A Reliable Output Gap Based on Imposing a Low Signal-to-Noise Ratio in a Beveridge-Nelson Decomposition of Simple Autoregressive Models *†

Güneş Kamber[1,3], James Morley[2,3], and Benjamin Wong[1,3]

[1]Reserve Bank of New Zealand
[2]University of New South Wales, Australia
[3]Centre for Applied Macroeconomic Analysis, The Australian National University

January 12, 2016

Abstract

The Beveridge-Nelson (BN) decomposition using an estimated autoregressive (AR) model of U.S. real GDP growth produces an output gap with small amplitude and a lack of persistence. We demonstrate such attributes are because the estimated AR coefficients imply a high signal-to-noise ratio for output growth. We show how to impose a low signal-to-noise ratio for an AR model, with the BN decomposition from this model producing an output gap that is both large in amplitude and persistent. Our estimated output gap also moves closely with the NBER-dated expansions and recessions and leads to better out-of-sample forecasts of output growth and inflation than output gaps based on other methods that impose a low signal-to-noise ratio, including deterministic detrending using a quadratic trend, the Hodrick-Prescott filter, and the bandpass filter. Most importantly, our estimated output gap is subject to smaller *ex post* revisions than other measures, yet maintains a stronger relationship to inflation for revised estimates, thus addressing the well known critique that output gap estimates are unreliable in real time.

JEL Classification: C18, E17, E32

Keywords: Beveridge-Nelson decomposition, output gap, signal-to-noise ratio

*Kamber: gunes.kamber@rbnz.govt.nz Morley: james.morley@unsw.edu.au Wong: benjamin.wong@rbnz.govt.nz

†The views expressed in this paper are those of the authors and do not necessarily represent the views of the Reserve Bank of New Zealand. We thank seminar and conference participants at the Reserve Bank of New Zealand, the Bank for International Settlements Asian office, University of Auckland, Otago University, the Australian National University, the University of Queensland, the Central Bank of Turkey, 2015 SNDE Symposium in Oslo, 2015 Melbourne Institute Macroeconomic Policy Meetings, 2015 Central Bank Macroeconomic Modelling Workshop and 2015 Workshop of the Australasia Macroeconomic Society, and especially George Evans, Adrian Pagan, Frank Smets, as well as our discussants, Michael Kouparitsas, Leif Anders Thorsrud and Glenn Otto, for helpful comments and suggestions. Any remaining errors are our own.

1 Introduction

The output gap is often conceived of as encompassing transitory movements in log real GDP. The Beveridge and Nelson (1981) (BN) definition of the trend of a time series is the long-run conditional forecast, so suggesting a natural approach to computing output gaps, as long as the researcher specifies a reasonable forecasting model. An autoregressive (AR) model for output growth is sensible place for a researcher seeking a reasonable forecasting model. The top panel of Figure 1 presents estimates of an output gap from a BN decompsition of an AR(1) forecasting model for the growth of U.S. real GDP.[1] The output gap from an AR(1) forecasting model is small in amplitude, lacks persistence, and does not match up well with the NBER chronology of expansions and recessions. The bottom panel of Figure 1 presents output gap estimates from the Greenbook and the output gap implied from the Congressional Budget's Office's estimate of potential output. Compared to the output gap obtained from a BN decomposition of an AR(1), these output gap estimates have higher persistence and are of larger amplitude. They are also pro-cyclical, if one matches these estimates against the NBER's chronology of recessions and expansions. While the estimates of the output gaps from the Greenbook and CBO do not neccessarily imply output gaps need to the described features, it does suggest that large amplitude and persistence are features that one may find consistent with a concept such as the output gap. Therefore, despite the BN decomposition providing a useful definition of the trend of a time series and an AR forecasting for output growth is a reasonable forecasting model, it is little wonder that one finds the output gap estimates of a BN decomposition of an AR forecasting model confronting. The output gap from the BN decomposition of a simple AR model possesses features such as a lack of persistence and has small amplitude because it has a high signal-to-noise ratio. That is, fluctuations of the trend are more variable than the cycle. The output gaps which features larger amplitudes will have a low signal-to-noise ratio, because the implied fluctuations of the trend are less variable than the cycle.

Our main contribution is to show that one can alter the signal-to-noise ratio of a simple AR forecasting model once one is armed with the insight that the signal-to-noise ratio of output gaps estimated from AR models is mechanically linked to the estimated AR coefficients. We therefore propose applying the BN decomposition based on Bayesian estimation of an autoregressive model with a pre-specified signal-to-noise ratio which corresponds with calibrating the sum of the autoregressive coefficients. Our approach is easy to implement in comparison to estimating UC models with an explicit prior on the signal-to-noise ratio (e.g., Harvey et al. (2007)). When applied to US quarterly log real GDP, our approach produces an output gap that is highly persistent, has large amplitude,

[1] An AR(1) for U.S. output growth is a natural reasonable forecasting model a researcher can choose, given it is obtained using the Bayesian Information Criteria, a common means of selecting a parsimonious forecasting model.

and match up with NBER-dated expansions and recessions. Our approach also reasonably well in out-of-sample forecasts of output growth and inflation at short to medium term horizons and is little revised compared to well known alternatives. Most importantly, our approach is reliable in a real time environment, to the extent that it is often little revised *ex post*. Our approach thus addresses a key critique by Orphanides and van Norden (2002) that common output gaps estimates in real time, like the HP filter, are often heavily revised and thus unreliable for real time analysis.

Being able to adequately address Orphanides and van Norden (2002) critique is a key result for the paper. That Orphanides and van Norden find common approaches to the output gap are unreliable in real time undermines the usefulness of the output gap in real time policy environments and forming an accurate gauge of current economic slack. The fact that our approach produces an output gap is little revised, and thus reliable, is not coincidental but a product of our choice of working with a simple AR model. In principle, the BN decomposition can be applied to any forecasting model, including multivariate time series models, such as Vector Autoregressions, structural models such as DSGE models, or nonlinear time series models with features such as Time Varying Parameters or Markov Switching. Our choice to alter the humble AR forecasting model is deliberate. Because the output gap from a BN decomposition are based on the estimated parameters of the forecasting model, it is thus mechanically that any instability in the estimated parameter in real time will produce output gaps that are heavily revised. The choice to alter the simple AR forecasting is precisely because the estimates of the AR coefficients are likely to be stable, unlike more complex forecasting models. Therefore, a natural outcome of our choice is an output gap that is little revised and thus reliable.

It is also worth noting that our approach does address the omission of multivariate information in the forecasting model and a structural breaks in the long-run growth rate, thus addressing issues with trend-cycle decomposition raised by Evans and Reichlin (1994) and Perron and Wada (2009). At the same time, because we use a BN decomposition, our approach takes explicit account of a stochastic trend in U.S. real GDP unlike existing methods like the HP filter or Bandpass filter, where addressing of the stochastic trend, if any, can be less obvious.

The rest of this paper is structured as follows. Section 2 presents our proposed approach and applies it to US real GDP. Section 3 presents analyses the revision properties of our approach and assesses the reliability of our approach, relative to other existing methods. Section 4 provides a thorough motivation for our approach, in particular why one might choose to lower the signal-to-noise ratio of the AR forecasting model. Section 5 presents a formal pseudo-real time forecast comparison for output growth and inflation. We then address some robustness issues and conceptual issues with our approach before concluding.

2 A Reliable Output Gap

2.1 The Beveridge-Nelson Decomposition and the Signal-to-Noise Ratio

Beveridge and Nelson (1981) define the trend of a time series as its long-run conditional expectation minus any a priori known (deterministic) future movements in the time series. In particular, letting $\{y_t\}$ denote a time series process with a trend that follows a random walk with drift, the BN trend, τ_t, at time t is

$$\tau_t = \lim_{j \to \infty} \mathbb{E}_t \left[y_{t+j} - j \cdot \mathbb{E} \left[\Delta y \right] \right].$$

The simple intuition behind the BN decomposition is that the long-horizon conditional expectation of a time series is the same as the long-horizon conditional expectation of its trend under the assumption that the long-horizon conditional expectation of the cycle is equal to zero. By removing the deterministic drift from the long-horizon forecast, the conditional expectation remains finite and becomes an optimal estimate of the random walk trend for a UC process (see Watson, 1986; Morley et al., 2003).

To implement the BN decomposition, it is typical to specify a forecasting model for the first differences $\{\Delta y_t\}$ of the time series. First differencing explicitly deals with the stochastic trend in the time series as it models for permanent shocks. We therefore will intepret the permanent component of the BN decomposition as the trend.

Based on the sample autocorrelations of many macroeconomic time series, including the first differences of log real GDP, it is natural when implementing the BN decomposition to consider an AR(p) forecasting model:

$$\Delta y_t = c + \sum_{j=1}^{p} \phi_j \Delta y_{t-j} + e_t, \tag{1}$$

where the forecast error $e_t \sim N(0, \sigma_e^2)$. For convenience when determining the signal-to-noise ratio below, let $\phi(L) \equiv 1 - \phi_1 L - \ldots - \phi_p L^p$ denote the autoregressive lag polynomial, where L is the lag operator.

Although the AR(1) might seem a reasonable forecasting model, we have already shown in Figure 1, and argued in the Introduction, a key misgiving with a BN decompositions based on a simple autoregressive model. In particular, the output gap does not match up at all with the NBER chronology of recessions and expansions. The output gap is also small in amplitude, suggesting that most of the fluctuations in U.S. real GDP have been driven by trend.

To understand why the BN decomposition based on Maximum Likelihood Estimation (MLE) for an AR(1) forecasting model of output growth produces an output gap with such features, it is useful to define a signal-to-noise ratio for output growth as the ratio of the variance of shocks to trend to the variance of the overall forecast error for output growth:

$$\delta \equiv \sigma^2_{\Delta\tau}/\sigma^2_e = \psi(1)^2, \tag{2}$$

where $\psi(1) \equiv \lim_{j\to\infty} \frac{\partial y_{t+j}}{\partial e_t}$ is the long-run multiplier for a forecast error and is the key summary statistic for a time series process when calculating the BN trend based on a forecasting model. In particular, $\Delta\tau_t = \psi(1)e_t$. For an AR(p) model, this long-run multiplier has the simple form $\psi(1) = \phi(1)^{-1}$ and, based on the estimated AR model for postwar US real output growth data, $\hat{\phi}(1)^{-1} > 1$.[2] Therefore, the implied signal-to-noise ratio δ is large and greater than unity, corresponding to the estimated trend being more volatile than the time series itself. The counterintuitive positive sign of the estimated output gap during NBER recessions reflects the positive persistence in the output growth, with a negative shock in a recession implying further below-average growth, meaning that the series is above the long-run conditional forecast minus deterministic drift.

The insight that the signal-to-noise ratio is mechanically linked to the sum of the AR coefficients is a powerful one. Many implementations of using a direct signal-to-noise prior require some form of posterior simulation using a Bayesian approach (see, e.g. Harvey et al., 2007). We now show that implementing a signal-to-noise ratio can be trivially achieved through a suitable manipulation of Equation 3 through the insight that the signal-to-noise ratio is linked to the sum of the AR coefficients. We first transform the AR(p) model to its Dickey-Fuller representation:

$$\Delta y_t = c + \rho \Delta y_{t-1} + \sum_{j=1}^{p-1} \phi_j^* \Delta^2 y_{t-j} + e_t \tag{3}$$

where $\rho \equiv \phi_1 + \phi_2 + \ldots + \phi_p = 1 - \phi(1)$ and $\phi_j^* \equiv -(\phi_{j+1} + \ldots + \phi_p)$. The Dickey Fuller representation reduces specification of the signal-to-noise ratio to a single parameter, ρ. Once one specifies, δ, this maps into a corresponding ρ. Equation 3 can then be estimated and a BN decomposition is performed to obtain the output gap.

Estimation is straightforward without the need of using complicated posterior simulators or complicated nonlinear restrictions. MLE estimation entails a single parametric restriction on ρ. While it is possible to implement our approach using MLE, we will opt for a Bayesian approach, mainly because this allows utilising a shrinkage prior on the higher lags to prevent overfitting of the higher lags, reducing the issue of specifying the lag order into an irrelevant one. Because Equation 3 is just a linear regression, we can implement estimation using a Bayesian approach by fixing a dogmatic prior on ρ through

[2] This is because the sum of the autoregressive coefficients of an AR model of real GDP growth is always estimated to be positive.

a pre-specified signal-to-noise ratio and a shrinkage prior which tightens around zero with each lag to avoid overfitting. Readers familiar with the Minnesota class of priors will recognise that the posterior distribution of Equation (3) has a closed form solution, and the posterior mode can be easily obtained without the need for a posterior simulator.

Regardless of one's modelling persuasion, our proposal allows computation of an output gap in a parsimonious and straightforward manner, which should appeal to practitioners. We relegate specifics about the implementation of our procedure, as well as details about the degree of shrinkage, to an appendix.

Figure 2 presents the output gap based on an estimated Bayesian AR(12) of U.S. real GDP growth with $\delta = 0.1$, which is comparable to an HP filter with $\lambda = 1600$. We refer to our method as BN Bayes. A cursory glance at the estimated output gap suggests that our approach is more successful than the standard BN decomposition in producing an output gap that is consistent with policymakers' beliefs. In particular, we are able to capture the turning points of all the NBER dated recessions. The amplitude of the output gap is also large and persistent, as per our expectation from imposing a low signal-to-noise ratio.

2.2 Revision Properties

Having demonstrated our approach of utilising the insight of linking the signal-to-noise ratio with the AR coefficients, we now assess the revision properties of our proposed estimated output gap. In motivating our approach, we had explicitly set out to address Orphanides and van Norden (2002) critique that methods of estimating the output gap are not reliable in real time. Orphanides and van Norden (2002) show most real time revisions of the output gap are due to the lack of information about the future, rather than data revisions. That is, it the the method of extracting the output gap in real time that is deficient, and not the ability of the models to predict revisions. We thus consider a pseudo-real-time exercise using final vintage data (2015Q3) rather than do a real time analysis. The pseudo-real time exercise is also appropriate because our approach is not designed to model data revisions, but rather a demonstration of a method of circumventing deficiencies in current methodological approaches.

We lay out several existing and widely-used alternatives. We first consider BN decomposition on various ARIMA forecasting models and Unobserved Components (UC) models estimated using MLE. As it is well known, the reduced form of a UC model is an ARIMA model, and thus under certain conditions, both will imply exactly the *same* decomposition for trend and cycle (see Morley et al., 2003). We consider BN decompositions of an AR(1), an AR(12) and an ARMA(2,2) estimated on real output growth. We also look at a UC model like the one estimated by Harvey (1985) and Clark (1987). The ARMA(1) is selected based on the Bayesian Information Criteria (BIC) on ARMA models of various AR and MA lag orders. The AR(12) is used to understand the effect of imposing a long lag order on an autoregressive model. Note that using standard information criteria will almost certainly preclude a researcher from selecting an AR(12) model ahead of more parsimonious specifications. The ARMA(2,2) is considered as it is shown by Morley et al. (2003) to be an unrestricted version of a popularly used UC model by Watson (1986). The Watson UC model features a random walk trend and an AR(2) cycle. As Morley et al. show, the restrictions implied by the Watson (1986) UC model are rejected by the data, suggesting an ARMA(2,2) is a good starting point when considering UC models that feature a random walk trend and an AR(2) cycle. We also consider the Harvey and Clark type UC model, similar to that considered by Orphanides and van Norden (2002). The Harvey-Clark UC model differs from the Watson (1986) model to the extent it features a time varying drift. To complete the set of alternative approaches we consider, we look at deterministic trends and non-parametric filters, as they are often used in practice. We therefore compare against a model with both a deterministic linear and quadratic trend, the Hodrick and Prescott (HP) (1997) filter, and the Bandpass (BP) filter by Baxter and King (1999) and Christiano and Fitzgerald (2003). For the HP filter, we set a smoothing parameter, $\lambda = 1600$, as is commonly done for quarterly data. For the BP filter, we target frequencies between 6 and 32 quarters, as is commonly done in business cycle analysis. It is worth noting whatever documented misgivings about the HP and BP filters (e.g. Cogley and Nason, 1995; Murray, 2003; Phillips and Jin, 2015), they remain perhaps the most commonly used approaches in practice.

Figure 3 compares the pseudo-real-time output gap estimate and the current measure of the output gaps as of the 2015Q3 vintage of real GDP data. In the top panel, we evaluate the revision properties of our proposed approach. The remaining subplot presents the other approaches. We first focus on the various BN decompositions. Of all the BN models considered, the AR(1) and ARMA(2,2) suggest a BN cycle which is be non-persistent, small in amplitude and not procyclical. Adding additional lags appears to help as the AR(12) appears to be pro-cyclical as it does appear to coincident with the NBER classification of business cycles. Even so, the AR(12) has low amplitude, consistent with our observation that AR forecasting models have a signal-to-noise ratio much larger than 1. One striking observation is that all the approaches using a BN decomposition are less revised. A key reason why real-time estimates of the BN approach are hardly revised is because the estimated coefficients of the forecasting model hardly change when an additional data point is added in real time. Our approach, being derived from a BN decomposition of an AR(p) model, is thus hardly revised. Even though approaches using BN decompositions appear little revised, a striking observation within the class of BN decompositions is that the more highly parameterised models like the AR(12) and ARMA(2,2) are more heavily revised relative to the AR(1) model. This observation suggests overparameterisation and overfitting compromises the real time reliability of BN approaches. This is a crucial reason why we specify an arbitrary long lag order as well as impose a shrinkage prior. The shrinkage prior prevents overfitting, but the long lag order allows for the modelling of richer dynamics, if present in the data. The revision properties of our approach suggests our modelling approach achieves a reasonable compromise between overfitting and considering richer dynamics.

We turn our attention to the other approaches. As mentioned, the non-BN approaches are considerably revised *ex post*, and thus less reliable. The HP and BP filters are two-sided filters, whose estimate of trend is contingent on future data. The arrival of additional data point thus heavily alters the gap, as shown in the Figure.

While the results from eyeballing Figure 3 suggests our approach should be appealing from a reliability perspective, we quantify these revision properties by calculating revision statistics, similar to the exercise by Orphanides and van Norden (2002) and Edge and Rudd (forthcoming). Therefore, similar to them, we calculate similar revision statistics. First, we calculate the correlation between the pseudo-real time estimate of the output gap and the final estimate of the output gap. We also consider two different measures of the revisions, namely the standard deviation and the root mean squre (RMS). The RMS approach is designed to penalise persistently large revisions more heavily relative to the standard deviation approach. Both the standard deviation and RMS measures are normalised by the standard deviation of the final estimate of the respectively output gaps to enable comparison as all the different methods produce output gap with differing amplitudes.[3] Finally, we also compute the number of times the pseudo-real time estimate produces an output gap which produces the same sign as the final estimate. We define for the final estimate of the output gap as the output gap estimated from 1970Q1-2012Q4 on the most recent vintage of real GDP data which we use for the main empirical analysis (i.e. 2015Q3) as the more recent estimates near the end of the sample may end up becoming more heavily revised in the future.

Table 1 reports the revision statistics. As one gathers from eyeballing Figure 3, the BN approaches produce a pesudo-real-time output gap which is highly correlated with the final estimate. Once again, the more highly parameterised the forecasting model used for the BN decomposition, the lower the correlation with the final estimate, suggesting once more that overfitting can be an issue. Our approach, which also utilises a shrinkage prior does relatively well with a correlation of 0.99 with the final estimate. The other approaches can be quite mixed, from a low of 0.55 for the HP filter to 0.75 for the Harvey-Clark UC model. Our proposed approach also does well in terms of size of revisions, with these revisions less than a quarter of a standard deviation of the final estimate of the output gap. Some of these approaches have revisions which are of the magnitudes comparable to one standard of the final estimate of the output gap. Even the highly parameterised BN approaches do

[3]Note these statistics are referred to by Orphanides and van Norden (2002) as noise-to-signal ratios, but apart from the labelling, have nothing to do with our approach of imposing a signal-to-noise ratio. We will use the terms standard deviation and RMS of the revisions in order to circumvent any confusion.

poorly on the revision metric, with the BN decomposition of an ARIMA(2,2) featuring revisions roughly about half a standard deviation of the final estimate of its respectively output gap. Finally, we turn our attention to whether the sign of the output gap changes, once one is endowed with future information. Once again, the BN decomposition of the various forecasting models tend to do much better. In fact, the BN decompositions as a whole do much better than alternatives, where an excess of 75% of the sign of the real time estimates are never revised. Within the class of BN decompositions, our approach still does relatively well, being able to infer the correct sign as the final estimate almost 90% of the time.

To summarise, BN decompositions of ARMA forecasting models can be fairly reliable in a pseudo-real time environment. This is because the addition of an additional data point does not drastically alter the estimates of the forecasting model. Therefore, they are more reliable than approaches such as HP and BP filters and deterministic detrending. Even so, within the class of BN decompositions of ARMA models, parsimony seems to be key for reliability of the estimated output gap. This is not so much a surprise, given models which are highly parameterised models, such as the AR(12), or feature Moving Average (MA) terms like the ARMA(2,1,2), will tend to feature coefficient estimates that can be more unstable with the addition of future data. The AR(12) is likely to overfit and the ARIMA(2,2), as we will argue later, is likely to suffer from weak identification due to the small sample size.

Our approach to impose a signal-to-noise ratio in a simple AR(p) model in addition to a shrinkage prior helps in producing a more reliable output gap. Within the class of models which feature a highly persistent output gap, with large amplitude and is pro-cyclical, our analysis suggest our approach is best. While our alteration of an AR model does not seem to compromise the reliability of the estimates of the output gap, he AR(1) does best on the reliability metrics which we have shown so far. Why then should one impose a signal-to-noise ratio, given the BN decomposition of an AR(1) model appears to be extremely reliability, albeit confronting as it lacks features one is familiar with, such as persistence, amplitude and pro-cyclicality? We address this issue in the next section of why one might find it desirable to impose a low signal-to-noise ratio.

2.3 Robustness

To explore the sensitivity of our approach of estimating the output gap, we first address Perron and Wada's (2009) claim that U.S. GDP should be modelled with a break in the trend growth rate in 1973Q1. In our approach, because we estimate an AR(p) model on output growth, we assume U.S. GDP evolves as a random walk with drift. We therefore we estimated our model, with the addition of a dummy in the drift term to account for a break in the trend in 1973Q1. The left panel of Figure 6 presents results with and without accounting for the structural break. There is a very slight shift upwards in the output gap estimates after the 2000s if one accounts for a break in 1973Q1, but the differences are trivial. We therefore conclude our approach of estimating the output gap is insensitive to a break in U.S. GDP growth in 1973Q1.

We also studying the sensitivity of our approach by varying the signal-to-noise ratio. The bottom panel of Figure 6 presents a moderate signal-to-noise ratio, $\delta = 0.3$, and a high signal-to-noise ratio, $\delta = 0.6$, alongside our benchmark choice of $\delta = 0.1$. Not surprisingly, increasing the signal-to-noise ratio mechanically reduces the amplitude of the output gap. Apart from a mechanical reduction in the amplitude, we note the shape of the estimated output gap is little changed, with the persistence profile virtually unaltered.[4] Because the profile of fluctuations in the estimated output gap are unaltered even as we increase the signal-to-noise ratio, observations of turning points, revision properties and real time forecast performance are insensitive to the choice of δ. Finally, we wish to understand if our choice to calibrate the signal-to-noise ratio has any effect, beyond that of just the change in amptitude. We therefore compare relative to the BN decomposition of an AR(12). Based on the coefficients estimated for an AR(12) for our sample, $\hat{\delta}$ is roughly 2. We can see that there are some differences in the profile of fluctuations in the output gap, beyond that of the different amplitude of the estimated output gap. In particular, the correlation is about 0.34, which is indicative that the imposition of a signal-to-noise ratio is altering the AR(12) forecasting beyond that that already in the standard.

3 General Discussion

3.1 Motivation for Imposing a Low Signal-to-Noise Ratio when Estimating the Output Gap

To recap, we have laid out a proposal for estimating output gaps by imposing a low signal-to-noise ratio. Our proposal output gap are reliable in the Orphanides and van Norden (2002) sense, in the sense that our approach of estimating the output gap features smaller data revisions. Our approach does well compared to many well known alternatives, apart form having marginally poorer revision properties compared to a BN decomposition of a simple AR(1) model for output growth. The BN decomposition of an AR(1) has a high signal-to-noise ratio. To the extent that one is agnostic about the signal-to-noise ratio, there is little reason to deviate from the BN decomposition of the AR(1) model, especially if real time reliability is the sole criteria of selecting an approach to estimate the output gap. In other words, one can only really justify using our approach to estimating a reliable output gap if there is reason to believe that a low signal-to-noise ratio represents the true

[4]We note correlation of the different estimated output gaps by varying the signal-to-noise ratio is more than 0.95.

state of the world. Whether a low or high signal-to-noise ratio represents the true state of the world remains unresolved. While considerable empirical research has found evidence for the presence of a volatile stochastic trend in real GDP (e.g. Nelson and Plosser, 1982; Morley et al., 2003), this view does not go unchallenged (e.g. Cochrane, 1994; Perron and Wada, 2009). Also, there is some evidence that transitory movements in real GDP are asymmetric (e.g. Kim and Nelson, 1999; Sinclair, 2009; Morley and Piger, 2012; Morley, 2014). To the extent that real GDP is generated univariate with a volatile stochastic trend represents the true state of the world, any imposition of a low signal-to-noise ratio must necessarily imply a deliberate model misspecification, and thus cannot be defended.

One can therefore only defend our approach of imposing a low signal-to-noise ratio if there is reason to believe that a low signal-to-noise ratio represents the true state of the world. A legitimate reason to believe the signal-to-noise ratio is much lower than that given by unrestricted AR models is that the forecasting model could be misspecified. It is known that the BN decompositions of a univariate AR model yield a high signal-to-noise ratio because estimated univariate ARMA model imply little forecastability for output growth (see, e.g. Nelson, 2008). It is known, at least since the work of Evans and Reichlin (1994), that the addition of multivariate information which can forecast output growth mechanically lowers the signal-to-noise ratio. In other words, the unrestricted AR model implies little forecastabilty for output because relevant information for forecasting output growth has been omitted. To the extent that the univariate ARMA forecasting model omits relevant variables which aid in forecasting output growth, the ARMA forecasting model is misspecified. If one takes this perspective, then the high signal-to-noise ratio one obtains from unrestricted ARMA forecasting models is a mere artefact of model misspecification. This line of argument suggests that if the world was *not* one where real GDP was generated univariate with a volatile stochastic trend, then the imposition of a low signal-to-noise ratio in a univariate model of real GDP merely corrects for the misspecification because the forecasting model is multivariate.

To the extent the underlying process generating output growth is multivariate offers the modeller three choices in terms of estimating the output gap using BN decompositions. First, if one knew the true nature of the multivariate model, the most straightforward solution must be to estimate that correctly specified multivariate model. This point is obvious if one ever knows the true state of the world. The second approach is to model MA terms. If the true multivariate process was multivariate, like a vector autoregression (VAR), then the omission of the multivariate information and modelling real GDP univariate would mean the VAR forecasting model can be cast into an ARMA forecasting model, with the MA terms proxy for the omitted information. The final option is our approach. Our approach is to recognise that relevant multivariate information lowers the signal-to-noise ratio, as per the insight by Evans and Reichlin (1994). Therefore, to model a univariate model requires one to lowers the signal-to-noise ratio as we have done.

We therefore conduct a Monte Carlo simulation to better understand how missing multivariate information may give rise to a low signal-to-noise ratio and whether our approach, vis-a-vis the approach of modelling MA terms can help recover the true state of the world. We consider a simple bi-variate VAR of lag order 4 featuring output growth and unemployment. Using U.S. data from 1948Q1-2015Q3, the null of unemployment does not Granger causes output growth can be easily rejected at 1%, suggesting unemployment as relevant multivariate information for forecasting output growth. As unemployment is relevant information to forecast output growth, the signal-to-noise ratio from an estimated AR(4) for output growth falls from 2.30 to 2.04 when the extra information is modelled as a bivariate VAR(4), as per the insight by Evans and Reichlin (1994). We therefore use a VAR(4) of output growth and unemployment as a Data Generating Process (DGP), with the parameters in the DGP set to that which is estimated from the sample. We obtain output gaps using BN decompositions using our simulated data first using our approach. We then compare against using an AR(1) as a forecasting model and a forecasting model which uses the BIC to choose the ARMA order. We do both Monte Carlo simulations of 2000 artifical datasets each with sample sizes of 250 and population analysis. As the DGP is a bivariate VAR, the true output gap is a multivariate BN decomposition, using a VAR(4) as the forecasting model.

The top panel of Table 2 first presents correlation of the various estimated output gap relative to the true output gap. Our proposed approach of calibrating the signal-to-noise ratio does reasonably well whether with repeated sampling in small samples, or in population. Note that these results are reasonably good even though the true signal-to-noise ratio is 2.04, but we calibrate ours to 0.1.[5] We observe the BN decomposition of a simple AR(1) bears little resemblance to the true output gap, suggesting that there is possible merit to the claim that the output gaps one gets from BN decompositions of low order AR models is perhaps due to model misspecification of omitting the relevant multivariate information. Relative to using an AR(1), allowing the possibility of MA terms helps mitigate multivariate information. This can be seen from the higher correlation with the true multivariate cycle once one considers the ARMA forecasting model. However, the gain from considering MA terms in the ARMA forecasting model seems only modest in small samples, but does very relatively well in population.

While it appears MA terms help mitigate the omission of multivariate information, as we expect theoretically, this seems to only work best in large samples. In small samples, our approach appears to still do better than modelling MA terms. A key reason for this is that there is weak identification of the MA terms in small samples. This is illustrated in the bottom panel of Table 2. In our repeated sampling of the small sample, whenever MA

[5]The main reason, as will be clear in the robustness analysis, is because fluctuations in the estimated output gap in our approach are very robust across different values of δ, therefore implying the only real difference is the amplitude of the cycle.

terms are chosen by the information criteria, over a quarter features near cancellation of the MA roots.[6] If we look at the proportion of Monte Carlo draws that fit no MA terms or features near cancellations, this ratio rises to 65%. This result explains why MA terms seem to do well in population, but not in small samples. The weak identification of MA terms means that in sample sizes typical for quarterly macroeconomic time series, the econometrician will find it challenging to fit meaningful MA terms.

The Monte Carlo simulation somewhat justifies the scepticism of BN decomposition based on MLE of univariate ARMA models and why small amplitude and non-persistent output gaps can be somewhat confronting. If the true economic environment were indeed multivariate, the BN decomposition based on MLE would produce very poor estimates of the permanent component, and thus the output gap. A direct solution to the problem illustrated above is to of course estimate a multivariate model. Indeed, if one knew the correct specification of the multivariate model, then one should directly estimate the said model. In practice, one will *not* be endowed with the knowledge of the true multivariate model. The precise nature of the multivariate information could be unobserved, unmeasured or mismeasured. A more subtle implication is that adding more multivariate information in empirical work also risks overfitting, and therefore compromise the real time reliability, as we have seen in the previous section. Moreover, the Evans and Reichlin (1994) result shows the inclusion of more (Granger Causing) variable will mechanically lower the signal-to-noise ratio, implying any choice variable set in multivariate system is an implicit calibration of the signal-to-noise ratio. Our approach of imposing a prior in the signal-to-noise ratio explicitly specifies the signal-to-noise ratio in a transparent manner and can also be seen as an attempt to incorporate multivariate information. Indeed, one can think of our approach as reducing the signal-to-noise ratio, as the Evans and Reichlin (1994) approach does, but bypassing the need to specify or choose the precise nature of the multivariate information. The latter point is especially relevant if one desires univariate detrending techonology without fully specifying an entire multivariate system.

Finally, we wish to make two comments on our results on weak identification and contrast our approach of working with a pure AR model relative to a UC model. Note the reduced form of a UC model is an ARMA forecasting model of output growth with non-trivial MA terms. First, as we have shown MA terms may suffer from weak identification, it is easy to reconcile why our previous results of the Harvey-Clark UC model and the ARMA(2,2), which is an unrestricted Watson UC model, are not reliable in real time. Second, from the perspective of UC models, we allow for correlation between the trend and cycle, which Morley et al. (2003), Dungey et al. (2015) and Chan and Grant (2015) have shown to be important empirically for U.S. real GDP. Without estimating the UC model, but directly estimating a long order AR(p) model, which is approximately the

[6] We classify a near cancellation if the absolute value of the largest inverted AR and MA root differs by less than 0.08.

reduced form of a UC model, we allow for this correlation without the need to specify the form of this correlation, as one would need to in the UC model. In choosing to work with a long order AR(p) model, we obviate the issues associated with weak identification of the MA terms, but allow the long AR lag order to mimic non-trivial MA dynamics, if they are indeed present in the data.

3.2 Pseudo-Real-Time Forecast Evaluation

We now conduct pseudo-real-time forecasting exercises. Naturally, if the output gap we propose has useful information to allow one to gain a good reading of the current state of the business cycle, then this useful information should manifest itself in its ability to forecast certain macroeconomic variables associated with the state of the business cycle.

We compare our proposed output gap measure by appealing to two forecast evaluation metrics previously established by the literature, namely future output growth and inflation. All of our forecasts are once again pseudo-real-time (i.e. we use the final vintage data). Our forecast evaluation starts in 1970Q1. We use an expanding window for estimation. The first estimate of an output gap we have is 1947Q2. We use the full extent of the data sample for our forecast evaluation after adjusting for the number of lags in the forecasting equation.

Output Growth Forecast Nelson (2008) argues for using future growth forecasts as a metric to evaluate competing estimates of the output gap. The underlying intuition is that if the output gap suggests output is below trend, this should imply faster future output growth as output returns towards the trend to close the gap. Conversely, if output is above trend, one should forecast slower output growth for output to return back towards the trend. The point is that the cycle of a time series must necessarily return to zero in the long run, and a good estimate of the output gap should be able to forecast this reversion. For a h period ahead output growth forecast, we consider a forecasting equation similar to Nelson (2008),

$$y_{t+h} - y_t = \alpha + \beta gap_t + \epsilon_t \tag{4}$$

where y is the natural log of real GDP, gap is the output gap measure under consideration, ϵ is a residual distributed under all the standard assumptions and α and β are coefficients estimated using least squares. Therefore, for a good measure of the output gap, we expect $\beta < 0$ and the inclusion of the output gap in the forecast equation to provide a better forecast for h period ahead growth.

We first confirm that β is the correct sign for all the different output gap measures, before proceeding to do our out of sample forecasting exercise. Figure 4 presents the pseudo out-of-sample forecasting results. The Relative Root Mean Squared Errors (RRMSE) are relative to our proposed approach. The results of the out-of-sample RRMSE includes 90% confidence bands obtained by inverting the Diebold and Mariano (1995) test.

We make two key observations. First, the output gaps constructed using the BN decomposition are better than using filters and UC models. This further vindicates our choice to work with a BN decomposition as our proposed approach inherits these good features. Second, within the class of BN models, similar to that of the revision statistics, parsimony seems to be the important. In particular, the AR(12) and ARMA(2,2) both do worse than our approach. Therefore, the growth forecasting exercise mimics many of the results we see for the revisions and reliability statistics that were presented earlier. In particular, BN approaches do well, and more parsimonious forecasting models using the BN decomposition do even better. It follows that our approach using a BN decomposition and a shrinkage prior does well on the output growth metric.

Inflation Forecast We also consider a Phillips Curve type inflation forecasting equation to evaluate our proposed output gap measure. Similar to, among others, Stock and Watson (1999, 2008) and Clark and McCracken (2006), we use a fairly standard specification in the inflation forecasting literature. We specify the following autoregressive distributed lag (ADL) representation for our psuedo-real- time h period ahead Phillips Curve inflation forecast

$$\pi_{t+h} - \pi_t = \gamma + \sum_{i=0}^{p} \theta_i \triangle \pi_{t-i} + \sum_{i=0}^{q} \kappa_i gap_{t-i} + \upsilon_t. \tag{5}$$

We choose the lag orders of the forecasting equation, namely p and q above, using the BIC.[7] As commonly done (see, e.g. Stock and Watson, 1999, 2008; Clark and McCracken, 2006), we apply the information criteria to the entire sample and run the pseudo real-time exercise using the same number of lags, implicitly assuming the inflation analyst *a priori* knows the optimal lag order. The set of lag orders we consider for our ADL forecasting equation are $p \in [0, 12]$ and $q \in [0, 12]$.

Figure 5 presents the out-of-sample RRMSE of the pseudo-real-time inflation forecasting exercise relative to our proposed output gap measure. Once again, like with the output growth forecasting, we compute bounds on the 90% interval of the Diebold and Mariano (1995) test. We note that our proposed output gap measure does well. In particular, calibrating the signal-to-noise ratio allows us to out-perform all other models based on the BN decomposition. We also generally do better than the HP filter, BP filter and deterministic trend. Even so, we state the differences in inflation forecast performance using the different output gap estimates are fairly similar other, with most RRMSE within the 1 to 1.05 range, indicating the gains in changing the output gap measure for forecast-

[7] The specification imposes a unit root in inflation and so implies an accelerationist view of the Phillips Curve. The forecasting equation though is standard, as per the references listed above.

ing inflation can be marginal. In particular, most of these RRMSE are not statistically significant. To some extent, this is not entirely surprising. Contributions such as Atkeson and Ohanian (2001) and Stock and Watson (2008) show that real-activity based Phillips Curve type forecast may not be useful for forecasting inflation. In some sense, our results are somewhat a manifestation of what is commonly found in the inflation forecasting literature. However, we note that our approach is still reasonably competitive and may be slightly better than competing options in terms of being a good real-time measure of economic slack. In particular, we do produce statistically significantly better Phillips Curve forecasts at some horizons relative to approaches such as the HP filter and the deterministic trend. It is noteworthy that none of the alternative output gap measures outperform our approach in a statistically significant way.

3.3 Are Revisions Useful in Understanding History?

If objection to the output gap is its real time unreliability, as is the central claim of the influential work by Orphanides and van Norden (2002), the ideas developed in this paper should suffice as a useful address of their critique. In this section, we wish briefly address whether an output gap that is little revised is necessarily better in understanding history.

However, in the presence of new information, is our output gap little revised because there is little incorporate beyond the real time view, or is it because the other methods are systematically incorporating new information to produce a more accurate *ex post* view of the output gap?

4 Conclusion

The Beveridge-Nelson trend-cycle decomposition, while conceptually elegant, is used infrequently because it produces an estimated output gap may appear confronting because it ascribes more variation to fluctuations in the trend than to the cycle because it has a high signal-to-noise ratio. In this paper, we use the insight that the signal-to-noise ratio is mechanically linked to the AR coefficients of the forecasting model and use this insight to lower the signal-to-noise ratio of the estimated output gap when applying applying the BN decomposition. This approach produces a more intuitive output gap, with larger amplitude and coherence with recessions and expansions. Crucially, in contrast to other methods that impose a low signal-to-noise ratio, this approach also produces an output gap with similar good forecasting properties to the standard BN decomposition. The approach is also simple to implement.

There is much room to expand our analysis. We plan to extend our analysis to real GDP for other economies and also to study the use of our approach in a real-time policy environment. It is also an open question what the optimal signal-to-noise ratio should be and how should one choose an optimal lag length for the forecasting model. These are all issue that we will address in future work.

Appendix

A1 Specification of the Prior

To recap our approach, we estimate the following AR(p) model with a dogmatic prior on the sum of the autoregressive coefficients:

$$\Delta y_t = c + \sum_{j=1}^{p} \phi_j \Delta y_{t-j} + e_t, \tag{A1}$$

where the forecast error $e_t \sim N(0, \sigma_e^2)$. For convenience when determining the signal-to-noise ratio below, let $\phi(L) \equiv 1 - \phi_1 L - \ldots - \phi_p L^p$ denote the autoregressive lag polynomial, where L is the lag operator. We wish to put a prior on the sum of the autoregressive coefficients $\phi(1)$. The signal-to-noise ratio, δ, is defined as the ratio of the variance of shocks to trend to the variance of the forecast error, $\sigma^2_{\Delta\tau}/\sigma^2_e$. As shown in the paper, the AR(p) model has a simple relation between the signal-to-noise ratio and the sum of the autoregressive coefficients where $\phi(1) = \delta^{-2}$. In our benchmark results, we set $\delta = 0.1$.

To also recap, estimating the AR(p) model directly often implies no close form solution. However, we can transform the above model to Dickey-Fuller form:

$$\Delta y_t = c + \rho \Delta y_{t-1} + \sum_{j=1}^{p-1} \phi_j^* \Delta^2 y_{t-j} + e_t \tag{A2}$$

where $\rho \equiv \phi_1 + \phi_2 + \ldots + \phi_p = 1 - \phi(1)$ and $\phi_j^* \equiv -(\phi_{j+1} + \ldots + \phi_p)$. Equation A2 is just a linear regression with the first lag of output growth and the lagged difference of output growth as covariates. If we treat the posterior variance of the forecasting model, ϵ_e, as fixed at the least square estimate, like how one would do in the case of a Minnesota prior, we can estimate Equation A2 without a posterior simulator by specifying a prior on ρ and ϕ_j^*.

We specify the prior consistent with our chosen prior on the signal-to-noise, δ and an AR(2) model with stationary dynamics. Consistent with AR(2) dynamics, the prior on $\phi_j^* = 0$, $j \in [2, 3, p-1]$. For an AR(2) model to be stationary, the three conditions are

$$\mid \phi_2 \mid \; < \; 1 \tag{A3}$$

$$\phi_1 + \phi_2 \; < \; 1 \tag{A4}$$

$$\phi_2 - \phi_1 \; < \; 1. \tag{A5}$$

It is easy to verify our prior on the signal-to-noise ratio cannot fall below 0.07 as this will violate (A3). We assume $\delta < 1$ as we wish to specify a prior where trend fluctuations are less viable than the forecast error itself. Given $\phi(1) = \delta^{-2}$, through our definition of a signal-to-noise ratio, it is easy to verify (A4) will be satisfied as long as $\delta < 1$. Finally given $\phi(1)$, we can solve for a prior on ϕ_2 to satisfy (A5). Defining ζ as an arbitrarily small positive constant. It is straightforward to verify $\phi_2 = (1 + \phi(1))/2 - \zeta$ is sufficient to satisfy (A5).

To specify the posterior variance, given we effectively calibrate the signal-to-noise ratio ρ, so the prior variance on ρ is zero. The prior variance on ρ_j^* shrinks tighter around zero with every lag to prevent overfitting and thus acts as a shrinkage prior. Specifying an arbitrary long lag order $p = 12$ allows the model to capture dynamics if they are present at longer lags, while at the same time allows the shrinkage prior to effectively prevent overfitting. At the same time, specifying a longer lag order in the forecasting model relative to the AR(2), as per the prior, the model is allowed to fit richer dynamics than an AR(2) if they are present in the data, but still retain the pre-specified signal-to-noise ratio as long as the sum of the autoregressive coefficients ρ is fixed. Finally, we keep the prior on the constant as uninformative as a normal distribution with an arbitrarily large variance centred around zero. We summarise the priors of the autoregressive coefficients as follows.

$$
\begin{aligned}
\rho &\sim \mathcal{N}(1 - \delta^{-2}, 0) \\
\phi_1^* &\sim \mathcal{N}(\frac{\rho + 1}{2} + 0.01, 0.5) \\
\phi_j^* &\sim \mathcal{N}(0, \frac{0.5}{j^2}) j \in [2, 3, \ldots p - 1]
\end{aligned}
$$

References

Atkeson A, Ohanian LE. 2001. Are Phillips curves useful for forecasting inflation? *Quarterly Review* : 2–11.

Baxter M, King RG. 1999. Measuring business cycles: Approximate band-pass filters for economic time series. *The Review of Economics and Statistics* **81**: 575–593.

Beveridge S, Nelson CR. 1981. A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. *Journal of Monetary Economics* **7**: 151–174.

Chan JC, Grant AL. 2015. A bayesian model comparison for trend-cycle decompositions of output. CAMA Working Papers 2015-31, Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University.

Christiano LJ, Fitzgerald TJ. 2003. The band pass filter. *International Economic Review* **44**: 435–465.

Clark PK. 1987. The cyclical component of U.S. economic activity. *The Quarterly Journal of Economics* **102**: 797–814.

Clark TE, McCracken MW. 2006. The predictive content of the output gap for inflation: Resolving in-sample and out-of-sample evidence. *Journal of Money, Credit and Banking* **38**: 1127–1148.

Cochrane JH. 1994. Permanent and transitory components of gnp and stock prices. *The Quarterly Journal of Economics* **109**: 241–65.

Cogley T, Nason JM. 1995. Effects of the Hodrick-Prescott filter on trend and difference stationary time series implications for business cycle research. *Journal of Economic Dynamics and Control* **19**: 253–278.

Diebold FX, Mariano RS. 1995. Comparing predictive accuracy. *Journal of Business & Economic Statistics* **13**: 253–63.

Dungey M, Jacob J, Tian J. 2015. The role of structural breaks in forecasting trends and output gaps in real GDP of the G-7 countries. *mimeo* .

Edge RM, Rudd JB. forthcoming. Real-time properties of the federal reserve's output gap. *Review of Economics and Statistics* .

Evans G, Reichlin L. 1994. Information, forecasts, and measurement of the business cycle. *Journal of Monetary Economics* **33**: 233–254.

Harvey AC. 1985. Trends and cycles in macroeconomic time series. *Journal of Business & Economic Statistics* **3**: 216–227.

Harvey AC, Trimbur TM, Van Dijk HK. 2007. Trends and cycles in economic time series: A Bayesian approach. *Journal of Econometrics* **140**: 618–649.

Hodrick RJ, Prescott EC. 1997. Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit and Banking* **29**: 1–16.

Kim CJ, Nelson CR. 1999. Friedman's plucking model of business fluctuations: Tests and estimates of permanent and transitory components. *Journal of Money, Credit and Banking* **31**: 317–34.

Morley J. 2014. Measuring economic slack: A forecast-based approach with applications to economies in asia and the pacific. BIS Working Papers 451, Bank for International Settlements.

Morley J, Piger J. 2012. The asymmetric business cycle. *The Review of Economics and Statistics* **94**: 208–221.

Morley JC. 2011. The two interpretations of the beveridgenelson decomposition. *Macroeconomic Dynamics* **15**: 419–439.

Morley JC, Nelson CR, Zivot E. 2003. Why are the Beveridge-Nelson and unobserved-components decompositions of GDP so different? *The Review of Economics and Statistics* **85**: 235–243.

Murray CJ. 2003. Cyclical properties of Baxter-King filtered time series. *The Review of Economics and Statistics* **85**: 472–476.

Nelson CR. 2008. The Beveridge-Nelson decomposition in retrospect and prospect. *Journal of Econometrics* **146**: 202–206.

Nelson CR, Plosser CI. 1982. Trends and random walks in macroeconmic time series : Some evidence and implications. *Journal of Monetary Economics* **10**: 139–162.

Orphanides A, van Norden S. 2002. The unreliability of output-gap estimates in real time. *The Review of Economics and Statistics* **84**: 569–583.

Perron P, Wada T. 2009. Let's take a break: Trends and cycles in US real GDP. *Journal of Monetary Economics* **56**: 749–765, doi: 10.1016/j.jmoneco.2009.08.001.

Phillips P, Jin S. 2015. Business cycle, trend elimination, and the HP filter. *Crwles Foundation Discussion Paper No. 2005* .

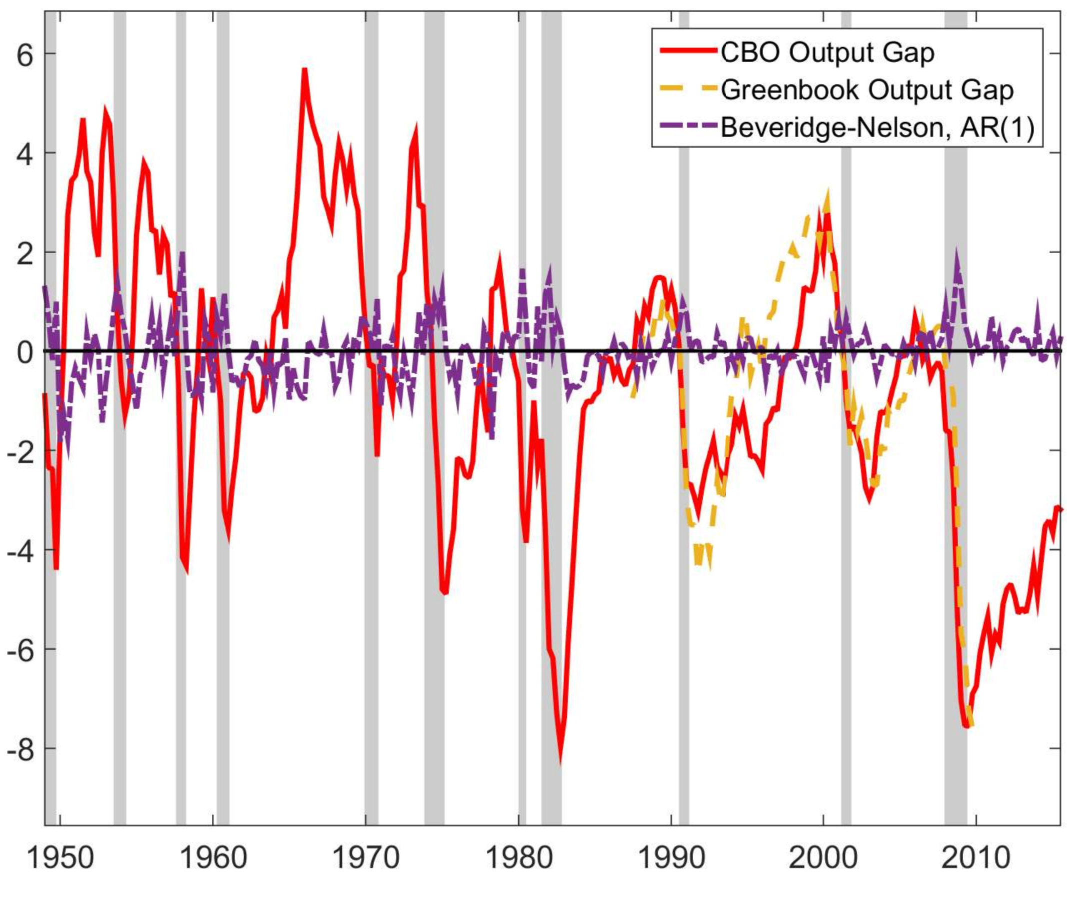
Sinclair TM. 2009. Asymmetry in the business cycle: Friedman's plucking model with correlated innovations. *Studies in Nonlinear Dynamics & Econometrics* **14**: 1–31.

Stock JH, Watson MW. 1999. Forecasting inflation. *Journal of Monetary Economics* **44**: 293–335.

Stock JH, Watson MW. 2008. Phillips curve inflation forecasts. NBER Working Papers 14322, National Bureau of Economic Research, Inc.

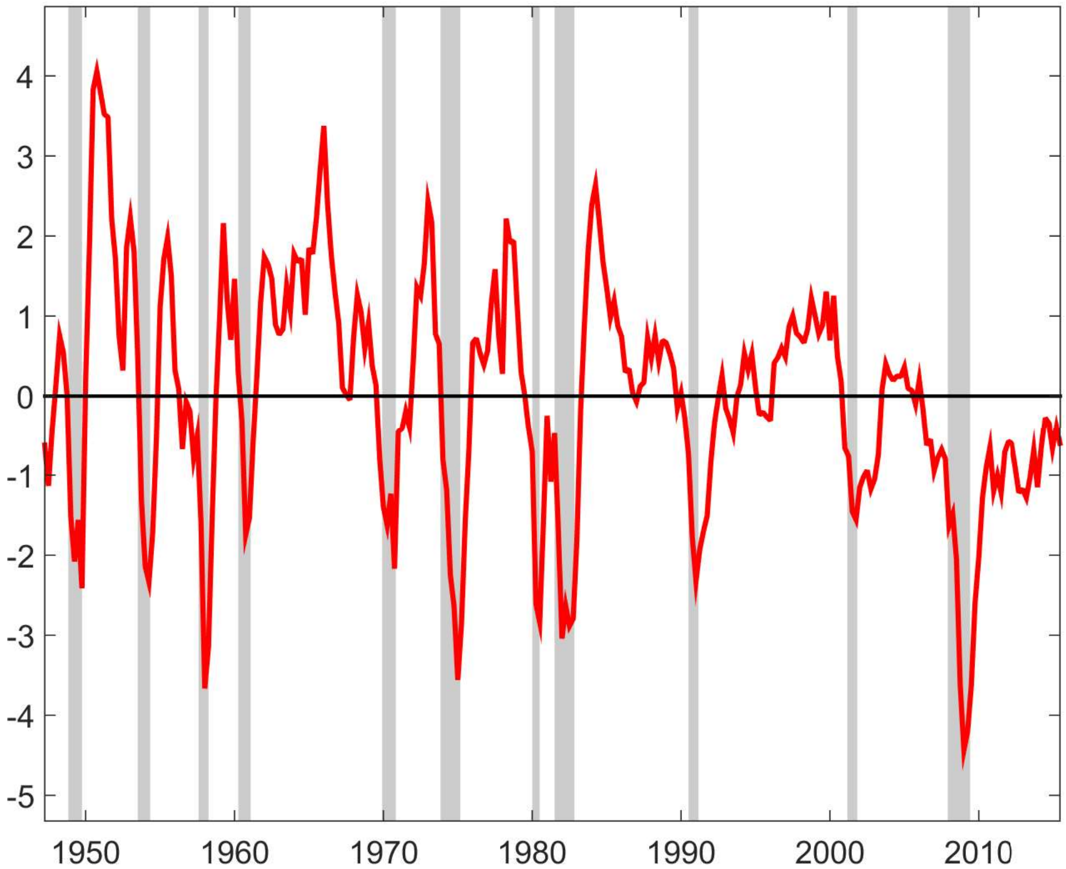
Watson MW. 1986. Univariate detrending methods with stochastic trends. *Journal of Monetary Economics* **18**: 49–75.

Figure 1: Estimated U.S. Output Gaps



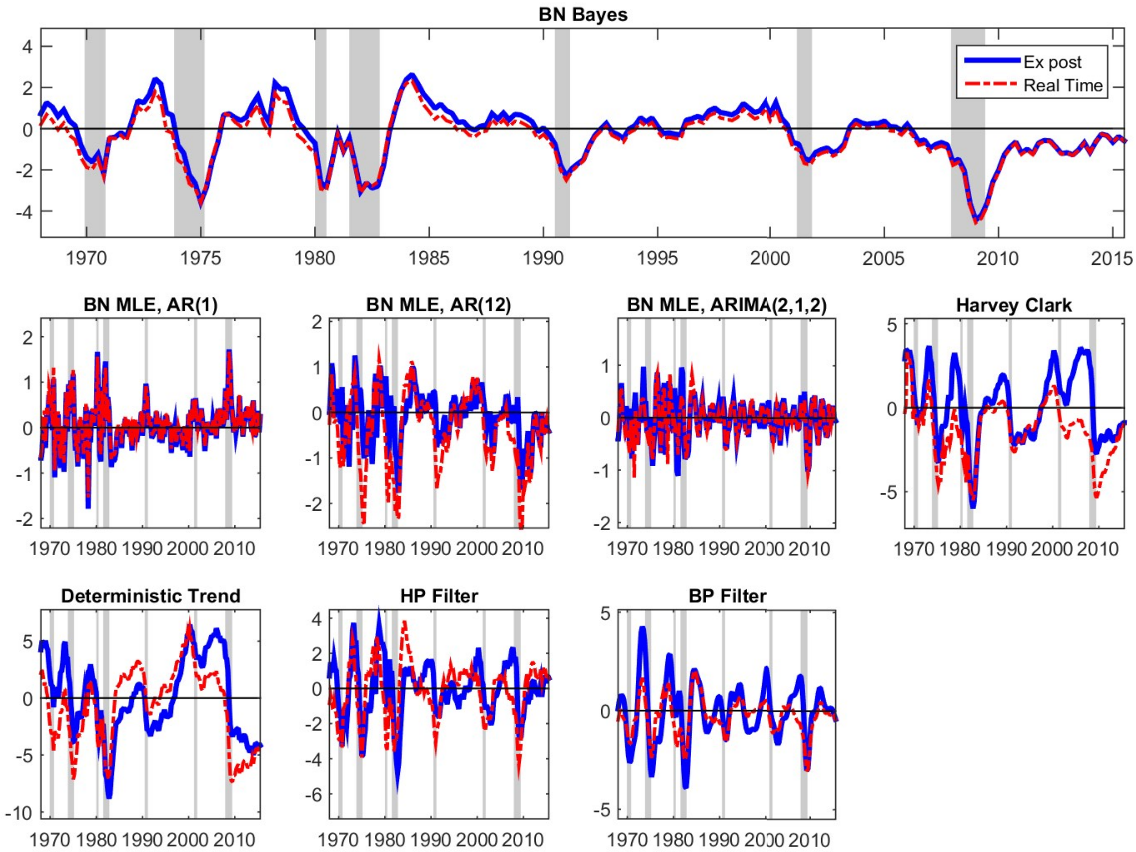
Notes: Percent deviation from trend. Shaded areas represent NBER recession dates.

Figure 2: Output Gap Estimated Using BN Bayes, $\delta = 0.1$



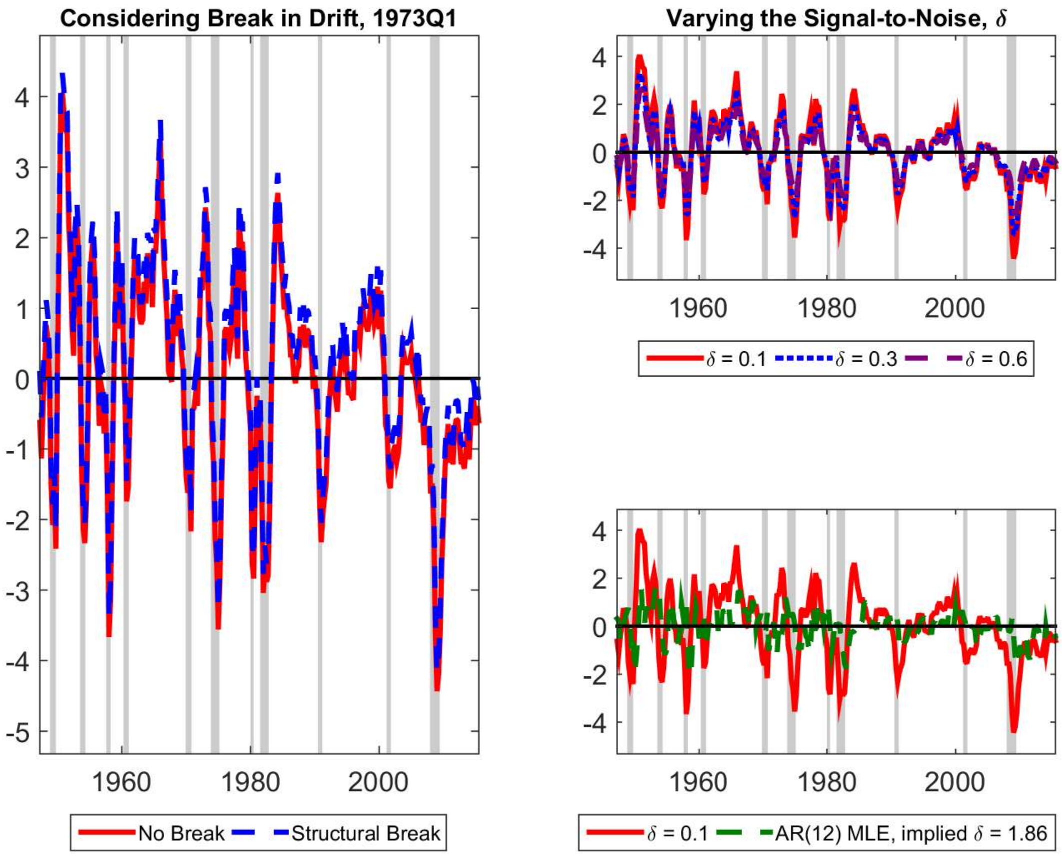
Notes: Percent deviation from trend. Shaded areas represent NBER recession dates. Forecasting model is an AR(12) process for U.S. GDP growth using a prior on the signal-to-noise ratio, $\delta = 0.1$.

Figure 3: Revision properties of estimated output gaps



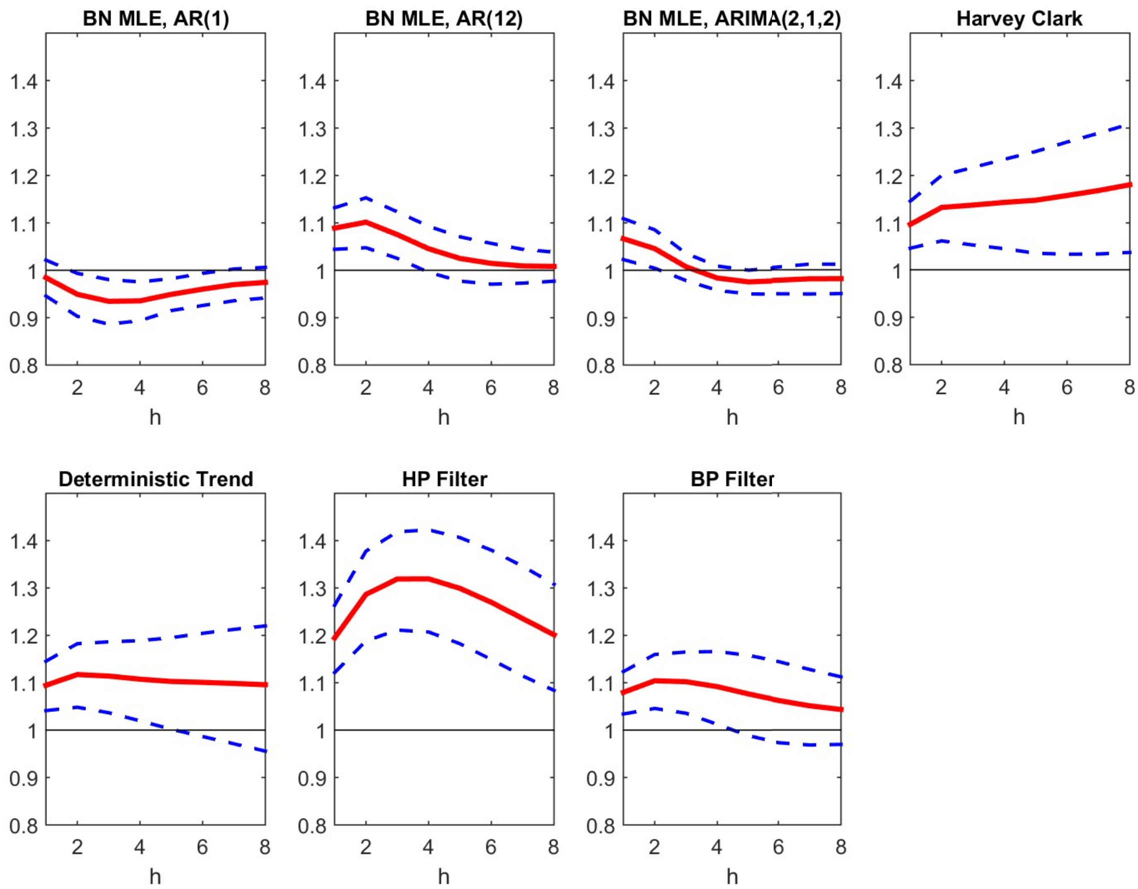
Notes: Percent deviation from trend. Shaded areas represent NBER recession dates.

Figure 4: Sensitivity Analysis



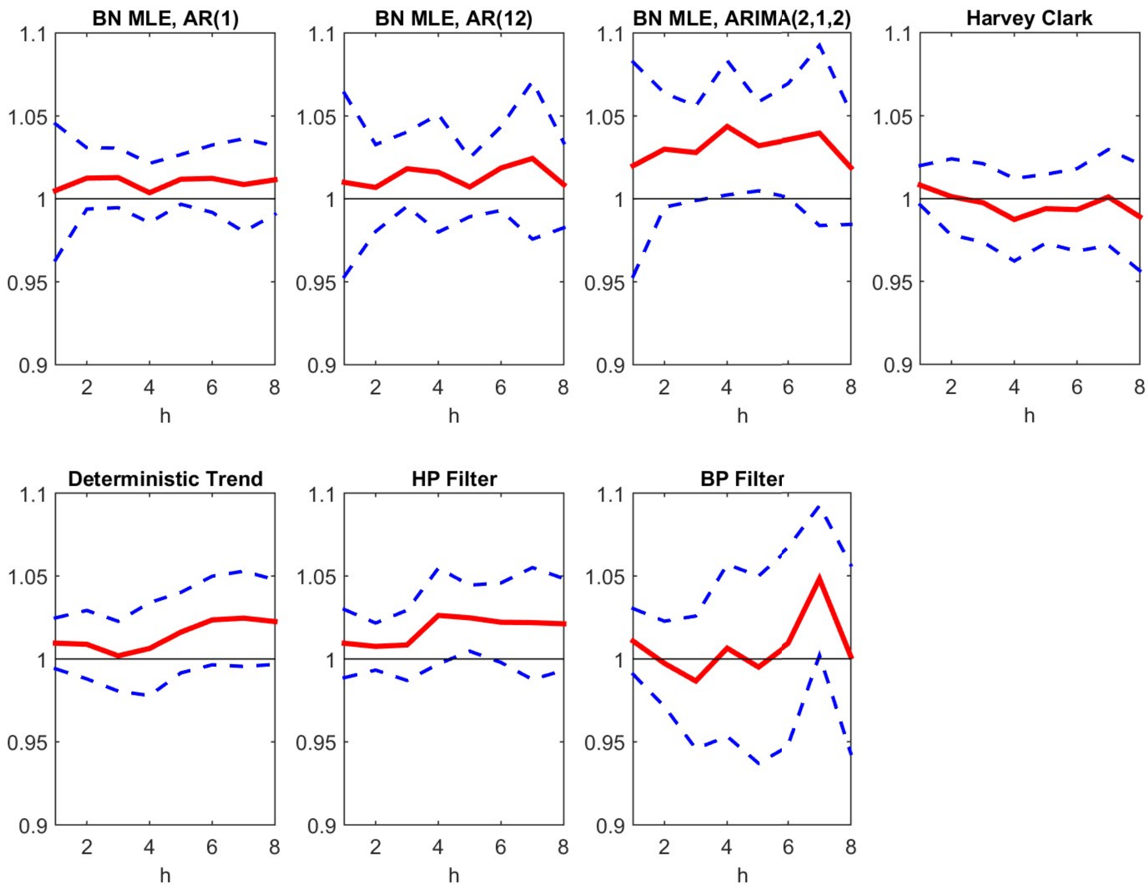
Notes: Percent deviation from trend. Shaded areas represent NBER recession dates.

Figure 5: Output growth forecasts relative to BN Decomposition with smoothing prior



Notes: The left graph presents in-sample Relative Root Mean Square Errors (RRMSE) relative to the BN decomposition with a smoothing prior. The right four graphs presents out-of-sample RRMSE relative to the BN decomposition with a smoothing prior. Out of sample evaluation begins in 1970Q1. The bands are 90% intervals from the Diebold and Mariano (1995) test of equal forecast accuracy.

Figure 6: Inflation forecasts relative to BN Bayes



Notes: Out-of-sample Relative Root Mean Squared Error relative to the BN decomposition with a smoothing prior. Out-of-sample evaluation begins in 1970Q1. The bands are 90% intervals from the Diebold and Mariano (1995) test of equal forecast accuracy.

Table 1: Revision Statistics, 1970Q1 - 2012Q4

	Correlation	**Size of Revisions**		**Same**
		Std	**RMS**	**Sign**
BN Bayes	0.99	0.14	0.23	0.88
BN, AR(1) MLE	1.00	0.12	0.16	0.95
BN, AR(12) MLE	0.74	0.96	1.11	0.75
BN, ARMA(2,2) MLE	0.81	0.59	0.60	0.84
Harvey-Clark	0.75	0.66	1.00	0.66
Deterministic Trend	0.70	0.78	0.81	0.70
HP Filter	0.55	0.97	0.97	0.59
BP Filter	0.77	0.65	0.67	0.74

Notes: Correlation refers to the correlation of the psuedo-real time estimate to the final estimate of the output gap. Std and RMSE refers respectively to the standard deviation and root mean square of revisions to the psuedo-real time estimate of the output gap, with both subsequently normalised by the standard deviation of the final estimate of the output gap. Same sign refers to the proportion of psuedo-real time estimates which shares the same sign as the final estimate of the output gap.

Table 2: Simulation Results

(a) Correlation with True Output Gap		
	Monte Carlo	**Population**
BN Bayes	0.32	0.48
BN, AR(1)	0.14	0.14
BN, ARMA(p,q)	0.22	0.54

(b) Roots	
Near Cancellation	26%
Redundant MA terms	65%

Note: (a) Monte Carlo refers to the mean correlation obtained from 2000 artificial datasets with a sample size of 250 generated from the DGP. Population is based on one long sample from the DGP. BN Bayes refers to our proposed approach to estimate the output gap. BN, AR(1) is based on a BN decomposition from an AR(1) forecasting model. BN, ARMA(p,q) is a BN decomposition from an ARMA model where the lag order p and q are chosen using the BIC. (b) Near cancellation is the proportion of Monte Carlo trials where there is a near cancellation of the MA roots. The AR and MA lag order are chosen using BIC. A Monte Carlo trial is classified as a near cancellation if the absolute value of the difference between the largest inverted AR and MA roots are within 0.08. Redundant MA terms is the proportion of Monte Carlo trials with near cancellations or if the BIC does not fit any MA terms.