UNSW Business School Research Paper No. 2016 ECON 09

Intuitive and Reliable Estimates of the Output Gap from a Beveridge-Nelson Filter

Gunes Kamber
James Morley
Benjamin Wong

Intuitive and Reliable Estimates of the Output Gap from a Beveridge-Nelson Filter *†

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

[1]Reserve Bank of New Zealand
[2]Bank for International Settlements
[3]University of New South Wales

June 15, 2016

Abstract

The Beveridge-Nelson (BN) trend-cycle decomposition based on autoregressive forecasting models of U.S. quarterly real GDP growth produces estimates of the output gap that are strongly at odds with widely-held beliefs about the amplitude, persistence, and even sign of transitory movements in economic activity. These antithetical attributes are related to the autoregressive coefficient estimates implying a very high signal-to-noise ratio in terms of the variance of trend shocks as a fraction of the overall quarterly forecast error variance. When we impose a lower signal-to-noise ratio, the resulting BN decomposition, which we label the "BN filter", produces a more intuitive estimate of the output gap that is large in amplitude, highly persistent, and typically positive in expansions and negative in recessions. Real-time estimates from the BN filter are also reliable in the sense that they are subject to smaller revisions and predict future output growth and inflation better than for other methods of trend-cycle decomposition that also impose a low signal-to-noise ratio, including deterministic detrending, the Hodrick-Prescott filter, and the bandpass filter.

JEL Classification: C18, E17, E32

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

*Kamber: gunes.kamber@bis.org 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 or the Bank for International Settlements. We thank seminar and conference participants at the Bank for International Settlements Asian Office, Bank of Canada, Central Bank of Turkey, Federal Reserve Bank of Cleveland, Federal Reserve Bank of San Francisco, Reserve Bank of New Zealand, Australian National University, Monash University, Otago University, University of Auckland, University of Queensland, 2015 SNDE Symposium in Oslo, 2015 Melbourne Institute Macroeconomic Policy Meetings, 2015 Central Bank Macroeconomic Modelling Workshop, 2015 Workshop of the Australasia Macroeconomic Society, and 2016 AJRC-HIAS Conference in Canberra, and especially Todd Clark, George Evans, Sharon Kozicki, Adrian Pagan, Frank Smets, Ellis Tallman, John Williams, as well as our discussants, Michael Kouparitsas, Glenn Otto, and Leif Anders Thorsrud 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 at business cycle frequencies. Because the Beveridge and Nelson (1981) (BN) trend-cycle decomposition defines the trend of a time series as its long-run conditional expectation after all forecastable momentum has died out (and subtracting off any deterministic drift), the corresponding cycle for log real GDP should provide a sensible estimate of the output gap as long as it is based on accurate forecasts over short and medium term horizons. Noting that standard model selection criteria suggest a low-order autoregressive (AR) model predicts U.S. quarterly real GDP growth better than more complicated alternatives, Figure 1 plots the estimate of the output gap from the BN decomposition based on an AR(1) model.[1] The estimated output gap is noticeably small in amplitude and lacking in persistence. It also does not match up well at all with the reference cycle of U.S. expansions and recessions determined by the National Bureau of Economic Research (NBER). For comparison, Figure 1 also plots estimates of the U.S. output gap based on the Federal Reserve (Fed) Board of Governors' Greenbook and the Congressional Budget Office (CBO) estimate of potential output. In contrast to the estimated output gap from the BN decomposition based on an AR(1) model, these output gap estimates have much higher persistence and larger amplitude. They are also strongly procyclical in terms of the NBER reference cycle. An important reason for these differences is that the estimated autoregressive coefficient for the AR(1) model used in the BN decomposition implies a very high signal-to-noise ratio in terms of the variance of trend shocks as a fraction of the overall quarterly forecast error variance, while the Fed and CBO implicitly assume a much lower signal-to-noise ratio when constructing their estimates.

Our main contribution in this paper is to show how to conduct a BN decomposition imposing a low signal-to-noise ratio on an AR model, an approach we refer to as the "BN filter". The BN filter is easy to implement in comparison to related methods that also seek to address the conflicting results in Figure 1, such as Bayesian estimation of an unobserved components (UC) model with a smoothing prior on the signal-to-noise ratio (e.g., Harvey et al., 2007). Notably, when we apply the BN filter to U.S. log real GDP, the resulting estimate of the output gap is persistent, has large amplitude, and matches up well with the NBER reference cycle. At the same time, real-time estimates are subject to smaller revisions and perform better in out-of-sample forecasts of output growth and inflation than estimates for other trend-cycle decomposition methods that also impose a low signal-to-noise ratio, including deterministic detrending using a quadratic trend, the

[1] The raw data for U.S. real GDP are taken from FRED for the sample period of 1947Q1-2015Q3. Real GDP growth is measured in continuously-compounded terms. Model estimation is based on least squares regression or, equivalently, conditional maximum likelihood estimation under the assumption of normality. Initial lagged values for AR(p) models are backcast using the sample average growth rate. Our specific choice of lag order p=1 is based on the Schwarz Information Criterion.

Hodrick-Prescott (HP) filter, and the bandpass (BP) filter. Thus, our proposed approach addresses a key critique by Orphanides and van Norden (2002) that popular methods of estimating the output gap are unreliable in real time.

That Orphanides and van Norden (2002) find output gap estimates unreliable in real time dramatically undermines their usefulness in policy environments and in forming a meaningful gauge of current economic slack more generally. Meanwhile, the fact that the BN filter estimates are not heavily revised is not coincidental, but stems from our choice to work within the context of AR models. In principle, the BN decomposition can be applied using any forecasting model, including multivariate time series models such as vector autoregressive (VAR) models, structural models such as dynamic stochastic general equilibrium (DSGE) models, or even nonlinear time series models such as time varying parameter (TVP) or Markov switching (MS) models. Our choice to work with AR models is deliberate. Because the estimated output gap for a BN decomposition directly reflects the estimated parameters of the model, it is mechanical that any instability in the estimated parameters in real time will produce estimates of the output gap that are heavily revised. However, estimates of autoregressive coefficients for AR models are relatively stable in real time, unlike with parameters for more complicated models. Therefore, a natural outcome of our modeling choice is output gap estimates that are reliable.

Our proposed approach is robust to the omission of multivariate information in the forecasting model and to structural breaks in the long-run growth rate, thus addressing important issues with trend-cycle decomposition raised by Evans and Reichlin (1994) and Perron and Wada (2009). Meanwhile, because we use the BN decomposition, our proposed approach takes account of a random walk stochastic trend in log real GDP and implicitly allows for correlation between movements in trend and cycle, unlike many popular methods that assume trend stationarity or that these movements are orthogonal. See Nelson and Kang (1981), Cogley and Nason (1995), Murray (2003), and Phillips and Jin (2015), amongst others, on the problem of "spurious cycles" in the presence of a random walk stochastic trend when using popular methods of trend-cycle decomposition such as deterministic detrending, the HP filter, and the BP filter. Meanwhile, see Morley et al. (2003), Dungey et al. (2015), and Chan and Grant (2015) on the importance of allowing for correlation between permanent and transitory movements.

The rest of this paper is structured as follows. Section 2 describes the BN filter and applies it to U.S. quarterly log real GDP, formally assessing its revision properties relative to other methods and considering some key robustness issues. Section 3 provides a theoretical justification for our proposed approach, in particular why one might choose to impose a low signal-to-noise ratio on an AR model, and then presents a forecast comparison with other methods and considers whether revisions are useful for understanding the past. Section 4 concludes.

2 The BN Filter

2.1 The BN 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 (i.e., deterministic) future movements in the time series. In particular, letting $\{y_t\}$ denote a time series process with a trend component that follows a random walk with constant 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 the trend component under the assumption that the long-horizon conditional expectation of the remaining cycle is equal to zero. By removing the deterministic drift, the conditional expectation remains finite and becomes an optimal estimate of the current trend component (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. Modeling the first differences explicitly allows for a random walk stochastic trend in the level of the time series because forecast errors for the first differences can be estimated to have permanent effects on the long-horizon conditional expectation of $\{y_t\}$.

Based on sample autocorrelation functions (ACFs) and partial autocorrelation functions (PACFs) for many macroeconomic time series, including the first differences of U.S. quarterly 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 $e_t \sim iidN(0, \sigma_e^2)$.[2] For convenience when defining the signal-to-noise ratio, let $\phi(L) \equiv 1 - \phi_1 L - \ldots - \phi_p L^p$ denote the autoregressive lag polynomial, where L is the lag operator. Then, assuming the roots of $\phi(z) = 0$ lie outside the unit circle, which corresponds to $\{\Delta y_t\}$ being stationary, the unconditional mean $\mu \equiv \mathbb{E}\left[\Delta y\right] =$

[2] The normality assumption is not strictly necessary for the BN decomposition. However, under normality, least squares regression for an AR model becomes equivalent to conditional maximum likelihood estimation and the Bayesian shrinkage priors used in our approach, as discussed below, become conjugate, making posterior calculations straightforward. Also, the forecast errors do not need to be identically distributed, as long as they form a martingale difference sequence. However, in terms of possible structural breaks in the variance of the forecast error, the key assumption we make in our proposed approach is that there are no changes in the signal-to-noise ratio as defined in this section, an assumption which is implicitly supported by the relative stability of the estimated sum of the autoregressive coefficients across possible variance regimes within the sample.

$\phi(1)^{-1}c$.

Although an AR(1) model of U.S. real GDP growth might seem reasonable given sample ACFs and PACFs and is supported by the Schwarz Information Criterion (SIC), we have already seen in Figure 1 that the estimated output gap from a BN decomposition based on an AR(1) model does not match well at all with widely-held beliefs about the amplitude, persistence, and sign of transitory movements in economic activity, as reflected, for example, in the Fed's Greenbook and CBO estimates of the output gap. Most noticeably, the estimated output gap is small in amplitude, suggesting that most of the fluctuations in economic activity have been driven by trend.

To understand why the BN decomposition produces output gap estimates with such features, it is helpful to note that the analytical formula for the cycle from the BN decomposition based on an AR(1) model is $-\phi(1-\phi)^{-1}(\Delta y_t - \mu)$ (see Morley, 2002). Therefore, the estimated output gap will only be as persistent as output growth, which is not very persistent given that $\hat{\phi}$ based on maximum likelihood estimation (MLE) is typically between 0.3 and 0.4 for U.S. quarterly data. Similarly, given that $\hat{\phi}(1-\hat{\phi})^{-1} \approx 0.5$, the amplitude of the estimated output gap will be small, as the implied variance will only be about one quarter that of output growth. Furthermore, given that $-\hat{\phi}(1-\hat{\phi})^{-1} < 0$, it is not surprising that the estimated output gap is generally positive in recessions when output growth is negative and vice versa in expansions. In terms of the intuition underlying the BN decomposition, the momentum in output growth implied by the AR(1) model means that when there is a negative shock in a recession, output growth is expected to remain below average in the quarters immediately afterwards before eventually returning back to its long-run average, with the converse holding for a positive shock in an expansion. Thus, log real GDP is initially above the BN trend defined as the long-run conditional expectation minus deterministic drift when a shock triggers a recession and below the BN trend when a shock triggers an expansion.

More generally, to understand the BN decomposition for an AR(p) model, it is useful to define a signal-to-noise ratio for a time series in terms of the variance of trend shocks as a fraction of the overall forecast error variance:

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

where $\psi(1) \equiv \lim_{j\to\infty} \frac{\partial y_{t+j}}{\partial e_t}$ is the "long-run multiplier" that captures the permanent effect of a forecast error on the long-horizon conditional expectation of $\{y_t\}$ and provides the key summary statistic for a time series process when calculating the BN trend based on a forecasting model given that $\Delta\tau_t = \psi(1)e_t$. For an AR($p$) model, this long-run multiplier has the simple form of $\psi(1) = \phi(1)^{-1}$ and, based on MLE for an AR(1) model of postwar U.S. quarterly real GDP growth, the signal-to-noise ratio appears to be quite high with $\hat{\delta} = 2.22$. That is, BN trend shocks are much more volatile than quarter-to-quarter forecast errors in log real GDP. Notably, $\hat{\delta} > 1$ holds for all freely estimated AR(p) models given that $\hat{\phi}(1)^{-1}$ is always greater than unity regardless of lag order p.

The insight that the signal-to-noise ratio δ is mechanically linked to $\phi(1)$ for an AR(p) model is a powerful one because it implies that we can impose a signal-to-noise ratio by fixing the sum of the autoregressive coefficients in estimation. To do so, we first transform the AR(p) model into 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 $\phi_j^* \equiv -(\phi_{j+1} + \ldots + \phi_p)$. Then, noting that $\delta = (1-\rho)^{-2}$ for an AR(p) model, equation (3) can then be estimated by fixing ρ as follows:

$$\bar{\rho} = 1 - 1/\sqrt{\bar{\delta}}, \tag{4}$$

In this way, the BN decomposition can be applied while imposing a particular signal-to-noise ratio $\bar{\delta}$.

If one were seeking to maximize the amplitude of the output gap, it turns out an AR(1) model would be a poor choice because the stationarity restriction $|\phi| < 1$ implies that $\delta > 0.25$, which is to say trend shocks must explain at least 25% of the quarterly forecast error variance. Higher-order AR(p) models allow lower values of δ (e.g., $\delta > 0.0625$ for an AR(2) model), although a finite-order AR(p) model would never be able to achieve $\delta = 0$ given that this limiting case would actually correspond to a non-invertible MA process with a unit MA root (i.e., $\{y_t\}$ would actually be (trend)-stationary, so $\{\Delta y_t\}$ would, in effect, be "over-differenced"). Consideration of a higher-order AR(p) model also allows for greater persistence in the output gap and different correlation with output growth than is possible for an AR(1) model.

Estimation of equation (3) imposing a particular signal-to-noise ratio is straightforward without the need of complicated nonlinear restrictions or posterior simulation. Conditional MLE entails a single parametric restriction $\rho = \bar{\rho}$, which can be done by bringing $\bar{\rho}\Delta y_{t-1}$ to the left-hand-side when running a least squares regression for equation (3). However, even though it is possible to implement our approach using conditional MLE, we opt for Bayesian estimation in practice because it allows us to utilize a shrinkage prior on the higher lags of the AR(p) model in order to prevent overfitting and to mitigate the challenge of how to specify the exact lag order beyond being large enough to accommodate small values of δ. We thus consider a prior on the second-difference coefficients as follows:

$$\phi_j^* \sim N(0, \frac{0.5}{j^2}), j \in [2, 3, \ldots p-1]$$

In practice, we consider an AR(12) model, although results are robust to consideration of higher lag orders given the shrinkage priors. Readers familiar with the Minnesota class of priors will recognize that the posterior distributions given the regression model in equation (3) and our priors will have a closed-form solution and can be easily calculated without the need for posterior simulation.[3]

All that remains is to specify a particular value $\bar{\delta}$ to impose. Given widely-held beliefs about the amplitude of transitory movements in economic activity reflected in, say, the Fed's Greenbook and CBO estimates of the output gap, we reduce δ from its freely estimated value of $\hat{\delta} > 1$ in order to increase the amplitude of the estimated output gap. However, for a given AR(p) model, it is not a given that decreasing δ will actually increase the amplitude of the estimated output gap. This is because, with reference to equation (2), decreasing δ can be achieved by either a fall in $\sigma^2_{\Delta\tau}$ or a rise in σ^2_e. In particular, setting δ below $\hat{\delta}$ worsens the fit of the AR(p) model, thus increasing the implied σ^2_e. Therefore, we consider a range of possible values for δ between 0 and 1, trading off a possible increase in amplitude with a decrease in fit. In particular, we do so by setting $\bar{\delta}$ to a small enough value such that, for a given decrease in δ, the percentage increase in the variance of the output gap becomes smaller than the percentage increase in the forecast error variance.

Figure 2 plots the relationships between δ and the standard deviation of the output gap, the RMSE of the forecasting model, and the percentage changes in the corresponding variances for a small decrease in the signal-to-noise ratio (i.e., $\Delta\delta = -0.01$) for the U.S. data. The impact of δ on the amplitude of the output gap is non-monotonic, with an inflection around 1.75, below which there is a steady increase in amplitude as δ gets smaller. Meanwhile, the impact of δ on the RMSE of the forecasting model is small unless δ is relatively close to zero, at which point decreasing δ has a deleterious impact on fit. From the third panel, it can be seen that $\bar{\delta} = 0.25$ (i.e., imposing that trend shocks account for 25% of the forecast error variance for quarterly output growth) optimizes the tradeoff between amplitude and fit given equal but opposite weights on percentage changes in the two variances.[4] Meanwhile, as we will show later, the shape of the output gap is reasonably robust to changes in δ around this value. Therefore, a researcher could impose an even lower value for $\bar{\delta}$, perhaps reflecting a dogmatic prior about a smaller role for trend shocks or, similarly, placing a higher positive weight on a percentage increase in the variance of the output gap than a negative weight on a percentage increase in the forecast error variance, with little impact on the estimated output gap apart from an

[3]It would, of course, also be possible to impose a low signal-to-noise ratio for a more general ARMA model. However, estimation would be less straightforward, there would be greater tendency to overfit the data, and the corresponding BN decomposition would be less reliable.

[4]Another way to assess fit is to consider the prior predictive density for our proposed model. We find that 82% of the postwar observations for U.S. real GDP growth lie within the equal-tailed 90% bands of the conditional prior predictive density. In this sense, our AR(12) model with $\bar{\delta} = 0.25$ and shrinkage priors on the second-difference coefficients, while different than what would be estimated by MLE, is not strongly at odds with the data.

increase in amplitude.

Reflecting the similar smoothing effect on the implied trend as for the HP and BP filters when imposing a low signal-to-noise ratio, we refer to our proposed approach as the "BN filter". Figure 3 plots the estimated U.S. output gap for the BN filter imposing $\bar{\delta} = 0.25$. A cursory glance at the figure suggests that the BN filter is much more successful than the traditional BN decomposition based on a freely estimated AR model at producing an estimated output gap that is consistent with widely-held beliefs about amplitude, persistence, and the sign of transitory movements in economic activity. In particular, the estimated output gap is large in amplitude, persistent, and successful at capturing the turning points in the NBER reference cycle.

2.2 Revision Properties

Having described the BN filter that imposes a low signal-to-noise ratio and applied it to U.S. log real GDP, we now assess its revision properties. As discussed previously, by working with an AR model, we seek to address the Orphanides and van Norden (2002) critique that popular methods of estimating the output gap are not reliable in real time. Orphanides and van Norden (2002) show that most real-time revisions of output gap estimates are due to the lack of information about the future, rather than data revisions. That is, it is the the method of extracting the output gap in real time that is deficient, not the ability of the models to predict data revisions. Thus, for simplicity, we consider a pseudo-real-time exercise using final vintage data (from 2015Q3) rather than a full-blown real-time analysis with different vintages of data.

To evaluate the performance of the BN filter, we compare it to several other methods of trend-cycle decomposition. First, we consider BN decompositions based on various freely estimated ARMA forecasting models of $\{\Delta y_t\}$ and Kalman filtering for a UC model of $\{y_t\}$. In particular, we consider BN decompositions based on an AR(1) model, an AR(12) model, and an ARMA(2,2) model, all estimated via MLE. We also consider a multivariate BN decomposition based on a VAR(4) model of U.S. real GDP growth and the civilian unemployment rate, also estimated via MLE. For the UC model, we consider a similar specification to Harvey (1985) and Clark (1987) estimated via MLE. The AR(1) model is chosen based on SIC for the whole set of possible ARMA models. The AR(12) model allows us to understand the effect of imposing a longer lag order, although we note that standard model selection criteria would generally lead a researcher to choose a more parsimonious specification in practice. Morley et al. (2003) show that the BN decomposition based on an ARMA(2,2) model is equivalent to Kalman filter inferences for an unrestricted version of the popularly used UC model by Watson (1986). In particular, the Watson UC model features a random walk with constant drift trend plus an AR(2) cycle, but, as Morley et al. (2003) show (also see Chan and Grant, 2015; Dungey et al., 2015), the zero restriction on the correlation between movements in trend and cycle can be rejected by the data, suggesting the BN decomposition based on an unrestricted ARMA(2,2) model is the appropriate approach when considering UC models that feature a random walk trend with drift plus an AR(2) cycle. Meanwhile, for completeness, we also consider the Harvey-Clark UC model, similar to that considered by Orphanides and van Norden (2002). The Harvey-Clark UC model differs from the Watson (1986) model in that it also features a random walk drift in addition to a random walk trend. For this model, we retain the zero restrictions on correlations between movements in drift, trend, and cycle.

We also consider some popular methods of deterministic detrending and nonparametric filters. In particular, we consider a deterministic 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 impose a smoothing parameter $\lambda = 1600$, as is standard 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 that, whatever documented misgivings about the HP and BP filters (e.g., Cogley and Nason, 1995; Murray, 2003; Phillips and Jin, 2015), they are relatively easy to implement, including often being available in many canned statistical packages, and are widely used in practice.

Figure 4 compares the pseudo-real-time and the ex post (i.e., full sample) estimates of the output gap for the different methods.[5] In the top panel, we evaluate the revision properties of the BN filter.[6] The remaining subplots present the results for the other methods. We first focus on the various BN decompositions based on freely estimated forecasting models. Both the AR(1) and ARMA(2,2) models produce output gap estimates that have little persistence, are small in amplitude, and are not procyclical in terms of the NBER reference cycle. Adding more lags impacts the persistence and sign of the estimated output gap, with the AR(12) model suggesting a more persistent output gap that is procyclical in terms of the NBER reference cycle. Even so, the estimated output gap for the AR(12) model still has relatively low amplitude, consistent with our earlier observation that AR forecasting models estimated via MLE always imply a relatively high signal-to-noise ratio of $\hat{\delta} > 1$ for U.S. data. The BN decomposition for the VAR model suggests an output gap that is more persistent, larger in amplitude, and moves with the NBER reference cycle, consistent with the point made by Evans and Reichlin (1994) that adding relevant information for forecasting of output growth mechanically

[5] We start the pseudo-real-time analysis with raw data from 1947Q1 to 1968Q1 for U.S. real GDP and 1948Q1 to 1968Q1 for U.S. civilian unemployment rate and add one observation at a time until we reach the full sample that ends in 2015Q3. Again, all raw data are taken from FRED. As with real GDP growth, the unemployment rate is backcast using its pseudo-real-time sample average to allow the estimation sample to always begin in 1947Q2.

[6] We re-calculate $\bar{\delta}$ for each pseudo-real-time sample. Encouragingly, the values that trade off model fit and amplitude of the output gap are quite stable, fluctuating between 0.25-0.28.

lowers the signal-to-noise ratio.

In terms of the revision properties of the output gap estimates, the first thing to notice in Figure 4 is that, regardless of the features in terms of persistence, amplitude, and sign, all of the estimates based on the BN decomposition, including in the case of the BN filter, are subject to relatively small revisions when compared to the other methods. A key reason why output gap estimates based on the BN decomposition are hardly revised is because the estimated parameters of the forecasting models appear to be relatively stable when additional observations are added in real time. Meanwhile, even though the output gap estimates using the BN decomposition appear relatively stable, the estimates for the more highly parameterized AR(12) and VAR(4) models are subject to more revisions than the simpler AR(1) and ARMA(2,2) models. This suggests overparameterization and overfitting can compromise the real-time reliability of the BN decomposition. This is the main reason we impose a shrinkage prior when considering the highly parameterized AR(12) model in our proposed approach. In particular, the shrinkage prior prevents overfitting, while the high lag order still allows for relatively rich dynamics. The revision properties of the BN filter estimates, which are more similar to those for the AR(1) model than for the AR(12) model based on MLE suggest that our proposed approach achieves a reasonable compromise between avoiding overfitting and allowing for richer dynamics. In terms of the other methods, they are heavily revised and thus unreliable in real time. In particular, the deterministic detrending is extremely sensitive to the sample period, while the HP and BP filters and the Harvey-Clark estimates all suffer from the endpoint problems of two-sided filters or smoothed inferences in the case of the Harvey-Clark model.

While eyeballing Figure 4 suggests the BN filter should be appealing from a reliability perspective, we formally quantify these revision properties by calculating revision statistics, similar to the analysis in Orphanides and van Norden (2002) and Edge and Rudd (forthcoming). First, to quantify the size of the revisions, we consider two measures, the standard deviation and the root mean square (RMS). The RMS measure is designed to penalize persistently large revisions more heavily relative to the standard deviation. Both the standard deviation and RMS measures are normalized by the standard deviation of the ex post estimate of the output gap for each method to enable a fair comparison as the different methods produce estimates with very different amplitudes.[7] Second, we calculate the correlation between the pseudo-real-time estimate of the output gap and the ex post estimate of the output gap. Third, we compute the frequency with which the pseudo-real-time estimate of the output gap has the same sign as the ex post estimate. We consider the evaluation sample of 1970Q1-2012Q4 to match with the starting point for the out-of-sample forecast comparison discussed in the next section and because the

[7]Note these statistics are referred to as "noise-to-signal ratios" by Orphanides and van Norden (2002). However, apart from the labelling, they have nothing to do with the signal-to-noise ratio in our proposed approach. Thus, we use the terms "standard deviation" and "RMS" of the revisions to avoid any confusion.

more recent estimates near the end of the full sample in 2015Q3 may end up becoming more heavily revised in the future.

Figure 5 reports the revision statistics. As was evident in Figure 4, the BN filter does well in terms of size of revisions, with revisions being less than one quarter of the standard deviation of the ex post estimate of the output gap. Some of the methods have revisions that are closer in magnitude to one standard deviation of the ex post estimate of the output gap. The BN decompositions based on the highly parameterized AR(12) and VAR(4) models do poorly in terms of the size of revisions, with revisions about one standard deviation or more of the ex post estimate of the output gap. In fact, apart from the BN filter and the BN decompositions based on freely estimated AR(1) and ARMA(2,2) models, all the alternative methods produce revisions that are well over half of one standard deviation of the ex post estimate of the output gap. In terms of accuracy, the BN decompositions tend to produce pseudo-real-time estimates of the output gap that are highly correlated with the ex post estimates, although the correlation is somewhat lower for the more highly parameterized AR(12) and VAR(4) models. The BN filter does well, with near perfect correlation between the pseudo-real-time and ex post estimates. The real-time accuracy of the other methods is quite mixed, with the HP filter performing the worst. Turning our attention to whether the sign of the estimated output gap changes once one is endowed with future information, we find that, again, the BN decompositions tend to do relatively well. Furthermore, within the class of BN decompositions, the pseudo-real-time estimates based on the BN filter perform particularly well, correctly identifying the same sign as the final estimate almost 90% of the time.

To summarize, the BN decomposition is reliable in a pseudo-real-time environment. This is because the addition of future data does not drastically alter the estimates of the forecasting model parameters under consideration. Therefore, it is more reliable than methods such as deterministic detrending, the HP filter, and the BP filter. Even so, within the class of BN decompositions, model parameter parsimony seems to be important for reliability of the estimated output gap. This is not much of a surprise given that models which are highly parameterized, such as the AR(12) model or the VAR(4) model, will tend to feature parameter estimates that can be more unstable with the addition of future data. Our proposed approach of imposing a low signal-to-noise ratio on a high-order AR(p) model estimated via Bayesian methods with a shrinkage prior on second-difference coefficients produces a very reliable estimate of the output gap. Amongst the various methods that implicitly or explicitly impose a low signal-to-noise ratio, including the HP and BP filters, the BN filter performs by far the best. At the same time, the BN decomposition based on an AR(1) model also performs very well in terms of reliability, perhaps begging the question of why we impose a low signal-to-noise ratio. After discussing some robustness issues next, we provide some justification for imposing a low signal-to-noise ratio in Section 3.

2.3 Robustness

To explore the robustness of our proposed approach, we first address Perron and Wada's (2009) claim that U.S. log real GDP should be modeled with a break in the trend growth rate in 1973Q1. In our approach, because we estimate an AR(p) model for output growth, we implicitly assume the trend in U.S. log real GDP is a random walk with constant drift. To check robustness, we re-estimate the AR(p) model with the addition of a dummy in the intercept to account for a break in trend growth in 1973Q1. The left panel of Figure 6 presents results with and without accounting for the structural break. The differences in the output gap estimates are trivial. This is because the magnitude of the estimated break in drift is relatively small in comparison to the unconditional variance of U.S. quarterly real GDP growth. Thus, as long as the autoregressive coefficient estimates are little impacted by allowing for a break, the estimated output gap from the BN decomposition will be largely unchanged.[8] We therefore conclude that the BN filter is robust to allowing for a structural break in long-run output growth in 1973Q1.[9]

We also consider the sensitivity of our approach to varying the exact value of the signal-to-noise ratio. The top-right panel of Figure 6 presents the estimated output gap given a somewhat higher signal-to-noise ratio, $\delta = 0.6$, and a lower signal-to-noise ratio, $\delta = 0.1$, alongside with our original results for $\bar{\delta} = 0.25$. Increasing the signal-to-noise ratio somewhat reduces the amplitude of the output gap, but the shape of the estimated output gap is little changed, with the persistence profile virtually unaltered.[10] Because the profile of fluctuations in the estimated output gap are similar even as we increase the signal-to-noise ratio, observations of turning points, revision properties, and real time forecast performance will be relatively robust to the exact choice of δ, at least as long as it is well below one. Meanwhile, given this apparent robustness to different values of δ, we check whether our results are actually being driven by the AR(12) specification rather than imposing a low signal-to-noise ratio. To do this, we compare the estimated output gap from the BN filter to that produced by the BN decomposition based on an AR(12) model freely estimated via MLE. The bottom-right panel Figure 6 presents the two output gap estimates and makes it clear that imposing a reasonably low signal-to-noise ratio is crucial. For the freely estimated AR(12) model, we obtain $\hat{\delta} = 1.86$ and the correlation between the two estimates is only 0.39.

[8] For example, thinking back the to analytical formula for the cycle from the BN decomposition based on an AR(1) model of $-\phi(1-\phi)^{-1}(\Delta y_t - \mu)$, note that it will only change significantly if the estimate of μ changes with the structural break by a large amount relative to the variance of Δy_t. This is not the case for the U.S. data.

[9] If there were evidence of a structural break in the persistence of U.S. real GDP growth, it might motivate consideration of a break in the imposed signal-to-noise ratio given the link between δ and $\phi(1)$ for an AR(p) model. However, we find no evidence for such a break.

[10] The correlation between the different estimated output gaps varying the signal-to-noise ratio is well in excess of 0.95.

3 Is Our Proposed Approach Reasonable?

3.1 Justification for Imposing a Low Signal-to-Noise Ratio

To recap, we have proposed a BN filter that imposes a low signal-to-noise ratio when conducting the BN decomposition. When applied to U.S. log real GDP, the resulting output gap estimates are reliable in the Orphanides and van Norden (2002) sense of being subject to small revisions over time. The BN filter does better than other popular methods, except for the BN decomposition based on an AR(1) model estimated via MLE, which does marginally better in terms of revisions, but corresponds to a much higher signal-to-noise ratio.

To the extent that one is agnostic about the true signal-to-noise ratio, there is little reason to deviate from the BN decomposition based on an AR(1) model, especially if real-time reliability is the sole criterion for choosing an approach to estimating the output gap. In other words, one can really only justify using the BN filter if there is a compelling reason to believe that a low signal-to-noise ratio represents the true state of the world. Whether a low or high signal-to-noise ratio represents the true state of the world remains unresolved in the empirical literature. While considerable empirical research has found support for the presence of a volatile stochastic trend in U.S. log real GDP (e.g., Nelson and Plosser, 1982; Morley et al., 2003), this view has not gone unchallenged (e.g., Cochrane, 1994; Perron and Wada, 2009).

One reason to believe the signal-to-noise ratio is much lower than that given by a freely estimated AR model is that $\{\Delta y_t\}$ may behave more like an MA process with a near unit root than a finite-order autoregressive process. In this case, the true signal-to-noise ratio would be small and the process would have an infinite-order AR representation. However, a finite-order AR(p) model would fail to capture the infinite-order AR dynamics and the estimated signal-to-noise ratio for such models could be biased upwards.

To demonstrate this possibility, we consider two empirically-plausible data generating processes (DGPs). In both cases, the observed time series is equal to a random walk with constant drift trend plus an AR(2) cycle. Furthermore, in both cases, the first difference of the time series follows the exact same ARMA(2,2) process with a low signal-to-noise ratio of $\delta = 0.50$.

For the first DGP, we parameterize the Watson (1986) UC model of $\{y_t\}$ with uncorrelated components as estimated for U.S. real GDP by Morley et al. (2003). We choose this DGP because it corresponds to a low signal-to-noise ratio, unlike the unrestricted UC model in Morley et al. (2003) that allows for correlation between permanent and transitory movements.[11] When considering model selection for possible ARMA specifications

[11]Morley et al. (2003) find that a zero correlation restriction can be rejected at the 5% level based on a likelihood ratio test. However, small values for the correlation cannot be rejected. Thus, we argue that this DGP is empirically plausible, if not necessarily probable in a Bayesian sense.

for $\{\Delta y_t\}$ given this DGP in finite samples, SIC will pick a low-order AR(p) model with reasonably high frequency, even though the true model has an ARMA(2,2) specification. Meanwhile, suppose there is some other observed variable $\{u_t\}$ that is related to the unobserved cycle $\{c_t\}$, but contains serially-correlated "measurement error". Tests for Granger causality will often suggest that $\{u_t\}$ has predictive power for $\{\Delta y_t\}$ beyond a low-order univariate AR(p) process. Based on this, a researcher might consider a multivariate BN decomposition, as argued for by Cochrane (1994) in such a setting. For this DGP, we consider how well different cases of the BN decomposition would do in estimating the true cycle $\{c_t\}$. Table 1 reports the results in a finite sample (T=250) and in population (T=500,000).

First, we find that the BN decomposition based on an AR(1) model does poorly in estimating the true cycle, both in a finite sample and in population. The estimated cycle is negatively correlated with the true cycle and its amplitude (as measured by standard deviation) is only about 20% that of the true cycle. So this is exactly the example of a true state of the world in which the BN decomposition based on an AR(1) model would be a bad approach to estimating the output gap, even though SIC might select a low-order AR model in a finite sample. Notably, when T=250 for this DGP, we find that SIC chooses a lag order for an AR(p) model of p=1 more than 95% of the time.

Next, we find that a multivariate BN decomposition based on a VAR(4) model of $\{\Delta y_t\}$ and the true cycle $\{c_t\}$ almost perfectly estimates the true cycle. This is not too surprising given the inclusion of the true cycle in the forecasting model corresponds to a highly unrealistic scenario in which there exists an observed variable that perfectly captures economic slack. Indeed, if such a variable really did exist, there would be little reason to estimate the output gap in the first place rather than just monitoring the observed variable when conducting any sort of current analysis. Instead, a more realistic scenario is one in which there exists an observed variable that is related to economic slack but is also affected by persistent idiosyncratic factors (e.g., the unemployment rate). In order to capture such a scenario, we generate an artificial time series $\{u_t\}$ which is linked to the true cycle $\{c_t\}$ up to a a persistent measurement error. When we estimate the cycle from a multivariate BN decomposition based on a VAR(4) model of $\{\Delta y_t\}$ and $\{u_t\}$, its correlation with the true cycle drops to around 0.5 and its amplitude is much less than that of the true cycle. These results hold in both a finite sample and in population.[12]

A natural question is what role does model misspecification play in the results for the BN decomposition. To consider this, we conduct the BN decomposition based on an ARMA(2,2) model estimated via MLE. Despite the fact that the model is correctly

[12]Furthermore, a researcher might not consider a multivariate BN decomposition in the first place given this DGP and a finite sample. In particular, when T=250, we find that a test of no Granger causality from $\{u_t\}$ to $\{\Delta y_t\}$ only rejects 30% of the time for a VAR(4) model. It should be noted, however, that this relatively low power clearly reflects the relative magnitude of the measurement error in $\{u_t\}$, as the empirical rejection rate is effectively 100% for a test of no Granger causality from $\{c_t\}$ to $\{\Delta y_t\}$.

specified, we can see that correlation between the estimated cycle and the true cycle is less than one and the amplitude is less than for the true cycle, even in population. This is similar to what was found in Table 1 of Morley (2011), where the BN decomposition based on the true model provides an unbiased estimator of the standard deviation of trend shocks for a UC process, but the estimate of the standard deviation of the cycle is downward biased. As long as the cycle is unobserved, there will always be a downward bias in estimating it.[13] Meanwhile, the finite sample results for the ARMA(2,2) model are much worse than the population results, with the correlation between the estimated cycle and the true cycle being close to zero. The relatively poor finite sample performance of the BN decomposition in this case likely reflects well-known difficulties with estimating ARMA parameters due to weak identification.

Turning to our proposed BN filter, we find that the estimated cycle shares the same relatively high correlation with the true cycle as a version of the BN decomposition that imposes the true signal-to-noise ratio (which, of course, is never known in practice) and the BN decomposition based on an AR(12) model. The AR(12) model does reasonably well given that it approximates the infinite-order AR representation of $\{\Delta y_t\}$. However, for this DGP, the BN decomposition based on the AR(12) model suffers from a larger downward bias in estimating the amplitude than our proposed approach. Indeed, the BN filter does even better in terms of amplitude than the BN decomposition imposing the true signal-to-noise ratio because it explicitly involves targeting $\bar{\delta}$ to maximize amplitude subject to a tradeoff with fit.

Following Morley (2011), we also consider a second DGP for which the BN trend based on the true model defines the trend rather than just provides an estimate of an unobserved random walk trend component, as was the case with the first DGP. In particular, we consider a single-source-of-error process (see Anderson et al., 2006) that is parameterized to imply the same ARMA(2,2) process for $\{\Delta y_t\}$ as the first DGP. Thus, the same signal-to-noise ratio and all the same tendencies for SIC to pick a low-order AR model hold for this DGP. The only difference in a univariate context is a conceptual one about whether forecast errors represent true trend shocks (i.e., they are the "single source of error" in the process for $\{y_t\}$) or they are linear combinations of unobserved trend and cycle shocks, as was the case in the first DGP. See Morley (2011) for a full discussion of this conceptual distinction.

Table 2 reports the results for the second DGP and they are fairly similar to before, except that the BN decomposition generally does a better job estimating the amplitude of the true cycle. However, there are a few key results to highlight. First, the BN decomposition based on the ARMA(2,2) model still does poorly in terms of the correlation

[13] The BN decomposition based on a VAR(4) model of $\{\Delta y_t\}$ and the true cyclical component $\{c_t\}$ does not suffer from a downward bias because the cyclical component is observed and the true DGP has a (restricted) VAR(2) representation.

with true cycle in finite samples, despite being correctly specified. The point here is that the estimation problems for ARMA models remain massive even for sample sizes as large as T=250. Imposing a low signal-to-noise ratio for an AR model appears to be a more effective way at getting at the true cycle than estimating the true model when estimation involves weak identification issues. Second, the BN decomposition based on a VAR(4) model of $\{\Delta y_t\}$ and $\{u_t\}$ does worse than for the first DGP, suggesting that measurement error in an observed measure of economic slack offsets the benefits of having a forecast error represent the true trend shock. Again, imposing a low signal-to-noise ratio appears to be a more straightforward and effective way to get the estimated cycle closer to the true cycle than adding multivariate information, even if the multivariate information also decreases the signal-to-noise ratio, as discussed in Evans and Reichlin (1994). Determining the appropriate multivariate information is a difficult econometric problem in itself, with finite-sample power issues and, at the same time, considerable danger of overfitting unless variable selection is carefully handled.[14] Meanwhile, even given the correct multivariate information, the practical issue of measurement error that effectively motivates the need to estimate the cyclical component in the first place means that a multivariate BN decomposition will suffer even in population. Third, although the BN decomposition based on an AR(12) model estimated via MLE does relatively well, especially for this DGP, we know from the analysis in the previous section that, like the BN decomposition based on a VAR model, it suffers from much larger reliability problems than our proposed approach.

The bottom line is that it is possible to think of a true state of the world in which standard model selection criteria and hypothesis testing would push a researcher towards an AR(1) model (based on parsimony), an ARMA(2,2) model (as estimation and testing eventually discovers the true model given enough data), or possibly a VAR model (based Granger causality tests), but the BN decomposition based on these models would do much worse at capturing the true cycle than our proposed approach. Although the ARMA(2,2) model is the correct specification, the BN decomposition for this model suffers in finite samples due to known estimation problems for such models. The BN decomposition for the AR(1) model performs poorly in large samples, as does the VAR(4) model when the multivariate information is measured with error. Meanwhile, even though the BN decomposition based on an AR(12) model does reasonably well, as would a VAR(4) model when the multivariate information is measured accurately, these versions of the BN decomposition still suffer from reliability issues. By contrast, the BN filter works well even in finite samples and is reliable.

[14]Interestingly, the finite-sample power of the Granger causality tests for $\{u_t\}$ to $\{\Delta y_t\}$ and $\{c_t\}$ to $\{\Delta y_t\}$ is lower for this DGP than the first one. In particular, when T=250, the respective empirical rejection rates for a VAR(4) model are only 8% and 51% compared to 30% and 100% for the first DGP. Thus, a researcher who only considered $\{u_t\}$ or $\{c_t\}$ as a possible predictive variable would be even less likely to consider a multivariate BN decomposition if this DGP represented the true state of the world.

Next, we consider whether the BN filter is also reliable in the sense of avoiding spurious cycles. Although we might worry about model selection criteria and hypothesis testing pushing us to consider models that lead to poor estimates of the output gap, we should at the same time worry that imposing a low signal-to-noise ratio could lead to a spurious cycle if the true state of the world for U.S. real GDP growth is more along the lines of an AR(1) model than the two DGPs considered above. In particular, if the BN filter produces a spurious cycle, then our estimated output gap should not perform as well as the BN decomposition based on a freely estimated AR(1) model in forecasting output growth and inflation out of sample. We check whether this is the case in practice.

3.2 Out-of-Sample Forecast Comparison

In this subsection, we evaluate different trend-cycle decomposition methods in terms of the ability of their pseudo-real-time output gap estimates to forecast future U.S. output growth and inflation. Our forecast evaluation sample starts in 1970Q1. We use an expanding window for estimation. The first estimate of an output gap we have is for 1947Q2. We use the full extent of the data sample for forecast evaluation after adjusting for the maximum number of lags in the forecasting equation.

U.S. Output Growth Forecast Nelson (2008) argues for using forecasts of future output growth as a way to evaluate competing estimates of the output gap. The underlying intuition is that if an estimated output gap suggests output is below trend, this should imply faster output growth in the future 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 will return to zero in the long run, and a good estimate of the output gap should be able to forecast the effects of this reversion. For an h-period-ahead output growth forecast, we consider a forecasting equation similar to Nelson (2008):

$$y_{t+h} - y_t = \alpha + \beta \hat{c}_t + \epsilon_{t+h,t} \tag{5}$$

where y is the natural log of real GDP, $\hat{c}$ is the estimated output gap, ϵ is a forecast error, and α and β are coefficients estimated using least squares. Therefore, for an accurate estimate of the output gap, we expect $\beta < 0$ and the inclusion of the estimated output gap in the forecast equation to help forecast h-period-ahead output growth.

Figure 7 presents the out-of-sample forecasting results. The Relative Root Mean Squared Errors (RRMSEs) are in comparison to forecasts using the BN filter estimate of the output gap. The figure includes 90% confidence bands obtained by inverting the Diebold and Mariano (1995) test of equal predictive accuracy.

We make two observations. First, the output gap estimates constructed using the

BN decomposition forecast better than those based on other methods, including the HP and BP filters. This further vindicates our choice to work with a BN decomposition. Second, within the class of BN decompositions, similar to with the revision statistics, parsimony or shrinkage priors seem to be key. In particular, the BN decomposition based on AR(12), ARMA(2,2), and VAR(4) models do worse than the BN decomposition based on an AR(1) model or the BN filter. Therefore, the results for forecasting output growth mimic many of the results we had for the revisions and reliability statistics.[15] Notably, despite imposing a low signal-to-noise ratio, the BN filter avoids producing a spurious cycle that would diminish the forecasting performance of an output gap estimate out of sample. It is true that the estimated output gap from the BN decomposition based on an AR(1) model does slightly better at short horizons. But the RRMSE is always close to one and is not significant at longer horizons. So, unlike other methods that produce intuitive estimates of the output gap by imposing a low signal-to-noise ratio, our approach does not seem to produce a spurious cycle.

U.S. Inflation Forecast We also consider a Phillips Curve type inflation forecasting equation to evaluate competing estimates of the output gap. Similar to, amongst others, Stock and Watson (1999, 2008) and Clark and McCracken (2006), we use a fairly standard specification from the inflation forecasting literature. In particular, we specify the following autoregressive distributed lag (ADL) representation for our pseudo-real-time h-period-ahead Phillips Curve inflation forecast:

$$\pi_{t+h} - \pi_t = \gamma + \sum_{i=0}^{p} \theta_i \triangle \pi_{t-i} + \sum_{i=0}^{q} \kappa_i \hat{c}_{t-i} + \epsilon_{t+h,t}. \tag{6}$$

where π is U.S. CPI inflation.[16] We choose the lag orders of the forecasting equation, namely p and q above, using the SIC. As is commonly done (see, for example, Stock and Watson, 1999, 2008; Clark and McCracken, 2006), we apply the information criteria to the entire sample and run the pseudo-real-time analysis using the same number of lags, implicitly assuming we know the optimal lag order *a priori*. The set of lag orders we consider for our ADL forecasting equation are $p \in [0, 12]$ and $q \in [0, 12]$.

Figure 8 presents the out-of-sample forecasting results for the U.S inflation. Once again, as with the results for output growth, we compute 90% confidence intervals by inverting the Diebold and Mariano (1995) test.

As in the case of forecasting output growth, the BN filter estimate of the output gap does relatively well. In particular, imposing a low signal-to-noise ratio allows us to

[15]Similar results hold in terms of forecasting changes in the unemployment rate too. These results are available from the authors upon request.

[16]The raw monthly data for the U.S. Consumer Price Index (CPI) for all urban consumers (seasonally adjusted) are again taken from FRED and are converted to the quarterly frequency for 1947Q1 to 2015Q3 by simple averaging.

outperform all other BN decompositions, although generally not significantly so. We also generally do better than the HP filter, BP filter, and deterministic detrending, although again not significantly so. In contrast to the results for forecasting output growth, the differences in inflation forecast performance using the different output gap estimates are fairly small, with most RRMSEs within the 1.00 to 1.05 range, indicating the gains in changing the output gap estimates for forecasting inflation can be marginal at best and are generally not significant. To some extent, this is not entirely surprising. Atkeson and Ohanian (2001) and Stock and Watson (2008) show that real-activity based Phillips Curve type forecasts may not be particularly useful for forecasting inflation. In some sense, then, our results are simply a manifestation of what is commonly found in the inflation forecasting literature. However, we note that our proposed approach is still competitive and may be slightly better than competing options in terms of providing a good real-time measure of economic slack. In particular, the BN filter estimate of the output gap produces statistically significantly better inflation forecasts at some horizons relative to approaches such as the HP filter and deterministic detrending. It is also noteworthy that none of the alternative methods outperform our proposed approach in a statistically significant way.

3.3 Are Large Revisions Useful for Understanding the Past?

In this section, we discuss whether heavily revised output gap estimates, although less useful for current analysis in real time, provide a better ex post understanding and interpretation of past economic activity.

Without ever knowing the true output gap, it can be a challenge to evaluate the historical accuracy of output gap estimates. However, we attempt to address the accuracy issue in two ways. First, we consider the ability of revised estimates to predict future inflation. Figure 9 presents a pseudo-out-of-sample U.S. inflation forecast comparison using revised output gap estimates instead of real-time estimates. Notably, we find only small differences in the forecasting results, with the relative performance of heavily-revised approaches often deteriorating in comparison to using the pseudo-real-time estimates.[17] Thus, the large revisions for deterministic detrending, the HP filter, the BP filter, and the Harvey-Clark model are clearly not providing any additional insights into the historical values of the output gap that are relevant for inflation. Second, we consider the relationship of the various revised estimates of the output gap with an alternative measure of U.S. economic activity that is constructed in a completely different way, namely the Chicago

[17] We do not report a similar exercise for predicting future output growth because the revised estimates for deterministic detrending, the HP filter, the BP filter, and the Harvey-Clark model will directly reflect future output growth. So it would be of little surprise, but not economically meaningful, that they would predict future output growth better than pseudo-real-time estimates. We thank Adrian Pagan for pointing this out.

Fed's National Activity Index (CFNAI). The CFNAI is a revised index of activity based on 85 data series. The top panel of Figure 10 plots the CFNAI and the revised estimate for the BN filter. Despite our proposed approach being based only on the U.S. quarterly real GDP data series, it displays a remarkable similarity to the CFNAI. Meanwhile, the bottom panel presents correlations between the CFNAI and the various revised estimates of the output gap for different methods. The correlations confirm a strong positive relationship between the CFNAI and the estimate based on the BN filter, with much weaker and sometimes even negative relationships for the other estimates. Again, this suggests that the large revisions for some of the other methods are not necessarily capturing anything about the true output gap. Also, this result confirms that our approach provides a convenient way to measure economic slack in that it provides a shortcut to a large-scale multivariate approach that would also lead to a lower signal-to-noise ratio, while avoiding the overfitting and instability issues that inevitably arise with such multivariate analysis.

4 Conclusion

In this paper, we have proposed a modification of the traditional BN decomposition to directly impose a low signal-to-noise ratio. In particular, rather than focusing solely on model fit by freely estimating a time series forecasting model, we propose a "BN filter" that trades off model fit and amplitude in order to determine a lower signal-to-noise ratio to impose in the Bayesian estimation of a univariate AR model. When applied to postwar U.S. quarterly log real GDP, the BN filter produces estimates of the output gap that are both intuitive and reliable, while estimates for other methods are, at best, either intuitive or reliable, but never both. In particular, the BN filter retains the general reliability of the BN decomposition based on freely estimated AR models, but the estimated output gap is much more intuitive in the sense of being relatively large in amplitude, persistent, and moving closely with the NBER reference cycle. Meanwhile, other methods that produce similarly intuitive estimates of the output gap are far less reliable in terms of their revision properties.

We motivate why it can be useful to impose a low signal-to-noise ratio. In particular, if the true state of the world is one in which there is an unobserved output gap that is large in amplitude and persistent, other methods can produce misleading estimates of the output gap in finite samples. By contrast, the BN filter performs relatively well in terms of correlation with the true output gap. At the same time, despite imposing a low signal-to-noise ratio, our proposed approach is also reliable in the sense that it does not appear to generate a spurious cycle in U.S. log real GDP. Specifically, the estimated output gap from the BN filter forecasts U.S. output growth and inflation similarly to estimated output gap from the BN decomposition based on a freely estimated AR(1) model and better than for other methods, especially those that also impose a low signal-to-noise ratio. The revised estimate from the BN filter also appears to be more accurate than more heavily-revised output gap estimates in terms of its relationships with inflation and a well-known revised measure of U.S. economic activity, the Chicago Fed's National Activity Index, that is constructed using a large number of economic variables.

References

Anderson HM, Low CN, Snyder R. 2006. Single source of error state space approach to the Beveridge Nelson decomposition. *Economics Letters* **91**: 104–109.

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.

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, Kang H. 1981. Spurious periodicity in inappropriately detrended time series. *Econometrica* **49**: 741–51.

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.

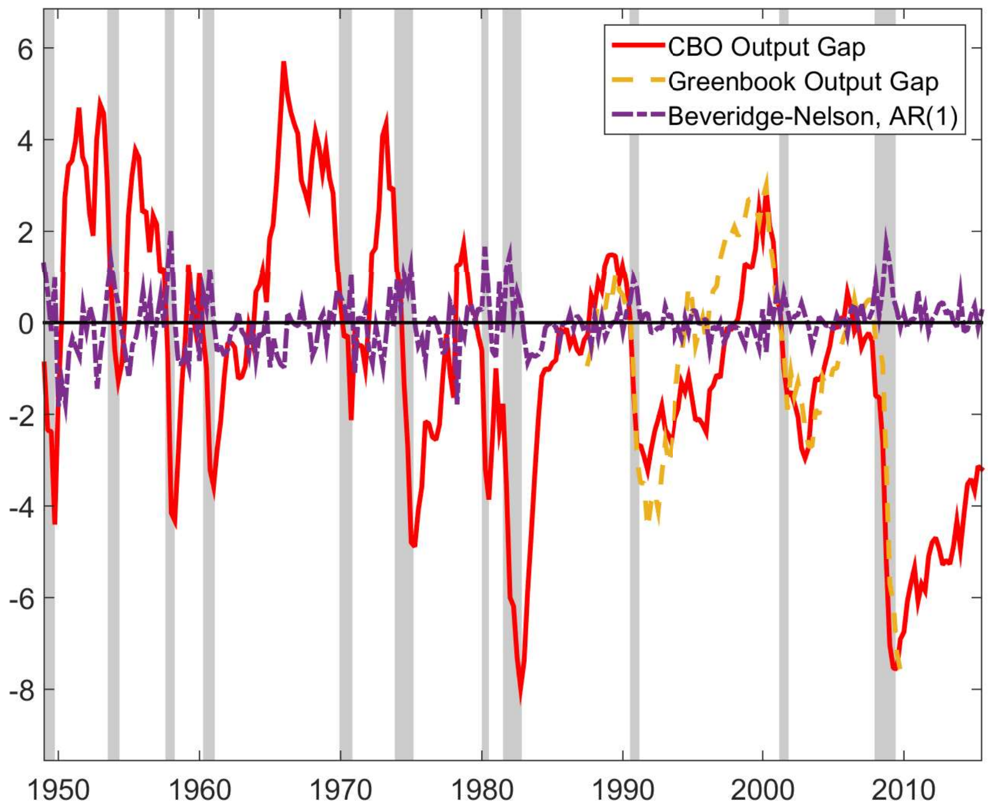
Phillips P, Jin S. 2015. Business cycle, trend elimination, and the HP filter. *Crwles Foundation Discussion Paper No. 2005* .

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.

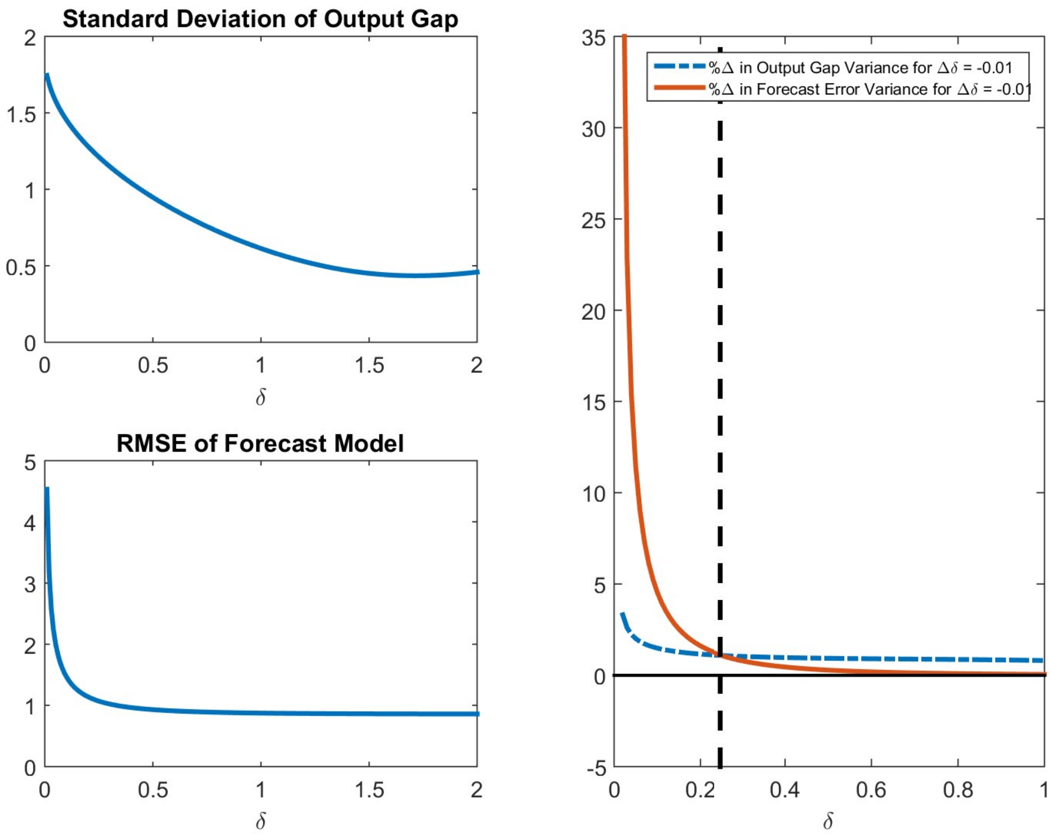
Watson MW. 1986. Univariate detrending methods with stochastic trends. *Journal of Monetary Economics* **18**: 49–75.

Figure 1: Three estimates of the U.S. output gap



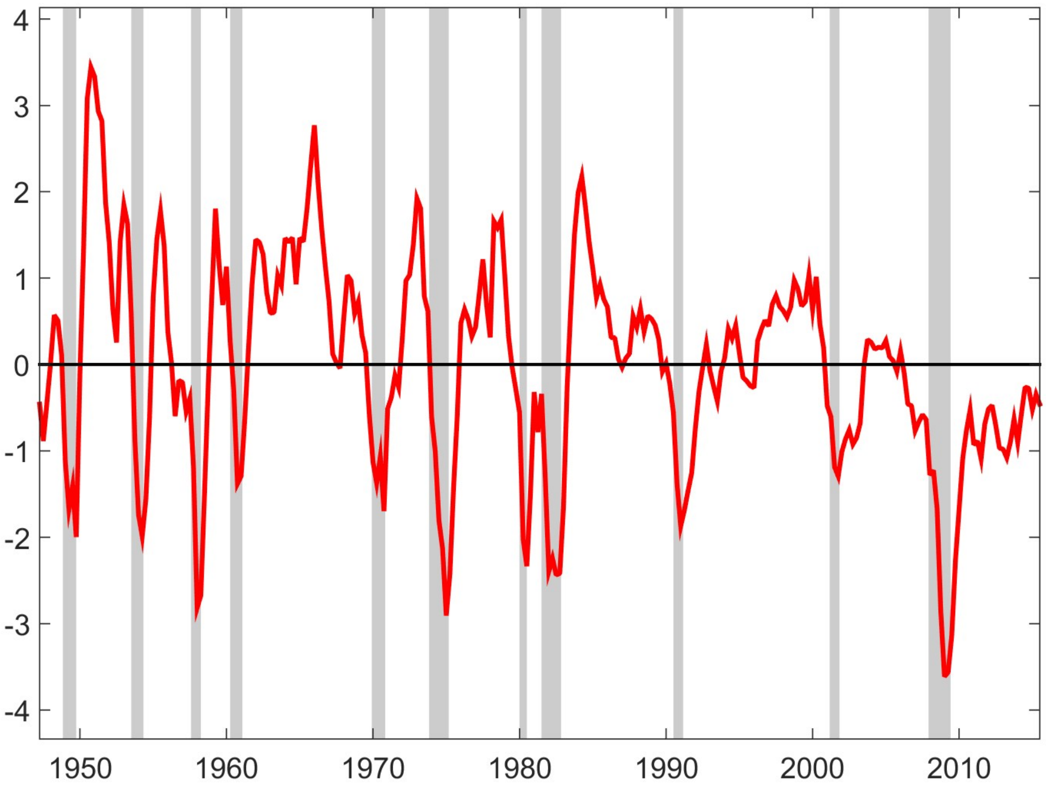
Notes: Percent deviation from trend. Shaded areas correspond to NBER recession dates. CBO output gap refers to an estimate based on the CBO's estimate of potential output. Greenbook output gap refers to the Fed's estimate. The Beveridge-Nelson decomposition estimate of the output gap is based on an AR(1) model of U.S. quarterly real GDP growth estimated via MLE.

Figure 2: Tradeoff across different signal-to-noise ratios between amplitude of estimated U.S. output gap based on the BN decomposition and forecasting model fit



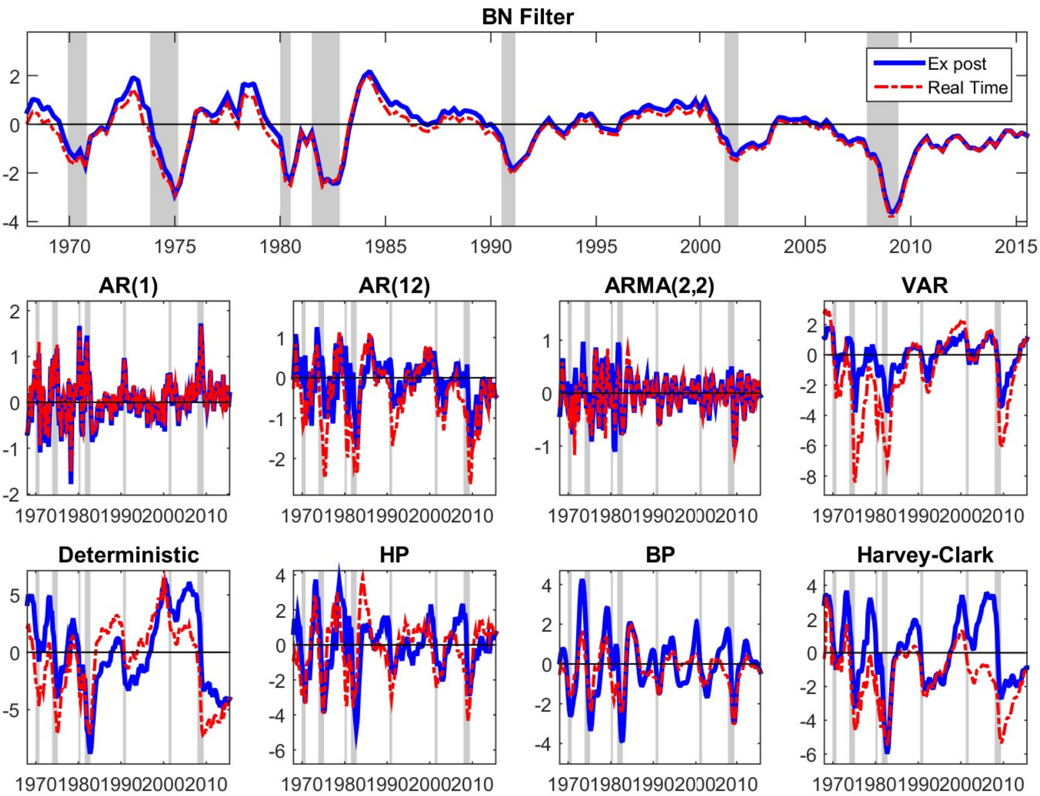
Notes: δ is the signal-to-noise ratio in terms of the variance of the trend shocks as a fraction of the overall quarterly forecast error variance. The estimated output gap is for a BN decomposition based on Bayesian estimation of an AR(12) model of U.S. quarterly real GDP growth with shrinkage priors on second-difference coefficients and different values of δ.

Figure 3: Estimated U.S. output gap based on the BN filter



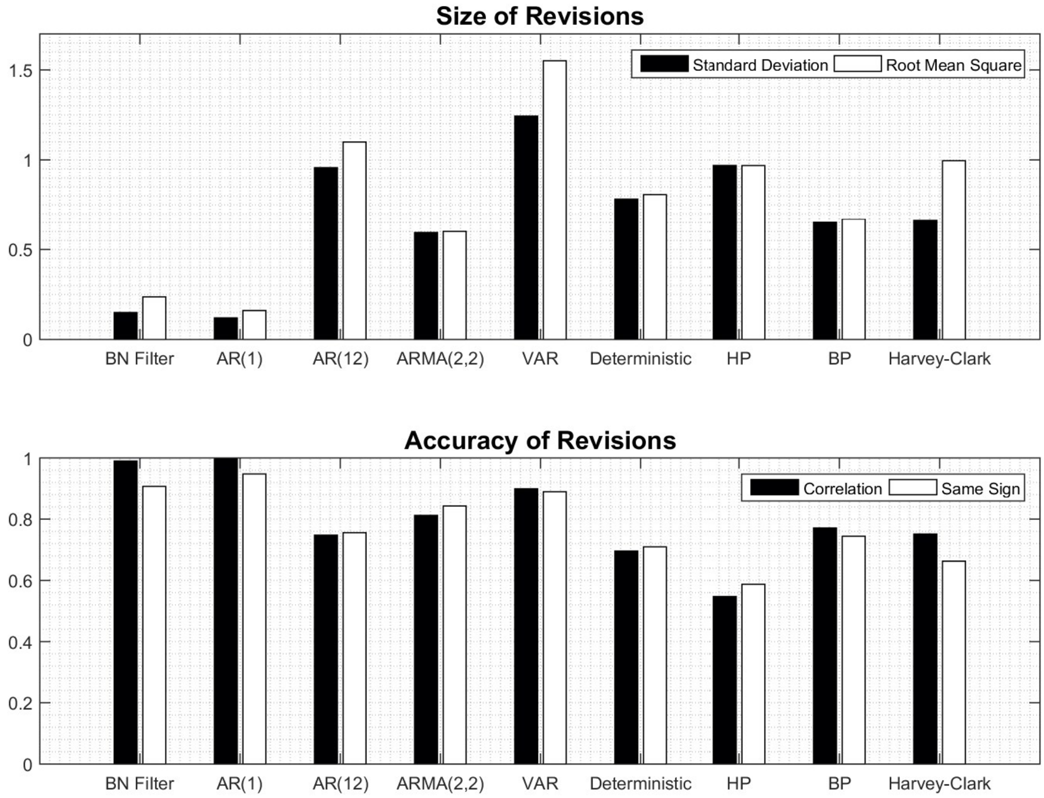
Notes: Percent deviation from trend. Shaded areas correspond to NBER recession dates. "BN filter" refers to our proposed approach of estimating the output gap using the BN decomposition based on Bayesian estimation of an AR(12) model of U.S. quarterly real GDP growth with shrinkage priors on second-difference coefficients and imposing the signal-to-noise ratio $\bar{\delta} = 0.25$ *that optimizes tradeoff between amplitude and fit.*

Figure 4: Revision properties of U.S. output gap estimates



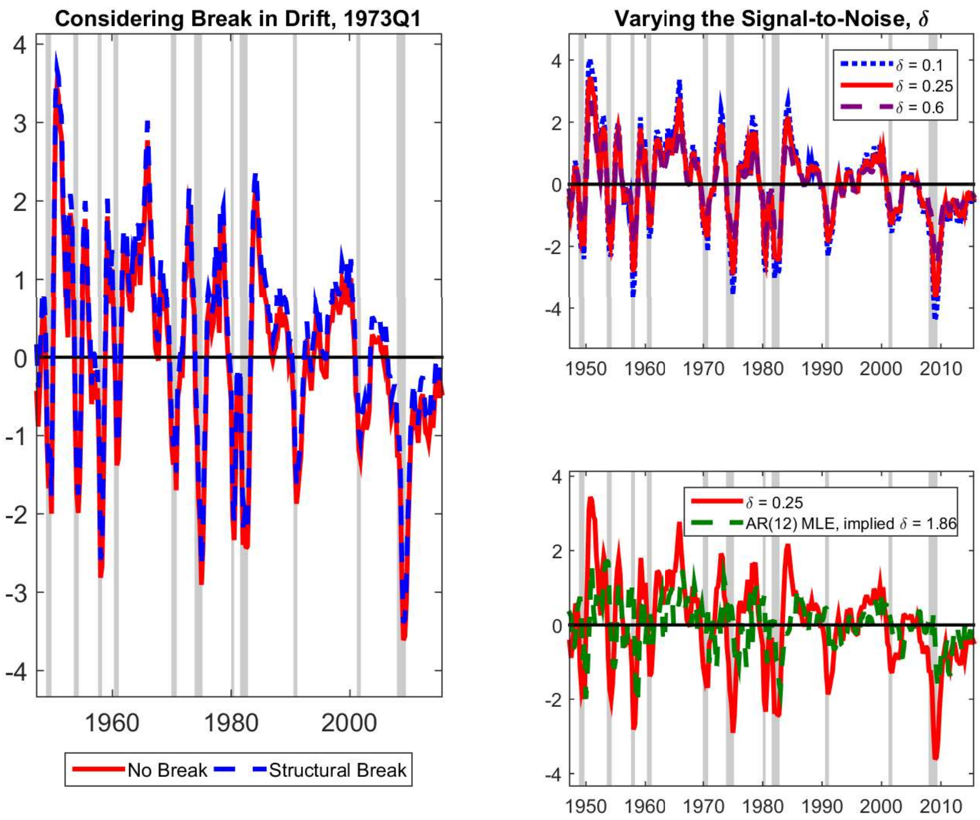
Notes: Percent deviation from trend. Shaded areas correspond to NBER recession dates. "BN filter" refers to our proposed approach of estimating the output gap using the BN decomposition based on Bayesian estimation of an AR(12) model of U.S. quarterly real GDP growth with shrinkage priors on second-difference coefficients and imposing a signal-to-noise ratio that optimizes tradeoff between amplitude and fit. "AR(1)", "AR(12)", and "ARMA(2,2)" refer to BN decompositions based on the respective models estimated via MLE. "VAR(4)" refers to the BN decomposition based on a VAR(4) model of output growth and the unemployment rate estimated via MLE. "Deterministic" refers to detrending based on least squares regression on a quadratic time trend. "HP" refers to the Hodrick and Prescott (1997) filter. "BP" refers to the bandpass filter of Christiano and Fitzgerald (2003). "Harvey-Clark" refers to the UC model as described by Harvey (1985) and Clark (1987).

Figure 5: Revision statistics for U.S. output gap estimates, 1970Q1 - 2012Q4



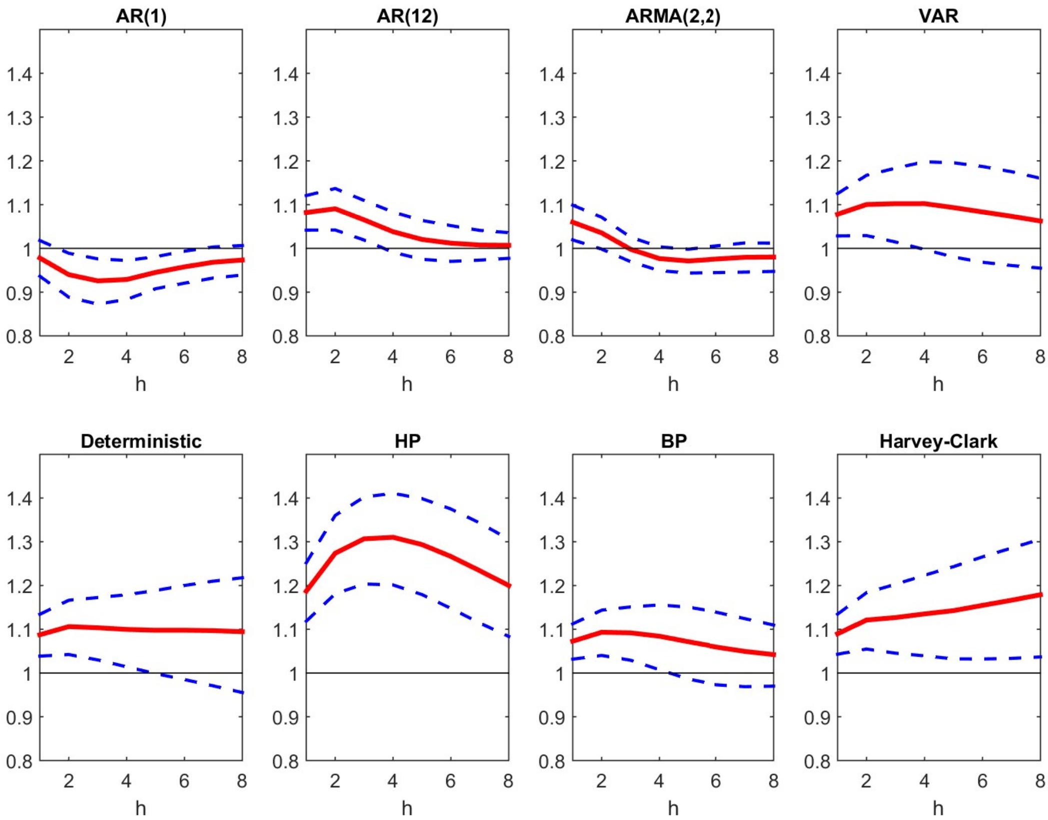
Notes: See notes in Figure 4 for description of methods. Standard deviation and root mean square of revisions to the pseudo-real-time estimate of the output gap are normalized by the standard deviation of the ex post estimate of the output gap. "Correlation" refers to the correlation between the pseudo-real-time estimate and the ex post estimate of the output gap. "Correct sign" refers to the proportion of pseudo-real-time estimates that share the same sign as the ex post estimate of the output gap.

Figure 6: Robustness analysis for U.S. output gap estimates from the BN filter



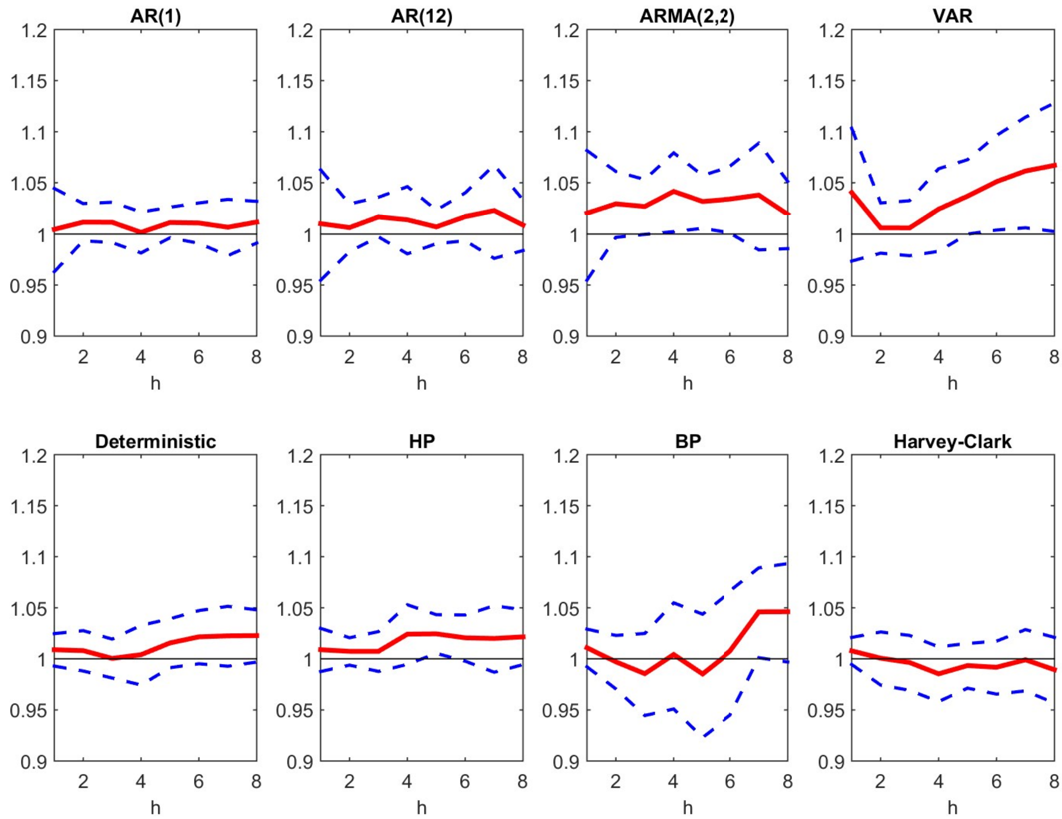
Notes: Percent deviation from trend. Shaded areas correspond to NBER recession dates. "BN filter" refers to our proposed approach of estimating the output gap using the BN decomposition based on Bayesian estimation of an AR(12) model of U.S. quarterly real GDP growth with shrinkage priors on second-difference coefficients and imposing a signal-to-noise ratio that optimizes tradeoff between amplitude and fit or is set to another particular value. "AR(12) MLE" refers to the BN decompositions based on an AR(12) model estimated via MLE.

Figure 7: Out-of-sample U.S. output growth forecast comparison relative to the BN filter benchmark using pseudo-real-time output gap estimates



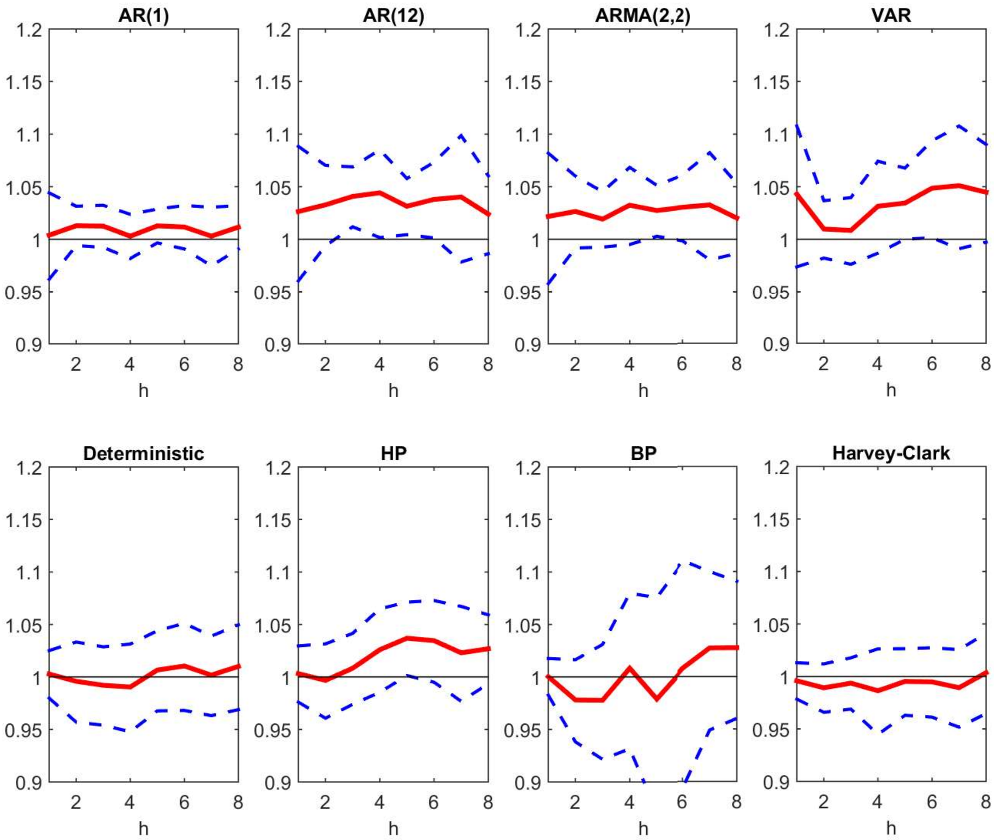
Notes: See notes in Figure 4 for description of methods. The graphs plot out-of-sample RRMSE compared to forecasts based on the BN filter estimated output gap. Out-of-sample evaluation begins in 1970Q1. The bands depict 90% confidence intervals from a two-sided Diebold and Mariano (1995) test of equal forecast accuracy.

Figure 8: Out-of-sample U.S. inflation forecast comparison relative to the BN filter benchmark using pseudo-real-time output gap estimates



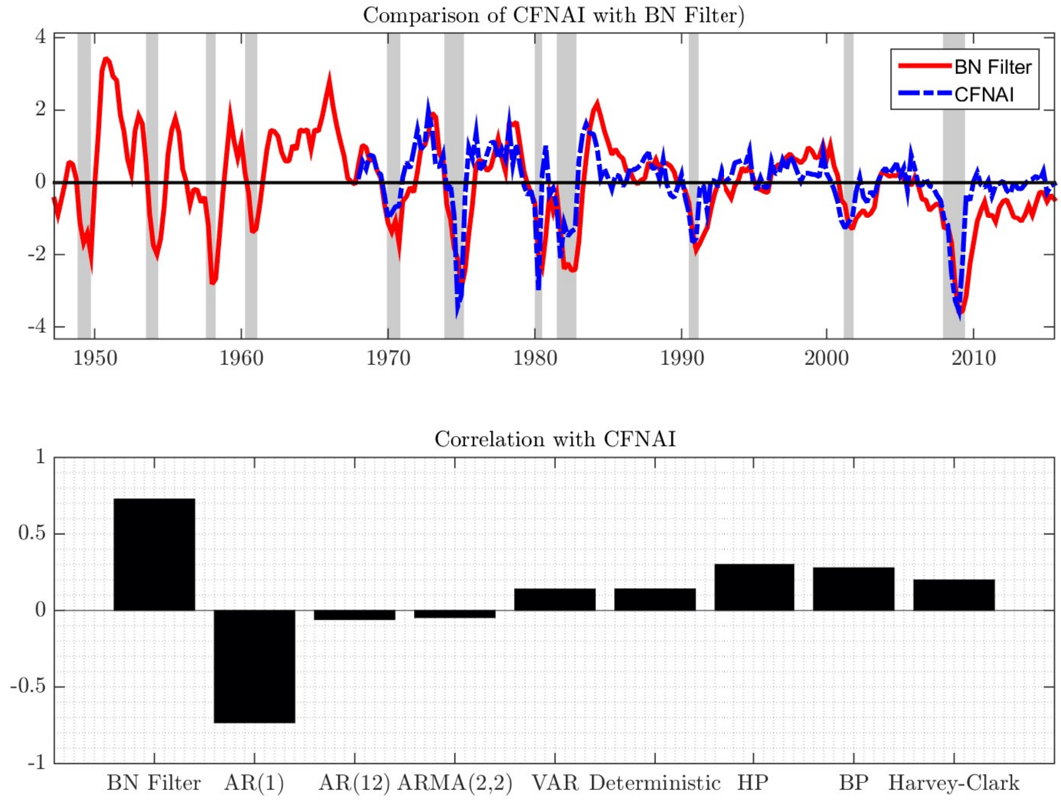
Notes: See notes in Figure 4 for description of methods. The graphs plot out-of-sample RRMSE compared to forecasts based on the BN filter estimated output gap. Out-of-sample evaluation begins in 1970Q1. The bands depict 90% confidence intervals from a two-sided Diebold and Mariano (1995) test of equal forecast accuracy.

Figure 9: Pseudo-out-of-sample U.S. inflation forecast comparison relative to the BN filter benchmark using revised output gap estimates



Notes: See notes in Figure 4 for description of methods. The graphs plot pseudo-out-of-sample RRMSE compared to forecasts based on the BN filter estimated output gap. Forecast evaluation begins in 1970Q1, but is only a pseudo-out-of-sample evaluation given the use of revised output gap estimates. The bands depict 90% confidence intervals from a two-sided Diebold and Mariano (1995) test of equal forecast accuracy.

Figure 10: Relationship of output gap estimates with the Chicago Fed's National Activity Index



Notes: Percent deviation from trend for estimated output gap and reported index units for the Chicago Fed's National Activity Index (CFNAI). Shaded areas correspond to NBER recession dates. See notes in Figure 4 for description of methods. Revised estimates are plotted and used in the calculation of correlations.

Table 1: Monte Carlo simulation for unobserved components process

	T = 250		**T = 500,000**	
	Correlation	**Amplitude**	**Correlation**	**Amplitude**
True Cycle		2.47		2.53
AR(1)	-0.12	0.51	-0.12	0.50
VAR(4)$[\triangle y_t, c_t]$	0.99	2.48	1.00	2.54
VAR(4)$[\triangle y_t, u_t]$	0.48	1.71	0.49	1.73
ARMA(2,2)	0.16	1.15	0.66	1.66
BN filter$[\delta = \bar{\delta}]$	0.49	1.27	0.59	1.97
BN filter$[\delta = 0.50]$	0.51	0.96	0.59	1.44
AR(12)	0.56	1.01	0.59	0.96

Notes: We consider the following DGP:

$y_t = \tau_t + c_t$

$\tau_t = 0.81 + \tau_{t-1} + \eta_t,$

$c_t = 1.53c_{t-1} - 0.61c_{t-2} + \epsilon_t,$

$u_t = -0.5c_t + v_t$

$v_t = 0.9v_{t-1} + \zeta_t,$

where $\eta_t \sim iidN(0, 0.69^2), \epsilon_t \sim iidN(0, 0.62^2), \zeta_t \sim iidN(0, 1),$ *and* η_t, ϵ_t, *and* ζ_t *are mutually uncorrelated.* $\delta = 0.50$ *is the true value. "Correlation" refers to correlation between the true cycle and the estimated cycle. "Amplitude" is in terms of the standard deviation of percent deviation from trend. All estimated cycles are derived from BN decompositions of the respective models.*

Table 2: Monte Carlo simulation single source of error process

	T = 250		**T = 500,000**	
	Correlation	**Amplitude**	**Correlation**	**Amplitude**
True Cycle		1.62		1.66
AR(1)	-0.17	0.51	-0.18	0.50
VAR(4)$[\triangle y_t, c_t]$	0.98	1.68	1.00	1.65
VAR(4)$[\triangle y_t, u_t]$	0.26	1.11	0.34	0.90
ARMA(2,2)	0.19	1.16	1.00	1.63
BN filter$[\delta = \bar{\delta}]$	0.73	1.27	0.89	1.96
BN filter$[\delta = 0.50]$	0.76	0.96	0.90	1.43
AR(12)	0.82	1.02	0.89	0.97

Notes: We consider the following DGP:

$y_t = \tau_t + c_t$

$\tau_t = 0.81 + \tau_{t-1} + \eta_t,$

$c_t = 1.53c_{t-1} - 0.61c_{t-2} + 0.42\eta_t - 0.18\eta_{t-1}$

$u_t = -0.5c_t + v_t$

$v_t = 0.9v_{t-1} + \zeta_t,$

$\eta_t \sim iidN(0, 0.69^2), \zeta_t \sim iidN(0, 1),$ *and* η_t *and* ζ_t *are mutually uncorrelated. "Correlation" refers to correlation between the true cycle and the estimated cycle.* $\delta = 0.50$ *is the true value. "Amplitude" is in terms of the standard deviation of percent deviation from trend. All estimated cycles are derived from BN decompositions of the respective models.*