OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 0305-9049
doi: 10.1111/obes.12524

Norges Bank Output Gap Estimates: Forecasting Properties, Reliability, Cyclical Sensitivity and Hysteresis

FRANCESCO FURLANETTO,†,‡ KÅRE HAGELUND,† FRANK HANSEN†,§ and ØRJAN ROBSTAD†

†Norges Bank, Bankplassen 2, PB 1179 Sentrum Oslo, 0107, Norway (e-mail: francesco.furlanetto@norges-bank.no)
‡BI Norwegian Business School, Nydalsveien 37, Oslo, 0484, Norway (e-mail: kare.hagelund@norges-bank.no)
§Gard AS, Kittelsbuktveien 31, Arendal, 4836, Norway (e-mail: frank.hansen@gard.no)

Abstract

This paper documents the suite of models (SoMs) used by Norges Bank to estimate the output gap. The models are estimated using data on GDP, unemployment, inflation, wages, investment, house prices and credit. We evaluate the estimated output gap series in terms of its forecasting properties, its reliability and its cyclical sensitivity to various measures of demand and supply shocks. A simple equally weighted average of estimates from different models features a better forecasting performance than each individual model. In addition, it helps predicting inflation in pseudo real time and exhibits limited variations when new data become available. The summary measure of potential output responds strongly and rapidly to permanent shocks and to narrative measures of technology shocks but, although to a more limited extent, also to demand shocks, thus partly capturing hysteresis effects.

I. Introduction

The output gap, defined as the difference between actual output and its potential level, is a key variable for central banks, in particular for those operating in a flexible inflation targeting regime, such as the Central Bank of Norway (Norges Bank henceforth). In fact, having an accurate measure of the state of the business cycle is crucial to appropriately evaluate trade-offs between different objectives and to fulfil the central bank's mandate.

JEL Classification numbers: C38, E17, E32.
This paper should not be reported as representing the views of Norges Bank. The views expressed are those of the authors and do not necessarily reflect those of Norges Bank. We would like to thank the editor Francesco Zanetti, two anonymous referees and one anonymous referee for the Norges Bank working paper series for constructive comments. We also thank Knut Are Aastveit, Ida Wolden Bache, Paul Beaudry, Julien Champagne, Tord Krogh, Per Espen Lilleås, Elmar Mertens, Alexander Rathke, Samad Sarferaz, Rodrigo Sekkel, Ingrid Solberg and seminar participants at KOF ETH in Zurich, CEF 2019 Conference in Ottawa, SNDE 2019 conference at Dallas Fed, CIRET/KOF/OECD/INSEE 2019 Workshop in Paris and Forecasting at Central Banks 2019 Conference at Bank of Canada.

Norges Bank bases its assessment of the output gap on a broad set of indicators and models that are revised and expanded over time. The objective of this paper is to document the suite of models (SoMs) used as a key input to produce output gap estimates at Norges Bank. In particular, we evaluate its forecasting performance and its reliability when new data become available. In addition, we also provide a novel analysis on the cyclical sensitivity of the output gap to various measures of supply and demand shocks identified with different methods, along the lines of Coibion, Gorodnichenko and Ulate (2018).

The output gap is not observable and its measurement is obviously challenging. Traditionally, simple univariate statistical filters, such as the Hodrick–Prescott (HP) filter, have been prominent. With sufficient information about the future, these simple methods provide a good measure of capacity utilization. Unfortunately, as emphasized in the seminal paper by Orphanides and van Norden (2002), they are less reliable in real time as they are often subject to substantial revisions. This is primarily because trend extraction becomes problematic at the end of a given sample period (see Hamilton (2018) for a comprehensive discussion of the HP filter and Canova (2020) for a broader discussion on the properties of several filters and their ability to recover the output gap generated by structural macroeconomic models). Despite these well-known issues, according to Coibion *et al.* (2018) simple univariate statistical filters still seem to play a central role in the production of output gap estimates at several policy institutions.

In order to estimate the output gap, Norges Bank uses multivariate models, which, in addition to GDP data, use data on other variables such as unemployment, wages, inflation, investment, credit growth and house price growth. Over time, a suite of multivariate models composed of unobserved component (UC) models and structural vector-autoregressive (SVAR) models has been developed. A simple equally weighted average of the estimates obtained from the different models included in the suite tracks well the official Norges Bank's output gap series, which is a judgemental measure representing the institutional view on capacity utilization in the Norwegian economy. Norges Bank's approach builds on Fleischman and Roberts (2011), Basistha and Startz (2008) and Jarocinski and Lenza (2018), among others, who find that using multivariate models is key to obtaining more precise and reliable estimates of the output gap for the United States and the euro area. The use of data on inflation in combination with GDP follows Kuttner (1994), who initiated the literature on the output gap and Phillips curve estimation in UC models. Clark (1989), Apel and Jansson (1999) and Sinclair (2009) model the relationship between the output gap and unemployment according to Okun's law. The use of data on investment follows Domenech and Gomez (2006), who emphasize the importance of the investment share in the economy as a key variable for identifying cyclical developments. Furthermore, Borio, Disyatat and Juselius (2017) argue that it is important to take into account data on financial variables when assessing potential output, a point also stressed by Furlanetto, Gelain and Taheri Sanjani (2021a) in the context of an estimated macroeconomic model with financial frictions and by Berger, Richter and Wong (2022b) in a Bayesian Vector Autoregression.

We show that the model suite delivers estimates of the output gap that perform relatively well in terms of forecasting and reliability when new data become available. Notably, a simple equally weighted average of the nine output gap estimates features better forecasting properties than the individual models. In addition, such a summary measure of the output gap provides useful information for forecasting purposes. First, a model using the pseudo real-time estimate of the output gap performs better at forecasting inflation than a univariate AR model in a direct forecasting exercise as in Orphanides and van Norden (2005). Second, when we compare the forecasting performance of bivariate models including the output gap estimate with simple univariate models for inflation, the bivariate models obtain on average a better performance. Significant gains are found, in particular at long horizons. Furthermore, we document that the summary measure of the output gap exhibits limited variation when new data become available, in particular when compared with an estimate obtained with a simple univariate HP filter. Therefore, we confirm the finding in Fleischman and Roberts (2011), that using data on other variables (and on unemployment in particular) is useful to obtain reliable estimates of the output gap for the US, also hold for Norway.

We evaluate the summary measure of the output gap also along a third, and more novel, dimension, which is its cyclical sensitivity to various measures of economic shocks. Coibion *et al.* (2018) show that estimates of potential output from the Federal Reserve Board, the International Monetary Fund (IMF) and the Organization for Economic Cooperation and Development (OECD) for various countries over-respond to shocks that have no long-run effects on the economy and under-respond to shocks that do have a long-run impact. More specifically, real-time estimates of potential output respond to transitory shocks obtained from a Blanchard–Quah decomposition and to various narrative measures of monetary and fiscal shocks. In contrast, real-time potential output estimates adjust only gradually to permanent shocks obtained from Blanchard–Quah decompositions and from narrative measures of supply shocks. Coibion *et al.* (2018) rationalize these results by showing that impulse responses of estimates of potential GDP from all organizations are nearly indistinguishable from the responses of a one-sided HP filter applied to real-time GDP data. We document that this is not the case for the summary measure of potential output obtained from Norges Bank's SoMs in pseudo real time. In fact, the summary measure responds as much as actual GDP (and substantially more than an HP filtered GDP series) to shocks with permanent effects obtained from a Blanchard-Quah decomposition and to shocks to a cyclically adjusted measure of total factor productivity (TFP) for Norway. Moreover, the summary measure of potential output responds less than actual output but more than an HP filtered GDP series to shocks with transitory effects obtained from a Blanchard–Quah decomposition and to monetary policy shocks derived using high frequency identification. All in all, our analysis shows that the summary measure of potential output responds rapidly and significantly to supply shocks but also to demand shocks, although to a more limited extent.

Finally, given its responsiveness to demand shocks, we study whether the summary measure of potential output captures the presence of hysteresis. We evaluate it against an SVAR model that allows for demand shocks with potentially permanent effects, thus generating hysteresis effects as in Furlanetto *et al.* (2021b). We find that the role of these shocks is non-negligible but not major. The summary measure of the output gap generated by the model suite aligns relatively well with our favourite measure of the output gap in the model with hysteresis, which, however, is more cyclical and features larger negative values in recessions.

Our paper integrates and extends previous analysis on Norwegian data by Bernhardsen *et al.* (2005), Bjørnland, Brubakk and Jore (2008) and Bjørnland *et al.* (2012), and contributes to a recent literature whose aim is to document and evaluate the performance of output gap estimates produced routinely at central banks. Two prominent examples are Edge and Rudd (2016) and Champagne, Poulin-Bellisle and Sekkel (2018) for the Board of Governors of the Federal Reserve and the Bank of Canada respectively. Edge and Rudd (2016) show that the real-time issues emphasized by Orphanides and van Norden (2002) become less severe when using a longer sample period. In particular, they find that revisions have become smaller since the mid-1990s and that there is no deterioration in forecast performance when inflation projections are conditioned on real time rather than on final estimates of the output gap. Champagne *et al.* (2018) present a similar analysis for Canada, but also include the Great Recession in their sample, arguably a period in which the measurement of potential output was particularly challenging. They also find that the staff gap estimates are subject to smaller revisions. In addition, they show that models that condition on the real-time estimates of the output gap perform slightly worse than those conditioned on the final estimate (although none of the models using output gap estimates outperform simple univariate models). It is important to stress that both Edge and Rudd (2016) and Champagne *et al.* (2018) evaluate the revision and forecasting properties of the *official estimates* of the output gap produced at the Fed and Bank of Canada, respectively. In contrast, Norges Bank, by choice, rarely revises historical estimates of the output gap further back than the current business cycle. This is because the purpose of the Norges Bank's official output gap estimate is also to reflect the view of the economy that was the basis for the monetary policy decisions at the time. Therefore, we focus our attention on *model-based* estimates that constitute a key input (combined with judgement and insights obtained from other models of the labour market) in the production of the official estimate, at least in recent years. Our distinctive contribution consists in investigating the cyclical sensitivity of these estimates. As far as we know, no other central bank has proposed an evaluation along this dimension. In addition, the evaluation of the summary measure of the output gap with respect to an estimate that accounts for hysteresis effects is also novel.

The rest of the paper is structured as follows. Section II presents the different models included in the suite. Section III evaluates the output gap estimates derived from the different models in the suite and a simple summary measure. Section IV discusses how the presence of hysteresis effects impacts the measurement of the output gap. Finally, section V concludes.

II. The suite of models

To estimate the output gap, Norges Bank relies on SoMs composed by several UC models and two SVAR models. Thus, all models in the suite belong to the 'aggregate (statistical) approach', following the classification used in Mishkin (2007).[1] The different models are described in the rest of this section.

[1]Models based on the 'production function approach' have been tested, but not implemented in the suite, being too data intensive and not sufficiently reliable in pseudo real time. Estimates derived from dynamic stochastic general equilibrium models ('DSGE approach') have also been produced. However, these estimates turned out to be

UC models

UC models are a key component of the suite. A UC model posits that GDP (y_t) can be decomposed into an output gap ($\hat{y}_t$) and potential GDP ($\bar{y}_t$), which are both unobservable:

$$y_t = \hat{y}_t + \bar{y}_t. \tag{1}$$

In addition, the model specifies how the unobserved variables evolve over time:

$$\hat{y}_t = \lambda_y \hat{y}_{t-1} + \epsilon_t, \tag{2}$$

$$\Delta\bar{y}_t = G_t + \eta_t. \tag{3}$$

The output gap (equation (2)) is modelled as an AR(1) process and thus depends on the output gap in the previous period and on a white noise shock (ϵ_t). The change in potential GDP (equation (3)) depends on potential growth (G_t) and on a white noise shock (η_t). Potential growth is also allowed to vary over time:

$$G_t = C_G + \lambda_G (G_{t-1} - C_G) + \psi_t. \tag{4}$$

In equation (4), ψ_t represents a white noise shock to potential growth whose persistence is governed by the parameter λ_G while C_G is a constant. When λ_G is set equal to 1, the process follows a random walk. The traditional HP filter measure of the output gap can be derived as a special case of the model above when λ_G is equal to 1 and the relative variances of ϵ_t and ψ_t are restricted appropriately (see Harvey and Trimbur, 2008). However, as discussed in the introduction, a more recent literature has shown that developments in variables other than GDP are important to obtain reliable estimates of the output gap. Therefore, seven multivariate UC models are included in the SoMs. In all models we use data on GDP for mainland Norway (thus excluding petroleum and ocean transport activities, as is standard in macroeconomic analysis for Norway). The models can be described as follows:

1. UC 1 uses annual data on GDP for mainland Norway, real wage growth as reported by Statistics Norway (Wages/CPI adjusted for taxes and energy) and registered unemployment (Norwegian Labour and Welfare Administration, NAV)) and is estimated over the period 1990–2019.
2. UC 2 uses annual data on GDP for mainland Norway, real wage growth and unemployment (Labour Force Survey, LFS) and is estimated over the period 1990–2019.
3. UC 3 uses annual data on GDP for mainland Norway, real wage growth, registered unemployment (NAV) and business investment as a percentage of GDP for mainland Norway and is estimated over the period 1990–2019.
4. UC 4 uses quarterly data on GDP for mainland Norway and change in domestic inflation and is estimated over the period 1990:Q1–2019:Q2.

extremely model dependent and exhibited an important low-frequency component in many instances, as shown in Furlanetto *et al.* (2021a). In addition, one important issue is that potential output responds more than actual output to technology and labour supply shocks, both important drivers of output fluctuations in DSGE models (cf. Bilbiie and Melitz, 2020).

5 UC 5 uses quarterly data on GDP for mainland Norway, registered unemployment (NAV) and domestic inflation and is estimated over the period 1990:Q1–2019:Q2.
6 UC 6 uses quarterly data on GDP for mainland Norway and four-quarter growth in total domestic credit and is estimated over the period 1990:Q1–2019:Q2.
7 UC 7 uses quarterly data on GDP for mainland Norway and four-quarter growth in house prices and is estimated over the period 1990:Q1–2019:Q2.

The models differ in terms of estimation frequency, data series, estimation period and modelling of potential output. Since the official quarterly series for wages is extremely noisy, the first three models in the suite are estimated using annual data. In a centralized wage negotiation system in which wages are changed once per year in a synchronized way as in Norway, the relevant wage measure is annual. Missing data within a year are extrapolated using simple AR forecasts, meaning that quarterly data are used to forecast the rest of the year and construct forecast for the annual observable variables. This forecast for the current year is treated as an observable in the estimation and filtering of the annual models. The estimates from the models using annual data are converted into quarterly data using the Denton algorithm (Denton, 1971). All models are estimated using Bayesian methods as in (Planas, Rossi and Fiorentini, 2008) and are described more in detail in the Appendix. The point estimates are obtained using the mode of the posterior distribution in the various models. For illustrative purposes, we focus now on UC 1. A Phillips curve for real wages and a relationship between unemployment and the output gap are specified (Okun's law):

$$\hat{W}_t = \lambda_W \hat{W}_{t-1} + \gamma \hat{y}_t + \upsilon_t, \tag{5}$$

$$\hat{u}_t = \lambda_u \hat{u}_{t-1} + \beta \hat{y}_t + \omega_t. \tag{6}$$

Both the wage gap ($\hat{W}_t$) and the unemployment gap ($\hat{u}_t$) are related to the output gap ($\hat{y}_t$).[2]

Figure 1 presents the output gap estimate obtained from model UC 1 (cf. yellow solid line). The blue solid line refers to an estimate of the output gap based on a univariate version of the model (i.e. when only the equations (1) to (4) are used). The dashed lines plot the evolution of unemployment and real wage growth over the sample period. Shaded areas indicate recessions as defined in Aastveit, Jore and Ravazzolo (2016). At the beginning of the 1990s, unemployment was high and real wage growth was low. When these data are taken into account in the estimation, the output gap is shown to be more negative at the beginning of the 1990s compared with an estimate that relies only on GDP data. The multivariate model also indicates higher capacity utilization at the end of the 1990s, while the estimation of the output gap in the period preceding the financial crisis is approximately the same. In the years following the financial crisis, the multivariate model indicates somewhat higher capacity utilization, both as a result of higher real wage growth and lower unemployment. In the wake of the fall in oil prices in 2014, the model indicates somewhat lower capacity utilization compared with the univariate model, primarily due to a sharp fall in real wage growth.

[2]A time-varying trend for real wages and unemployment are estimated, following the same specification used for GDP (see equation (3)). The full model specification is described in the Appendix.

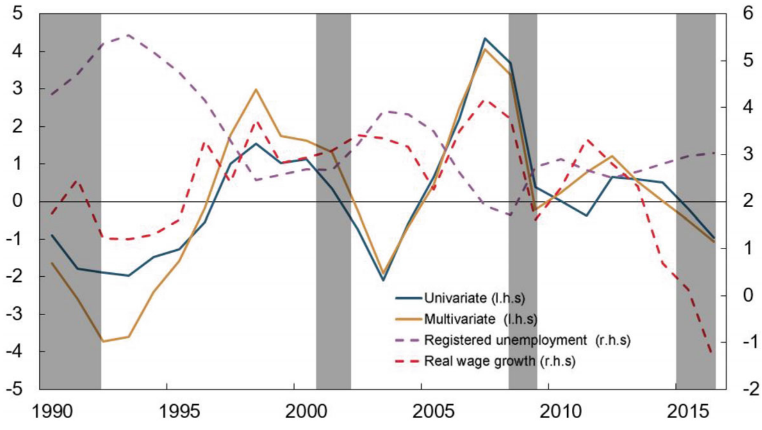

FIGURE 1. Estimated output gap in percent from unobserved component model UC 1 and from a univariate model

SVAR models

The model suite also includes structural VAR models. We estimate the following system:

$$y_t = \mu_0 + A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t, \tag{7}$$

where y_t is a vector of endogenous variables, μ_0 is a vector of constant terms and u_t is a vector of reduced form residuals. A_l for $l \in [1, p]$ are matrices of coefficients on lagged variables while p refers to the number of lags for the endogenous variables included in the system.

The use of SVAR models to estimate the output gap has been advocated recently by Coibion *et al.* (2018). Two SVAR models have been part of the model suite since the beginning of its development. The first model (SVAR 1) builds on Blanchard and Quah (1989) and uses data on GDP growth for mainland Norway and unemployment (NAV). The second model (SVAR 2) follows Cerra and Saxena (2000) and also includes data on domestic inflation. In Blanchard and Quah (1989), it is assumed that GDP growth is driven by two types of shocks: demand shocks and (permanent) supply shocks. Demand shocks are identified as shocks that do not affect GDP in the long term. Potential output is thereafter defined as GDP in the absence of the identified demand shocks. Cerra and Saxena (2000) identify in addition a temporary supply shock that does not have long-term effects on domestic inflation or GDP. As in Blanchard and Quah (1989), potential output is given by the counterfactual level of GDP driven only by permanent supply shocks. Both models are described in more detail in the Appendix.

Output gap estimates from the different models

Figure 2 shows output gap estimates from the different models together with Norges Bank's assessments of the output gap (black solid line) as presented in Monetary Policy Report 3/19. It is important to stress that the solid black line represents the official Norges Bank view on the state of resource utilization. In contrast to, e.g. the United States and

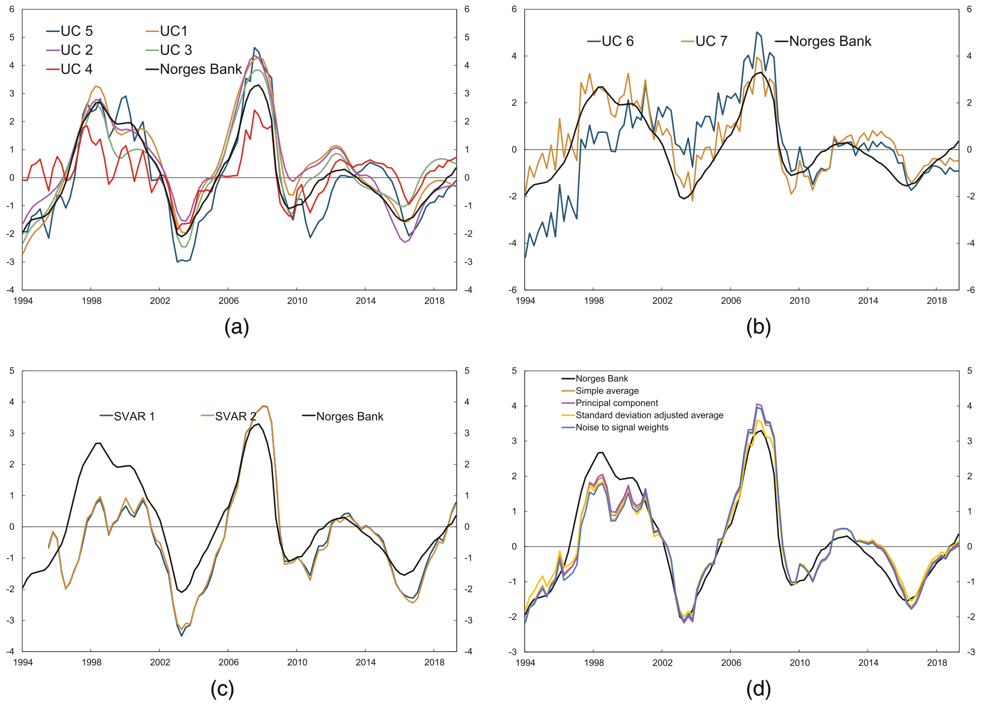

FIGURE 2. Output gap estimates from the different models and summary measures, 1994Q1–2019Q2. (a) UC models, (b) financial UC models, (c) SVAR models, (d) summary measures

Canada, there is no separate staff assessment of the output gap. The Norges Bank output gap series has been published since 2003, well before the development of the SoMs that started around 2012 and reached the state reported here only around 2017. Therefore, the SoMs capture the institutional view over the first part of the sample, while it influences the institutional view in the second part of the sample. Notably, the official Norges Bank series is rarely revised further back than the current business cycle.

Panel (a) shows the estimates based on the first five UC models. Norges Bank's official measure of the output gap is closely in line with the estimates from the different models. However, in retrospect, some of the models indicate that capacity utilization in the pre-Great Recession period was somewhat higher than in Norges Bank's assessment. In the period following the financial crisis, the different models provide a somewhat different view on the path of the output gap. In particular, the model using LFS unemployment data (UC 2) identifies a sharp drop in capacity utilization in 2014. In this period LFS unemployment increased substantially more than registered unemployment in response to the large decline in the oil price. This episode highlights the benefits of including alternative labour market indicators into the suite.

The output gap estimates from UC models using financial variables as observables (panel (b)) show somewhat different paths. The model using data on credit growth indicates that the output gap was substantially lower in the mid-1990s. This is because credit growth was low in this period compared with its historical average, thus driving GDP below its potential level. In the 2000s, credit growth accelerated. This may indicate that financial imbalances were building up and that output was well above its potential level. In fact, UC 6 indicates that the output gap was substantially higher during the downturn in 2003–2004 and in the period preceding the financial crisis. In the period following the financial crisis, the model is more closely in line with Norges Bank's assessments, although financial factors sustain a positive output gap until 2015. The model that includes data on house price inflation (UC 7) does not present the same picture as the gap based on credit growth, most likely because house price fluctuations correlate to a greater extent with GDP fluctuations. A partial exception is identified around 2014, when the housing sector was booming in the context of a stagnating economy.

The two structural VAR models (panel (c)) deliver similar estimates and are closely in line with Norges Bank's official assessment of the output gap in the second part of the sample. They indicate a substantially lower output gap at the end of the 1990s and a more negative gap during the downturn in 2003-2004 if compared with UC models and the assessment of Norges Bank.

Panel (d) presents four ways to weight together the different estimates of the output gap: a simple equally weighted average, the average with weights adjusted for the standard deviation of each individual model, the average with weights based on the noise-to-signal ratio (presented in Table 5) and a principal component.[3] Overall, these simple combinations of different models are closely in line with Norges Bank's official output gap assessment over time. Note that more sophisticated methods to combine the different models have been tested. Since they were leading to very similar results, the simple (and easy to communicate) equally weighted average has been preferred as a summary measure in the practical implementation of the model suite. The emphasis on model averaging follows a long tradition in Norges Bank, as documented in Bjørnland *et al.* (2012) and Aastveit *et al.* (2014).

All in all, labour market variables are widely used and are important drivers of the output gaps in the model suite. In practice, models relying on unemployment data have proven to be reliable, in particular when new information arrives. This is confirmed by recent evidence for the United States that was not available when the suite was implemented. In fact, Barbarino *et al.* (2020) document that models relating the unemployment rate to the output gap produce stable real-time output gap estimates since labour market data reduce the end-point problem of statistical models. In addition, Morley and Wong (2020) estimate the output gap with a Bayesian VAR comprising a large dataset of 138 variables and find that the unemployment rate is the single most important variable that contributes to the cycle. The extensive use of data on wage inflation builds on Phillips (1958) who specified his original analysis in terms of wage inflation and unemployment. In a small open economy like Norway where the exchange rate and import prices in general play a major role in explaining variation in consumer price inflation, the wage Phillips curve provides a clearer signal on the state of capacity utilization in the economy. Finally, a strong focus on the labour market is also justified under the mandate assigned to Norges

[3]A normalization by the standard deviation of each individual output gap series is performed in order to avoid those models with large variance or amplitude mechanically have a large impact on the summary measure of the output gap. This adjustment has limited impact since the nine series have similar amplitude. A normalization in terms of the signal-to-noise ratio is performed in order to attach a lower weight to models that are more likely to be revised.

Bank stating that the central bank should (i) target a rate of inflation close to 2% over time, (ii) contribute to *high and stable employment* and iii) counteract the build-up of financial imbalances.

III. Evaluation of estimated output gap measures

Since the output gap is not observable, there is no direct way of evaluating the model estimates. Nevertheless, three criteria have been discussed in the literature:

1. The estimated output gap should have reasonable forecasting properties for inflation and GDP.
2. Ex-post revisions should not be too extensive. When new data become available, the output gap estimate should not change substantially.
3. The estimated measure of potential output should respond substantially and rapidly to shocks that have a permanent (or at least a very persistent) effect on GDP and respond mildly or not at all to shocks that have only transitory effects.

It is important to stress here that all models are estimated using the final vintage of data (from 2019:Q2). We therefore ignore the role of data revisions in real-time revisions of output gap estimates.[4] Put differently, we conduct a pseudo real-time analysis that allows us to investigate the role of parameter instability and end-point problems of the statistical filters. Notably, Orphanides and van Norden (2002) show that most real-time revisions of output gap estimates are due to the filtering methods rather than data revisions, a point also emphasized more recently by Barbarino *et al.* (2020), with a special focus on the role of labour market data for real-time stability.

Forecasting properties

One important and challenging dimension to evaluate output gap estimates is the ability to forecast inflation. In fact, Orphanides and van Norden (2005) show that models using real-time estimates of the output gap perform worse at forecasting inflation than models using the final estimates of the gap as well as univariate AR models of inflation. To evaluate the forecasting properties of our estimates, we follow Orphanides and van Norden (2005) and perform a direct forecasting exercise. We estimate the following equation:

$$\pi_{t+h} = \alpha + \sum_{s=0}^{4} \lambda_s \pi_{t-s} + \sum_{s=0}^{4} \beta_s \hat{y}_{t-s} + \varepsilon_{t+h}, \quad (8)$$

where π_t is four-quarter growth in domestic CPI inflation and $\hat{y}_t$ represents the pseudo real-time output gap estimate. Equation (8) is a direct forecasting Phillips curve, where inflation h periods ahead depends on current and lagged values of inflation and the output gap. We estimate a separate regression for each forecast horizon (up to eight quarters

[4]Ideally, it is desirable to take data revisions into account. Unfortunately, a sufficiently long historical dataset of real-time vintages for all variables included in the suite is not available.

ahead). More specifically, we estimate the output gaps and equation (8) in pseudo real time and then project inflation up to eight quarters ahead.[5] For models estimated at annual frequency we use pseudo-real-time information to forecast the rest of the current year, before estimation and out-of-sample forecasting. We also estimate a bivariate VAR model with inflation and the output gap as an alternative to the direct forecasting approach in equation (8). In both cases, we investigate whether the forecasting accuracy improves using pseudo-real-time information on a summary measure of the output gap based on the simple equally weighted average of the models belonging to the suite.

In Table 1 we show how the simple summary measure of the output gap helps predicting domestic CPI inflation in pseudo real-time at different horizons when the benchmark for comparison is a simple AR model with four lags, arguably a model that generates forecasts that are hard to improve upon (cf. Champagne *et al.*, 2018). More specifically, values above (below) 1 mean that the model has a larger (smaller) average forecast error than the AR model. As an alternative for comparison, we also evaluate the output gap obtained from an HP filter with λ equal to 40,000.[6] In both exercises, the models using pseudo real-time estimates of the output gap perform better than the AR(4) model and largely better than the HP filter based measure at all horizons with the exception of horizon 1. The difference in forecasting performance between the AR model and the summary measure in Table 1 is not statistically significant in a Diebold-Mariano test. However, the HP-filter based measure's forecasting performance is significantly worse than the summary measure at horizon 3–8 at the 5% level. Therefore, our results are somewhat more rosy than Orphanides and van Norden (2005) and suggest that the output gap helps forecasting inflation, at least to some extent. One simple explanation for this result relies on the fact that the Phillips curve has not flattened in Norway. In particular, Brubakk, Hagelund and Husabø (2018) show that wage inflation has become more sensitive to fluctuations in unemployment over the last two decades, unlike in other countries.

In an additional forecasting exercise we use the domestic CPI price level instead of inflation to assess the output gaps ability to forecast cumulative inflation over longer horizons. For the first four quarters this is almost identical to forecasting annual inflation, but with a slightly different model specification. Results are presented in Table 2. In general, the forecast errors are lower in the level specification for all models, including the AR model, thus suggesting that the level specification is preferable for forecasting purposes. In relative terms, the largest improvements are featured by the models including the summary measure of the output gap, in particular at longer horizons. We remark that the better performance with respect to the AR model is statistically significant at horizons three to eight in a Diebold–Mariano test, in many instances at the 1% level. The summary

[5] First, the output gaps and equation (8) are estimated using data up to 2000:Q1 and then forecasts are produced for 2000:Q2 until 2002:Q1. Thereafter, the estimation period is extended by one quarter at a time up to 2019:Q2. Equation (8) is similar to the Phillips curve used in Edge and Rudd (2016) and Champagne *et al.* (2018). Since they can rely on real-time forecasts for the output gap, these papers can use conditional forecasts to evaluate longer horizons. For the UC models the estimates from the Kalman smoother are used.

[6] A value of λ equal to 40,000 is commonly used on quarterly Norwegian GDP data (Statistics Norway, 2018), although it is relatively high compared to the standard value of 1600 for quarterly data. Historically, a lower value would have implied a far too shallow recession after the banking crisis in the late 1980s, in particular when compared with other indicators such as the unemployment rate.

TABLE 1

Inflation forecast evaluation in comparison to an AR(4) model

	Direct forecasting		*Bivariate VAR*	
Horizon	*Average model suite*	*HP filter*	*Average model suite*	*HP filter*
1	1.02	1.03	1.02	1.03
2	0.95	1.02	0.96	1.05
3	0.95	1.09	0.90	1.07
4	0.94	1.15	0.86	1.08
5	0.88	1.21	0.86	1.09
6	0.90	1.29	0.88	1.12
7	0.98	1.34	0.93	1.15
8	1.09	1.37	0.98	1.19

Notes: Forecast error (RMSFE) from equation (8) and a bivariate VAR with pseudo real-time output gap estimates combined with domestic inflation relative to the RMSFE of AR(4). Numbers below one indicate RMSFE that is lower than the AR model. Estimation period is from 1990:Q1 using expanding windows from data vintage 2019:Q2 and pseudo-out-of-sample evaluation is 2000:Q2–2019:Q2.

TABLE 2

Price level forecast evaluation in comparison to an AR(4) model

	Direct forecasting		*Bivariate VAR*	
Horizon	*Average model suite*	*HP filter*	*Average model suite*	*HP filter*
1	0.95	0.98	0.95	0.98
2	0.95	0.95	0.89	0.93
3	0.89	0.91	0.82	0.90
4	0.86	0.87	0.77	0.86
5	0.83	0.85	0.72	0.84
6	0.80	0.86	0.69	0.82
7	0.75	0.85	0.66	0.80
8	0.68	0.85	0.63	0.79

Notes: Forecast error (RMSFE) from equation (8), replacing inflation with the price index, and a bivariate VAR with 4 lags, relative to the RMSFE of an AR(4) model. Average model suite is the simple average of the output gap models. The hp filter is estimated using $\lambda = 40{,}000$. Estimation period is from 1990:Q1 using expanding windows from data vintage 2019:Q2 and pseudo out-of-sample evaluation is 2000:Q2–2019:Q2.

measure of the output gap performs consistently better than the HP-filter based measure also in this exercise, in particular at long horizons.

We now evaluate alternative summary measures of the output gap and the individual models in the exercise based on a bivariate VAR model presented earlier in Tables 1 and 2.[7] In fact, so far we have focused our attention only on a simple equally weighted average of the individual models as summary measure of the output gap. In the last three columns in Table 3, we compare its performance against two alternative measures (both plotted in Figure 2, panel (d)), the first with weights adjusted for the standard deviation of each individual model and the second with weights based on the noise-to-signal ratio. We observe that the three summary measures obtain a very similar forecasting performance, thus justifying our focus on the simple equally weighted average when combining the individual models.

[7] Similar results are obtained in the direct forecasting exercise and are available upon request.

TABLE 3

Price level forecast evaluation: Individual models and weighted estimates

	Individual models									*Weighted estimates*		
H.	*UC1*	*UC2*	*UC3*	*UC4*	*UC5*	*UC 6*	*UC 7*	*SVAR1*	*SVAR2*	*Avg*	*std adj.*	*NSR*
1	1.02	1.01	1.05	0.98	1.05	1.04	**0.97**	1.03	1.03	0.95	0.95	0.95
2	**0.90**	0.94	0.92	0.93	1.00	1.04	0.93	0.98	0.99	0.89	0.88	0.90
3	**0.84**	0.89	0.86	0.88	0.95	1.01	0.89	0.93	0.94	0.82	0.81	0.83
4	**0.79**	0.84	0.81	0.84	0.90	1.00	0.85	0.88	0.89	0.77	0.74	0.77
5	**0.74**	0.91	0.76	0.79	0.86	1.00	0.82	0.84	0.86	0.72	0.69	0.71
6	**0.70**	0.78	0.73	0.75	0.83	1.00	0.80	0.82	0.84	0.69	0.65	0.68
7	**0.67**	0.76	0.70	0.72	0.80	0.99	0.79	0.81	0.84	0.66	0.62	0.65
8	**0.63**	0.74	0.66	0.68	0.76	0.97	0.78	0.79	0.84	0.63	0.59	0.61

Notes: Forecast error (RMSFE) from a bivariate VAR with 4 lags, relative to the RMSFE of an AR(4) model. Avg is a simple equally weighted average of the individual models. std adj. is the average of the individual models adjusted for their relative standard deviation. NSR weights the individual models based on the pseudo real-time noise-to-signal ratio (NSR(SD)) reported in Table 5, where models with small NSR are given higher weight. Estimation period is from 1990:Q1 using expanding windows from data vintage 2019:Q2 and pseudo out-of-sample evaluation is 2000:Q2–2019:Q2. Single model with lowest RMSFE at each horizon in bold.

In Table 3 we evaluate also the forecasting performance of bivariate VAR models using output gap estimates from each model belonging to the suite. Two remarks are in order. First, the simple equally weighted average of the models features a better forecasting performance than each individual model.[8] Second, the bivariate VAR models using output gap estimates from each individual model perform all better than the AR(4) model with one exception. In fact, the model using data on domestic credit growth (UC 6) performs on average worse than the AR(4) model, thus supporting the disconnect between the credit cycle and inflation dynamics, as discussed in Christiano *et al.* (2010), Borio *et al.* (2017) and Furlanetto, Ravazzolo and Sarferaz (2019).

In addition to inflation, one can also evaluate how well the output gap forecasts future GDP as in Kamber, Morley and Wong (2018). The idea is that a positive output gap should indicate that the economy operates above potential and therefore signal low GDP growth going forward. In Table 4 we repeat the forecast evaluation exercise performed above replacing inflation with log level of GDP. Also in this exercise, the model suite performs markedly better than the HP filter. However, the AR model perform approximately as well as the model suite when using the VAR specification and slightly better in the direct forecasting exercise. This is not necessarily surprising. In fact, it is conceivable that the AR model also captures that periods of high GDP growth are followed by lower growth.

Pseudo real-time evaluation

The second criterion is that a good estimate of the output gap, as measured in pseudo real time, should not change substantially when new data become available. We compare the performance of the summary measure obtained from the suite against estimating the output gap estimate derived from a simple HP filter. In Figure 3, we plot the difference

[8]The difference between the best individual model (UC 1) and the summary measure is not statistically significant. The forecasting performance needs to be approximately 20%–25% better in terms of RMSFE in order to be significant in a Diebold-Mariano test for this evaluation period.

TABLE 4

GDP forecast evaluation in comparison to an AR(4) model

	Direct forecasting		*Bivariate VAR*	
Horizon	*Average model suite*	*HP filter*	*Average model suite*	*HP filter*
1	0.89	1.07	0.89	1.07
2	1.03	1.15	0.97	1.10
3	1.08	1.22	1.00	1.13
4	1.07	1.26	1.01	1.13
5	1.14	1.29	1.04	1.16
6	1.21	1.41	1.05	1.20
7	1.20	1.47	1.06	1.22
8	1.24	1.50	1.06	1.25

Notes: Forecast error (RMSFE) from equation (8), replacing inflation with GDP (log level), and a bivariate VAR with 4 lags, relative to the RMSFE of an AR(4) model. Average model suite is the simple average of the output gap models. The hp filter is estimated using $\lambda = 40{,}000$. Estimation period is from 1990:Q1 using expanding windows from data vintage 2019:Q2 and pseudo out-of-sample evaluation is 2000:Q2–2019:Q2.

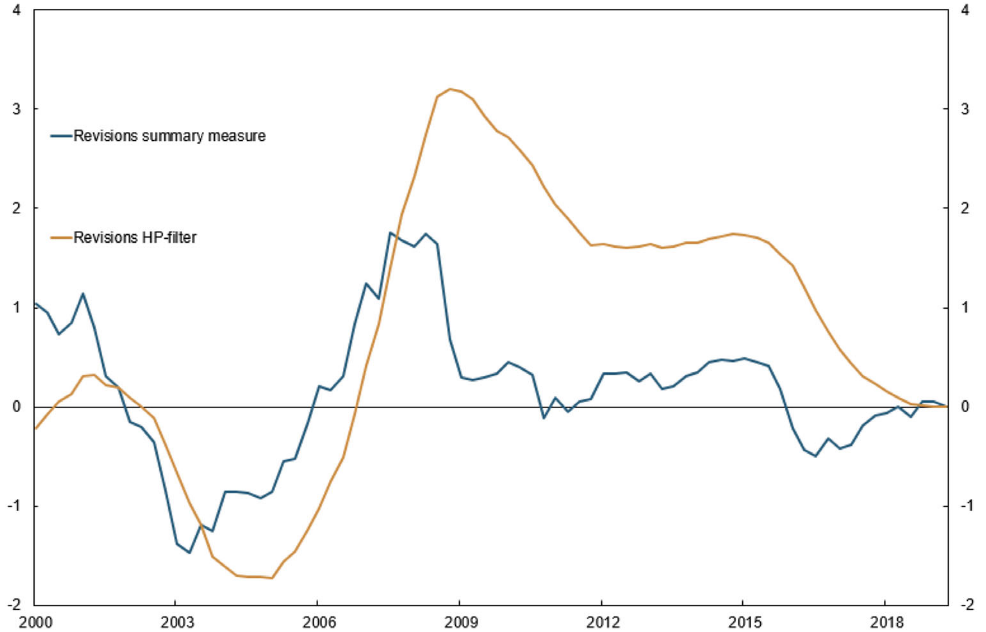

FIGURE 3. Final output gap estimate (2019 Q2) relative to the first estimate. Summary measure and an HP filter measure ($\lambda = 40{,}000$)

between the ‘final’ output gap estimates and the pseudo real-time estimates obtained by estimating the model recursively. The difference between the ‘final’ output gap estimate and the pseudo real-time estimate is considerably smaller for the summary measure than for the one based on the HP filter. Not surprisingly, the difference between the two estimates is substantially reduced when the entire data set is used.

Following Edge and Rudd (2016) and Champagne *et al.* (2018), we evaluate in Table 5 the statistical properties of output gap revisions for each model belonging to the suite and for the summary measure. Notwithstanding the important caveat that the role of data

TABLE 5

Summary statistics of output gap revisions

Model	*Mean*	*std dev*	*RMSR*	*Corr.*	*Sign agree*	*NSR (SD)*	*NSR (RMSR)*
UC 1	0.08	0.60	0.61	0.94	0.91	0.39	0.40
UC 2	0.50	0.70	0.86	0.90	0.76	0.45	0.55
UC 3	−0.19	0.59	0.62	0.90	0.89	0.42	0.45
UC 4	−0.03	0.48	0.48	0.85	0.81	0.53	0.53
UC 5	0.33	1.59	1.62	0.44	0.65	0.90	0.92
UC 6	0.11	0.59	0.60	0.94	0.86	0.37	0.38
UC 7	0.26	1.22	1.25	0.55	0.68	0.86	0.88
SVAR 1	0.01	0.63	0.63	0.94	0.84	0.37	0.37
SVAR 2	0.13	0.66	0.68	0.94	0.81	0.40	0.40
Average	0.13	0.66	0.67	0.92	0.87	0.47	0.48
HP filter	0.76	1.37	1.57	0.75	0.70	0.70	0.80

Notes: This table reports summary statistics of output gap revisions for each model included in the suite. We use data from the GDP vintage from 2019:Q2 and have pseudo real-time estimates from 2000:Q1–2019:Q2. The first two columns report the mean revision and the standard deviation. RMSR is the root of the mean squared revision. Corr is the correlation between the final estimate (2019:Q2) and the pseudo real-time estimate. Sign agree indicates how often the sign of the final estimate and the first estimate is the same. Noise-to-signal ratio (NSR) is the standard deviation/RMSR of the revision relative to the final estimate.

revisions is not accounted for by our estimates, the mean revision is rather small for most models and for the summary measure in particular with a value of 0.13 percentage point. Such a value for the mean revision compares favourably with recent estimates for Canada and the United States. The standard deviation and the root mean squared revision (RMSR) are slightly lower than 0.7, a value perhaps large in absolute terms but substantially smaller than the corresponding numbers for the HP filter-based measure. The correlation between the final and the pseudo real-time estimates is larger than 0.9 for most models and equal to 0.92 for the summary measure. In addition, the two estimates share the same sign in almost 90% of cases for the summary measure. Finally, we report two measures for the noise-to-signal ratio, based on the standard deviation and on the root mean squared error as in Edge and Rudd (2016). Both values are slightly lower than 0.5 for the summary measure of the output gap.

Cyclical sensitivity

We now evaluate the summary measure of the output gap along a third (and more novel) dimension, i.e. its cyclical sensitivity to various measures of economic shocks in keeping with the analysis in Coibion *et al.* (2018).

We first consider a measure of technology shocks proxied by quarterly changes in cyclically adjusted total factor productivity (TFP), following Fernald (2014). The cyclical adjustment, however, is different. Since hours per worker, the indicator used by Fernald (2014), are not closely related to the cycle in Norway, we rely on the Norges Bank's Regional Network measure of capacity utilization for the Norwegian economy over the period 2005-2018. We extend the series backward to 2000 with Statistics Norway's capacity utilization measure for manufacturing.

Figure 4 shows how the summary measure of potential GDP obtained from the model suite and the HP filter measure in pseudo real time respond to an exogenous increase in TFP using local projection methods (cf. Jordà, 2005) as in Coibion *et al.* (2018).

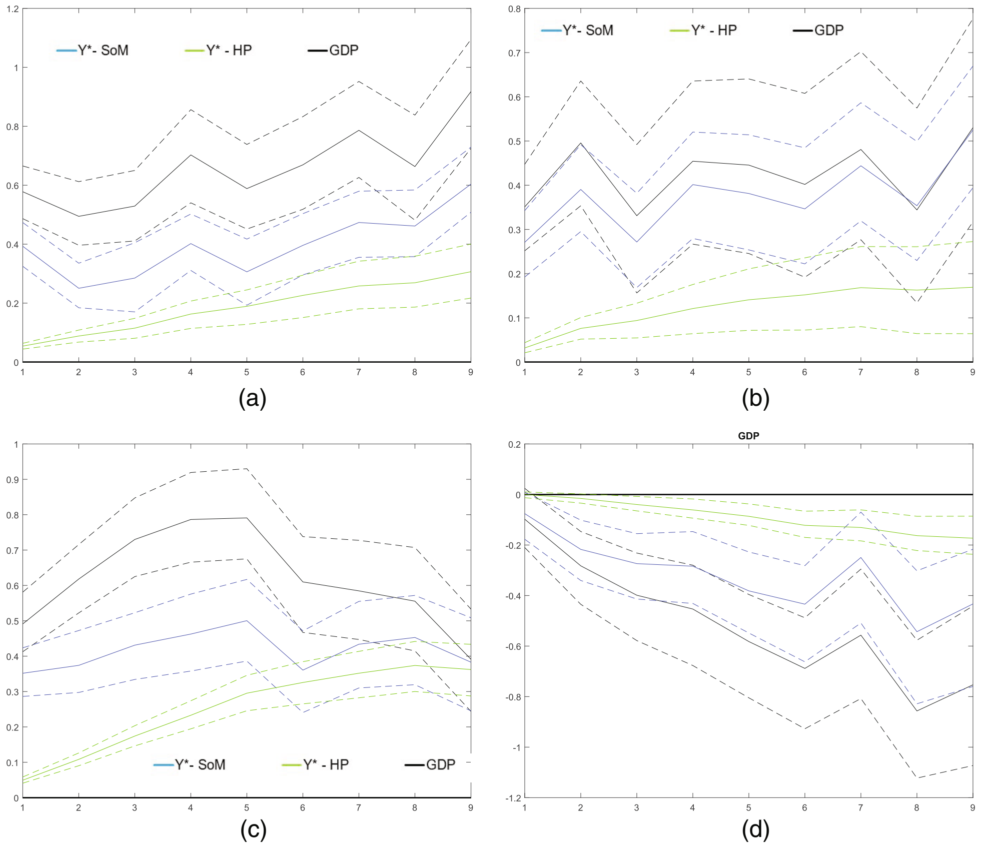

FIGURE 4. Impulse responses of the summary measure of potential output from the suite of models (SoMs) and the HP filter measure ($\lambda = 40{,}000$). 2000:Q1 - 2019:Q2. (a) TFP, (b) Blanchard–Quah supply, (c) Blanchard–Quah demand, (d) monetary policy shock

The summary measure of potential output (blue solid line) responds substantially more than the HP filtered GDP series (green solid line) to an expansionary TFP shock. We conclude that the summary measure of potential output reacts strongly and persistently to technology shocks in real time, although not as much as actual output (solid black line).

We repeat the same exercise in response to shocks derived from a Blanchard–Quah decomposition. As in Coibion *et al.* (2018), we interpret shocks with potentially permanent effects as supply shocks and shocks with transitory effects as demand shocks.

The summary measure of potential output responds strongly to the permanent shock, as much as actual GDP. This confirms that the Norges Bank's summary measure of potential output behaves differently than the estimates of potential output provided by Federal Reserve Board, IMF and OECD reviewed by Coibion *et al.* (2018). The summary measure of potential output also responds to some extent to demand shocks obtained from the Blanchard–Quah decomposition. It reacts more than the measure of potential output based on the HP filter in the short run (see panel (c) in Figure 4). Note that such a response to demand shocks is common in dynamic stochastic general equilibrium models where potential output is defined as the counterfactual level of output in the absence of sticky prices and sticky wages. In this kind of model, potential output responds to demand shocks like government spending shocks, investment specific shocks or discount factor shocks.

As Blanchard (2018) notes, even monetary policy shocks have an effect on potential output if one assumes in the counterfactual that nominal rigidities are removed only from the date of the shock on, thus taking history as given up to the date of the shock. We consider the effects of monetary policy shocks on potential output by using a series of shocks derived by Brubakk *et al.* (2022) using high frequency financial data around monetary policy meetings. In panel (d) in Figure 4 we present impulse responses to a contractionary monetary policy shock. As in the case of demand shocks, the summary measure of potential output reacts more than the HP filter-based measure and less than actual output. Somewhat intriguingly, monetary policy shocks seem to have persistent effects on both potential and actual output, in line with recent evidence using local projection methods presented in Jordà, Singh and Taylor (2020).

All in all, we conclude that the pseudo real-time summary measure of potential output responds to shocks substantially more than a measure based on the HP filter, in particular when it comes to shocks with potentially permanent effects but to some extent also to demand shocks.

IV. New developments: Hysteresis and output gap

The responsiveness of the summary measure of potential output to demand shocks discussed above may be related to hysteresis, i.e. the presence of demand shocks that have a long-run impact on the productive capacity of the economy. The link between hysteresis and the measurement of the output gap is of special interest since Norges Bank should contribute to achieve high and stable employment, in addition to its traditional price stability objective, under the new mandate approved in 2018. While particularly relevant for Norway, the topic is of general interest. In fact, the long-lasting effects of the Great Recession in the US, presumably a large negative demand shock, on employment and real economic activity have revived the interest in hysteresis. Several recent papers (see Cerra, Fatas and Saxena (2022) for a review of the literature) using a broad set of methodologies have found evidence of long-run effects of demand-driven recessions (but also booms).

One framework in which it is possible to discuss the link between hysteresis and output gap is a simple extension of the SVAR models already included in the suite. Following Blanchard and Quah (1989), both SVAR 1 and SVAR 2 assume the presence of one (and only one) shock with long-run effects on GDP. Furlanetto *et al.* (2021a) have relaxed this assumption by allowing two shocks to have long-run effects: a traditional supply shock and a more novel demand shock. Here, we propose a simplified version of their model using the same observables included in SVAR 2, i.e. output growth, inflation and the unemployment rate. We assume the presence of three shocks: a transitory shock (with zero long-run effect on output), a demand shock with potentially permanent effects on output (unrestricted in the long-run and driving inflation and output in the same direction on impact) and a supply shock (unrestricted in the long-run and driving inflation and output in the opposite direction on impact). The model (SVAR 3 henceforth) is estimated over the same sample period of SVAR 2 (i.e. 1990Q1–2019Q2) and is identified using a combination of sign and long-run restrictions.

We find that demand shocks do have long-run effects on GDP. They explain a large share of output fluctuations in the short run and a solid 20% in the medium to long run. Although these effects are considerably smaller than the ones estimated by Furlanetto *et al.* (2021b) for the United States, we conclude that the model assigns a non-negligible role to shocks generating hysteretic effects in Norway. In light of this result, one may wonder whether the measurement of the output gap is affected in presence of hysteresis effects. We stress at the outset that this topic has been discussed only marginally in the literature (see section V-C in Cerra *et al.* (2022)). Our contribution here is to frame the question and provide some alternative options. The key question is how to define potential output in a world where demand shocks have non-marginal long-run effects. One could maintain a strict 'trend' definition, in keeping with Blanchard and Quah (1989), and assume that all shocks with long-run effects have a direct and immediate effect on potential output. According to this definition (that we label 'trend concept'), potential output is the counterfactual level of output driven by all shocks except the transitory ones. Naturally, this definition implies a measure of potential output that is rather volatile and the output gap will be, in general, relatively small. A second possibility is to assume that all demand factors do not affect potential output, neither in the short run nor in the long run. In this case, potential output is defined as the counterfactual level of output driven only by permanent supply shocks. Such a measure of potential output (that we label 'supply-only concept') is clearly less cyclical. However, since we are removing a permanent component from the definition of potential output, we expect a less stationary, and thus less interpretable, measure of the output gap. We believe that both definitions are rather extreme. Let us take the Great Recession in the United States as an example and suppose that the output gap is closed on impact of the recession. The 'trend concept' would imply an immediate and protracted decline in potential output leading to a small negative output gap, and thus signalling limited negative slack in the United States during the largest recession since World War II. Such a measure of the output gap may make policy makers hesitant to use their stabilization tools precisely when they are most effective (Fernández-Villaverde *et al.*, 2019).[9]

In contrast, 'the supply-only concept' would signal a large negative output gap even when the negative effects of the recession on skills, labour force participation and productivity have become irreversible.[10] We argue that a more meaningful definition for practical purposes allows potential output to be affected by permanent demand shocks in the medium to long run but not in the short run. In our empirical application, we set the cut-off at a 12-quarter horizon. Imagine a negative demand shock with permanent effects: potential output would be affected by the shock only from quarter 12 onwards. We label this third alternative as 'medium-run hysteresis concept' since hysteresis impact potential output only in the medium run.

[9]In the context of our Great Recession example, the distinction between 'trend concept' and 'supply-only concept' vanishes in a scenario in which the decline of trend growth in the US economy predates the Great Recession and is driven exclusively by supply-side factors. Eo and Morley (2022) provide evidence in favour of this scenario using a univariate Markov-switching model of real GDP growth.

[10]Note that potential output (defined as the flex-price, flex-wage counterfactual) in DSGE models with endogenous growth mechanisms is unaffected by demand shocks generating hysteresis effects in line with the supply-only concept. In fact, both the liquidity shock in Anzoategui *et al.* (2019) and the discount factor shock in Garga and Singh (2021) do not propagate under flexible prices and wages.

We plot the three proposed estimates of the output gap in Norway using SVAR 3 in Figure 5. Notably, peaks and troughs are synchronized across the three different estimates that, however, feature different properties. As expected, the ‘trend concept’ (green line) is the least volatile and signals a positive gap of only 2% at the peak of the oil boom in the 2000s. In this case, permanent demand shocks (that explain at least in part the boom) are affecting potential output at any horizon. The ‘supply-only concept’ refers to the orange line and, as expected, is less stationary and features a positive mean. GDP growth is high in the first part of the sample but potential output is little responsive since it is unaffected by permanent demand shocks. Such a difference in the level with respect to the ‘trend concept’ persists over time and is further amplified during the expansion of the mid-2000s before closing around 2015. The ‘medium-run hysteresis concept’ (blue line) shares the same dynamics as the ‘trend concept’ but is more volatile, in particular around peaks and troughs. The difference between the blue and the green line is more visible in the second part of the sample, and in the two most recent recessions in particular. It captures the importance of short-run effects of permanent demand shocks and highlights how these shocks amplify the cycle. We believe that the ‘medium-run hysteresis concept’ reflects a reasonable compromise between the two more extreme alternatives. Clearly, alternative cut-off horizons, or alternative degrees of pass-through, could be considered.

In Figure 6, we compare the output gap associated with the ‘medium-run hysteresis concept’ of potential output (blue line) against the summary measure of the output gap obtained from the suite of models (green line) and the measure obtained from SVAR 2 (yellow line) that shares the same observables with SVAR 3. The three measures co-move and share synchronized peaks and troughs. The summary measure and the one obtained

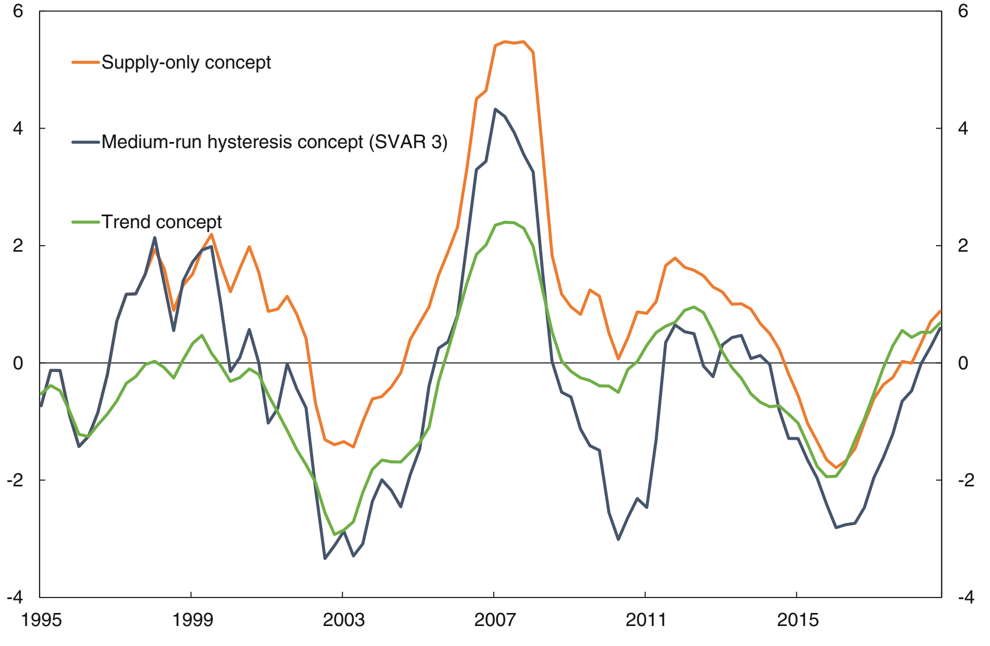

FIGURE 5. Different output gap concepts from SVAR 3

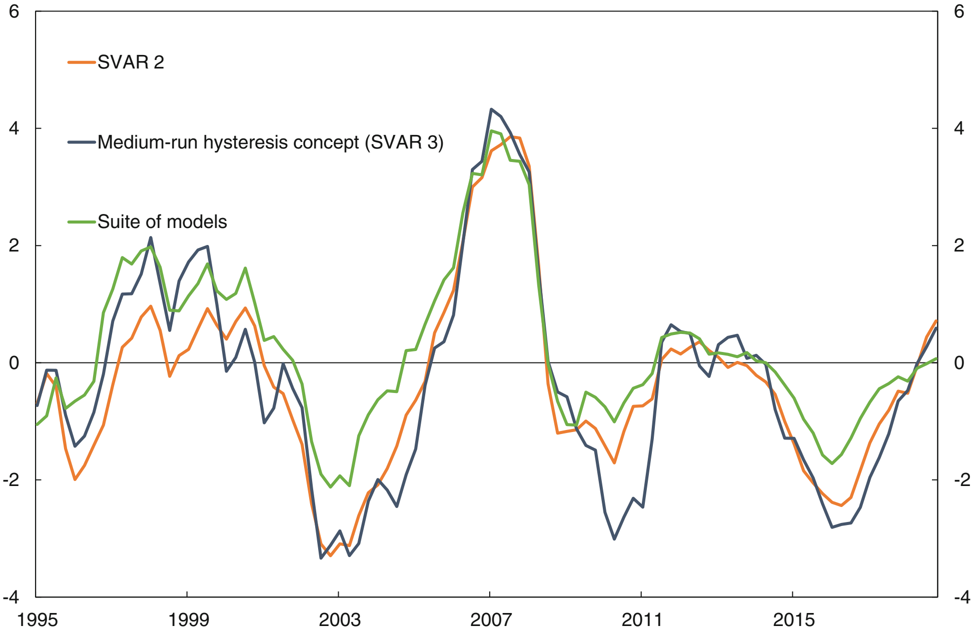

FIGURE 6. Comparison of alternative output gap estimates. The medium-run hysteresis concept indicates deeper recessions

from SVAR 3 are similar in general, thus showing that considering hysteresis effects does not overturn the view on capacity utilization for the Norwegian economy. However, we remark one notable difference: the model that accounts explicitly for hysteresis implies a more negative output gap during the three recessions in the sample. Arguably, permanent demand shocks are particularly important in recessions. SVAR 3 generates a more negative output gap during the last two recessions also when compared with SVAR 2.

This reflects the fact that in SVAR 2 the permanent component is driven only by supply shocks that have full pass-through on potential output. In the model with hysteresis, permanent demand shocks limit somewhat the explanatory power of both permanent supply and transitory shocks and affect potential output only in the medium to long run. Finally, we check how the summary measure of the output gap responds in pseudo real time to demand shocks with permanent effects, as identified by SVAR 3, by repeating the cyclical sensitivity analysis performed in the previous section. In Figure 7, we see that the summary measure responds substantially to the shock, almost as much as GDP at horizon 9. However, it responds also in the short run (and somewhat more than the HP-based measure, as in previous examples). This is consistent with SVAR 3 featuring a more negative output gap in recessions, arguably periods when permanent demand shocks are important. In fact, the summary measure of potential output responds to demand shocks in the short run while the 'medium-run hysteresis concept' of potential output in SVAR 3 does not.

All in all, it seems that the current SoMs align relatively well with a model explicitly considering hysteresis effects. This is also consistent with the fact that the summary measure of potential output exhibits some sensitivity to demand shocks (although

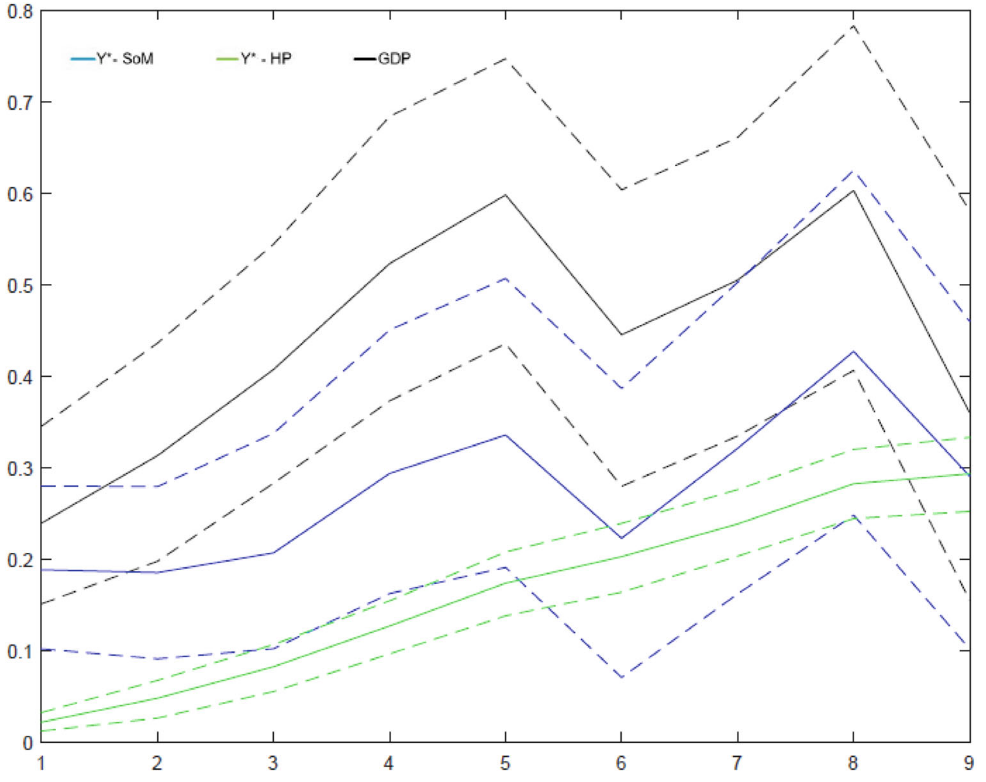

FIGURE 7. Impulse responses of the summary measure of potential output from the suite of models (SoM) and the HP filter to the permanent demand shock from SVAR 3

substantially less than to supply shocks), as shown in section III. Nonetheless, our exercise reveals that accounting for hysteresis may be particularly important in recessions. To take this insight into account, it is certainly conceivable to include SVAR 3 in the SoMs in the future.

V. Conclusion

We documented the SoMs currently used to estimate the output gap at Norges Bank. The models are estimated using data on GDP, unemployment, inflation, wages, investment, credit and house prices. A simple equally weighted average of the models outperforms individual models and helps predict domestic CPI when compared with a simple AR model. In addition, the summary measure of the output gap shows relatively little variation when new information arrives, unlike simple trend estimates based only on GDP data. Notably, it responds strongly and rapidly to permanent shocks and to narrative measures of technology shocks but also, although to a more limited extent, in response to transitory shocks and to monetary policy shocks. While the main goal of the paper is to document the methods used at Norges Bank to estimate the output gap (as it has been done at the Federal Reserve and at Bank of Canada), we believe that our key results (like the importance of using multivariate methods, the usefulness of labour market variables and the value of testing for how potential output responds to structural shocks) have some external validity that goes beyond the specificities of the Norwegian economy.

The SoMs have been expanded and refined over time and we find it encouraging to document that its performance has proved satisfactory, even in a dimension like the cyclical sensitivity that has so far been untested for measures of output gaps produced by other central banks. Nonetheless, it is fair to say that the suite could (and should) be extended in several dimensions to account for recent developments in the literature. This includes the use of models featuring stochastic volatility (Mertens, 2020 and Clark and Ravazzolo, 2015), models using mixed-frequency data to nowcast the output gap (Berger, Morley and Wong, 2022a), dynamic factor models (Aastveit and Trovik, 2014 and Jarocinski and Lenza, 2018), large Bayesian Vector Autoregressions (Morley and Wong, 2020), models taking explicitly into account the effective lower bound and other non-linearities (Liu *et al.*, 2019), the use of alternative benchmarks for comparison (Kamber *et al.*, 2018) and the use of different assumptions to evaluate the role of the oil sector (Bersch and Sinclair, 2011. The application of some of these techniques to the case of Norway is in our agenda for future research.

Final Manuscript Received: January 2021

References

Aastveit, K. A., Gerdrup, K. R., Jore, A. S. and Thorsrud, L. A. (2014). ‘Nowcasting GDP in real time: A density combination approach’, *Journal of Business & Economic Statistics*, Vol. 32, pp. 48–68.

Aastveit, K. A., Jore, A. S. and Ravazzolo, F. (2016). ‘Identification and real-time forecasting of Norwegian business cycles’, *International Journal of Forecasting*, Vol. 32, pp. 283–292.

Aastveit, K. A. and Trovik, T. (2014). ‘Estimating the output gap in real time: A factor model approach’, *Quarterly Review of Economics and Finance*, Vol. 54, pp. 180–193.

Anzoategui, D., Comin, D., Gertler, M. and Martinez, J. (2019). ‘Endogenous technology adoption and R&D as sources of business cycle persistence’, *American Economic Journal: Macroeconomics*, Vol. 11, pp. 67–110.

Apel, M. and Jansson, P. (1999). ‘System estimates of potential output and the nairu’, *Empirical Economics*, Vol. 24, pp. 373–388.

Barbarino, A., Berge, T. J., Chen, H. and Stella, A. (2020). *Which output gap measures are stable in real time and why? Finance and Economics Discussion Series 2020/102*, Washington DC: Board of Governors of the Federal Reserve System.

Basistha, A. and Startz, R. (2008). ‘Measuring the nairu with reduced uncertainty: a multiple-indicator common-cycle approach’, *Review of Economics and Statistics*, Vol. 90, pp. 805–811.

Berger, T., Morley, J. and Wong, B. (2022a). ‘Nowcasting the output gap’, *Journal of Econometrics* (forthcoming). https://doi.org/10.1016/j.jeconom.2020.08.011

Berger, T., Richter, J. and Wong, B. (2022b). ‘A unified approach for jointly estimating the business and financial cycle, and the role of financial factors’, *Journal of Economic Dynamics and Control*, Vol. 136, pp. 104–315.

Bernhardsen, T., Eitrheim, Ø., Jore, A. S. and Røisland, Ø. (2005). ‘Real-time data for norway: challenges for monetary policy’, *North American Journal of Economics and Finance*, Vol. 16, pp. 333–349.

Bersch, J. and Sinclair, T. M. (2011). ‘Mongolia: Measuring the output gap’, *Working Paper*, Vol. 11/79, p. IMF.

Bilbiie, F. O. and Melitz, M. J. (2020). Aggregate-Demand Amplification of Supply Disruptions: The Entry-Exit Multiplier. Technical report, National Bureau of Economic Research.

Bjørnland, H. C., Brubakk, L. and Jore, A. S. (2008). 'Forecasting inflation with an uncertain output gap', *Empirical Economics*, Vol. 35, pp. 413–436.

Bjørnland, H. C., Gerdrup, K., Jore, A. S., Smith, C. and Thorsrud, L. A. (2012). 'Does forecast combination improve Norges Bank inflation forecasts?', *Oxford Bulletin of Economics and Statistics*, Vol. 74, pp. 163–179.

Blanchard, O. (2018). 'Should we reject the natural rate hypothesis?', *Journal of Economic Perspectives*, Vol. 32, pp. 97–120.

Blanchard, O. J. and Quah, D. (1989). 'The dynamic effects of aggregate demand and supply disturbances', *American Economic Review*, Vol. 79, pp. 655–673.

Borio, C., Disyatat, P. and Juselius, M. (2017). 'Rethinking potential output: Embedding information about the financial cycle', *Oxford Economic Papers*, Vol. 69, pp. 655–677.

Brubakk, L., Hagelund, K. and Husabø, E. (2018). The Phillips Curve and Beyond-Why has Wage Growth been so Low? Technical report, Norges Bank Staff Memo 10/2018.

Brubakk, L., ter Ellen, S., Robstad, Ø. and Xu, H. (2022). 'The macroeconomic effects of forward communication', *Journal of International Money and Finance*, Vol. 120, 102536.

Canova, F. (2020). FAQ: How do I extract the output gap? Working paper 386, Sveriges Riksbank.

Cerra, V., Fatas, A. and Saxena, S. (2022). 'Hysteresis and business cycles', *Journal of Economic Literature* (forthcoming).

Cerra, V. and Saxena, S. (2000). Alternative Methods of Estimating Potential Output and the Output Gap: an Application to Sweden. IMF Working paper 59.

Champagne, J., Poulin-Bellisle, G. and Sekkel, R. (2018). 'The real-time properties of the bank of Canada's staff output gap estimates', *Journal of Money, Credit and Banking*, Vol. 50, pp. 1167–1188.

Christiano, L., Ilut, C. L., Motto, R. and Rostagno, M. (2010). Monetary Policy and Stock Market Booms. Working paper 16402, National Bureau of Economic Research.

Clark, P. K. (1989). 'Trend reversion in real output and unemployment', *Journal of Econometrics*, Vol. 40, pp. 15–32.

Clark, T. E. and Ravazzolo, F. (2015). 'Macroeconomic forecasting performance under alternative specifications of time-varying volatility', *Journal of Applied Econometrics*, Vol. 30, pp. 551–575.

Coibion, O., Gorodnichenko, Y. and Ulate, M. (2018). 'The Cyclical Sensitivity in Estimates of Potential Output', *Brookings Papers on Economic Activity*, Vol. 49, pp. 343–441.

Denton, F. (1971). 'Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization', *Journal of the American Statistical Association*, Vol. 66, pp. 99–102.

Domenech, R. and Gomez, V. (2006). 'Estimating potential output, core inflation, and the nairu as latent variables', *Journal of Business and Economic Statistics*, Vol. 24, pp. 354–365.

Edge, R. M. and Rudd, J. B. (2016). 'Real-time properties of the federal reserve's output gap', *Review of Economics and Statistics*, Vol. 98, pp. 785–791.

Eo, Y. and Morley, J. (2022). 'Why has the us economy stagnated since the great recession?', *Review of Economics and Statistics*, Vol. 104, pp. 246–258.

Fernald, J. (2014). A Quarterly, Utilization-Adjusted Series on Total Factor Productivity. Working paper 2012-19, Federal Reserve Bank of San Francisco.

Fernández-Villaverde, J., Mandelman, F., Yu, Y. and Zanetti, F. (2019). Search Complementarities, Aggregate Fluctuations, and Fiscal Policy. Technical report, National Bureau of Economic Research.

Fleischman, C. A. and Roberts, J. M. (2011). From Many Series, One Cycle: Improved Estimates of the Business Cycle From a Multivariate Unobserved Components Model. Working paper 2011-46, Federal Reserve Board.

Furlanetto, F., Gelain, P. and Taheri Sanjani, M. (2021a). 'Output gap, monetary policy trade-offs, and financial frictions', *Review of Economic Dynamics*, Vol. 41, pp. 52–70.

Furlanetto, F., Lepetit, A., Robstad, Ø., Rubio Ramirez, J. and Ulvedal, P. (2021b). Estimating Hysteresis Effects. Working Paper DP 16558, CEPR.

Furlanetto, F., Ravazzolo, F. and Sarferaz, S. (2019). 'Identification of financial factors in economic fluctuations', *Economic Journal*, Vol. 129, pp. 311–337.

Garga, V. and Singh, S. R. (2021). ‘Output hysteresis and optimal monetary policy’, *Journal of Monetary Economics*, Vol. 117, pp. 871–886.
Hamilton, J. D. (2018). ‘Why you should never use the Hodrick-Prescott filter’, *Review of Economics and Statistics*, Vol. 100, pp. 831–843.
Harvey, A. and Trimbur, T. M. (2008). ‘Trend estimtimation and the hodrick-prescott filter’, *Journal of the Japan Statistical Society*, Vol. 38, pp. 41–49.
Hjelm, G. and Jonsson, K. (2010). In Search of a Method for Measuring the Output Gap of the Swedish Economy. Working paper 115, National Institute of Economic Research.
Jarocinski, M. and Lenza, M. (2018). ‘An inflation predicting measure of the output gap in the euro area’, *Journal of Money, Credit and Banking*, Vol. 50, pp. 1189–1224.
Jordà, Ò. (2005). ‘Estimation and inference of impulse responses by local projections’, *American Economic Review*, Vol. 95, pp. 161–182.
Jordà, Ò., Singh, S. R. and Taylor, A. M. (2020). The Long-Run Effects of Monetary Policy. Working paper 2020-01, Federal Reserve Bank of San Francisco.
Kamber, G., Morley, J. and Wong, B. (2018). ‘Intuitive and reliable estimates of the output gap from a beveridge-nelson filter’, *Review of Economics and Statistics*, Vol. 100, pp. 550–566.
Kuttner, K. N. (1994). ‘Estimating potential output as a latent variable’, *Journal of Business and Economic Statistics*, Vol. 12, pp. 361–368.
Liu, P., Theodoridis, K., Mumtaz, H. and Zanetti, F. (2019). ‘Changing macroeconomic dynamics at the zero lower bound’, *Journal of Business & Economic Statistics*, Vol. 37, pp. 391–404.
Mertens, E. (2020). On the reliability of output gap estimates in real time (unpublished manuscript).
Mishkin, F. (2007). Estimating Potential Output. In Speech Delivered at the Conference on Price Measurement for Monetary Policy, Sponsored by the Federal Reserve Bank of Dallas, Dallas, Texas, Volume 24.
Morley, J. and Wong, B. (2020). ‘Estimating and accounting for the output gap with large Bayesian vector autoregressions’, *Journal of Applied Econometrics*, Vol. 35, pp. 1–18.
Orphanides, A. and van Norden, S. (2002). ‘The unreliability of output gap estimates in real time’, *Review of Economics and Statistics*, Vol. 84, pp. 569–583.
Orphanides, A. and van Norden, S. (2005). ‘The reliability of inflation forecasts based on output gap estimates in real time’, *Journal of Money, Credit and Banking*, Vol. 37, pp. 583–601.
Phillips, A. W. (1958). ‘The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957’, *Economica*, Vol. 25, pp. 283–299.
Planas, C., Rossi, A. and Fiorentini, G. (2008). ‘Bayesian analysis of the output gap’, *Journal of Business and Economic Statistics*, Vol. 26, pp. 18–32.
Sinclair, T. M. (2009). ‘The relationships between permanent and transitory movements in us output and the unemployment rate’, *Journal of Money, Credit and Banking*, Vol. 41, pp. 529–542.
Statistics Norway, A. (2018). Konjunkturtendensene, Report 4/18.

Appendix

UC models 1 and 2: GDP mainland Norway, unemployment (NAV or LFS) and real wage growth

Definition of GDP:

$$y_t = \hat{y}_t + \bar{y}_t. \tag{A1}$$

Process for the output gap:

$$\hat{y}_t = \lambda_y \hat{y}_{t-1} + \epsilon_t. \tag{A2}$$

Process for the growth in potential GDP:

$$\Delta \bar{y}_t = G_t + \eta_t. \tag{A3}$$

Process for potential growth:

$$G_t = C_G + \lambda_G(G_{t-1} - C_G) + \psi_t. \tag{A4}$$

Definition of Real wage:

$$W_t = \hat{W}_t + \overline{W}_t. \tag{A5}$$

Process for the real wage gap:

$$\hat{W}_t = \lambda_W \hat{W}_{t-1} + \gamma \hat{y}_t + \upsilon_t. \tag{A6}$$

Process for the trend in real wages:

$$\Delta\overline{W}_t = C_W + \lambda_{\overline{W}}(\Delta\overline{W}_{t-1} - C_W) + \mu_t. \tag{A7}$$

Definition of unemployment:

$$u_t = \hat{u}_t + \bar{u}_t. \tag{A8}$$

Process for the unemployment gap:

$$\hat{u}_t = \lambda_u \hat{u}_{t-1} + \beta \hat{y}_t + \omega_t. \tag{A9}$$

Process for the change in unemployment trend (NAIRU):

$$\Delta\bar{u}_t = \lambda_{\bar{u}} \Delta\bar{u}_{t-1} + \nu_t. \tag{A10}$$

The models are set up in state space form. For the observable variables, we use the demeaned changes in log GDP, log real wage and unemployment rate. The estimation uses the Kalman filter in combination with a minimization algorithm (fmincon) to search for the posterior mode. We use the estimated posterior mode and the Kalman smoother to recover the output gap estimate. Tables A1 and A2 report the priors and posterior mode and standard deviation. All UC models follow this estimation routine.

UC model 3: GDP mainland Norway, unemployment (NAV), real wage growth and investment share

In addition to equations (A8 to A10), the investment share (business investment as share of GDP for mainland Norway) is related to the output gap, following Domenech and Gomez (2006):

$$x_t = \lambda_x x_{t-1} + (1 - \lambda_x)\bar{x}_t + \gamma_x \hat{y}_t + e_t, \tag{A11}$$

$$\Delta\bar{x}_t = z_t, \tag{A12}$$

where x_t is the investment share and $\overline{x_t}$ is the trend component of the investment share. e_t is a shock to the investment share, while z_t represents a shock to the level of potential investment. The observable variables are the same as in UC 2 with the demeaned investment share as an additional observable (Table A3).

TABLE A1

UC 1. Estimated and calibrated parameters. Annual data. Estimation period: 1990-2019

Parameter	*Prior*	*Prior distribution*	*Posterior mode*
λ_y	0.7 (0.2)	Gamma	0.79 (0.12)
λ_G	0.9 (0.2)	Gamma	0.77 (0.14)
λ_W	0.75 (0.25)	Gamma	0.89 (0.12)
$\lambda_{\overline{W}}$	0.6 (0.2)	Gamma	0.61 (0.14)
λ_u	0.5 (0.2)	Gamma	0.43 (0.07)
$\lambda_{\bar{u}}$	0.9 (0.1)	Gamma	0.87 (0.09)
γ	0.29	Calibrated	
β	−0.29	Calibrated	
σ_ϵ	2 (10)	Inverse Gamma	1.20 (0.18)
σ_η	2 (10)	Inverse Gamma	0.71 (0.18)
σ_ψ	1 (10)	Inverse Gamma	0.32 (0.15)
σ_υ	0.5 (10)	Inverse Gamma	0.17 (0.10)
σ_μ	2 (10)	Inverse Gamma	0.72 (0.11)
σ_ω	0.4 (10)	Inverse Gamma	0.09 (0.03)
σ_ν	0.2 (10)	Inverse Gamma	0.06 (0.03)

Note: Standard deviation in parantheses.

TABLE A2

UC 2. Estimated and calibrated parameters. Annual data. Estimation period: 1990-2019

Parameter	*Prior*	*Prior distribution*	*Posterior*
λ_y	0.7 (0.2)	Gamma	0.78 (0.12)
λ_G	0.9 (0.2)	Gamma	0.77 (0.13)
λ_W	0.75 (0.25)	Gamma	0.91 (0.13)
$\lambda_{\overline{W}}$	0.6 (0.2)	Gamma	0.59 (0.17)
λ_u	0.5 (0.2)	Gamma	0.49 (0.07)
$\lambda_{\bar{u}}$	0.9 (0.1)	Gamma	0.86 (0.10)
γ	0.29	Calibrated	
β	−0.29	Calibrated	
σ_ϵ	2 (10)	Inverse Gamma	1.12 (0.25)
σ_η	2 (10)	Inverse Gamma	0.77 (0.29)
σ_ψ	1 (10)	Inverse Gamma	0.32 (0.16)
σ_υ	0.5 (10)	Inverse Gamma	0.17 (0.12)
σ_μ	2 (10)	Inverse Gamma	0.75 (0.12)
σ_ω	0.4 (10)	Inverse Gamma	0.23 (0.06)
σ_ν	0.2 (10)	Inverse Gamma	0.05 (0.02)

Note: Standard deviation in parentheses.

UC model 4: GDP mainland Norway and change in domestic inflation

Definition of GDP:

$$y_t = \hat{y}_t + \overline{y}_t. \tag{A13}$$

Process for the output gap:

$$\hat{y}_t = \lambda_y \hat{y}_{t-1} + \epsilon_t. \tag{A14}$$

Process for the change in potential GDP:

$$\Delta \overline{y}_t = \Delta \overline{y}_{t-1} + \eta_t. \tag{A15}$$

TABLE A3

UC 3. Estimated and calibrated parameters. Annual data. Estimation period: 1990–2019

Parameter	*Prior*	*Prior distribution*	*Posterior*
λ_y	0.7 (0.2)	Gamma	0.73 (0.11)
λ_G	0.9 (0.2)	Gamma	0.77 (0.12)
λ_W	0.75 (0.25)	Gamma	0.87 (0.14)
$\lambda_{\overline{W}}$	0.6 (0.2)	Gamma	0.67 (0.19)
λ_u	0.5 (0.2)	Gamma	0.49 (0.06)
$\lambda_{\bar{u}}$	0.9 (0.1)	Gamma	0.85 (0.09)
λ_x	0.7 (0.2)	Gamma	0.42 (0.07)
γ	0.29	Calibrated	
β	−0.29	Calibrated	
γ_x	0.5	Calibrated	
σ_ϵ	2 (10)	Inverse Gamma	1.18 (0.16)
σ_η	2 (10)	Inverse Gamma	0.64 (0.17)
σ_ψ	1 (10)	Inverse Gamma	0.43 (0.27)
σ_υ	0.5 (10)	Inverse Gamma	0.17 (0.09)
σ_μ	2 (10)	Inverse Gamma	0.77 (0.13)
σ_ω	0.4 (10)	Inverse Gamma	0.10 (0.04)
σ_ν	0.2 (10)	Inverse Gamma	0.06 (0.03)
σ_e	1 (10)	Inverse Gamma	0.34 (0.12)
σ_z	1 (10)	Inverse Gamma	0.49 (0.30)

Note: Standard deviation in parentheses.

TABLE A4

UC 4. Estimated and calibrated parameters. Quarterly data. Estimation period: 1990Q1-2019Q2

Parameter	*Prior*	*Prior distribution*	*Posterior*
λ_y	0.9 (0.2)	Gamma	0.90 (0.07)
γ	[0,1]	Uniform	0.03 (0.01)
δ	0.3 (0.2)	Gamma	0.12 (0.06)
σ_ϵ	0.7(10)	Inverse Gamma	0.47 (0.09)
σ_η	0.1(10)	Inverse Gamma	0.12 (0.06)
σ_υ	0.2(10)	Inverse Gamma	0.06 (0.01)

Note: Standard deviation in parentheses.

Process for the change in domestic inflation:

$$\Delta\pi_t = \gamma\hat{y}_t + \upsilon_t + \delta\upsilon_{t-1}, \tag{A16}$$

$$\upsilon_t = \eta_t. \tag{A17}$$

The model above is based on Kuttner (1994) (see also Hjelm and Jonsson, 2010) and relates the output gap ($\hat{y}$) to the change in domestic inflation ($\Delta\pi_t = \pi_t - \pi_{t-1}$). Modelling inflation in first differences implies that potential GDP corresponds to the level of GDP consistent with constant inflation (Table A4).

UC model 5: GDP mainland Norway, domestic inflation and unemployment (NAV)
Definition of GDP:

$$y_t = \hat{y}_t + \bar{y}_t. \tag{A18}$$

Process for the output gap:

$$\hat{y}_t = \hat{y}_{t-1} + \beta \hat{u}_t + \epsilon_t. \tag{A19}$$

Process for the growth in potential GDP:

$$\Delta \bar{y}_t = \Delta \bar{y}_{t-1} + \eta_t. \tag{A20}$$

Process for the unemployment gap:

$$\hat{u}_t = \lambda_u \hat{u}_{t-1} + \omega_t. \tag{A21}$$

Process for the trend in unemployment (NAIRU):

$$\Delta \bar{u}_t = \Delta \bar{u}_{t-1} + \nu_t. \tag{A22}$$

Process for domestic inflation (Table A5):

$$\pi_t = \lambda_\pi \pi_{t-1} + \gamma \hat{u}_{t-2} + \upsilon_t. \tag{A23}$$

UC models 6 and 7: GDP mainland Norway and growth in credit or in house prices
Definition of GDP:

$$y_t = \hat{y}_t + \bar{y}_t. \tag{A24}$$

Process for the output gap:

$$\hat{y}_t = \lambda_y \hat{y}_{t-1} + \gamma x_t + \epsilon_t. \tag{A25}$$

Process for the change in potential GDP:

$$\Delta \bar{y}_t = \Delta \bar{y}_{t-1} + \eta_t. \tag{A26}$$

TABLE A5

UC 5. Estimated and calibrated parameters. Quarterly data. Estimation period: 1990Q1-2019Q2

Parameter	*Prior*	*Prior distribution*	*Posterior*
λ_y	0.8 (0.1)	Gamma	0.9 (0.06)
β	−3.45	Calibrated	
λ_u	0.9 (0.2)	Gamma	0.9 (0.04)
γ	[−1,0]	Uniform	0.49 (0.22)
λ_π	0.5 (0.2)	Gamma	0.39 (0.12)
σ_ϵ	0.7 (10)	Inverse Gamma	0.41 (0.04)
σ_η	0.1 (10)	Inverse Gamma	0.04 (0.04)
σ_ω	0.2 (10)	Inverse Gamma	0.1 (0.01)
σ_ν	0.1 (10)	Inverse Gamma	0.07 (0.01)
σ_υ	0.2 (10)	Inverse Gamma	0.21 (0.02)

Note: Standard deviation in parentheses.

Process for the financial variable (x):

$$x_t = x_{t-1} + \upsilon_t, \tag{A27}$$

where x represents four-quarter growth in total credit (i.e. the sum of credit to households and firms) in UC 6 and house prices in deviation from their historical average in UC 7. In both models, the output gap's standard deviation (ϵ) is set equal to the standard deviation of change in the output gap estimated using an HP filter with λ equal to 1600, i.e. $\sigma_\epsilon = std(\Delta \bar{y}_t^{HP})$ where $\bar{y}^{HP}$ is the output gap estimated using a simple HP filter. The standard deviation of potential output growth is then scaled up by a factor z to ensure that the relative output gap variation is the same as for a normal HP filter, i.e. $\sigma_\eta = (1/z)\sigma_\epsilon$. Note that in this model the gap of the financial variable is calculated directly as the demeaned four-quarter growth rate, with no time variation. This implies no time variation in the trend growth of the financial variables (Tables A6 and A7).

SVAR 1: Mainland GDP and unemployment (NAV)

The model is based on Blanchard and Quah (1989) and includes two lags of the endogenous variables. A demand shock and a supply shock are identified. The demand shock is identified as a shock that has no long-term effects on GDP.

SVAR 2: Mainland GDP, unemployment (NAV) and domestic inflation

The model is based on (Cerra and Saxena, 2000) and includes two lags of the endogenous variables. A demand shock, a temporary supply shock and a permanent supply shock are identified. The demand shock is identified as a shock that has no long-term effects on GDP. The temporary supply shock is identified as a shock that has no long-term effects on domestic prices.

TABLE A6

UC 6. Estimated and calibrated parameters. Quarterly data. Estimation period: 1990Q1-2019Q2

Parameter	*Prior*	*Prior distribution*	*Posterior*
z	4.44	Calibrated	
λ_y	0.8 (0.2)	Gamma	0.71 (0.11)
γ	0.1 (0.2)	Gamma	0.15 (0.06)
σ_υ	1.2 (10)	Inverse Gamma	1.17 (0.07)

Note: Standard deviation in parentheses.

TABLE A7

UC 7. Estimated and calibrated parameters. Quarterly data. Estimation period: 1990Q1-2019Q2

Parameter	*Prior*	*Prior distribution*	*Posterior*
z	7.2	Calibrated	
λ_y	0.8 (0.2)	Gamma	0.80 (0.13)
γ	0.1 (0.2)	Gamma	0.05 (0.012)
σ_υ	3 (10)	Inverse Gamma	3.1 (0.18)

Note: Standard deviation in parentheses.

SVAR 3: Mainland GDP, unemployment (NAV) and domestic inflation

The model is based on (Furlanetto *et al.*, 2021b) and includes two lags of the endogenous variables. A temporary shock, a (potentially) permanent demand shock and a potentially permanent supply shock are identified. The temporary shock is identified as a shock that has no long-term effects on GDP. The permanent shocks are separated using sign restriction on the correlation between gdp and inflation. A positive demand shock increases inflation and output, while a positive supply shock increases gdp and lowers inflation.