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

Güneş Kamber, James Morley, and Benjamin Wong

The Beveridge-Nelson decomposition based on autoregressive models produces estimates of the output gap that are strongly at odds with widely-held beliefs about transitory movements in economic activity. This is due to parameter estimates implying a high signal-to-noise ratio in terms of the variance of trend shocks as a fraction of the overall forecast error variance. When we impose a lower signal-to-noise ratio, the resulting Beveridge-Nelson filter produces a more intuitive estimate of the output gap that is large in amplitude, highly persistent, and typically increases in expansions and decreases in recessions. Notably, our approach is also reliable in the sense of being subject to smaller revisions and predicting future output growth and inflation better than other trend-cycle decompositions that impose a low signal-to-noise ratio.

JEL Classification: C18, E17, E32

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

---

*This draft: March 12, 2017. Kamber: Bank for International Settlements, gunes.kamber@bis.org Morley: University of New South Wales, james.morley@unsw.edu.au Wong: Reserve Bank of New Zealand, benjamin.wong@rbnz.govt.nz. The views expressed in this paper are those of the authors and do not necessarily represent those of the Reserve Bank of New Zealand or the Bank for International Settlements. We thank the Editor, Yuriy Gorodnichenko, four anonymous referees, seminar and conference participants at various institutions, and especially Todd Clark, George Evans, Sharon Kozicki, Adrian Pagan, Barbara Rossi, Frank Smets, Ellis Tallman, Tugrul Vehbi, John Williams, as well as our discussants, Michael Kouparitsas, Glenn Otto, and Leif Anders Thorsrud, for helpful comments, suggestions, and more. Any 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-horizon conditional expectation (minus any deterministic drift) after all forecastable transitory momentum has died out, the corresponding cycle for log real GDP should provide a sensible estimate of the output gap so long as it is based on reasonably accurate forecasts over short to medium term horizons associated with the business cycle. Noting that standard model selection criteria suggest a low-order autoregressive (AR) model predicts quarterly output growth better at these horizons than more complicated alternatives, Figure 1 plots the estimate of the U.S. output gap from the BN decomposition based on an AR(1) model.[1] What is immediately apparent about the estimated output gap is its small amplitude and lack of persistence. Its movements also do 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 an estimate of the U.S. output gap based on the Congressional Budget Office (CBO) estimate of potential output. In contrast to the estimate from the BN decomposition, the CBO output gap has much higher persistence and larger amplitude. Its movements are also strongly procyclical in terms of the NBER reference cycle. An important reason for these differences is that the estimate of the 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 for output growth, while the CBO implicitly assume a much lower signal-to-noise ratio.

Our main contribution in this paper is to show how to conduct the BN decomposition imposing a low signal-to-noise ratio on an AR model, an approach we refer to as the "BN filter". The

[1] U.S. real GDP data are from FRED for the sample period of 1947Q1-2016Q2. Output growth is measured in continuously-compounded terms. Model estimation is based on least squares regression or, equivalently, conditional maximum likelihood estimation (MLE) under normality. Throughout the paper, initial lags are backcast using the sample mean growth rate. Our choice of lag order p=1 is based on the Schwarz Information Criterion (SIC).

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, as shown below, when we apply the BN filter to U.S. log real GDP, the resulting estimate of the output gap is persistent and has large amplitude, while its movements match up well with the NBER reference cycle. At the same time, real-time estimates are subject to smaller revisions and appear to be more accurate in the sense of performing better in out-of-sample forecasts of output growth and inflation than real-time estimates for other trend-cycle decomposition methods that also impose a low signal-to-noise ratio, including deterministic detrending using a quadratic trend, the Hodrick-Prescott (HP) filter, and the bandpass (BP) filter. Thus, our proposed approach directly addresses a key critique by Orphanides and van Norden (2002) that popular methods of estimating the output gap are unreliable in real time.

The fact that the BN filter estimates are not heavily revised stems directly from our choice to work with AR models. In principle, the BN decomposition can be applied using any model that generates a long-horizon forecast, including multivariate vector autoregressive (VAR) models. However, 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. Notably, estimates of autoregressive coefficients for univariate AR models of output growth are particularly stable in comparison to parameters for more complicated models such as VAR models. Therefore, a natural outcome of our modeling choice is output gap estimates that are more reliable in the sense of being subject to small revisions. Meanwhile, the out-of-sample forecasting results are suggestive of reliability in the sense of being more accurate than other methods.

Our proposed approach is robust to the omission of multivariate information in the forecasting model and can be adapted to accommodate 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 explicitly 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) 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), Chan and Grant (forthcoming), and Dungey et al. (forthcoming) on the importance of allowing for correlation between permanent and transitory movements. Application of the BN filter to data for other countries confirms its ability to produce intuitive estimates of output gaps and suggests strong Okun's Law relationships when also considering unemployment gaps.

The rest of this paper is structured as follows. Section 2 presents our proposed approach and applies it to U.S. log real GDP, formally assessing its revision properties relative to other methods. Section 3 provides a justification for our approach, in particular why one might choose to impose a lower signal-to-noise ratio on an AR model than would be implied by sample estimates, and then presents forecast comparisons with other methods. Section 4 suggests how to account for structural breaks and applies our approach to other data series. Section 5 concludes.

2 Our Approach

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-horizon 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 at time t, τ_t^{BN}, is

$$\tau_t^{BN} = \lim_{j \to \infty} \mathbb{E}_t \left[y_{t+j} - j \cdot \mathbb{E} \left[\Delta y_t \right] \right]. \tag{1}$$

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 conditional expectation of the remaining cyclical component goes to zero at long horizons. By removing the deterministic drift, $\mathbb{E}[\Delta y_t]$, the conditional expectation in (1) remains finite and becomes an optimal (minimum mean squared error) estimate of the current trend component (see Watson, 1986; Morley et al., 2003).

To implement the BN decomposition, it is typical to specify a stationary forecasting model for the first differences $\{\Delta y_t\}$ of the time series. Modeling the first differences in this way directly 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 and partial autocorrelation functions 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:

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

where the forecast error $e_t \sim iidN(0, \sigma_e^2)$.[2] For convenience when defining 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. Then, assuming the roots of $\phi(z) = 0$ lie outside the unit circle, which corresponds to $\{\Delta y_t\}$ being stationary, the unconditional mean $\mu \equiv \mathbb{E}[\Delta y_t] = \phi(1)^{-1}c$ exists. Using the state-space approach to calculating the BN decomposition in Morley (2002), the BN

[2]A normality assumption is not strictly necessary for the BN decomposition. However, under normality, least squares regression for an AR model becomes equivalent to conditional MLE, while the Bayesian shrinkage priors used in our approach, as discussed in the next subsection, become conjugate, making posterior calculations straightforward. Also, forecast errors need not be identically distributed, so long as they form a martingale difference sequence. However, in terms of possible structural breaks in the forecast error variance, the key assumption we make is that there are no changes in the signal-to-noise ratio, as defined in this subsection below, 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.

cycle at time t, c_t^{BN}, for this model is

$$c_t^{BN} = -[1 \quad 0 \quad \cdots \quad 0] F(I - F)^{-1} X_t, \tag{3}$$

where $X_t = (\Delta\tilde{y}_t, \Delta\tilde{y}_{t-1}, ..., \Delta\tilde{y}_{t-p+1})'$, with $\Delta\tilde{y}_t \equiv \Delta y_t - \mu$ denoting the deviation from the unconditional mean, and F is the companion matrix for the AR(p) model:

$$F = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_p \\ 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 \end{bmatrix}.$$

Although an AR(1) forecasting model might seem reasonable for output growth given sample autocorrelations and partial autocorrelations and is supported by 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 transitory movements in economic activity as reflected, for example, in the CBO 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 see why the BN decomposition based on an AR(1) model produces an estimated output gap with such features, note that from (3) the BN cycle is simply $-\phi(1-\phi)^{-1}\Delta\tilde{y}_t$. Therefore, by construction, the estimated output gap will only be as persistent as output growth itself, which is not very persistent given that $\hat{\phi}$ based on 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, with the implied variance only about one quarter that of output growth itself. Furthermore, given that $-\hat{\phi}(1-\hat{\phi})^{-1} < 0$, the estimated output gap will increase in recessions when output growth becomes negative and vice versa in expansions.

More generally, to understand the BN decomposition for an AR(p) model, it is useful to define the following 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: $\delta \equiv \sigma^2_{\Delta\tau}/\sigma^2_e$. Given a Wold representation for $\{\Delta y_t\}$,

$$\delta = \psi(1)^2, \tag{4}$$

where $\psi(1) \equiv \lim_{j\to\infty} \frac{\partial y_{t+j}}{\partial e_t}$ is the "long-run multiplier" corresponding to the sum of the Wold coefficients that captures the permanent effect of a forecast error on the long-horizon conditional expectation of $\{y_t\}$ and relates to the BN trend as follows: $\Delta\tau_t^{BN} = \psi(1)e_t$.[3] 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 U.S. quarterly output 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, leading to an estimated output gap with small amplitude and counterintuitive sign. Notably, however, $\hat{\delta} > 1$ holds for all freely estimated AR(p) models given that $\hat{\phi}(1)^{-1}$ is always greater than unity regardless of lag order p. Thus, many of the surprising results for a BN decomposition based on an AR(1) model carry over to higher-order AR(p) models, although the estimated output gap no longer has to have the same persistence as output growth and will not be perfectly correlated with $\Delta\tilde{y}_t$ given that, as can be seen from (3), the BN cycle would depend on a linear combination of current and lagged values of output growth, rather than just the current value, as is the case with the AR(1) model.

2.2 Imposing a Low Signal-to-Noise Ratio

The insight that the signal-to-noise ratio δ is mechanically linked to $\phi(1)$ for an AR(p) model is important because it implies that we can impose a low signal-to-noise ratio by fixing the sum of the autoregressive coefficients when estimating an AR(p) model. To do this, we transform the AR(p) model in (2) into its Dickey-Fuller representation, which we also write in terms of

[3] An implicit equivalence between the variance of trend shocks, $\sigma^2_{\Delta\tau}$, and the variance of the changes in the BN trend follows from the equivalence of the spectral density at frequency zero for $\{\Delta y_t\}$ based on the reduced-form forecasting model and a more structural representation that separates out the true permanent and transitory shocks, but implies the same autocovariances as the reduced-form model.

deviations from the unconditional mean for convenience when implementing our approach:

$$\Delta\tilde{y}_t = \rho\Delta\tilde{y}_{t-1} + \sum_{j=1}^{p-1}\phi_j^*\Delta^2\tilde{y}_{t-j} + e_t, \tag{5}$$

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, (5) can be estimated imposing a particular signal-to-noise ratio $\bar{\delta}$ by fixing ρ as follows:

$$\bar{\rho} = 1 - 1/\sqrt{\bar{\delta}}. \tag{6}$$

The BN decomposition can then be applied imposing a particular signal-to-noise ratio $\bar{\delta}$ by first solving the restricted estimates of $\{\phi_j\}_{j=1}^p$ when inverting the Dickey-Fuller transformation given $\bar{\rho}$ and estimates of $\{\phi_j^*\}_{j=1}^{p-1}$ and then calculating the BN cycle as in (3).

Before discussing estimation of the model in (5) and how we choose $\bar{\delta}$ in practice, it is helpful to explain why we need to consider a higher-order AR(p) model in order to impose a low signal-to-noise ratio. In particular, 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 $\delta > 0.25$, 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 level or trend stationary, so $\{\Delta y_t\}$ would, in effect, be "over-differenced"). Also, as noted above, consideration of a higher-order AR(p) model allows for different persistence in the output gap and different correlation with output growth than is possible for an AR(1) model.

Given a higher-order AR(p) model, estimation of (5) imposing a particular signal-to-noise ratio is straightforward without the need of complicated nonlinear restrictions or posterior simulation.[4] Conditional MLE entails a single parametric restriction $\rho = \bar{\rho}$, which can be imposed

[4]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 far less straightforward, there would also be greater tendency to overfit the data given well-known problems of weak identification

by bringing $\bar{\rho}\Delta\tilde{y}_{t-1}$ to the left-hand-side when conducting least squares regression. However, even though it is possible to implement our approach using 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 "Minnesota" shrinkage prior on the second-difference coefficients $\{\phi_j^*\}_{j=1}^{p-1}$ as follows:

$$\phi_j^* \sim N(0, \frac{0.5}{j^2}).$$

In practice, we consider an AR(12) model, although our results are robust to consideration of higher lag orders given the shrinkage. With the Minnesota prior and conditional on σ_e^2, the posterior distribution for $\{\phi_j^*\}_{j=1}^{p-1}$ has a closed-form solution and can be easily calculated without the need for posterior simulation. For simplicity, we condition on the least squares estimate for σ_e^2 using the sample mean for μ, which is equivalent to assuming a flat/improper prior for these parameters.[5]

All that remains is to choose the particular value of $\bar{\delta}$ to impose. We see this choice as a dogmatic prior based on beliefs about large transitory movements in economic activity as reflected in, say, the CBO output gap in Figure 1. This dogmatic prior is analogous to the imposition of a low signal-to-noise ratio by fixing λ=1600 when implementing the HP filter for quarterly data (see Harvey and Jaeger (1993)).[6]

Imposing a dogmatic prior could be as simple as setting $\bar{\delta}$ to a particular low value such as,

and near-cancellation of roots, and the corresponding BN decomposition would be less reliable.

[5]It is still possible to solve the posterior distribution analytically given an informative prior on σ_e^2 by using dummy observations. However, we find different priors on σ_e^2 have very little impact on BN filter estimates, so we use the improper prior for convenience.

[6]The HP filter can also be interpreted as an approximate high-pass frequency-domain filter of a stationary process, with the choice of λ=1600 for quarterly data isolating movements at particular frequencies that are often associated with the business cycle. In the online appendix, we examine the sample periodogram for the estimated output gap to determine what frequencies are being highlighted and confirm that they correspond to those highlighted by the HP filter.

for example, $\bar{\delta} = 0.05$, which would correspond to the strict belief that only 5% of the quarterly forecast error variance for output growth is due to trend shocks. However, we recognize that any such precise choice might appear somewhat arbitrary in practice. Therefore, we propose an automatic selection of $\bar{\delta}$ based on the following implied amplitude-to-noise ratio: $\alpha(\delta) \equiv \sigma_c^2(\delta)/\sigma_e^2(\delta)$, where, noting the implicit dependence on the signal-to-noise ratio δ through ρ and $\{\phi_j^*\}_{j=1}^{p-1}$, $\sigma_c^2(\delta)$ is the variance of the corresponding BN cycle in (3) and $\sigma_e^2(\delta)$ is the variance of the forecast error for the corresponding restricted version of the AR(p) model in (5).[7] For the automatic selection, $\bar{\delta}$ is chosen to maximize the signal-to-noise ratio as follows:

$$\bar{\delta} = \inf\{\delta > 0 : \alpha'(\delta) = 0, \alpha''(\delta) < 0\}.$$

This is still a dogmatic prior in the sense that, even in large samples, $\bar{\delta}$ will generally be smaller than if it were freely estimated.[8] However, with the use of a local maximizer for values of δ close to zero, the prior is now framed in terms of the amplitude-to-noise ratio being "large" for a "low" signal-to-noise ratio rather than in terms of the signal-to-noise ratio taking on a particular low value.[9] Meanwhile, as we show below, the shape of the output gap is highly

[7]The analytical expression for the variance of the BN cycle is given in the online appendix. However, in practice, when implementing our approach to selecting $\bar{\delta}$, we simply calculate the sample variance of the estimated output gap. Also, we use the sample variance of the forecast errors from the restricted model rather than the posterior estimate of σ_e^2, which actually corresponds to the least squares estimate for the unrestricted model given our priors.

[8]We confirm this with a Monte Carlo simulation in the online appendix, with the downward bias much larger when the true signal-to-noise ratio is large than when it is small. Note that this approach of imposing a dogmatic prior to induce a downward bias on the signal-to-noise ratio is related to, but different than the suggestion in Bewley (2002) of using Bayesian estimation to offset biases in least squares estimates of autoregressive parameters.

[9]We provide a visual perspective on our proposed approach to selecting $\bar{\delta}$ in the online appendix. The visual analysis makes it clear that it is the imposition of a low δ, not the use of shrinkage priors on the second-difference coefficients, that produces a large amplitude-to-noise ratio. It also explains why we focus on a local maximizer of the amplitude-to-noise ratio.

robust to imposing other low values of δ. Therefore, a researcher with a particular dogmatic prior about the value of the signal-to-noise ratio could simply impose a low value such as $\bar{\delta} = 0.05$ and the estimated output gap would remain similarly intuitive and reliable.

2.3 The BN Filter and the Estimated Output Gap

Reflecting the similar smoothing effect on the implied trend as for the HP filter when imposing a low signal-to-noise ratio, we refer to our proposed approach as the "BN filter".[10] To summarize, it has three steps:

1. Set a low $\bar{\delta}$. We employ an automatic selection based on the local maximum of the implied amplitude-to-noise ratio for values of δ closest to zero. This is done by repeating steps 2 and 3 below for proposed incremental increases in $\bar{\delta}$ from an initial increment just above zero until the estimated amplitude-to-noise ratio $\hat{\sigma}_c^2(\bar{\delta})/\hat{\sigma}_e^2(\bar{\delta})$ decreases.

2. Given $\bar{\delta}$, estimate the Dickey-Fuller transformed AR(p) model in (5) imposing $\bar{\rho} = 1 - 1/\sqrt{\bar{\delta}}$. We conduct Bayesian estimation assuming implicit flat/improper priors for μ and σ_e^2 and a Minnesota shrinkage prior for $\{\phi_j^*\}_{j=1}^{p-1}$. We set $p = 12$, but our results are robust to higher values of p given the shrinkage.

3. Given $\bar{\rho}$ and estimates of $\{\phi_j^*\}_{j=1}^{p-1}$, solve for restricted estimates of $\{\phi_j\}_{j=1}^{p}$ by inverting the Dickey-Fuller transformation and then apply the BN decomposition as in (3).

Figure 2 plots the estimated U.S. output gap along with 95% confidence bands for the BN filter, with $\bar{\delta} = 0.24$ determined by automatic selection based on maximizing the amplitude-to-noise ratio. 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

In particular, because of non-monotonicities in the relationships of δ with amplitude and fit, a global maximum does not exist, while the local maximizer for values of δ close to zero imposes a low signal-to-noise ratio by construction.

[10]We note that a few previous studies have referred to a "Beveridge-Nelson filter", but always as an alternative terminology for the traditional BN decomposition, which does not impose a smoothing effect on the trend.

estimated output gap that is consistent with widely-held beliefs about transitory movements in economic activity. In particular, the estimated output gap is large in amplitude, persistent, and moves procyclically with the NBER reference cycle. In contrast to the small and negative correlation of -0.30 for the BN decomposition based on an AR(1) model and the CBO output gap that was evident in Figure 1, the correlation of the BN filter output gap with the CBO output gap is reasonably high at 0.75. Meanwhile, as detailed in the online appendix, the 95% confidence bands are based on inverting a simple z-test that the true output gap, c_t, is equal to a hypothesized value based on the unbiasedness, variance, and assumed normality of the BN cycle. According to these confidence bands, the estimates are reasonably precise and significantly different from zero at many points of time, especially during recessions.[11]

Before examining the reliability of our approach, we consider the sensitivity of the estimated output gap to varying the signal-to-noise ratio. The top panel of Figure 3 plots the estimated output gap for $\delta \in \{0.05, 0.8\}$ compared to our approach where $\delta = 0.24$. The shape of the estimated output gap is little changed, with the persistence profile virtually unaltered. Indeed, the correlations between the different estimated output gaps varying the signal-to-noise ratio is always greater than 0.95. Meanwhile, because the estimated output gap is so similar as we change the signal-to-noise ratio, the revision properties and real time forecast performance will also be highly robust to $\delta \in \{0.05, 0.8\}$, as we show in the online appendix.

Given this apparent robustness to different values of δ, we check whether our results are actually 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 panel of Figure 3 plots the two output gap estimates and makes it clear that imposing a low signal-

[11] All of the methods considered here effectively impose a mean of zero on the estimated output gap by construction. However, the significantly negative estimates from the BN filter during recessions are consistent with the findings in Morley and Piger (2012) and Morley and Panovska (2017) using weighted averages of model-based trend-cycle decompositions for linear and nonlinear time series models that the output gap is asymmetric in the sense of being generally close to zero during expansions, but large and negative during recessions.

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. Also, as shown below, the revision properties and forecast performance are not nearly as good when using the BN decomposition based on a freely estimated AR(12) model as they are for the BN filter.

2.4 Revision Properties of the BN Filter and Other Methods

Having proposed a 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 in the introduction, 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 unreliable in real time. Because Orphanides and van Norden (2002) show that most real-time revisions of output gap estimates are due to the filtering method rather than data revisions for real GDP, we consider "pseudo-real-time" analysis using the final vintage of data (from 2016Q2) in order to focus on the revision properties of the filtering method. However, as shown in the online appendix, our results are generally robust to consideration of data revisions too.[12]

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

[12]The only major change in a fully real-time environment with different vintages of data is that the BN filter clearly outperforms the traditional BN decomposition based on an AR(1) model, which is the one other method that does comparatively well in the pseudo-real-time environment. We speculate that the reason for the improved relative performance of the BN filter when also allowing for data revisions is that its BN cycle is a weighted average of 12 quarters of output growth, while the BN cycle based on an AR(1) model only reflects the current quarter. To the extent that most data revisions apply most heavily to the current quarter or even recent quarters, the estimated output gap for the BN decomposition based on an AR(1) model will be more heavily revised with each data revision. Meanwhile, the size of revisions are, of course, somewhat larger across the board given data revisions. However, the size of revisions for the BN filter remain well below even the best results in the fully real-time environment in Orphanides and van Norden (2002).

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. quarterly output 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 effects 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 popular 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, forthcoming), the zero restriction on the correlation between movements in trend and cycle can be rejected by statistical tests, 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. However, 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 filtering. In particular, we consider a deterministic quadratic trend, the HP filter by Hodrick and Prescott (1997), and the 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 with periods 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. We report results for standard implementations, although we note that the results are virtually unchanged when padding the data with AR(4) forecasts to try to address endpoint problems, as done in Edge and Rudd (2016), which directly reflects the difficulty of forecasting future output growth.

Figure 4 plots the pseudo-real-time and the ex post (i.e., full sample) estimates of the output gap from the BN filter and the various other methods.[13] The first result to notice is that all of the features of the output gap for the BN filter in Figure 2 carry over to the pseudo-real-time environment.[14] Meanwhile, in contrast to the BN filter, both the AR(1) and ARMA(2,2) models produce output gap estimates that have little persistence, are small in amplitude, and do not move procyclically with 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 moves procyclically with 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 and larger in amplitude, consistent with the point made by Evans and Reichlin (1994) that adding relevant information for forecasting of output growth mechanically lowers the signal-to-noise ratio. Finally, the estimates for the other popular methods of trend-cycle decomposition are all reasonably intuitive in the sense of being persistent, large in amplitude, and generally moving procyclically in terms of the NBER reference cycle. However, it is notable how different they are from each other and from the estimates based on the BN filter and the BN decompositions for the AR(12) model and the

[13] 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 2016Q2. Again, all raw data are taken from FRED. As with output growth, the unemployment rate is backcast using its pseudo-real-time sample average to allow the estimation sample to always begin in 1947Q2.

[14] We re-calculate $\bar{\delta}$ for each pseudo-real-time sample. Encouragingly, the values that maximize the amplitude-to-noise ratio are quite stable, fluctuating between 0.21-0.26.

VAR model. Thus, being "intuitive" is clearly not a sufficient condition for choosing amongst competing methods. Hence, we also consider reliability, which is why, like Orphanides and van Norden (2002), we look at revision properties of the estimates, although we will also examine other aspects of reliability in the next section.

In terms of the revision properties of the output gap estimates, the main finding in Figures 4 is that, regardless of the features in terms of persistence, amplitude, and sign, all of the estimates for various BN decompositions, including the BN filter, are subject to relatively small revisions in comparison to the other methods. The key reason why output gap estimates for the various BN decompositions are hardly revised is because the estimated parameters of the forecasting models turn out to be relatively stable when additional observations are considered 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 somewhat larger 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.[15] 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

[15]Notably, as shown in the online appendix, the pseudo-real-time estimates for the AR(12) and VAR(4) models lie outside the ex post 95% confidence bands much more than 5% of the time, raising serious doubts about their reliability in terms of accuracy. At the same time, even the ex post estimates are not particularly precise, especially with the BN cycle for the VAR(4) model, which is significantly different than zero less than 5% of the time. The BN cycle for the ARMA(2,2) model is more accurate in the sense that the pseudo-real-time estimates lie within the ex post 95% confidence bands 100% of the time, but the ex post estimates are almost never statistically different than zero. Meanwhile, the pseudo-real-time estimates for the BN filter and the AR(1) model appear accurate, with the estimates within the ex post 95% confidence bands 100% of the time and the ex post estimates are relatively precise with the cycle being significantly different than zero about 40% and 30% of the time, respectively.

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 potential overfitting and allowing for richer dynamics. Finally, all of the estimates for the other popular methods in Figure 4 are heavily revised and clearly highly unreliable in real time.[16] 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 relatively 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), Edge and Rudd (2016), and Champagne et al. (2016). First, to quantify the size of the revisions, we consider two measures, the standard deviation and the root mean square (RMS), with the RMS measure designed to penalize a bias in revisions more heavily than the standard deviation measure. 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 given that the different methods produce estimates with very different amplitudes.[17] 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

[16]As shown in the online appendix, the pseudo-real-time estimate for the Harvey-Clark UC model is inaccurate in the sense that it lies outside the 95% confidence bands about 25% of the time, similar to the BN decomposition for the AR(12) model. Meanwhile, the cycle is only significantly different than zero about 15% of the time, so it is relatively imprecise. For deterministic detrending, the pseudo-real-time estimate always lies within what are very wide 95% confidence bands, with the cycle is significantly different than zero less than 5% of the time. Unlike with the model-based methods, we do not consider confidence bands for the HP and BP filters.

[17]Note that 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 confusion.

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 more recent estimates near the end of the full sample in 2016Q2 may end up becoming more heavily revised in the future.

Figure 5 presents the revision statistics. As was visually apparent in Figure 4, the BN filter does quite well in terms of size of revisions, with both the standard deviation and RMS statistics being less than one quarter of a standard deviation of the ex post estimate of the output gap. The BN decompositions based on freely estimated AR(1) and ARMA(2,2) models also do well in terms of size of revisions, with the standard deviation and RMS statistics slightly better than the BN filter for the AR(1) model and somewhat worse for the ARMA(2,2) model. By contrast, the BN decompositions based on the highly parameterized AR(12) and VAR(4) models do not do as well, with the standard deviation and RMS statistics about one standard deviation or more of the ex post estimates. Meanwhile, all of the other popular methods produce revisions that are well over half of one standard deviation of the ex post estimates, implying very large revisions in absolute terms given the relatively large amplitude of their output gap estimates. In terms of correlation of pseudo-real-time estimates with the ex post estimates, the BN decompositions tend to do well, although the correlation is somewhat lower for the more highly parameterized AR(12) and VAR(4) models. Notably, the BN filter and BN decomposition based on an AR(1) model have near perfect correlation between the pseudo-real-time and ex post estimates. The correlation for other popular methods is generally quite low, with the HP filter performing the worst. Finally 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 well, with the pseudo-real-time estimates for the BN filter and the BN decomposition based on an AR(1) model performing best and correctly identifying the same sign as the final estimate about 90% of the time.

To summarize, the BN decomposition appears more reliable in a pseudo-real-time environment than other popular methods. This is because the consideration of future observations does not drastically alter the estimates of the forecasting model parameters. Even so, among the different 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 larger changes in parameter estimates. 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 what appears to be a highly reliable pseudo-real-time 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 revision properties, perhaps begging the question of why we impose a low signal-to-noise ratio. We discuss this issue next.

3 Is Our Approach Reasonable?

3.1 Justification for Imposing a Low 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 using the BN decomposition based on an AR(1) model, especially if standard model selection and revision properties are the main criteria for choosing an approach to estimating the output gap. We 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. Yet whether it actually does so remains unresolved in the empirical literature. Indeed, 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), although this view has not gone unchallenged (e.g., Cochrane, 1994; Perron and Wada, 2009).

One possible reason to believe the signal-to-noise ratio is lower than that given by a freely estimated AR model is that $\{\Delta y_t\}$ may behave much 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 heavily 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 trend plus an AR(2) cycle. Also, in both cases, the first difference of the time series follows the identical ARMA(2,2) process with a near unit MA root and a signal-to-noise ratio of δ=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 unobserved components process 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.[18] When considering model selection for possible ARMA specifications for $\{\Delta y_t\}$ given this DGP in finite samples, SIC picks 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 for the BN decomposition would do in estimating the true cycle $\{c_t\}$.

The top half of Table 1 reports the results for this first DGP in a finite sample (T=250) and in population (T=500,000). The immediate result to notice is 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, we find that SIC chooses a lag

[18]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.

order for an AR(p) model of p=1 more than 95% of the time.

The next result to notice is 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 particularly surprising given the inclusion of the true cycle in the forecasting model, corresponding to the highly unrealistic scenario in which there exists an observed variable that perfectly captures economic slack. Of course, 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. Instead, the 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 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.[19]

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 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 Morley (2011), where the BN decomposition based on the true model provided an unbiased estimator of the standard deviation of trend shocks for a UC process, but the estimate of the standard deviation of the cycle was downward biased. Indeed, as long as the true cycle is unobserved, there will generally be a bias in estimating its standard deviation using the BN cycle.[20] Meanwhile,

[19]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 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 $\{\Delta y_t\}$.

[20]The BN decomposition based on a VAR(4) model of $\{\Delta y_t\}$ and the true cyclical compo-

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 in 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 the BN decomposition based on an AR(12) model and a version of the BN decomposition that imposes the true signal-to-noise ratio (which, of course, is never known in practice). 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. Meanwhile, 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 model 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

nent $\{c_t\}$ does not suffer from a downward bias because the cyclical component is observed and the true DGP has a (restricted) VAR(2) representation. Also, it is important to emphasize that a bias in the estimate of the standard deviation of the cycle is not the same as a bias in the estimate of the cycle. In particular, the BN cycle provides an unbiased estimator of the true cycle given the correct model specification. It is just that a property of the estimated cycle – in this case its standard deviation – is different than the property of the true cycle. This is somewhat analogous to least squares residuals providing unbiased estimates of the true regression errors given the correct model, but the sample variance of the least squares residuals providing a downwardly biased estimate of the variance of the true errors. In general, it is always possible for an optimal estimate to have different properties than the object being estimated. An obvious and relevant example is filtered and smoothed inferences from the Kalman filter and smoother, which are both optimal subject to different information sets, but which inevitably have different properties in terms of their variability, as alluded to by their labels.

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 for the first DGP. See Morley (2011) for a full discussion of this conceptual distinction.

The bottom half of Table 1 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 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 to get close to 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 close 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 also a difficult econometric problem in practice, with finite-sample power issues and, at the same time, considerable danger of overfitting unless variable selection is handled carefully.[21] Meanwhile, even given the

[21]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

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 is less reliable than our proposed approach.

The bottom line is that it is quite 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 suffer more from reliability issues. By contrast, the BN filter works relatively well even in finite samples and is more reliable in terms of its revision properties.

Next, we consider whether the BN filter is also reliable in the sense of minimizing a spurious cycle that is unrelated to future output growth or other macroeconomic variables. 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 also worry that imposing a low signal-to-noise ratio might produce a spurious cycle if the true state of the world is more along the lines of an AR(1) model than the two DGPs considered above. In particular, if the BN filter produces a large 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

represented the true state of the world.

and inflation out of sample. We check whether this is the case in practice.

3.2 Out-of-Sample Forecast Comparisons

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. Given the first estimate of an output gap in 1947Q2 and the forecast evaluation sample starting in 1970Q1, we use an expanding window estimation of a forecasting relationship based on the available sample for each 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 output is currently below trend, this should imply faster output growth at some point in the future as output adjusts back up to trend. Conversely, if output is currently above trend, one should forecast slower output growth at some horizon for output to revert back down to trend. In particular, because the true cycle will cross its unconditional mean of zero at some point, a good estimate of the output gap should be able to forecast the effects of this reversion to mean. Thus, for an h-period-ahead output growth forecast, we consider a forecasting equation similar to Nelson (2008):

$$y_{t+h} - y_t = \alpha + \beta \hat{c}_t + \epsilon_{t+h,t} \tag{7}$$

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. For an accurate estimate of the output gap, we expect $\beta < 0$ at some horizon h and the inclusion of the estimated output gap in the forecast equation to help predict h-period-ahead output growth.

Figure 6 presents the out-of-sample forecasting results, with relative root mean squared errors (RRMSEs) in comparison to forecasts using the BN filter estimate of the output gap. The first result to notice is that the output gap estimates constructed using the BN decomposition do better at all horizons than those based on other methods, including the HP and BP filters. This further vindicates our choice to work with a BN decomposition. Meanwhile, among the different BN decompositions, similar to with the revision statistics, parsimony or shrinkage priors seem to be key to good performance. 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 statistics.

Notably, despite imposing a low signal-to-noise ratio, the BN filter appears to avoid producing much of a spurious cycle that would diminish the forecasting performance of its 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 this could be due to momentum in output growth initially following a transitory shock. In particular, the BN cycle for an AR(1) model is proportional to output growth, so its strong forecast performance could just be capturing positive serial correlation at short horizons rather than an ability to capture reversion to trend. The RRMSE is close to one and not significant at longer horizons, suggesting that an AR model does no better than the BN filter 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 in the sense of having less of a link to future output growth at longer horizons than a basic AR(1) model.[22]

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, 2009) 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:

[22] Also, supportive of the greater accuracy of the BN filter compared to the other more heavily-revised methods in particular, we show in the online appendix that the revised estimate from the BN filter is much more positively correlated with the Chicago Fed's National Activity Index based on 85 data series than are any of the other revised estimates. We also show that the pseudo-real-time estimate from the BN filter is more (typically negatively) correlated with future revisions in other estimates than other pseudo-real-time estimates are correlated with future revisions in the estimate from the BN filter.

$$\pi_{t+h} - \pi_t = \gamma + \sum_{i=0}^{p} \theta_i \triangle \pi_{t-i} + \sum_{i=0}^{q} \kappa_i \hat{c}_{t-i} + \epsilon_{t+h,t}. \quad (8)$$

where π is U.S. CPI inflation.[23] We choose the lag orders of the forecasting equation, namely p and q above, using SIC. As is commonly done (see, for example, Stock and Watson, 1999, 2009; Clark and McCracken, 2006), we apply the information criterion 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 7 presents the out-of-sample forecasting results for the U.S inflation. 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 outperforms all other BN decompositions, although generally not significantly so. The BN filter also generally does 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 an extent, these results are not entirely surprising. Atkeson and Ohanian (2001) and Stock and Watson (2009) 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 other methods in terms of providing a good real-time measure of economic slack as it pertains to inflation. 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

[23] 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 2016Q2 by simple averaging.

of the alternative methods outperform our proposed approach in a statistically significant way.[24]

4 Extensions

4.1 Accounting for Structural Breaks

The traditional BN decomposition assumes that the trend component of $\{y_t\}$ follows a random walk with constant drift. One potential concern then is, if there has been a sufficiently large change in the long-run growth rate, the assumption of constant drift will lead to biased estimates of the output gap. For example, if there is a large reduction in the long-run growth rate, a forecasting model that fails to account for it will keep anticipating faster growth than actually occurs after the break, leading to a persistently negative estimate of the output gap based on the BN decomposition. More generally, Perron and Wada (2009, 2016) argue that estimates of the output gap from different methods can be highly sensitive to accounting for structural breaks. Therefore, we consider the effect of structural breaks on our estimates and propose a simple way to address their possible presence.

When we test for breaks in the long-run growth rate for U.S. real GDP using Bai and Perron (2003) procedures, we find one break in 2006Q1.[25] We therefore adjust the data for the break in the long-run growth rate in 2006Q1 and apply the BN filter to the adjusted data. If there were evidence of a structural break in the persistence of U.S. output growth, we would have considered a break in the imposed signal-to-noise ratio given the link between δ and $\phi(1)$ for an

[24]Furthermore, in our analysis in the online appendix of whether large revisions are useful for understanding the past, we show that revised estimates are not capturing anything about the true output gap in terms of a relationship with inflation that was not already captured in pseudo-real-time estimates. In particular, the inflation forecast comparisons are almost identical using revised estimates instead of pseudo-real-time estimates.

[25]We use 15% trimming of the sample between potential breaks. One concern is that a possible break has occurred since the Great Recession, a period which falls within the 15% trimming window at the end of the sample. However, when we reduce the trimming window to 5% or 10%, we still find evidence of only one break in 2006Q1.

AR(p) model. However, we find no evidence for a break in persistence and assume a constant $\bar{\delta}$ for the whole sample.

The estimated output gap allowing for a break in long-run growth in 2006Q1 is shown in the top panel of Figure 8. Because we capture an apparent slowdown in long-run growth, the estimated output gap turns positive around 2010 and remains slightly positive from then on. However, a notable finding is that the estimated output gap prior to the break is virtually identical to the benchmark results when not allowing for a break.

There are two practical issue to deal with when considering structural breaks. First, it would clearly be difficult to detect breaks in real time. Our empirical example suggests that one can date a possible break in 2006Q1, but this is only with hindsight given an addition of 10 extra years of data.[26] Therefore, allowing for breaks as done above with a Bai and Perron (2003) test is ultimately an ex post exercise that requires a long span of data. Second, even though we can allow for breaks ex post, there can remain a concern that break date estimates are not particularly robust. For example, when we use data up to 2016Q2, the break date is estimated at 2006Q1, but using data up to 2016Q1, the break date is estimated at 2000Q2.

To help address these practical concerns, we propose a way to guard against possible structural breaks in the long-run growth rate in real time, while still being robust to different possible break dates. In particular, instead of testing for breaks using Bai and Perron (2003) procedures and adjusting the data to their regime specific mean, we propose dynamically demeaning the data using a backward-looking rolling 40-quarter average growth rate.[27] That is, we construct deviations from mean as follows:

$$\Delta\tilde{y}_t = \Delta y_t - \frac{1}{40}\sum_{i=0}^{39}\Delta y_{t-i}.$$

[26] Andrews (2003) proposes a generalized test for a structural break that is applicable at the end of a sample. However, not surprisingly, the test has limited power unless the magnitude of the break in mean is very large relative to the error variance.

[27] One potential issue is what to choose as the appropriate window for estimating a changing long-run growth rate. We use 40 quarters in order to smooth over the effects of most business cycle fluctuations on average growth.

We then apply the BN filter using $X_t = (\Delta\tilde{y}_t, \Delta\tilde{y}_{t-1}, ..., \Delta\tilde{y}_{t-p+1})'$ in (3) for the dynamically demeaned data. This procedure loses some precision in the estimate of the mean compared to knowing the exact break date. However, it is robust to multiple or gradual breaks and it can be easily applied in real time.[28]

In the middle panel of Figure 8, we plot the estimated output gap given dynamic mean adjustment and compare it to the results when allowing for a break in 2006Q1. As can be seen, the BN filter with dynamic mean adjustment does quite well at approximating the BN filter when allowing for a break detected by Bai and Perron (2003) procedures.[29] Meanwhile, in the bottom panel of Figure 8, we compare pseudo-real-time and ex post estimates of the output gap using dynamic mean adjustment. Encouragingly, the pseudo-real-time estimates appear reasonably reliable in terms of their revision properties.

4.2 Application to Other Data Series

We apply the BN filter to log real GDP and unemployment rate data for the G7, Australia, and New Zealand in order to estimate both output and unemployment gaps.[30] Figure 9 plots the estimated gaps, noting the signal-to-noise ratio that maximizes the amplitude-to-noise ratio for the output gap and an implied Okun's Law coefficient in terms of the percentage change in the output gap per percentage point change in the unemployment gap. All of output growth rates, except for the U.S. data, are adjusted for breaks in the long-run growth rate found by Bai and Perron (2003) procedures. The negative of the unemployment gap is plotted to enhance the

[28]For the first 40 quarters of the sample, we use the average growth rate for that 10 year period. However, after that, the estimate of the mean is completely backward looking, making it suitable for our pseudo-real-time analysis given our evaluation samples that start in 1970Q1.

[29]Furthermore, in the online appendix, we show that the BN filter with dynamic mean adjustment is still comparatively reliable in terms of its revision properties and out-of-sample forecast performance. We also show that dynamic mean adjustment also does quite well at approximating results when allowing for breaks detected from Bai and Perron (2003) procedures for the non-U.S. G7 countries, Australia, and New Zealand.

[30]The data are sourced from the IMF Outlook and we apply the BN filter to each series separately.

visual comparison with the output gap.

For the U.S. data, the estimated unemployment gap is highly (negatively) correlated with estimated output gap. When we regress the output gap against the unemployment gap, we obtain a coefficient of -1.4, which is slight lower than the consensus estimate of Okun's Law, but in agreement with a comparable analysis by Sinclair (2009), who estimates output and unemployment gaps with a multivariate UC model. For the other countries, the δ that maximizes the amplitude-to-noise ratio is comparable to $\bar{\delta} = 0.24$ for the U.S. data, but generally a bit smaller, ranging from as low as 0.08 for New Zealand to 0.21 for Canada. As with the U.S. data, there is a clear negative relationship between the estimated output and unemployment gaps for the other countries. The Okun's Law coefficient ranges from -0.3 for Germany to -2.5 for Italy. In some cases, such as for Australia, Canada, and the United Kingdom, the negative correlation between the estimated output and unemployment gaps is visually quite clear, while in other cases, such as for Italy and Japan, it is less so, perhaps reflecting the varying degrees of labour market rigidities across countries.

Overall, the main conclusion from these results for other data series is that the BN filter is able to produce intuitive estimates of output and unemployment gaps in the sense of concordance of movements in the gaps over the business cycle, not just for the United States, but for other countries as well.

5 Conclusion

We have proposed the application of a restricted BN decomposition that imposes a low signal-to-noise ratio. In particular, rather than focusing solely on model fit by freely estimating a time series forecasting model, we develop a "BN filter" that trades off amplitude and model fit by maximizing the amplitude-to-noise ratio in order to determine a low signal-to-noise ratio that we impose in 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 at the same time. Notably, the BN filter retains the relative reliability of the traditional

BN decomposition based on freely estimated AR models, but the estimated output gap is much more intuitive in the sense of being large in amplitude, persistent, and moving closely with the NBER reference cycle. Other methods that produce similarly intuitive estimates of the output gap are far less reliable in terms of their revision properties.

We consider 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 will tend to produce highly 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 also appears reliable in the sense of not generating a large spurious cycle when applied to U.S. log real GDP. Specifically, the estimated output gap from the BN filter forecasts U.S. output growth and inflation essentially as well as the 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.

Finally, we show how to account for potential structural breaks in long-run growth rates when implementing our approach and we apply our approach to real GDP and unemployment rate data for the United States and other countries, producing intuitive estimates of output gaps that have strong Okun's Law relationships with estimated unemployment gaps.

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.

Andrews DW. 2003. End-of-sample instability tests. *Econometrica* **71**: 1661–1694.

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

Bai J, Perron P. 2003. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics* **18**: 1–22.

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.

Bewley R. 2002. Forecasting accuracy, coefficient bias and Bayesian vector autoregressions. *Mathematics and Computers in Simulation* **59**: 163–169.

Champagne J, Poulin-Bellisle G, Sekkel R. 2016. The real-time properties of the Bank of Canada's staff output gap estimates. Bank of Canada Staff Working Papers 16-28.

Chan JC, Grant AL. forthcoming. A Bayesian model comparison for trend-cycle decompositions of output. *Journal of Money, Credit and Banking* .

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, Jacobs J, Tian J. forthcoming. Forecasting output gaps in the G-7 countries: The role of correlated innovations and structural breaks. *Applied Economics* .

Edge RM, Rudd JB. 2016. Real-time properties of the Federal Reserve's output gap. *Review of Economics and Statistics* **98**: 785–791.

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, Jaeger A. 1993. Detrending, stylized facts and the business cycle. *Journal of Applied Econometrics* **8**: 231–247.

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 J, Panovska I. 2017. Is business cycle asymmetry intrinsic in industrialized economies? UNSW Business School Research Paper 2016-12a, UNSW.

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

Morley JC. 2002. A state-space approach to calculating the Beveridge-Nelson decomposition. *Economics Letters* **75**: 123–127.

Morley JC. 2011. The two interpretations of the Beveridge-Nelson 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 macroeconomic 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. *Cowles Foundation Discussion Paper No. 2005* .

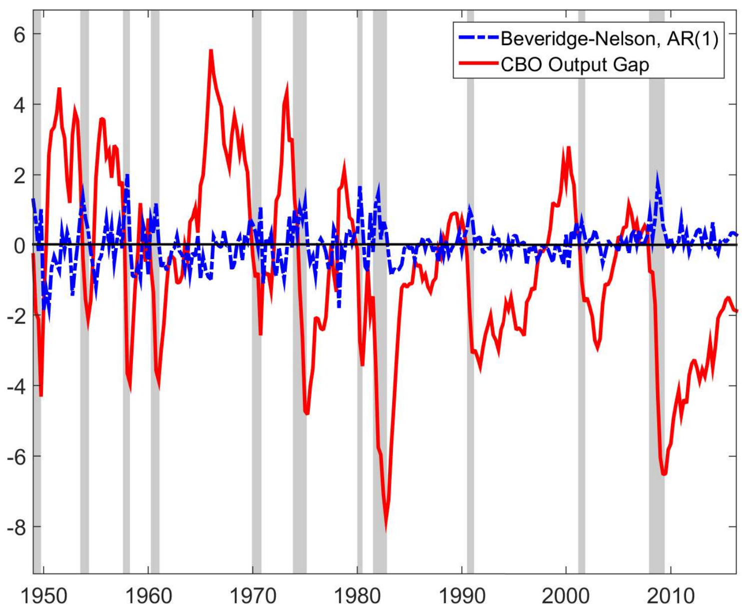
Sinclair TM. 2009. The Relationships between Permanent and Transitory Movements in U.S. Output and the Unemployment Rate. *Journal of Money, Credit and Banking* **41**: 529–542.

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

Stock JH, Watson MW. 2009. Phillips curve inflation forecasts. In Fuhrer J, Kodrzycki Y, Little J, Olivei G (eds.) *Understanding Inflation and the Implications for Monetary Policy, a Phillips Curve Retrospective*. Cambridge: MIT Press, 99–186.

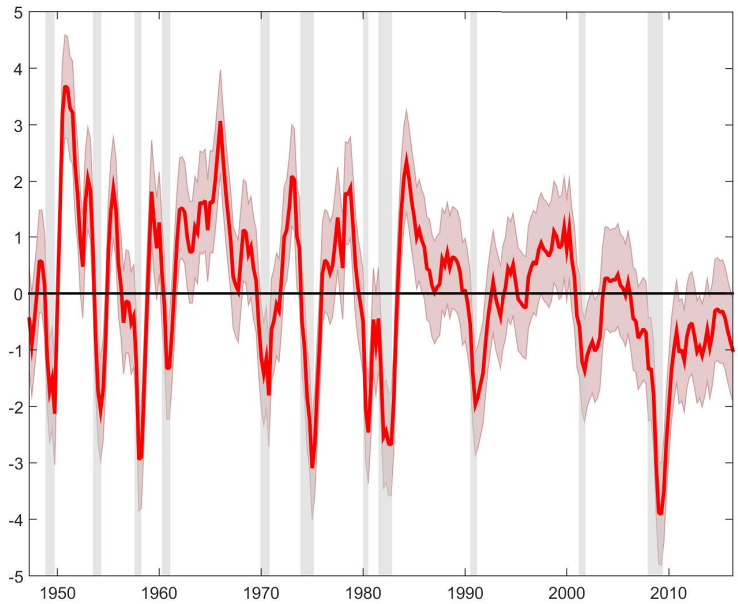
Watson MW. 1986. Univariate detrending methods with stochastic trends. *Journal of Monetary Economics* **18**: 49–75.

Figure 1: Two contrasting estimates of the U.S. output gap



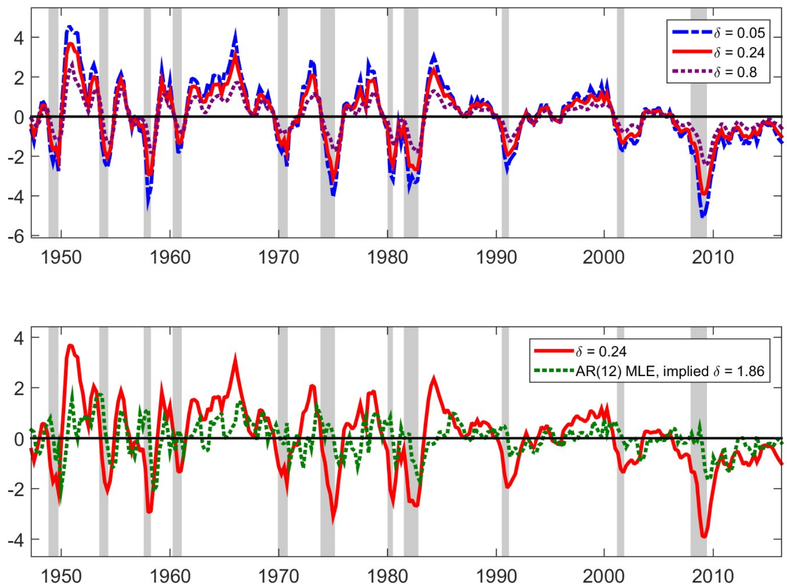
Notes: Units are 100 times natural log deviation from trend. The Beveridge-Nelson decomposition estimate of the output gap is based on an AR(1) model of U.S. quarterly output growth estimated via MLE. The CBO output gap is derived from the natural log of real GDP minus the natural log of the CBO's estimate of potential output. Shaded bars correspond to NBER recession dates.

Figure 2: Estimated U.S. output gap from the BN filter with 95% confidence bands



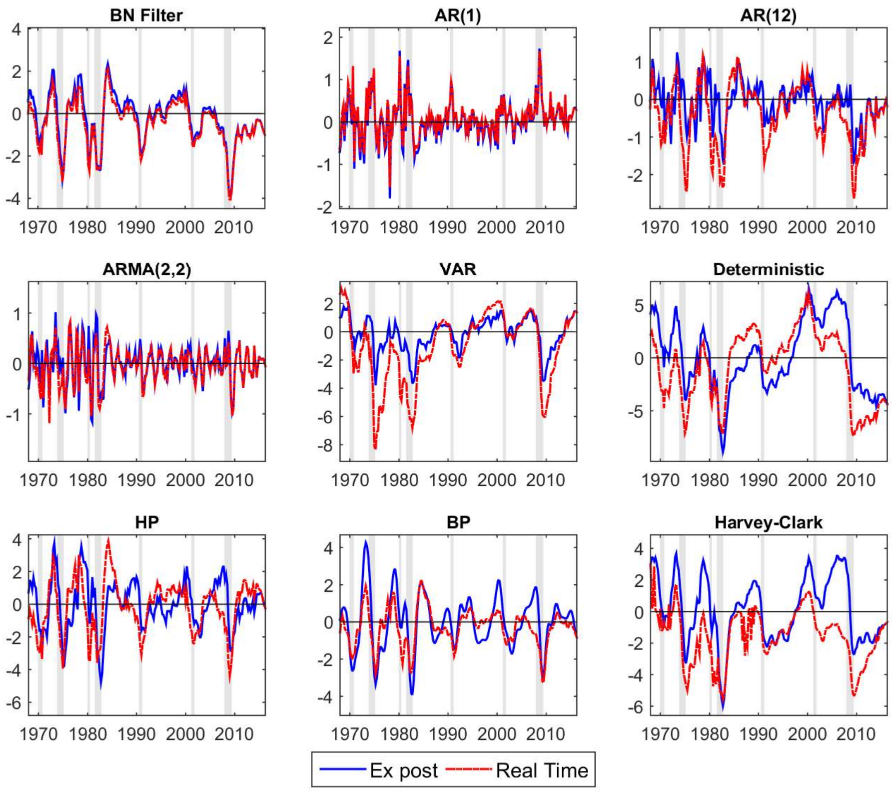
Notes: Units are 100 times natural log deviation from trend. "BN filter" refers to our proposed approach of estimating the output gap using the BN decomposition based on Bayesian estimation of an AR(12) model of U.S. quarterly output growth with shrinkage priors on second-difference coefficients and imposing the signal-to-noise ratio that maximizes the amplitude-to-noise ratio. The solid line is the estimate. Shaded bands around the estimate correspond to a 95% confidence interval from inverting a z-test that the true output gap is equal to a hypothesized value using the standard deviation of the BN cycle. Shaded bars correspond to NBER recession dates.

Figure 3: U.S. output gap estimates based on the BN decomposition when varying the signal-to-noise ratio



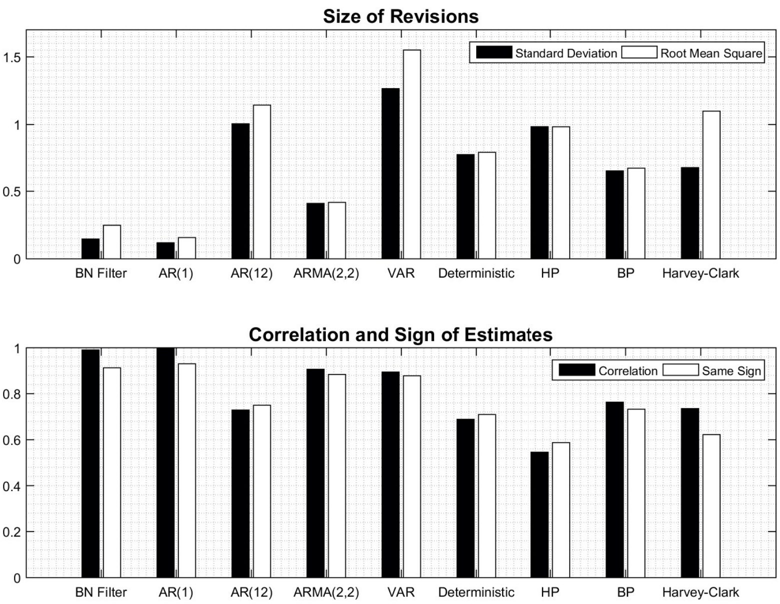
Notes: Units are 100 times natural log deviation from trend. The different lines in the top panel are for a BN decomposition based on Bayesian estimation of an AR(12) model of U.S. quarterly output growth with shrinkage priors on second-difference coefficients imposing different signal-to-noise ratios. In the bottom panel, "AR(12) MLE" refers to the BN decomposition based on an AR(12) model estimated via MLE. Shaded bars correspond to NBER recession dates.

Figure 4: Pseudo-real-time and ex post U.S. output gap estimates for various methods



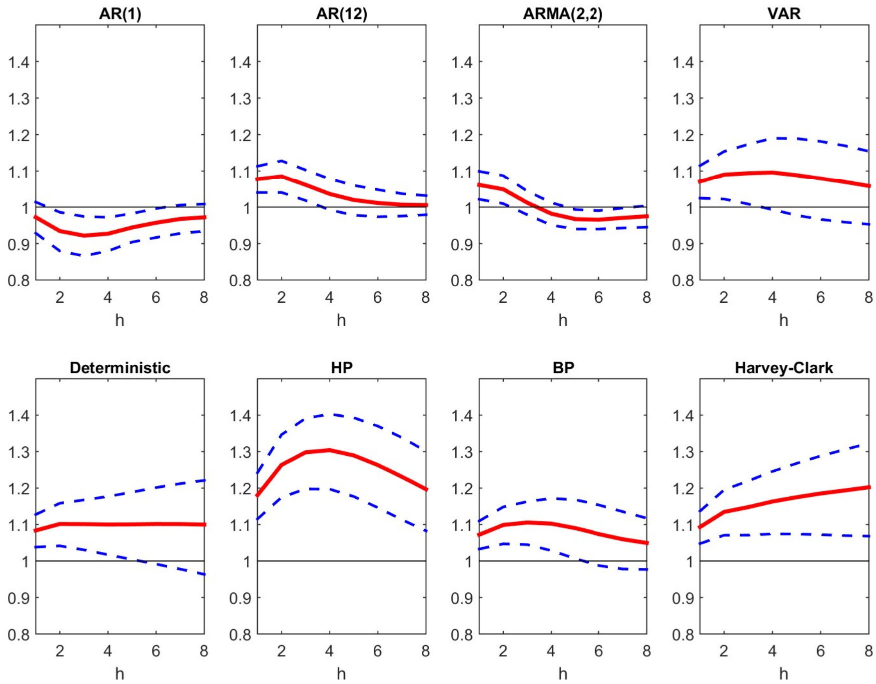
Notes: Units are 100 times natural log deviation from trend. "BN filter" refers to our proposed approach of estimating the output gap using the BN decomposition based on Bayesian estimation of an AR(12) model of U.S. quarterly output growth with shrinkage priors on second-difference coefficients and imposing a signal-to-noise ratio that maximizes the amplitude-to-noise ratio. "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). Shaded bars correspond to NBER recession dates.

Figure 5: Revision statistics for U.S. output gap estimates



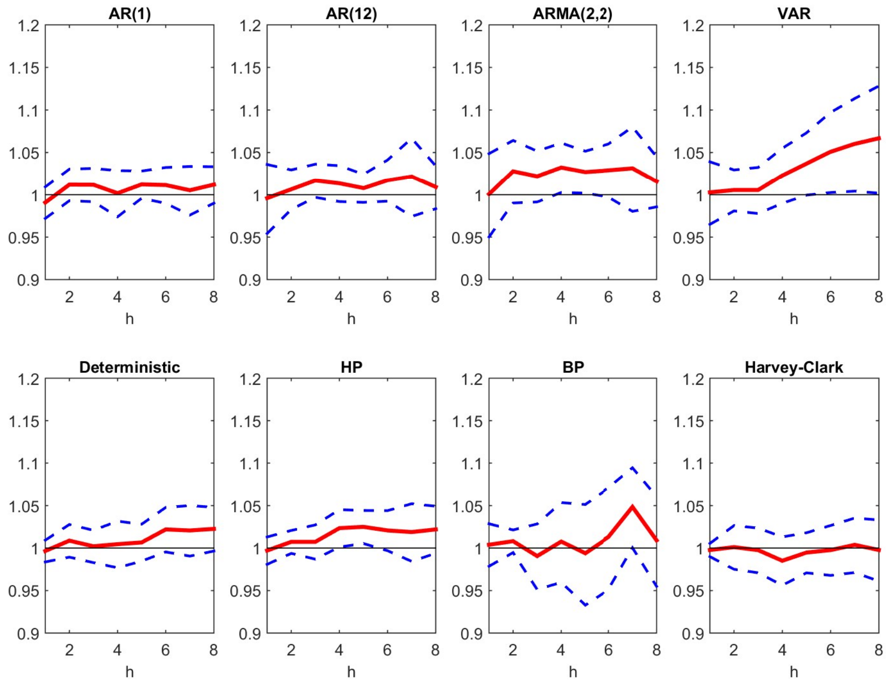
Notes: See notes for Figure 4 for descriptions of labels of methods. Standard deviation and root mean square of revisions to the pseudo-real-time estimate of the output gap are normalized by the standard deviation of the ex post estimate of the output gap. "Correlation" refers to the correlation between the pseudo-real-time estimate and the ex post estimate of the output gap. "Same sign" refers to the proportion of pseudo-real-time estimates that share the same sign as the ex post estimate of the output gap. The sample period for calculation of revision statistics is 1970Q1-2012Q4.

Figure 6: Out-of-sample U.S. output growth forecast comparison relative to the BN filter benchmark using pseudo-real-time output gap estimates



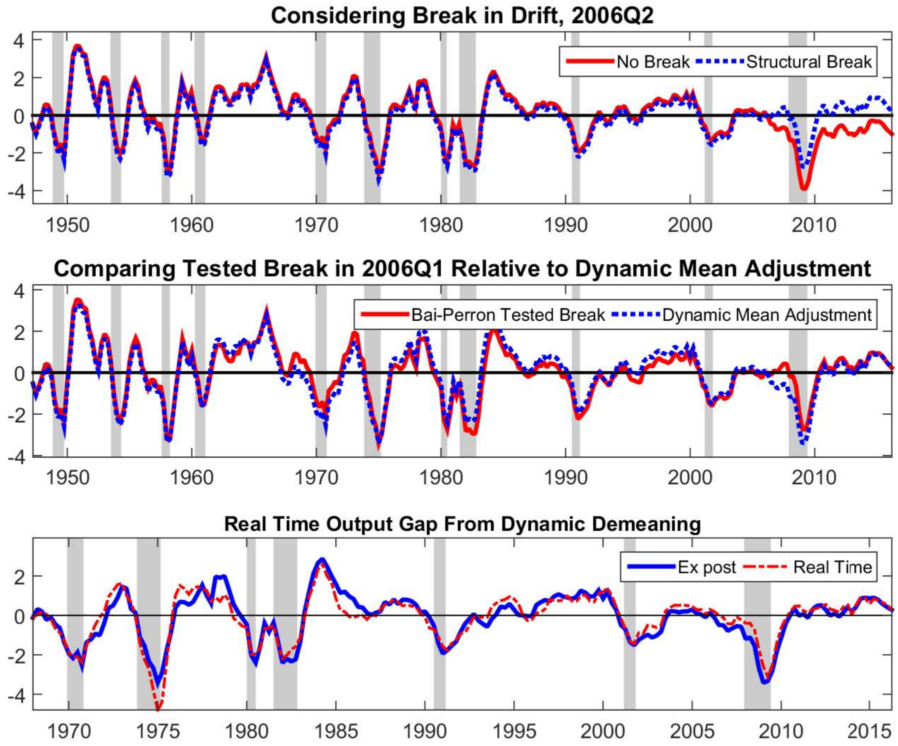
Notes: See notes for Figure 4 for descriptions of labels of methods. The graphs plot out-of-sample RRMSEs compared to forecasts based on the BN filter estimated output gap. Out-of-sample evaluation begins in 1970Q1. The bands depict 90% confidence intervals from a two-sided Diebold and Mariano (1995)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

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



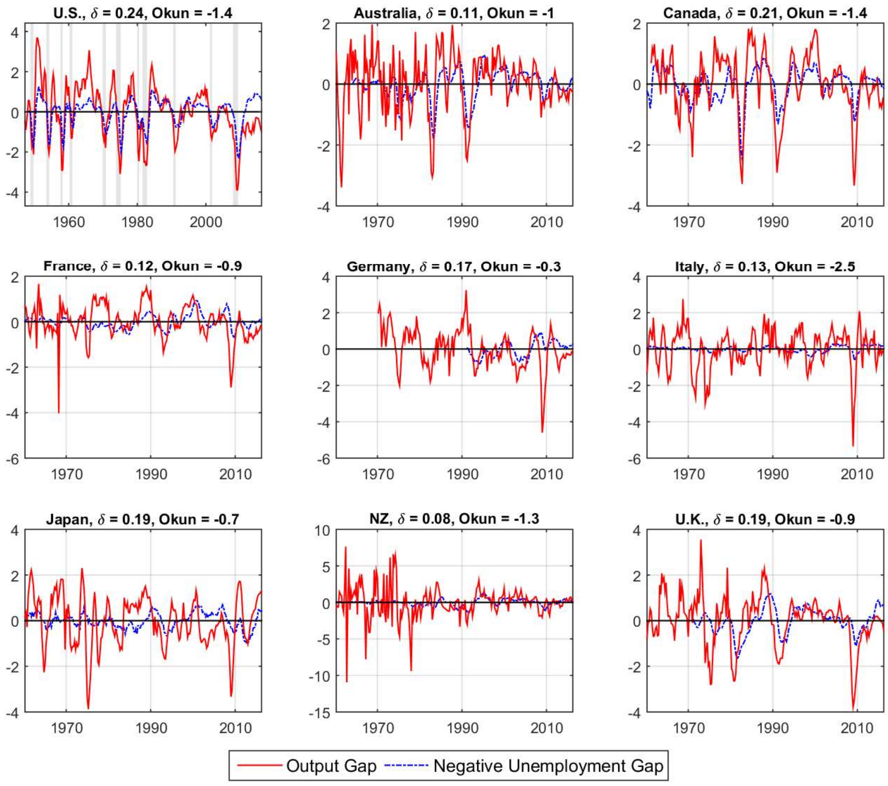
Notes: See notes for Figure 4 for descriptions of labels of methods. The graphs plot out-of-sample RRMSEs compared to forecasts based on the BN filter estimated output gap. Out-of-sample evaluation begins in 1970Q1. The bands depict 90% confidence intervals from a two-sided Diebold and Mariano (1995)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

Figure 8: U.S. output gap estimates from the BN filter when allowing for structural breaks



Notes: Units are 100 times natural log deviation from trend. The top panel compares the BN filter estimated output gap in Figure 2 to a BN filter estimated output gap allowing for a break in long-run growth in 2006Q2 found using the Bai and Perron (2003) test. The middle panel compares the BN filter estimated output gap allowing for a break in long-run growth in 2006Q2 to a BN filter estimated output gap when dynamically demeaning the data relative to a backward-looking rolling 40 quarter average. The bottom panel compares the ex post and pseudo-real-time BN filter estimates of the output gap when dynamically demeaning the data. Shaded bars correspond to NBER recession dates.

Figure 9: Output and unemployment gap estimates from the BN filter for the G7, Australia, and New Zealand



Notes: Units are 100 times natural log deviation from trend. The estimated output gap is from the BN filter and δ is the corresponding signal-to-noise ratio that maximizes the amplitude-to-noise ratio. The estimated unemployment gap is from the BN filter and "Okun" refers to the implied Okun's law coefficient in terms of the percentage point change in the output gap per percentage point change in the unemployment gap. Shaded bars correspond to NBER recession dates.

Table 1: Monte Carlo simulations

	T = 250		*T = 500,000*	
	Correlation	Amplitude	Correlation	Amplitude
	DGP 1 – an unobserved components process			
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
	DGP 2 – a single source of error process			
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: For both DGPs, $y_t = \tau_t + c_t$, $\tau_t = 0.81 + \tau_{t-1} + \eta_t$, $u_t = -0.5c_t + v_t$, $v_t = 0.9v_{t-1} + \zeta_t$, where $\eta_t \sim iidN(0, 0.69^2)$ and $\zeta_t \sim iidN(0, 1)$. For DGP 1, $c_t = 1.53c_{t-1} - 0.61c_{t-2} + \epsilon_t$, where $\epsilon_t \sim iidN(0, 0.62^2)$ and η_t, ϵ_t, and ζ_t are mutually uncorrelated. For DGP 2, $c_t = 1.53c_{t-1} - 0.61c_{t-2} + 0.42\eta_t - 0.18\eta_{t-1}$, where η_t and ζ_t are mutually uncorrelated. $\delta = 0.50$ is the true value for both DGPs. "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 for the respective models.