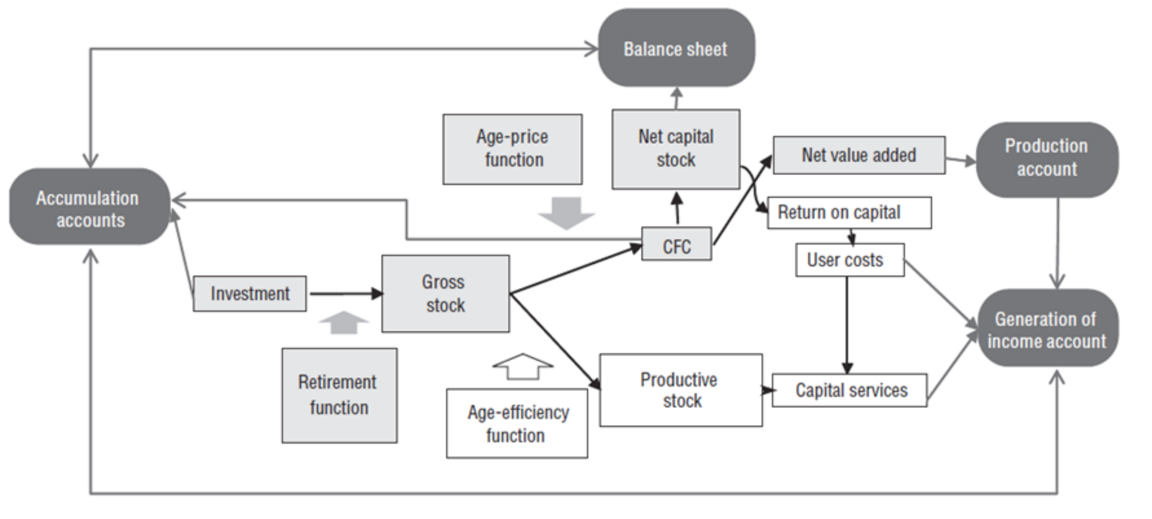
$$M_t = \frac{\partial Y_t^*}{\partial L_t} = \alpha \frac{Y_t^*}{L_t}, \qquad y_t^* = l_t + m_t - \alpha, \tag{45}$$

where small letters denote the log-transformed variables, and m_t is the marginal product of labour. The economic assumptions underpinning this particular decomposition are that the production technology is CD in labour and in all other inputs and that markets are perfectly competitive. If these assumptions are violated, the variable m_t might be interpreted as a *scaled average product of labour*, rather than the marginal product of labour (Butler 1996).

Box 4. Which capital?

In terms of the measurement of the capital stock, the perpetual inventory method (PIM) is often used. The problem of the CAM in applying the method is a wide range of rough assumptions and conceptual issues, which are at odds with the notion of the potential output. All essential issues are broadly discussed in the OECD manual on Measuring capital (OECD 2009). **Figure 7** depicts an integrated system of links between the key capital stocks: gross, net and productive; capital flows: investments, value and efficiency depreciation, capital services; and national accounts: accumulation, production, income and balance sheet.

Figure 7 – capital measures in the system of national accounts



Source: Measuring capital (OECD 2009)

Conceptually, any capital asset represents two *complementary* economic aspects: (1) the wealth or market value of the asset and (2) the contribution of capital to production. First, the *net* or *wealth* capital stock measures the market value of the existing physical capital, while the *gross* capital stock measures the hypothetical value of accumulated assets being treated as new. In the system of national accounts, most attention is paid to the assessment of gross and net capital stocks, which are used to derive the loss of an asset value (depreciation) as it ages. This is typically associated with the age-price profile of the asset. This definition of an asset's value is needed to consistently define the national wealth and the income accounts. However, it is not necessarily a relevant concept for a production function approach. In particular, the value of an asset may not be relevant for the actual production derived from that asset. Nevertheless, data limitations and the complexity of deriving more appropriate measures mean that the less meaningful capital stock measure is typically relied on in production function models such as the CAM (Havik, et al. 2014) and in similar methodologies used by other institutions.

On contrary, the *productive* stock expressed in *efficiency* units allows the practitioners to impute the flow of *capital services* used in production. Past investment in every group of assets is accumulated after correcting for the efficiency loss of the asset, which is linked to the age-efficiency profile. Conventionally, unobserved productive capital services are assumed to be proportional to the total productive capacity of the asset, bearing in mind the intensity with which the productive capital was actually used for the production purposes.

These two primary concepts differ in the chain-linked aggregation across different assets, where *market prices* are used to chain-link the net capital stocks and derive depreciation, while *the user cost of capital* is applied to aggregate the productive capital and obtain the flow of capital services. Despite their differences, **Figure 7** shows the complementarity of the approaches in the system of national accounts. Both capital services and capital stock are key inputs for the household consumption: services determine the value of the capital services rental flow – a part of the other income, and stock – the market value of capital assets, the main part of non-human wealth component. Unfortunately, the estimates of the productive stock and capital services are still an optional part of the ESA 2010. Besides, a practitioner may question the PIM assumptions used by the Eurostat, national statistical agencies or the European Commission to estimate the net capital stock. Therefore, it is operational to impute the whole set of capital measures and make them country-specific (Celov 2015).

Following the system of national accounts, let us assume that the total capital formation could be divided into dwellings, structures, machinery and equipment, transport and intangible assets. The estimation of capital measures for each of these assets by PIM then requires a particular set of country-specific parameters and data. First, the average service lifetime of the assets T_a, which could be based on an empirical research or calibrated to the studies in similar countries. Second, a retirement (survival) pattern $F_{a,\tau}$, which describes how assets of age τ are withdrawn from the service (scrapped, discarded). This survival function is associated with the mortality function $f_{a,s}$, which is assumed to follow a specific distribution (e.g., a log-normal as in France, Estonia or Finland). Practical considerations suggest to choose the standard deviation from $[T_a/4\,, T_a/2]$, for example, $T_a/4$ will show a more peaked mortality functions around the average service life-time of the assets. The long-normal distribution mean then is $\mu_a = \log(4 \cdot T_a) - 0.5 \cdot \sigma^2$, while standard deviation $\sigma = \sqrt{\log(1 + (1/4)^2)}$ is the same for all assets. The assumed retirement pattern and gross-fixed capital formation for each of assets are the only inputs required by PIM to derive the *gross capital stock*:

$$K^G_{a,t} = \sum_{\tau=0}^{T} F_{a,\tau} \cdot \frac{I^N_{a,t-\tau}}{P^I_{a,t-\tau,0}}, \tag{46}$$

where T is the maximum service life of asset type a (a practitioner may assume the maximum cut-off to be 1000 years or less for all assets); $I^N_{a,t}$ is the gross nominal expenditure on asset a in period t deflated by the prices, when the asset was new $P^I_{a,t,0}$.

Third, the loss in productive capacity of a capital good over time is shown in its age-efficiency profile or the rate at which the physical contributions of a capital good to production decline over time, as a result of wear and tear. For this purpose, a practitioner may use hyperbolic age-efficiency profiles adjusted for the chosen log-normal mortality pattern:

$$h_{a,\tau} = \sum_{s=\tau}^{T} g_{a,s} \cdot f_{a,s} = \sum_{s=\tau}^{T} \frac{s-\tau}{s-\beta_a \tau} \cdot f_{a,s}, \tag{47}$$

where $\beta_a \in [0, 1]$ is the efficiency discount factor, which determines a form where assets lose little of their productive capacity during the early stages, but experience rapid loss of productive capacity towards the final stage of their service lives. Now, using (47) the *productive capital stock* for the asset a is defined by PIM as:

$$K_{a,t}^{P} = \sum_{\tau=0}^{T} h_{a,\tau} \cdot \frac{I_{a,t-\tau}^{N}}{P_{a,t-\tau,0}^{I}} \Rightarrow K_{a,t}^{P} \cong K_{a,t-1}^{P}\left(1-\delta_{a,t}\right) + I_{t}^{R}, \quad (48)$$

where I_t^R is the real expenditure on asset a in period t, $\delta_{a,t}$ is the time-varying depreciation rate, and the evolution of productive capital is approximated by a geometric profile. If there is no efficiency loss $g_{a,s} = 1$ the chain-linked volumes of productive and gross (as new) stocks are identical. Market values of both are obtained by multiplying with the purchase prices of the new assets $P_{a,t,0}^{I}$.

Fourth, the age-price profile $z_{a,\tau}$ could be recursively obtained from the age-efficiency profile $h_{a,\tau}$ assuming a country-specific discount rate r for investments:

$$z_{a,\tau} = \frac{P_{a,\tau,0}^{I}}{P_{a,t,0}^{I}} = \frac{\sum_s h_{a,s+\tau} \cdot (1+r)^{-(1+s)}}{\sum_s h_{a,s} \cdot (1+r)^{-(1+s)}}. \quad (49)$$

The derived age-price profile again could be approximated by the geometric or declining balance model of depreciation, where the key procedure is to assess the depreciation rates from $z_{a,\tau} = \left(1-\delta_{a,t}\right)^{\tau}$. Using the age-price profile a practitioner then gets the net capital stock at current prices (the residual market value):

$$K_{a,t}^{N} = \sum_{\tau=0}^{T} P_{a,\tau,0}^{I} \cdot \frac{I_{a,t-\tau}^{N}}{P_{a,t-\tau,0}^{I}} = P_{a,t,0}^{I} \sum_{\tau=0}^{T} z_{a,\tau} \cdot \frac{I_{a,t-\tau}^{N}}{P_{a,t-\tau,0}^{I}}. \quad (50)$$

Comparing (48) and (50) shows that the two different concepts of capital are the same if and only if the age-price and age-efficiency profiles are identical. This is a strong simplifying assumption often violated in empirical tests.

Fifth, to initialize the PIM a practitioner needs to find valid initial values for the capital stocks. For the economies in transition, which are from the steady-state growth, it is not recommended to apply the approximate formula $K_{a,0} = I_{a,0}^{R} / (\gamma_a + \delta_a)$ using Harberger suggestion to account for in-sample average output growth γ_a and depreciation rate δ_a. It is more operational to assume that the investments prior to the sample where the same as $I_{a,0}^{R}$ and allow the PIM to start from say 100 years before the initial year of the sample.

Finally, to obtain the *capital services* – the relevant capital input into a production function – a practitioner first computes the real rate of return on capital with capital gains less imputed gains from the housing (with depreciation) and the depreciation of the public sector capital stock. Denoting the capital income generated by the private sector without housing services as $Y_{-d,t}^{K}$ a practitioner then gets the *imputed rate of return on capital*:

$$R_t^{imp} = \frac{Y_{-d,t}^{K} - \sum_a P_{a,t-1}^{I}\left(\delta_{a,t}\left(1+\pi_{a,t}\right) - \pi_{a,t}\right) K_{a,t-1}^{P}}{\sum_a P_{a,t-1}^{I} K_{a,t-1}^{P}}, \quad (51)$$

where $\pi_{a,t}$ is the (smoothed) inflation of the asset prices $P_{a,t}^{I}$, $\delta_{a,t}$ denotes depreciation rate, $K_{a,t}^{P}$ is the productive capital stock defined in (48). Then, the user cost of capital for an asset a is:

$$P_{a,t}^{K} = P_{a,t}^{I}\left(1 + R_t^{imp} - \left(1+\pi_{a,t}\right)\left(1-\delta_{a,t}\right)\right), \quad (52)$$

which are used to determine the capital shares $\omega_{a,t} = P_{a,t}^{K} K_{a,t-1}^{P} / \sum_a P_{a,t}^{K} K_{a,t-1}^{P}$ for Törnqvist aggregation to obtain the sought capital services:

$$\Delta \log K_t^S = \sum_a \Delta \log K_{a,t}^{P} \left(\omega_{a,t} + \omega_{a,t-1}\right)/2. \quad (53)$$

The last step completes the roadmap depicted in **Figure 7**.

We now turn to the labour input in the context of the PF approach. Labour input is typically defined in terms of total hours worked. The trend of labour input L_t consists of several subcomponents:

$$L_t = POP_t \cdot PR_t \cdot (1 - NAWRU_t) \cdot H_t, \tag{54}$$

where POP_t is the actual population of working age (20–74 age group). The working age population is not de-trended implying a zero gap. The $NAWRU_t$ is the Non-Accelerating Wage Rate of Unemployment; PR_t is a trend of participation rate; and H_t denotes a trend of hours worked. The contribution from labour inputs and the largest portion of criticism is often focused on the identification of the NAWRU (Casey 2018). The trend unemployment rate (NAWRU) in the CAM is based on an "accelerationist" Philips curve (Havik, et al. 2014). The basic idea originates from the application of multivariate unobserved components model, where the signal equation is a particular modification of the (new) Keynesian Philips curve (36) expressed in terms of real or nominal unit labour costs.

Despite the more structural and detailed decomposition of the output gap, there are also disadvantages of the method. This is relevant in, but not limited to, applications such as the EU CAM (Havik, et al. 2014). For example, the method shares the end-of-sample biases that affect the underlying de-trending techniques that are used for detrending the subcomponents (Cerra and Saxena 2000). The potential output estimates are affected by the measurement errors in primary inputs and their pro-cyclicality due to the problems assessing the NAWRU and unsmoothed investments being used as inputs to PIM. The approach may also suffer from the omitted variable bias if the value-added input is not corrected for the imperfectly competitive market's mark-ups (Willman 2002). Fortunately, the disadvantage of the a-theoretic nature of TFP as Solow residual in CD production function case turns into a feature allowing to absorb all abovementioned deviations in definitions, assuming that the long-run imperfect markets' mark-ups, the rates of effective use of primary inputs are constant or slowly varying with exponential trends.

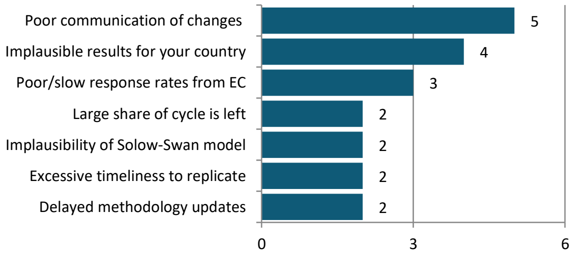
The next two subsections shed more light on what are the key problems envisaged by the EU IFIs when replicating the EU CAM. Some practical solutions are offered to help overcome difficulties faced with the still prevalent production function approach.

The Commonly Agreed Methodology (CAM)

Production function approaches like the CAM share some important theoretical features. First, there is the key institutional role of the CAM's estimates for the assessment of national fiscal policy under the Stability and Growth Pact. Second, there is the method's superior reliability relative to that of other estimates produced under approaches by international institutions such as the OECD and IMF, as found empirically by Mc Morrow, et al. (2015). Third, there is a fair ability for CAM-based estimates to co-move with economic cycles. For example, Atanas, Raciborski and Vandermeulen (2017) find that the CAM's output gap estimates are significantly correlated with a number of economic and business cyclical indicators for most EU member states (applying the so-called "plausibility tool"). Fourth, there is the PF's advantage in terms of yielding greater economic meaning, especially when compared to univariate statistical techniques. These and other features discussed in Havik, et al. (2014) justify the practical use of the CAM's estimates for benchmarking and/or their reproduction under a different scenario. At the same time, the noted practical shortcomings remain as does the need for tailor-made country-specific alternatives.

Other attempts, within the production function framework, include formulation and estimation of models with alternative production functions. For example, the multi-sector model with imperfect competition by Willman (2002) advanced further within the New Multi-Country model, developed at the ECB (Anderton, et al. 2014, Celov 2015). The former assumes two alternative production functions: a 'Cobb-Douglas' and a 'Constant Elasticity of Substitution' (CES). The latter builds on a CES production function and allows for a flexible Box-Cox functional form for the technological progress. Though theoretically sound, the analytical complexity and moderate empirical performance of these approaches sets limits on the extent to which wider use of these methods across the Network of EU IFIs may arise.

Figure 8 – Difficulties replicating the methodology used by the European Commission



Source: the Network of EU IFIs

Many EU IFIs try to replicate the CAM. According to the survey**Error! Reference source not found.Error! Bookmark not defined.** of the Network of EU IFIs, 9 of 20 respondents have tried to replicate the production function approach based on the CAM (Havik, et al. 2014) using information and software obtained from the CIRCABC website[20]. In some cases, this has proven challenging. The key problem cited by EU IFIs is one of the frequent and substantial methodological changes that are poorly communicated. In addition, the results are often found to be implausible for a number of countries. Many cite the fact that revisions to output gap and potential output estimates are often pro-cyclical **(Fig. 8)**. Ódor and Kucserová (2014), Casey (2018) among others highlighted the increased computational complexity of the CAM. Also cited is the high degree of sensitivity of the estimates to the initial choice of model parameters, which are set by the European Commission staff and are often changed in an opaque way. Another issue is the overall increased uncertainty of the estimated output gaps. Other problems cited relate to the timing of output gap publication by the European Commission and, of course, end-point issues.

Ódor and Kucserová (2014) provided a comprehensive list of weaknesses of the CAM approach in the context of the small and open economies:

- small sample sizes used to estimate long-term trends with many structural breaks;
- a high uncertainty regarding data inputs, especially the notion and assessment of capital data;
- an empirical implausibility of the neoclassical Solow-Swan growth model and its assumptions (e.g., a perfect competition in the product and labour markets);
- downplaying the role of international capital and labour mobility as well as the impacts of current account and financial account imbalances.

[20] https://circabc.europa.eu/faces/jsp/extension/wai/navigation/container.jsp

All of these aspects may lead one to conclude that the CAM eventually requires addressing country-specific issues in a diverse EMU. Therefore, IFIs should adopt country-specific methods based on the best practices of the Network of EU IFIs and other international institutions (Ódor and Jurašekova Kucserová 2014). It is worth noting that the EC's researchers have reflected criticisms of the CAM to various extents and have sought to improve the methodology in recent years by moving to an incorporation of more country-specific measures of potential output and the output gap.

Frale and De Nardis (2017) emphasized the complexity to determine the required data inputs and to apply the CAM in the course of the recent global economic crisis. The long-lasting recession and divergent recovery may have permanently damaged the potential productive capacity[21] of the EU economies implying a hysteresis *a la* Blanchard and Summers (1987) and increasing the difficulty of distinguishing between the temporary (cyclical) component and the trend (structural) component of GDP growth. However, an option in the CAM that is often applied is to force the output gap to close at the end of the forecast period (Casey 2018). This, by design, can rule out the possibility of persistent output gaps and may give an overly benign impression of the path of the economy over the forecast horizon, which may not always be appropriate. There may, for example, be strong reasons to believe that the trajectory of the economy over the immediate forecast horizon will not tend towards a closed output gap. Besides, Frale and De Nardis (2017) noted that methods characterised by standard economic conditions and historic cyclical frequencies, might not fit the anomalous cyclical environment seen after 2009. Finally, the business survey indicators like capacity utilisation rate or CUBS used as a supplementary cyclical variable may have become less informative in the last few years, as the recent crisis may have changed the assessments and expectations of economic agents about what is considered a normal level of economic activity. Therefore, more adaptive and flexible semi-structural approaches like MUC or suite of models are useful to consider.

Casey (2018) discussed country-specific issues related to the applicability of the production function methodology for small and/or open economies. The author considered the appropriate measurement and modelling of the production function inputs, i.e., the imputed values for capital, trend-cycle decomposition of labour inputs (in particular, pro-cyclical estimates of NAWRU) and the treatment of the total factor productivity as a Solow residual. For example, external trade balance or current account balance data provide essential information when analysing open economies. The ability to absorb excess demand shocks should not be disregarded. Darvas and Simon (2015) showed that the size of revisions to CAM-based output gap estimates is correlated with the variability of the current account balances prompting about omitted absorption cycle variables in the CAM. Concerning capital measurement, Anderton, et al. (2014) pointed to the importance of the distinction between housing and non-housing capital in the case of countries with construction booms and busts, not to mention the conceptually opaque definition of the capital inputs applying the capital data series presented in AMECO. However, adding analogous economy-wide considerations to the production function approach leads inevitably to the formulation of a small-scale Dynamic Stochastic General Equilibrium (DSGE) models (Ódor and Jurašekova Kucserová 2014). Similarly to the CAM, the use of DSGE models is limited due to their increased analytical complexity and moderate empirical performance, which again advocates using more adaptive semi-structural methods.

All the findings regarding pro-cyclicality in inputs are supported by IFIs colleagues from other institutions. For instance, Huovari, Jauhiainen and Kekäläine (2017) scrutinizing the CAM estimates

[21] In some recent papers the slowdown in productivity growth in the euro area is explained by the misallocation of primary production factors (Dias, Marques and Richmond 2015, Gamberoni, Giordano and Lopez-Garcia 2016).

for Finland concluded that the Finnish output gap estimates vary greatly with assumptions, especially with those made on the trend unemployment (pro-cyclical NAWRU estimates) and the participation rate. All these variations are then reflected in the cyclically adjusted budgets (CAB) implying that the CAM does not ensure plausibility and robustness criteria for the CAB to play a role as a binding rule. Fioramanti (2016) analysing similar problems for Italy has found that a more complicated model with a lot of possible sources of microscopic differences (e.g., data and projections revisions, changes in model assumptions) leads to considerable consequences in terms of policy implication.

Best practices of the network

The critical issues discussed above have led to a number of papers studying the uncertainty underlying the European Commission's (EC) official estimates of the particular economy's potential output and examining their sensitivity to alternative deviations from the official implementation framework. Axioglou (2017) studies the application of the CAM to the Greek economy and points to considerable uncertainties, which stem from parameter estimation uncertainty as well as the accumulating impact of a sequence of methodological choices and technical adjustments underlying the EC's final estimates.

More specifically, Axioglou (2017) finds substantial estimation uncertainty of both inputs of production that are subject to estimation, i.e., potential labour and the trend in TFP. This reflects the lack of a strong empirical 'Phillips Curve' relationship between unemployment and wage inflation (in case of potential labour) and a weak relationship between capacity utilization and total factor productivity. In the absence of strong identifying relationships, parameter estimation shows sensitivity to the imposition of parameter restrictions and/or the choice of Bayesian priors. Both effects, though individually small, may add up in the final estimates, as the aforementioned production inputs are estimated individually.

Post-estimation adjustments of the official methodology due to structural unemployment 'anchoring' and/or 'demeaning' of the NAWRU were also found to affect the Greek vintage. However, anchoring was found advantageous to reduce the revisions and pro-cyclicality of the structural labour estimates in other countries. On contrary, some elements of the official methodology were found with negligible impacts, such as the choice of the smoothing parameter in the implementation of the HP filter on the working hours or the participation rate series. This is due to the empirical behaviour of these series in the data and their small contribution to the evolution of the potential labour series (Axioglou 2017).

The choice of the parameter α of the CD production function had only a minor impact on the cyclicality of the estimated total factor productivity. This mainly reflects the low degree of the cyclicality of the labour/capital ratio, conjectured in the study, based on visual inspection of the relevant data, not only for Greece, but also for other EU countries. This is an anticipated result considering that the TFP growth is conventionally re-assessed conditional on the changes in the parameter α. If a practitioner makes the technological growth endogenous and fixes the growth rate, the changes in the parameter α then results in the rotation of the level of the potential output around the mid-sample point and a substantial revision of the output gap estimates at the sample ends. Consequently, Axioglou (2017) shows that the effect of α on the potential output estimate mainly depends on the deviation between potential and actual labour. In turn, this suggests that the effect of α can be rather different for countries and periods with large deviations between observed and potential labour, such as in Greece over the recent years.

Murray (2014), describing the PF approach, points to two crucial judgements associated with the estimates of the production function. The first is the choice of aggregated production technology – an economic theory that relates the outputs to inputs. The second is the choice of trend-cycle decomposition of primary factor inputs and TFP, which are to be aggregated using the specified production technology. Following the simplest majority choice, Murray (2014) applies the CD technology (43), the functional form of which is attractive in terms of computation and robustness to omitted variable problems. An alternative model (Lunsing 2011, Celov 2015) is to consider CES technology that allows a practitioner to test the relevance of CD restrictions, the neutrality cases of the technology and resolve medium-run debate between substitutionary (capital fundamentalism inspired endogenous growth) or complementary (Acemoglu-augmented technological process) effects varying the primary inputs (Celov 2015). However, there is a little empirical evidence that supports the use of one technology over the other when the objective is to forecast potential GDP or factor shares, provided the labour income share is stable (Miller 2008). The same typically applies to the multi-sectoral decomposition of the gross value-added, where the deeper sectoral story does not necessarily imply better predictions of the potential output at the aggregate level.

Murray (2014) claims that it is reasonable to assume that the non-market sector (a large portion of which corresponds to the public sector) does not generate the gap as such. Then the output gap associated with the remaining market sector will be, in general, of a greater magnitude. On the other hand, Celov (2015) shows that, for Lithuania's case, the gap in the non-market sector actually exists and is different in phase and amplitude from the market sector gap. In fact, the assessed market sector gap closed faster (around 2012) after the crisis of 2009, while the non-market sector's gap was still negative at the end of 2013 due to the discretionary pro-cyclical measures of fiscal consolidation. The aggregate level result, however, was almost identical to the one sector assessment. The idea that some subsectors of the economy could experience no relevant gap is considered in Casey (2018), where the author excludes the sectors of the Irish economy, where the value-added is produced by large foreign-owned multinational enterprises. This warrants a focus on the domestic sectors of the economy.

When the production technology is selected the further non-structural choice is to pick a particular trend-cycle decomposition method for the factor inputs and TFP. The vast majority of practitioners do this by using filters of some sort. This choice varies from already discussed univariate filters and also include their multivariate extensions. The multivariate approach allows a practitioner to include additional structural information describing the evolution of the TFP and the primary input factors of the production (Butler 1996, Murray 2014). Besides, the decomposition may involve the long-run (steady-state) targets (anchors) to which the corresponding elements converge in the long-run. The use of anchors is appealing since they stabilize the potential output revisions as they make the de-trending procedures less sensitive to the data updates and new data inputs. In the case of TFP such anchors could be linked to the empirically testable hypothesis of sub-club convergence of the EU countries which is consistent with the multi-speed EU economic growth hypothesis and would result in country-specific convergence dynamics of the member states, rather than the deterministic (mechanical) output gap closure rules as in the EC's CAM.

Murray (2014) shows that the trend-cycle decomposition of the subcomponents in the UK's case, in general, is equivalent or rival to the direct decomposition of the actual output. The advantage of the PF approach is that a practitioner can decompose the output gap into contributions from the labour input[22] and TFP gaps. There are two possible contribution decompositions (Murray 2014). The first breaks the output gap down into contributions from employment, average hours worked and labour productivity deviations from their trends. The more detailed decomposition additionally provides a practitioner with an indication of the TFP and capital intensity contributions to the output gap.

Much of the criticism of the CAM and of applications of the PF approach comes from the pro-cyclicality of the NAWRU estimates – essentially the level of unemployment keeping wage inflation constant (Casey 2018). Currently, the CAM production function obtains the implied trend unemployment rate based on a version of an accelerationist Phillips curve. Combining this with trend labour force levels gives trend employment levels, which together with trend average hours worked delivers the total potential level of factor inputs from the labour side (i.e., trend total hours worked).

The estimation of the NAWRU has been a focal point for recent criticism of the production function approach employed under the CAM (Fioramanti 2016, Darvas and Simon 2015). The plausibility of results can, in some cases, be questionable in the absence of clear wage pressures. However, without observing the actual rate of unemployment that would be consistent with constant inflation, it is difficult to dispute the validity of such estimates. Perhaps more concerning is the extent to which the estimates can tend to track actual unemployment for some economies. Rather than identifying a persistent trend unemployment rate, the NAWRU appears to more closely approximate the actual unemployment rate (Casey 2018).

Casey (2018) also points to the possibility that the standard Phillips curve approach (used in the CAM) on which NAWRU estimates are grounded could be extended. Recent decades have seen inflation become less sensitive to unemployment changes. One reason for this may be that inflation expectations have become better anchored. The presence of credible inflation-targeting central banks is an often cited reason for this anchoring.

Finally, a further issue is the influence of migration flows on estimates of potential output (Casey 2018). Net inward migration can boost labour inputs and hence potential output estimates in the PF approach. However, these flows can also dampen the traditional Phillips curve relationship between output (or unemployment) and inflation. This dampening effect arises due to the additional labour supply prompted by migration, which can serve to limit the expected inflationary pressures that might arise when unemployment is low. In turn, this can add to difficulties in discerning a stable level of unemployment at which inflation does not change (the NAWRU) and, hence, in distinguishing between cyclical and trend developments.

In conclusion, it seems reasonable to practitioners in EU IFIs that the role of the production function may be more suited to long-run issues rather than short/medium term issues. This could prompt a two-way methodology: long-term potential output estimates based on the production function approach (more affine to growth accounting) and medium-term output gap estimates based on semi-structural multivariate time series models. The first approach goes from the distant future (e.g., 20 years or more ahead) to the present while the second one goes from the present to the medium-term (e.g., up to 5 years) targets.

[22] Although capital and population gaps are conventionally assumed to be zero, a practitioner may relax these assumptions. Indeed, since the capital and labour for small open economies are highly mobile these results in investments and labour force migration waves to actually impact the changes in the potential output. This is evident in the recent boom-bust cycle case for the most of the EU economies.

Mixing things up

Fiscal policy formulation and assessment in real-time requires forecasting the output gap at the end-of-sample and onwards. This task can be performed in several ways, depending on the suite of models available to estimate the output gap and the range of available forecasts of the variables used as inputs by the different models. A model (or a procedure) designed to estimate the output gap can be considered as a transformation of some observable variables (inputs) that generate a decomposition of the observable GDP into a trend component (the potential output) and a cyclical component (the output gap), which could be described by a generic function:

$$OG_{t,j} = \Psi_j(y_t, X_{t,j}), \tag{55}$$

where t denotes time ranging from 1 to T; Ψ_j is a model (or a particular method) used to estimate the output gap ($j = 1, 2, \dots, J$); $y_t$is the observed actual output; $X_{t,j}$ is a subset of k_j variables that are used as inputs of the procedure, the size and composition of which may vary according to the model j, e.g. for univariate methods the subset is empty; $OG_{t,j}$ is the output gap estimated by the method j.

In the context of equation (55), forecasting the output gap h steps ahead means generating estimates above T, including the nowcasts for the current year T:

$$OG_{f,j} = \left(OG_{T,j}, OG_{T+1,j}, \dots, OG_{T+h,j}\right), \tag{56}$$

where f corresponds to a particular forecasting scenario ($f = 1, 2, \dots, F$). For any fixed method j a practitioner could actually apply two alternative forecasting procedures each of which generates a *within uncertainty range*, or the uncertainty associated with the model inputs:

(a) An explicit *model-based* or *input-based* perspective (e.g., a MUC model) and external projections of the inputs $\hat{y}_t$ and $\hat{X}_{t,j}$, where $t = T, T+1, \dots, T+h$. For example, such external projections could follow from an independent economic development scenario used to prepare budgetary plans of the general government. Then, the model-based procedure yields standard errors of the output gap forecasts as a by-product of running the Kalman filter (Kalman 1960).

(b) An implicit *output-based* perspective, when using alternative projections of the inputs (e.g. alternative GDP projections provided by the Consensus) a practitioner generates a set of different output gap forecasts. Then any statistic (e.g., mean, median, middle of the range) define the point estimate for the output gap and their standard deviation (or median absolute deviation) can be used to set an uncertainty range. In other, words the Consensus uncertainty is translated into within uncertainty range for a particular method.

Assume that in total F different input scenarios exist defined by either (a) or (b) and each of the alternative scenarios is applied for a suite of J models. Then, for each alternative set of inputs been fixed a practitioner could compute a *between uncertainty range*, or the uncertainty associated with the model selection, including the parameter uncertainty (e.g., the choice of a penalty parameter for the HP filter). Both types of uncertainty ranges then could be placed into a two-way uncertainty table (**Table 2)**, the core of which is a $F \times J$ matrix of alternative point estimates of the output gap for a given year t. **Table 2** quantifies at the same time the model uncertainty forming a between uncertainty range (reading it column-wise) and uncertainty about the inputs forming the within uncertainty range (reading it row-wise). The summary estimates, $\overline{OG}_{\cdot,\cdot}$and $\sigma\left(OG_{\cdot,\cdot}\right)$, are computed using all the entries of the table.

Table 2 – Uncertainty table for a given year t.

		Model, j					
		1	2	…	J	Mean	Std. dev.
Alternative inputs, f	1	$OG_{1,1}$	$OG_{1,2}$	…	$OG_{1,J}$	$\overline{OG}_{1,\cdot}$	$\sigma(OG_{1,\cdot})$
	2	$OG_{2,1}$	$OG_{2,2}$	…	$OG_{2,J}$	$\overline{OG}_{2,\cdot}$	$\sigma(OG_{2,\cdot})$
	…	…	…	…	…	…	…
	F	$OG_{F,1}$	$OG_{F,2}$	…	$OG_{F,J}$	$\overline{OG}_{F,\cdot}$	$\sigma(OG_{F,\cdot})$
	Mean	$\overline{OG}_{\cdot,1}$	$\overline{OG}_{\cdot,2}$	…	$\overline{OG}_{\cdot,J}$	$\overline{OG}_{\cdot,\cdot}$	
	Std. dev.	$\sigma(OG_{\cdot,1})$	$\sigma(OG_{\cdot,2})$	…	$\sigma(OG_{\cdot,J})$		$\sigma(OG_{\cdot,\cdot})$

Source: the Network of EU IFIs.

Computing the table can be quite cumbersome in practice, especially if the model-based approach (a) is used because numerical methods (e.g. Monte Carlo) are required to implement it. Therefore, in a few examples, where the suite of models is used, the practical solution is to consider just the between uncertainty range that, without loss of generality, is computed for the first row of alternative inputs.

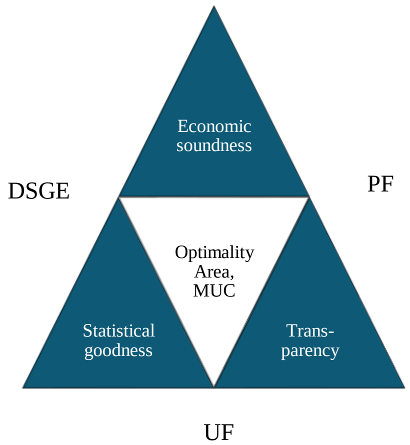
The first approach suggested in Murray (2014), Ódor and Kucserová (2014) is to consider just the arithmetic averages of the between uncertainty range ($\overline{OG}_{1,\cdot}$) for each given year. Another simple adjustment of the point estimates is to change the mean statistic into any other parametric or non-parametric estimate of the central point of the between uncertainty range. For example, such statistics could be a median, a mid-range – an arithmetic average of minima and maxima values of estimates produced under each method in each period (Casey 2018), a middle of the interquartile range, etc. The statistic suggested by Casey (2018) is useful for several reasons: (i) it is robust to overuse of a particular method (e.g. a production function approach as in Ódor and Kucserová (2014), or various modifications of the MUC models); (ii) it is simple to compute and (iii) it has an intuitive interpretation. On the other hand, mid-range estimates, as they are determined by estimates at the upper- and lower-ends of the range of models selected, may miss out on important dynamics of models inside of the range. Outliers may also have an undue bearing on mid-range estimates. However, neither a simple arithmetic mean, nor the mid-range statistic incorporates a model selection criteria (for instance, in terms of necessary and sufficient conditions for the "beauty contest" in **Table 1**) or some likelihood notions similar to Bayesian model averaging.

Ultimately, any model combination method will have advantages and disadvantages. It's important that IFIs keep some sight and expert judgement of the individual models as well, at least in their internal deliberations.

Nominating the winner

In the recent years, the view of different EU IFIs seems to converge that it is best to use many alternative methodologies jointly for comparison and assessment. This underlies, for example, the ‘Suite of Models’ approach by Casey (2018) or the ‘Estimate Combination Approach’ by Ódor and Kucserová (2014) or a recent schematic representation provided by Cuerpo, Cuevas and Quilis (2018). In the latter study, the authors classify the existing methods and their main challenges in the form of a trilemma: on the three sides of the triangle, they respectively set three limiting modelling cases, namely DSGE models, Univariate Filters and Production Function Methods. On the corners of a triangle in **Figure 9**, the authors set three optimality criteria: statistical goodness, economic soundness and transparency from a user perspective. An inner area of the triangle, called ‘optimality area’, represents modelling options that can combine the merits from all three aforementioned approaches. The authors propose a multivariate unobserved components (MUC) model described by (40) and (41) that, they believe, falls within this area. It serves as a computationally less complex alternative to the ‘suite of models’, where the specific theory is selected and validated through the proposed ‘beauty contest’ approach.

Figure 9 – Optimality necessary requirements



Source: Cuerpo, Cuevas and Quilis (2018)

Following Cuerpo, Cuevas and Quilis (2018), **Table 3** collects a number of studies that fall into the categories classified by these authors, adding some extra optimality criteria. Given the many variants of models falling in the same category and the country-specific context they are often applied to, it is difficult to reach quantitative results. Comparisons are therefore rather qualitative, summarizing a general view of the relative performance of each category across the various criteria.

Table 3 reveals that there is no uniformly the best method and none of the methods takes priority other the rest in all aspects. Indeed, different approaches or their groups employ explicitly or implicitly different assumptions, level of complexity, the stability of estimates and balance between theoretical and empirical adequacy. The trade-offs between Cuerpo, Cuevas and Quilis (2018) defined trilemma imply that a practitioner has to choose the subjective combination of features he would like a preferred method or the suite of methods to possess. The least demanding univariate methods are advantageous in terms of transparency but suffer in terms of end-point biases and lack of theoretical story behind the scenes. A richer story typically comes at the cost of increased complexity and not necessarily resolves the end-of-sample issues as in the case of production function approach applied following the CAM. Whilst statistical goodness in terms of lower revisions and better detection of essential growth cycle’s turning points comes at the cost of reduced transparency and theoretical adequacy. On top of that, a sufficient condition for the cycle is given by the ‘smell test’ as judged by the policy maker (Cuerpo, Cuevas and Quilis 2018).

Table 3 – Comparison of different estimation approaches.

	UF	MUF	PF	DSGE
Indicative Studies	Ódor and Kucserová (2014), Murray (2014), Casey (2018)	Frale and De Nardis (2017), Casey (2018), Cuerpo, Cuevas and Quilis (2018)	Havik, et al. (2014), Ódor and Kucserová (2014), Murray (2014), Casey (2018)	Vetlov, et al. (2011)
Approach	HP filter, PC filter, CF filter, BN filter, UC model and TC filter	Bivariate PC filter, MUC, PCA, a suite of models	Production function (the CAM)	General equilibrium model
Economic soundness				
Theory/empirical adequacy balance	Statistical	Statistical – theory (Okun's law, (N)KPC, absorption cycle, financial cycle, etc.)	Neoclassical supply-side driven growth theory – statistical	Micro-founded general equilibrium theory – statistical
Consistency with stylized facts	Low	Medium	Medium – high	High
Statistical goodness				
State-space representation	Yes, but for CF filter	Yes for MUC, no for PCA and suite of models	Yes (at components level)	No
Stability of Estimates	Low (HP, PC, end-point, omitted variable biases) – medium (UC)	Medium	Low (end-point, omitted variable biases)	–
Conformity with Business Cycles Chronologies	Low – medium	Medium	Low – medium	–
Transparency				
Computational complexity	Low (HP, PC, BP filters) – medium (BN filter, UC, TC filter)	Medium – high (a suite of models)	High	High
Analytical complexity	Low	Medium	High	High
Data requirements	Low	Medium – high (a suite of models)	High	High
Robustness to Pre-processing (seasonal adjustments, outliers)	Low	Medium due to structural constraints	Low at components level – medium at an aggregate level	High
Easy to include forecasted inputs	Yes	Yes	Yes	Yes

Source: the Network of EU IFIs.

To sum up, it seems prudent to consider various methods; and assess the robustness of inference to the chosen detrending method on a country-specific basis (Mazzi, Ozylidririm and Mitchell 2017). While timely decisions require policymakers to focus on a particular measure, the credibility of the decisions gains in solidity as different measures confirm the same message (Álvarez and Gómez-Loscos 2017). This is especially important in the context of the existing fiscal framework.

Concluding remarks

The objective of this paper was to review the conceptual and methodological issues of the potential output and the output gap estimation from the perspective of a fiscal authority and an independent fiscal institution.

Conceptually, the focal point of the potential output and the output gap estimation is twofold. First, we are interested in the growth cycle – the difference between actual output and potential output expressed as a per cent of potential output, which is then used to cyclically adjust various general government's fiscal indicators. Second, we are interested in long-term projections of potential output growth: an essential input into public debt sustainability analysis. This can also assess, for instance, the impacts of population ageing on the general government budget and on the real economy. In designing a suite of models, a practitioner must be aware of both goals. This calls for a consideration of structural and statistical methods as complementary tools rather than as substitutes in the "horse race" fashion.

The non-self-correcting Keynesian theory that allows for an active countercyclical policy intervention is a mainstay in terms of modern macroeconomics and the driving forces of growth cycles. Implicitly, the mere existence of IFIs as fiscal watchdogs implies a view in which fiscal issues (and fiscal policy) matters a lot. This view is close to a Keynesian perspective, in a broad sense. On the other hand, the production function approach follows the neoclassical growth theory grounded in the fiscal policy ineffectiveness and is primarily concerned about long-term growth. Fiscal authorities setting their budgetary plans are (or have to be) more concerned about sustainability. This means focusing on the medium- to long-term concept of fiscal space rather than elimination of short-run inflationary pressures. Blindly mimicking the central bank approaches may mislead fiscal authorities and IFIs in reaching their own objectives, not to mention the available range of fiscal policy instruments: tax rates, expenditure components (both as final consumption and productive investment) and various transfers.

The data used for output gap estimation can come from several sources. It may cover external, internal, fiscal, and financial variables that correspond to different sources of imbalances in the economy. A key concern of IFIs is an appropriate macroeconomic aggregate of actual output that is often country-specific. The prevalent real GDP could be changed to GNI or GVA net of particular subcomponents not related to the domestic cyclical fluctuations or that experience significantly lower cycles (oil production, multinationals, non-market sector, etc.).

The output gap estimates are surrounded by considerable uncertainty originating from both the observed data revisions and estimates of the unobserved potential output. The uncertainty stems from three main sources: model uncertainty both within a particular method and between 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.

Performance assessment is based on the trade-off between the stability of the real-time estimates in the sense of small ex-post revisions at the end-of-sample; and the plausibility of the estimates with hindsight. To maintain the credibility of the output gap estimates in the fiscal framework context, it is recommended to avoid hikes in the revisions of the output gap estimates, especially in the short-term horizon. This is why the end-point stability of the revisions has to be a primary concern for the EU countries. On top of that, IFIs have to judge the methods by considering the general plausibility of the output gap estimates – an intuitive "smell" test or country-specific narrative approach, keeping an eye on the descriptive statistics of their revisions. A comprehensive selection method is a "beauty contest" that detects an optimal model as the one balancing the set of the statistical-based necessary and the economically and policy-oriented sufficient conditions.

There are many different methods through which one can conduct a trend-cycle decomposition and none of them are necessarily superior to others. Trend-cycle decomposition methods can be grouped into several overlapping categories highlighting their particular features: (a) parametric, semi-parametric and non-parametric; (b) univariate and multivariate; (c) statistical, semi-structural and structural; (d) linear and non-linear. All these categories reflect particular trade-offs that a practitioner faces while selecting a country-specific method that wins a domestic "beauty contest" or a "horse race". Besides, in a diverse economic world, any commonly agreed approach eventually requires addressing country-specific issues (Ódor and Jurašekova Kucserová 2014). 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.

In solving the trilemma problem, an agreeable method should achieve three necessary conditions: economic soundness, statistical goodness and transparency. On top of this, a sufficient condition for the final estimate of the cycle is given by the smell test, often implemented by policymakers and a must-have for IFIs (Cuerpo, Cuevas and Quilis 2018). A practitioner has to choose the subjective combination of features he would like in a preferred method or suite of methods. The least demanding univariate methods are advantageous in terms of transparency but suffer in terms of end-point biases and a lack of theoretical narrative. However, more advanced methods do not necessarily resolve the end-of-sample problem. Since every cycle is different keeping analysis simple and with a clear narrative is problematic in a complex world. Designing a "least bad" solution among a host of mediocre choices might be the only realistic goal for the problem of estimating potential output (Blagrave, et al. 2015). The suite of models approach seems a reasonable response to these considerations.

There are several interesting directions for IFI practitioners to consider as takeaways for future research, following Ódor and Kucserová (2014).. First, one could explore an asymmetric loss function of fiscal authorities and IFIs, resulting from unequal costs of falsely launching correction mechanisms when these should not be activated (and not activating them when there is, in fact, a significant deviation). Second, one could consider more sophisticated ways to combine the estimates from the suite of models. Third, further research on how best to incorporate within model uncertainty could be explored, when producing uncertainty ranges. Fourth, the inclusion of other output gap estimation methods (wavelets, regime switching models, real business cycle and DSGE models) into the suite of models that are not currently applied by the Network of EU IFIs is worth exploring. Fifth, one could consider bias correction mechanisms that may follow from the analysis of one-sided and two-sided estimates. In the case of systematic bias, it is worth considering how this may create additional fiscal space to absorb negative surprises.

References

Álvarez, L.J., and A. Gómez-Loscos. "A menu on output gap estimation methods." *Bank of Spain, Working Paper No. 1720*, 2017.

Anderton, R., et al. "Potential output from Euro area perspective." *ECB Occasional Paper No. 156*, 2014.

Atanas, H., R. Raciborski, and V. Vandermeulen. "Assessment of the Plausibility of the Output Gap Estimates." *European Commission, European Economy Economic Brief 023*, 2017.

Axioglou, Ch. "The Commission's approach to assessing the output gap of the Greek economy: A Sensitivity Analysis." *Hellenic Fiscal Council Working Paper No. 3*, 2017.

Bailey, N., and A. Okun. *'The battle against unemployment and inflation: Problems of the modern economy.'.* New York:: W.W. Norton & Co, 1965.

Baxter, M., and R.G. King. "Measuring business cycles: Approximate bandpass filters." *The Review of Economics and Statistics No. 81(4)*, 1999.

Benetrix, A., and P.R. Lane. "Financial Cycles and Fiscal Cycles." *EUI-IMF conference on Fiscal Policy, Stabilization and Sustainability*, 2011.

Beveridge, S., and C. Nelson. "A new approach to the decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle." *Journal of Monetary Economics, Volume 7*, 1981.

Blagrave, P., R. Garcia-Saltos, D. Laxton, and F. Zhang. "A Simple Multivariate Filter for Estimating Potential Output." *IMF Working Paper 15/79*, 2015.

Blanchard, O.J., and L.H. Summers. "Hysteresis in unemployment." *European Economic Review vol. 31(1-2)*, 1987.

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.

Bouthevillain, C., et al. "Cyclically adjusted budget balances: an alternative approach." *ECB Working Paper No. 7*, 2001.

Burns, A.F., and W.C. Mitchel. *Measuring Business Cycles.* New York: NBER, 1946.

Butler, L. "A Semi-Structural Method to Estimate Potential Output: Combining Economic Theory with a Time-Series Filter. The Bank of Canada's New Quarterly Projection Model, Part 4." *Bank of Canada Technical Report No. 77*, 1996.

Casey, E. "Inside the "Upside Down": Estimating Ireland's Output Gap." *Irish Fiscal Advisory Council Report*, 2018.

Celov, D. "LEMPA - a Quarterly Lithuanian Economy Model for Projections and Analysis." *Bank of Lithuania, mimeo*, 2015.

Cerra, V., and S.C. Saxena. “Alternative Methods of Estimating Potential Output and the Output Gap: An Application to Sweden.” *IMF Working Paper 00/59*, 2000.

Christiano, L., and T.J. Fitzgerald. "The Band Pass Filter’,." *NBER Working Paper No. 7257*, 1999.

Clark, T. "The Cyclical Component of U.S. Economic Activity." *Quarterly Journal of Economics, vol. 102(4)*, 1987.

Constantinescu, M., and A.D.M. Nguyen. “Unemployment or Credit: Who Holds the Potential? Results from a Small-Open Economy.” *Bank of Lithuania, Discussion Paper Series No. 4*, 2017.

Cuerpo, C., Á. Cuevas, and E.M. Quilis. “Estimating Output Gap: a Beauty Contest Approach.” *AIReF Working Paper Number 18/02*, 2018.

Cuevas, A., and E. Quilis. "Non-Linear Modelling of the Spanish GDP." *AIReF Working Paper No. 17/01*, 2017.

Darvas, Z., and A. Simon. “Filling the gap: open economy considerations for more reliable potential output estimates.” *Bruegel Working Paper No. 2015/11*, 2015.

Dias, D.A., C.R. Marques, and Ch. Richmond. "Misallocation and Productivity in the Lead Up to the Eurozone Crisis." *Board of Governors of the Federal Reserve System, International Finance Discussion Papers No. 1146*, 2015.

Durbin, J., and S. J. M. Koopman. *Time Series Analysis by State Space Methods.* Oxford university press, 2012.

Fioramanti, M. "Potential output, output gap and fiscal stance: Is the EC estimation of the NAWRU too sensitive to be reliable?" *MPRA repository*, 2016.

Frale, C., and S. De Nardis. “Which gap? Alternative estimation of potential output and the output gap in the Italian economy.” *Parliamentary Budget Office Working Paper No. 2*, 2017.

Gali, J., and T. Monacelli. “Monetary policy and exchange rate volatility in a small open economy.” *Review of Economic Studies Vol. 72*, 2005.

Gamberoni, E., C. Giordano, and P. Lopez-Garcia. “Capital and labour (mis)allocation in the euro area: some stylized facts and determinants.” *ECB Working Paper No. 1981*, 2016.

Gómez, V. "The use of Butterworth filters for trend and cycle estimation in economic time series." *Journal of Business and Economic Statistics No. 19(3)*, 2001.

Gómez, V., and A. Maravall. “Programs TRAMO (Time Series Regression with Arima noise, Missing observations, and Outliers) and SEATS (Signal Extraction in Arima Time Series). Instruction for the User.” *Bank of Spain Working Paper No. 28*, 1996.

Hamilton, J.D. "Why You Should Never Use the Hodrick-Prescott Filter." *NBER Working Papers No. 23429*, 2017.

Harvey, A.C. "Trends and Cycles in Macroeconomic Time Series." *Journal of Business and Economic Statistics, vol. 3*, 1985.

Havik, K., et al. “The production function methodology for calculating potential growth rates & output gaps.” *Economic Papers 535, Directorate-General for Economic and Financial Affairs, European Commission*, 2014.

Hodrick, R.J., and E.C. Prescott. "Postwar U.S. Business Cycles: An Empirical Investigation." *Journal of Money, Credit, and Banking, vol. 29, No. 1*, 1997.

Holly, S., and K. Wade. "UK Exports of manufactures: the role of supply side factors." *Scottish Journal of Political Economy* 38, no. 1-18 (1991).

Huovari, J., S. Jauhiainen, and A. Kekäläine. "Assessment of the EC cyclical adjustment methodology for Finland – impact on budget balances." *PTT Working Papers 191*, 2017.

IHS Global Inc. *Eviews User Guide I.* 2016.

Jarociński, M., and M. Lenza. "An inflation-predicting measure of the output gap in the euro area." *ECB Working Paper No. 1966*, 2016.

Juglar, C. *Des Crises commerciales et leur retour periodique en France, en Angleterre, et aux Etats-Unis (Nouvelle édition Lyon : ENS Éditions, 2014).* Paris: Guillaumin, 1862.

Justiniano, A., G. Primiceri, and A. Tambalotti. "Is There a Trade-Off between Inflation and Output Stabilization?" *American Economic Journal: Macroeconomics, American Economic Association, vol. 5(2)*, 2013.

Kalman, R. "A New Approach to Linear Filtering and Prediction Problems." *Journal of Basic Engineering Vol. 82 (1)*, 1960.

Kamber, G., J. Morley, and B. Wong. "Intuitive and Reliable Estimates of the Output Gap from a Beveridge-Nelson Filter." *The Review of Economics and Statistics*, 2017.

Keynes, J.M. *The General Theory of Employment, Interest and Money.* London: Macmillan, 1936.

King, R.G., and S. Rebelo. "Low Frequency Filtering and Real Business-Cycles." *Journal of Economic Dynamics and Control No. 17*, 1993.

Koopmans, L.H. *The Spectral Analysis of Time Series.* New: Academic Press, 1974.

Laxton, D., and R. Tetlow. "Simple Multivariate Filter for the Measurement of Potential Output." *Bank of Canada*, 1992.

Laxton, D., P. Isard, H. Faruquee, E. Prasad, and B. Turtelboom. "Multimod Mark III. The Core dynamic and steady-state models." *IMF Occasional Paper No 164*, 1998.

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.

Lunsing, H. "Potentiële groei en evenwichtswerkloosheid." *CPB Achtergronddocument*, 2011.

Maravall, A., and A. del Rio. "Temporal aggregation, systematic sampling, and the Hodrick-Prescott filter." *Coputational Statistics & Data Analysis, Elsevier, vol. 52(2)*, 2007.

Martti, H. "The grounds for the euro crisis and the future of the euro-zone (in Finnish)." *Finnish Economic Papers No. 1*, 2015.

Mazzi, G.L., A. Ozylidririm, and J. Mitchell. "Chapter 9. Alternative Detrending Methods." In *Handbook on Cyclical Composite Indicators: for Business Cycle Analysis (2017 edition)*, by Eurostat. Publications Office of the European Union, 2017.

Mazzi, G.L., and A. Ozylidririm. “Chapter 2. Business Cycles Theories: An Historical Overview.” In *Handbook on Cyclical Composite Indicators: for Business Cycle Analysis (2017 edition)*, by Eurostat. Publications Office of the European Union, 2017.

Mc Morrow, K., W. Roeger, V. Vandermeulen, and K. Havik. "An assessment of the relative quality of the Output Gap estimates produced by the EU's Production Function Methodology." *European Economy Discussion Papers, Directorate General Economic and Financial Affairs (DG ECFIN), European Commission*, 2015.

Melolinna, M., and M. Tóth. "Output gaps, inflation and financial cycles in the United Kingdom." *Bank of England Staff Working Paper No. 585*, 2016.

Miller, E. "An Assessment of CES and Cobb-Douglas Production Functions." *CBO Working Paper No. 5*, 2008.

Mintz, I. “Dating postwar business cycles: Methods and their application to western Germany, 1950-67.” *NBER Occasional Paper No. 107*, 1969.

Moauro, F. “Chapter 11. The Stock and Watson Approach to Composite Indexes of CLIs.” In *Handbook on Cyclical Composite Indicators: for Business Cycle Analysis (2017 edition)*, by Eurostat. Publications Office of the European Union, 2017.

Mohr, M. "A Trend-Cycle(-Season) filter." *ECB Working Paper No. 499*, 2005.

Morley, J.C., C.R. Nelson, and E. Zivot. "Why Are the Beveridge-Nelson and Unobserved Components Decompositions of GDP so Different?" *Review of Economics and Statistics No. 85*, 2003.

Murray, J. “Output gap measurement: judgement and uncertainty.” *Office for Budget Responsibility Working Paper No. 5*, 2014.

Nilsson, R., and G. Gyoma. "Cycle Extraction: A Comparison of the Phase-Average Trend Method, the Hodrick-Prescott and Christiano-Fitzgerald Filters." *OECD Statistics Working Papers No. 2011/04*, 2011.

Ódor, L., and J. Jurašekova Kucserová. “Finding Yeti: More Robust Estimates of Output Gap in Slovakia.” *Council for Budget Responsibility Working Paper No. 2*, 2014.

OECD. *Measuring Capital.* OECD, 2nd edition, 2009.

—. *Measuring Productivity: Measurement of Aggregate and Industry-Level Productivity Growth.* OECD, 2001.

Okun, A.M. *Potential GNP: Its Measurement and Significance.* Proceedings of the Business and Economic Statistics Section of the American Statistical Association, 1964.

Pedersen, T.M. "Alternative Linear and Non-Linear Detrending Techniques: A Comparative analysis based on Euro-Zone Data." *Copenhagen: Ministry of Economic and Business Affairs*, November 2002.

Pybus, T. "Estimating the UK’s Historical Output Gap." *Office for Budget Responsibility, Working Paper No. 1*, 2011.

Pollock, D.S.G. "Trend estimation and de-trending via rational square-wave filters." *Journal of Econometrics*, 1999.

Ravn, M., and H. Uhlig. "On adjusting the HP-filter for the frequency of observations." *Review of Economics and Statistics No. 84*, 2002.

St-Amant, P., and S. Van Norden. "Measurement of the Output Gap: A Discussion of Recent Research at the Bank of Canada." *Bank of Canada Technical Report No. 79*, 1997.

Turner, D, M.Ch. Cavalleri, Y. Guillemette, A. Kopoin, P. Ollivaud, and E. Rusticelli. “An investigation into improving the real-time reliability of OECD output gap estimates.” *OECD Working Paper No. 1294*, 2016.

Van den Brakel, J., et al. “Chapter 4. Data Availability, Frequency and Adjustment Techniques.” In *Handbook on Cyclical Composite Indicators: for Business Cycle Analysis (2017 edition)*, by Eurostat. 2017.

Vetlov, I., T. Hl'edik, M. Jonsson, H. Kucsera, and M. Pisani. *Potential output in DSGE models.* ECB Working Paper Nr. 1351, 2011.

Willman, A. "Euro area production function and potential output: a supply-side system approach." *ECB Working Paper No. 153*, 2002.

Annexes

Appendix A. Data collection

Method	Data required	Category	Notes
Univariate filters	GDP / GVA data	GDP and output	
	Domestic GVA	GDP and output	Other similar measures could be used in cases where a long time-series is not available
	GVA excluding volatile sectors	GDP and output	In the UK output gap models are currently based on a measure of GVA excluding the volatile North Sea oil sector
Multivariate filters	Real credit growth	Financial markets and asset prices	
	Real interest rates	Financial markets and asset prices	
	Real effective exchange rate	Financial markets and asset prices	
	House prices	Financial markets and asset prices	
	Housing completions	GDP and output	
	Inflation	Prices	
	Domestically generated inflation	Prices	
	Inflation expectations	Prices	Useful if wanted to estimate Phillips curve as part of this work
	Unemployment rate	Labour market	
	Current account	GDP and output	
	Asset/commodity prices	Financial markets and asset prices	Oil prices in resource-dependent countries, or asset prices in economies with large financial sectors
	Money supply	Financial markets and asset prices	
	Gross national savings	GDP and output	
	Public borrowing and debt	Fiscal	Fiscal variables could be useful in determining which version of the output gap is most important for the public finances
	Taxes and	Fiscal	

Method	Data required	Category	Notes
	benefits		
	Investment in equipment / machines	GDP and output	
	Investment in construction	GDP and output	
Production function	Net capital stock	GDP and output	
	Net capital stock excluding dwellings	GDP and output	In the UK dwellings are excluded: questionable whether it should be counted as productive capital, but also the issue raised in Casey (2018) about house price bubbles distorting the levels
	Trend employment	Labour market	Could be obtained by a combination of trend participation and an estimate of the NAIRU, or directly, e.g. by applying a filter to the actual employment rate
	Trend average hours	Labour market	
	Net migration	Labour market	Just adding in case we wanted to look into the extent to which changes in migration flows affect some of these models.
Principal components analysis	Capacity utilisation measures	GDP and output	
	Recruitment difficulties indicators	Labour market	
	Earnings growth	Labour market	
	New car registrations	GDP and output	
Sector-weighted cyclical indicators analysis	Sectoral shares	GDP and output	Used to in some models to weight the different sectors when forming a view of the output gap for the whole economy

Appendix B. Kalman filter and Kalman smoother

Filtering is one of the methods applied to assess the cyclical information from actual data – it directly extracts the cyclical component of the series by removal of trends and isolation of other recurring cycles.[23] In particular, state-space models and Kalman filter techniques play a special role in trend-cycle decompositions. In these approaches, the output gap is treated as a *latent variable* (a variable that can be estimated but not observed) and many of them use the unobserved components model following Harvey (1985). Similar to (1) consider a generic decomposition that represents the dependent time-series variable y_t into a set of the unobserved components in the (log-) additive form:

$$y_t = y_t^* + c_t + \gamma_t + \nu_t + \varepsilon_t, \; \forall t = 1, \dots, T, \qquad (57)$$

where y_t^* is a slowly-varying unobserved component (trend); c_t is a periodically-recurring unobserved component (cycle); γ_t denotes a periodic unobserved component (season); ν_t is an unobserved autoregressive component (inertia); and ε_t is an unobserved irregular component (disturbance).

In a structural time-series (STS) or unobserved component (UC) model, all right-hand side components are modelled exceptionally as stochastic unobserved processes having a direct (semi-structural) interpretation. By their nature, following BN decomposition, these components can be further split into a deterministic function of time (e. g., polynomials with dummy variables) or stochastic processes. Each component of the UC model given by the equation is modelled in state-space form and estimated using Kalman filter. For instance, the trend component,y_t^* , can be modelled either deterministically as
$y_t^* = \tau \cdot t + \varepsilon_t$, with $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$, or stochastically by random walk plus noise, giving rise to the so-called local level or random walk with noise model (Rummel, 2015).

Stochastic filters may be best used within a production function framework, rather than as stand-alone. Positive examples of how to design flexible economic approaches to potential output are:

1) Estimating time varying NAIRUs through Kalman filters;
2) Filtering total factor productivity and participation rates.

Filtering such individual components of potential output should help avoid the kind of intense endpoint biases observed when applied directly to ‘politically charged’ GDP series.

A wide range of time series models, including the classical linear regression model and ARIMA models, can be written and estimated as special cases of a state-space specification. The Kalman filter algorithm has been used, among other things, to compute exact, finite sample forecasts for Gaussian ARMA models, multivariate (vector) ARMA models, MIMIC (multiple indicators and multiple causes), and time-varying (random) coefficient models.

The meaning of filter is to transform a function of a time series into another series. This representation of the model is also known as a state-space system, with the first equation representing the signal equation (the equation of the observable variables) and the second representing the state equation (the equation of the unobservable variables). The Kalman filter (Kalman 1960) is an algorithm for generating minimum mean squared error forecasts in a state-space model. It is a recursive algorithm

[23] Even after the separation of trend from cycle is accomplished, additional steps may be necessary to isolate cycles according to the frequency of their recurrence (e.g., patterns of fluctuation that recur at classical-cycle frequencies and other) (Rummel, 2015).

for sequentially updating the *one-step-ahead* estimate of the state mean and variance are given the new information.

Consider the MUC represented in the multivariate Gaussian form (40) and (41) based on Durbin and Koopman (2012). Recall, that:

Signal: $y_t = H(\alpha)S_t + v_t, v_t \sim NID(0, \Sigma_v)$,

State: $S_t = F(\phi)S_{t-1} + u_t, u_t \sim NID(0, \Sigma_u), S_1 \sim N(a_1, P_1)$,

where $t = 1, 2, \ldots, T$ denotes time, y_t is a $p \times 1$ vector of p *observed* endogenous variables and is called the *observation* (measurement or signal) vector; S_t is a $m \times 1$ vector of m *unobserved* states and is called the *state* (transition) vector, the dynamics of which is represented as a first-order autoregression. The disturbance vectors v_t and u_t are assumed to be serially independent, with contemporaneous variance structure.

Let us stack all model parameters $\alpha, \phi, \Sigma_v, \Sigma_u$ into one vector θ. Then, conditional on a particular realisation of the initial state vector S_1 and assuming that the values of vector θ are known, the Kalman filter can be used to estimate the state vector and its corresponding standard error.

In practice, the vector θ is unknown and must be estimated from the sample. Fortunately, the state space format and the Kalman filter provide a feasible way to evaluate the likelihood function and, using numerical methods, to maximize it. The parameters θ could be also assessed by applying Bayesian methods and/or calibrated.

Once the θ parameters have been estimated, the Kalman filter is run to derive new initial conditions by means of backcasting (i.e., forecasting observations prior to the first observation). This process of backcasting can be done just by projecting forward the model using the reversed time series. In this way, a new set of initial conditions exerting a limited influence on the estimation of the state vector is derived by means of the Kalman filter. The complete algorithm can be stated as follows.

- *Initialization 1:* Set initial parameters: θ_0.
- *Initialization 2:* Set initial conditions: $S_{1,0}$. Initial conditions for the state vector are provided using a diffuse prior centred on zero with an arbitrarily large VCV matrix.
- *Likelihood computation:* Conditioned on the initial parameters and the initial conditions, we run the Kalman filter to compute the likelihood.
- *Likelihood maximization:* The maximum likelihood estimation (MLE) is implemented. The definition of the objective function incorporates the constraints that ensure the non-negativity of the variances and the stationary nature of the AR(2) parameters for the stochastic cycle.
- *Re-initialization:* The use of diffuse initial conditions to run the Kalman filtering is a simple device to start its algorithm but may generate some sensitivity in the estimates of the state vector. To reduce the sensitivity for these estimates, first generate backcasts (e.g., forecasts of observations prior to the first observation). This process of backcasting is done just by projecting forward the model using the reversed time series. In this way, a new set of initial conditions $S_{1,1}$ is obtained that exerts a limited influence on the estimation of the state vector as derived by means of the Kalman filter.

A Kalman filter produces one-sided or concurrent estimates of the state vector, while a Kalman smoother is associated with two-sided or historic estimates of the state vector.

One-sided (concurrent) estimates of the state vector

The one-sided (or concurrent) estimates of the state vector are obtained running recursively the Kalman filter from $t = 1, 2, \ldots, T$ one-step forward in time. To estimate the state vector at time $t = h$ the method considers only the information available from $t = 1$ to $t = h$ and is very useful to analyse the state of the system on a real-time basis.

In the nutshell, using the observed signal variables and some initial assumptions about state mean and variance values, the Kalman filter first calculates one-step-ahead estimates of state values $S_{t|t-1}$ and variances $P_{t|t-1}$:

$$S_{t|t-1} = E_{t-1}(S_t), \tag{58}$$

$$P_{t|t-1} = E_{t-1}\left[\left(S_t - S_{t|t-1}\right)\left(S_t - S_{t|t-1}\right)'\right], \tag{59}$$

that gives an introductory (*a priori*) projection of the state variable and variances, when y_t is not known. The observable data for the next period is then used to update the projections (*a posteriori*) from the first step, giving more weight to components with lower variances and considering y_t is observed. This *a posteriori* estimate is a linear combination of an a priori estimate of the state and the measurement residual (innovation) from the corresponding signal equation:

$$\tilde{S}_t = S_{t|t-1} + K_t\left(y_t - H(\alpha)S_{t|t-1}\right), \tag{60}$$

where K_t is a gain (blending factor) that minimizes a posteriori error covariance:

$$\tilde{P}_t = E_t\left[\left(S_t - \tilde{S}_t\right)\left(S_t - \tilde{S}_t\right)'\right] \Rightarrow K_t = P_{t|t-1}H(\alpha)'\left(H(\alpha)P_{t|t-1}H(\alpha)' + \Sigma_u\right)^{-1}, \tag{61}$$

from which follows that a posteriori update is heavier the smaller the variance of signal equations Σ_u and lighter when the $P_{t|t-1}$ approaches zero. The *a posteriori* estimates are then used to compute the *a priori* projections for the next period:

$$S_{t+1|t} = F(\phi)\tilde{S}_t, \tag{62}$$

$$P_{t|t-1} = F(\phi)\tilde{P}_t F(\phi)' + \Sigma_v. \tag{63}$$

And the process is recursively iterated until the end-of-sample T is reached.

Two-sided (historical, smoothed) estimates of the state vector

The Kalman filtered estimates of the past data can be updated if new data becomes available (Van den Brakel, et al. 2017). This procedure is referred to as Kalman smoothing. In addition to one-sided Kalman forward filtering step, the two-sided (or historical) estimates of the state vector are obtained running recursively the Kalman filter from $t = T$ to $t = 1$ (backward in time), using as initial conditions the terminal concurrent estimates obtained in the previous step. This process considers all the information available from $t = 1$ to $t = T$ to estimate the state vector at any time $t = h, 1 \leq h \leq T$. In other words, the smoothed estimate for the state vector for period h also accounts for the future information after time period h.

Smoothing uses all of the information in the historical time series to provide smoothed estimates of the states and smoothed estimates of the state variances. From an econometric view, one-sided and two-sided estimates play a complementary role. The first one serves as the starting point for the second and provides a benchmark to quantify the additional precision that the full sample introduces. Note that two-sided estimates are more precise because they incorporate all the available information from $t = 1$ up to time $t = T$ to estimate the state vector in any intermediate point and, due to their symmetric nature. Note that this symmetry is due to the fact that the filter runs backwards from estimates derived forward. In this way, two-sided filtering does not introduce any form of phase-shift in the estimates.

However, a two-sided estimate is not useful for real-time analysis since it incorporates information not available a $t = h$ to evaluate the state of the system at that time and hence introduces some form of hindsight bias – a direct consequence of the fact that the ends-of-sample however the one-sided and two-sided estimates are identical. This is particularly important when dealing with output gap estimation because its main use is related to the assessment of the fiscal policy stance. In practice, fiscal policy at time t is primarily determined using only information available up to time t. Therefore, in a real-time application, it is sufficient to run Kalman filter only.

Of course, this pre-eminence does not imply that two-sided estimates are irrelevant. Quite the contrary, they serve to produce useful measures of uncertainty and to gauge the impact of the full sample on the estimates of the output gap, especially around the turning points.