Crawford School of Public Policy

CAMA

Centre for Applied Macroeconomic Analysis

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

---

CAMA Working Paper 3/2017
January 2017

Gunes Kamber
Reserve Bank of New Zealand
Bank for International Settlements and
Centre for Applied Macroeconomic Analysis, ANU

James Morley
University of New South Wales and
Centre for Applied Macroeconomic Analysis, ANU

Benjamin Wong
Reserve Bank of New Zealand and
Centre for Applied Macroeconomic Analysis, ANU

Abstract

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

Keywords

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

JEL Classification

C18, E17, E32

Address for correspondence:

(E) cama.admin@anu.edu.au

ISSN 2206-0332

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

Güneş Kamber[1,2,4], James Morley[3,4], and Benjamin Wong[1,4]

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

Abstract

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

JEL Classification: C18, E17, E32
Keywords: Beveridge-Nelson decomposition, output gap, signal-to-noise ratio

---

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

†The views expressed in this paper are those of the authors and do not necessarily represent the views of the Reserve Bank of New Zealand or the Bank for International Settlements. We thank the Editor, Yuriy Gorodnichenko, four anonymous referees, seminar and conference participants at the Bank for International Settlements Asian Office, Bank of Canada, Bank of Korea, Central Bank of Turkey, Federal Reserve Bank of Cleveland, Federal Reserve Bank of San Francisco, Reserve Bank of New Zealand, Australian National University, Deakin University, Monash University, Otago University, University of Auckland, Universitat Pompeu Fabra, University of Queensland, 2015 SNDE Symposium in Oslo, 2015 Melbourne Institute Macroeconomic Policy Meetings, 2015 Central Bank Macroeconomic Modelling Workshop, 2015 Workshop of the Australasia Macroeconomic Society, and 2016 AJRC-HIAS Conference in Canberra, and especially Todd Clark, George Evans, Sharon Kozicki, Adrian Pagan, Barbara Rossi, Frank Smets, Ellis Tallman, John Williams, as well as our discussants, Michael Kouparitsas, Glenn Otto, and Leif Anders Thorsrud for helpful comments and suggestions. We also thank Tugrul Vehbi for help with some of our computations. Any remaining errors are our own.

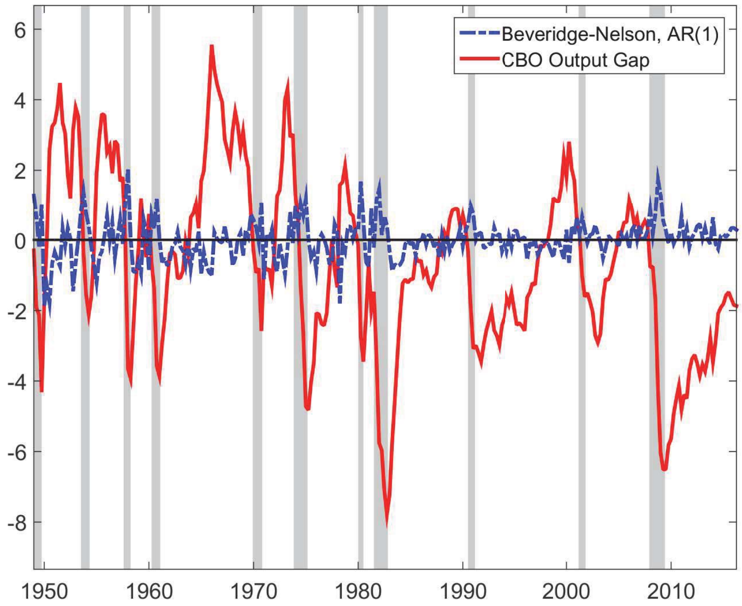
1 Introduction

The output gap is often conceived of as encompassing transitory movements in log real GDP at business cycle frequencies. Because the Beveridge and Nelson (1981) (BN) trend-cycle decomposition defines the trend of a time series as its long-run conditional expectation after all forecastable momentum has died out (and subtracting off any deterministic drift), the corresponding cycle for log real GDP should provide a sensible estimate of the output gap as long as it is based on accurate forecasts over short and medium term horizons. Noting that standard model selection criteria suggest a low-order autoregressive (AR) model predicts U.S. quarterly real GDP growth better than more complicated alternatives, Figure 1 plots the estimate of the output gap from the BN decomposition based on an AR(1) model.[1] What is immediately noticeable 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, while the CBO implicitly assume a much lower signal-to-noise ratio when constructing its estimate.

Our main contribution in this paper is to show how to conduct a BN decomposition imposing a low signal-to-noise ratio on an AR model, an approach we refer to as the "BN filter". The BN filter is easy to implement in comparison to related methods that also seek to address the conflicting results in Figure 1, such as Bayesian estimation of an unobserved components (UC) model with a smoothing prior on the signal-to-noise ratio (e.g., Harvey et al., 2007). Notably,

[1]The raw data for U.S. real GDP are taken from FRED for the sample period of 1947Q1-2016Q2. Real GDP growth is measured in continuously-compounded terms. Model estimation is based on least squares regression or, equivalently, conditional maximum likelihood estimation under the assumption of normality. Initial lagged values for AR(p) models are backcast using the sample average growth rate. Our specific choice of lag order p=1 is based on the Schwarz Information Criterion.

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 real GDP 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.

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.

That Orphanides and van Norden (2002) find output gap estimates unreliable in real time dramatically undermines their usefulness in policy environments and in forming a meaningful gauge of current economic slack more generally. Meanwhile, the fact that the BN filter estimates are not heavily revised is not coincidental, but stems from our choice to work within the context of AR models. In principle, the BN decomposition can be applied using any forecasting model, including multivariate time series models such as vector autoregressive (VAR) models, structural models such as dynamic stochastic general equilibrium (DSGE) models, or even nonlinear time series models such as time varying parameter (TVP) or Markov switching (MS) models. Our choice to work with AR models is deliberate. Because the estimated output gap for a BN decomposition directly reflects the estimated parameters of the model, it is mechanical that any instability in the estimated parameters in real time will produce estimates of the output gap that are heavily revised. However, estimates of autoregressive coefficients for AR models of real GDP growth turn out to be relatively stable in real time, unlike with parameters for more complicated models. Therefore, a natural outcome of our modeling choice is output gap estimates that are reliable 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. Also supportive of greater accuracy, we find that the revised estimate from the BN filter is much more positively correlated with the Chicago Fed's index of economic activity based on 85 data series than are the other more heavily-revised output gap estimates, while the real-time estimate from the BN filter is more (typically negatively) correlated with future revisions in other estimates than other real-time estimates are correlated with future revisions in the estimate from the BN filter.

Our proposed approach is robust to the omission of multivariate information in the forecasting model and can account for 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), amongst others, on the problem of "spurious cycles" in the presence of a random walk stochastic trend when using popular methods of trend-cycle decomposition such as deterministic detrending, the HP filter, and the BP filter. Meanwhile, see Morley et al. (2003), Dungey et al. (2015), and Chan and

Grant (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. quarterly 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 is implied by sample estimates, and then presents forecast comparisons with other methods and considers whether large revisions are useful for understanding the past. Section 4 considers how to modify our approach to account for structural breaks and applies it 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-run conditional expectation minus any *a priori* known (i.e., deterministic) future movements in the time series. In particular, letting $\{y_t\}$ denote a time series process with a trend component that follows a random walk with constant drift, the BN trend, τ_t^{BN}, at time t is

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

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 long-horizon conditional expectation of the remaining cycle component goes to zero. By removing the deterministic drift, the conditional expectation in equation (1) remains finite and becomes an optimal estimate (in a minimum mean squared error sense) 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 functions (ACFs) and partial autocorrelation functions (PACFs) for many macroeconomic time series, including the first differences of U.S. quarterly log real GDP, it is natural when implementing the BN decomposition to consider an AR(p) forecasting model:

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

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 $\mu \equiv \mathbb{E}[\Delta y] = \phi(1)^{-1} c$ exists. Using the state-space approach to calculating the BN decomposition in Morley (2002), the BN cycle, c_t^{BN}, at time t for this model is

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

where $X_t = (\Delta \tilde{y}_t, \Delta \tilde{y}_{t-1}, \ldots, \Delta \tilde{y}_{t-p+1})'$, with $\Delta \tilde{y}_t \equiv \Delta y_t - \mu$ denoting the deviation from the un-

[2]The normality assumption is not strictly necessary for the BN decomposition. However, under normality, least squares regression for an AR model becomes equivalent to conditional maximum likelihood estimation, while the Bayesian shrinkage priors used in our approach, as discussed in the next subsection, become conjugate, making posterior calculations straightforward. Also, the forecast errors do not need to be identically distributed, as long as they form a martingale difference sequence. However, in terms of possible structural breaks in the variance of the forecast error, the key underlying assumption we make in our proposed approach 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.

conditional 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, in particular, seem reasonable for U.S. real GDP growth given sample ACFs and PACFs and is supported by the Schwarz Information Criterion (SIC), we have already seen in Figure 1 that the estimated output gap from a BN decomposition based on an AR(1) model does not match well at all with widely-held beliefs about the amplitude, persistence, and sign of transitory movements in economic activity as reflected, for example, in the 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 understand why the BN decomposition based on an AR(1) model in Figure 1 produces an estimated output gap with such features, it is helpful to note from equation (3) that the BN cycle in this case 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 maximum likelihood estimation (MLE) is typically between 0.3 and 0.4 for U.S. quarterly data. Similarly, given that $\hat{\phi}(1-\hat{\phi})^{-1} \approx 0.5$, the amplitude of the estimated output gap will be small, as the implied variance will only be about one quarter that of output growth itself. Furthermore, given that $-\hat{\phi}(1-\hat{\phi})^{-1} < 0$, it is not surprising that the estimated output gap generally increases in recessions when output growth becomes negative and vice versa in expansions. In terms of the intuition underlying the BN decomposition, the momentum in output growth implied by the AR(1) model means that when there is a negative shock in a recession, output growth is expected to remain below average in the quarters immediately afterwards before eventually returning back to its long-run average, with the converse holding for a positive shock in an expansion. Thus, log real GDP will initially be above the BN trend defined as the long-run conditional expectation minus deterministic drift when a shock triggers a recession and below the BN trend when a shock triggers an expansion.

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

$$\delta \equiv \sigma_{\Delta\tau}^2 / \sigma_e^2 = \psi(1)^2, \tag{4}$$

where $\psi(1) \equiv \lim_{j\to\infty} \frac{\partial y_{t+j}}{\partial e_t}$ is the "long-run multiplier" that captures the permanent effect of a forecast error on the long-horizon conditional expectation of $\{y_t\}$ and provides the key summary statistic for a time series process when calculating the BN trend based on a forecasting model given that $\Delta\tau_t^{BN} = \psi(1)e_t$.[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 postwar U.S. quarterly real GDP growth, the signal-to-noise ratio appears to be quite high with $\hat{\delta} = 2.22$. That is, BN trend shocks are much more volatile than quarter-to-quarter forecast errors in log real GDP, leading to an estimated output gap with small amplitude and generally 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 equation (3), the BN cycle depends 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 a powerful one because it implies that we can impose a low signal-to-noise ratio by fixing

[3] An implicit equivalence between the variance of trend shocks, $\sigma_{\Delta\tau}^2$, 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 autocovariance structure as the reduced-form model. Note that, as found in Morley (2011) and discussed below in Section 3, a similar equivalence does not hold for the variance of the transitory component and the variance of the BN cycle.

the sum of the autoregressive coefficients when estimating an AR(p) model. To do so, we transform the AR(p) model in equation (2) into its Dickey-Fuller representation, which we also write in terms of 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, equation (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 solving for the restricted estimates of $\{\phi_j\}_{j=1}^{p}$ by inverting the Dickey-Fuller transformation given $\bar{\rho}$ and estimates of $\{\phi_j^*\}_{j=1}^{p-1}$ and calculating the BN cycle as in equation (3).

Before discussing estimation of the model in equation (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 equation (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 by bringing $\bar{\rho}\Delta\tilde{y}_{t-1}$ to the left-hand-side when running a least squares regression for equation (5). 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"-type 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 prior. Readers familiar with the Minnesota class of priors will recognize that, conditional on σ_e^2, the posterior distribution for $\{\phi_j^*\}_{j=1}^{p-1}$ will have 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] However, it should be noted that the estimated output gap is highly robust to conditioning on different values for σ_e^2.[6]

All that remains is to choose a 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 $\lambda = 1600$ when implementing the HP filter for quarterly data under the assumption $\{y_t\}$ follows a UC process with a stochastic trend (see

[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 be greater tendency to overfit the data given well-known problems of weak identification and near-cancellation of roots, and the corresponding BN decomposition would be less reliable.

[5]In the case of a time series with no drift, such as is the case with U.S. unemployment rate data, it is possible to impose $\mu = 0$ by simply setting $\Delta\tilde{y}_t = \Delta y_t$.

[6]The estimated output gap is virtually identical when halving or doubling σ_e^2 compared to the least squares estimate. Meanwhile, it would be possible to place an informative prior on σ_e^2 and still solve the posterior for the autoregressive coefficients analytically by using dummy observations. However, given that this would require eliciting an explicit prior in any particular application, while having virtually no impact on the results for the BN decomposition, we employ the flat/improper prior for simplicity.

Harvey and Jaeger (1993)).[7]

Imposing a dogmatic prior could be as simple as setting $\bar{\delta}$ to a particular low value such as, 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 particular choice for $\bar{\delta}$ might appear somewhat arbitrary in practice. Therefore, we propose an automatic selection of $\bar{\delta}$ based on the 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 equation (3) and $\sigma_e^2(\delta)$ is the variance of the forecast error for the corresponding restricted version of the AR(p) model in equation (5).[8] For the automatic selection, $\bar{\delta}$ is chosen to maximize the signal-to-noise ratio as follows:

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

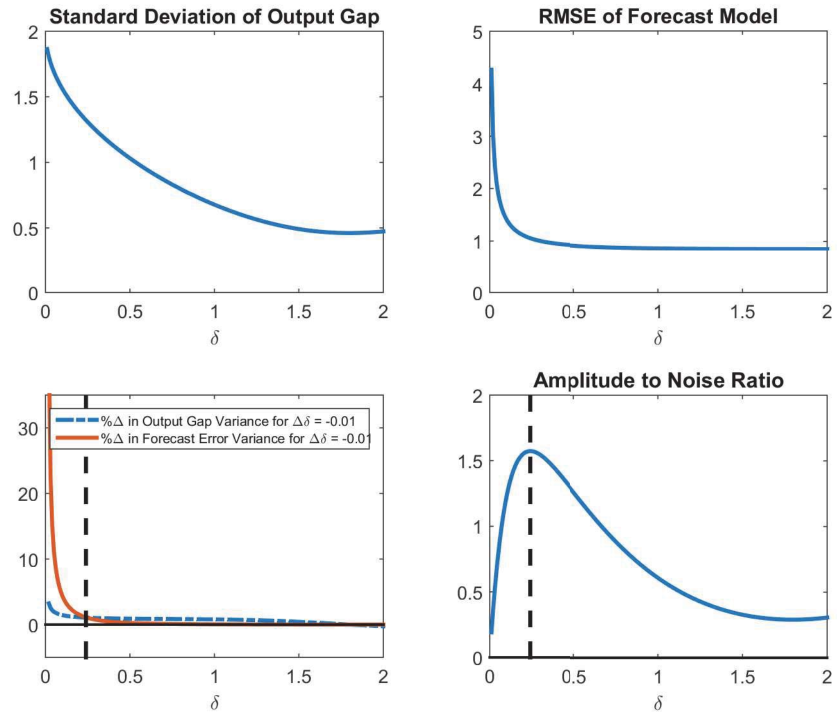
This is still a dogmatic prior in the sense that, even in large samples, $\bar{\delta}$ will generally be smaller than if it were freely estimated (e.g., if, instead, $\bar{\delta} = \operatorname{argmin} \sigma_e^2(\delta)$).[9] However, given the use of a local maximizer for values of δ close to zero, 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

[7] Under the assumption that $\{y_t\}$ is level or trend stationary, the HP filter can be interpreted as an approximate high-pass frequency-domain filter and the choice of $\lambda = 1600$ for quarterly data has an interpretation of isolating transitory movements at particular frequencies that are often associated with the business cycle. In applying the BN decomposition, we take it as a given that $\{y_t\}$ can contain a stochastic trend. Thus, we do not frame the choice of $\bar{\delta}$ in terms of what frequencies it isolates under the assumption that $\{y_t\}$ is stationary. However, we examine the sample periodogram for the estimated output gap to determine what frequencies are being highlighted and confirm that they correspond closely to those highlighted by the HP and BP filters and that are often associated with business cycles. We thank a referee for suggesting this comparison and the results are available in the online appendix.

[8] 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. Likewise, it is important to note that 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.

[9] We have confirmed 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.

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



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

the signal-to-noise ratio taking on a particular low value. To be clear, this is not the same as simply maximizing the amplitude of the estimated output gap. It will still rule out even lower values of δ that would increase amplitude, but not enough compared to worsening the fit of the forecasting model. Specifically, maximizing the amplitude-to-noise ratio will balance this tradeoff between amplitude and fit by equating the percentage increase in the variance of the BN cycle with the percentage decrease in the variance of the forecast error.

To provide a visual perspective on our approach, Figure 2 plots the relationships for the U.S. data between δ and the standard deviation of the estimated output gap, the RMSE of the restricted forecasting model, the percentage changes in the corresponding variances for a decrease in the signal-to-noise ratio of $\Delta\delta = -0.01$, and the implied amplitude-to-noise ratio. In the first panel, the relationship between δ and the amplitude of the output gap is non-monotonic, with an inflection around 1.75, below which there is a steady increase in amplitude as δ gets smaller. Meanwhile, the relationship between δ and the RMSE of the forecasting model considered in the second panel is also non-monotonic (the minimum, not displayed, is at $\delta = 2.25$, which reflecting the impact of the shrinkage priors on the second-difference coefficients is closer to the MLE for an AR(1) model than for an AR(12) model), but is relatively flat except when δ is very close to zero, at which point decreasing δ has a deleterious impact on fit. In the third panel, it can be seen that $\bar{\delta} = 0.24$ (i.e., imposing that trend shocks account for 24% of the forecast error variance for quarterly output growth) optimizes the tradeoff between amplitude and fit in the sense that a decrease in δ below $\bar{\delta} = 0.24$ worsens fit more in percentage terms than it increases amplitude. The fourth panel shows directly that $\bar{\delta} = 0.24$ corresponds to the local maximum in the amplitude-to-noise ratio for values of δ closest to zero.

Because of the non-monotonicities of the relationships of δ with amplitude and fit, we focus on a local maximizer of the amplitude-to-noise ratio. In particular, for very large values of δ (much larger than the freely-estimated value), the amplitude-to-noise ratio actually becomes quite large and increasing with δ, meaning that a global maximum does not exist. However, because we are interested in imposing a dogmatic prior of a "low" signal-to-noise ratio, the local maximum for values of δ closest to zero provides a sensible basis for selecting $\bar{\delta}$.[10] Meanwhile, as we show below, the shape of the output gap is highly robust to imposing other low values of $\bar{\delta}$. 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".[11] To summarize,

[10]As a check on the reasonableness of restricting fit to increase amplitude in this way, we also consider the prior predictive density for the restricted model. We find that 82% of the postwar observations for U.S. real GDP growth lie within the equal-tailed 90% bands of the conditional prior predictive density. In this sense, our restricted AR(12) model with $\bar{\delta} = 0.24$ and shrinkage priors on the second-difference coefficients, while clearly different than what would be freely estimated by MLE, is not strongly at odds with the data.

[11]We note that a few previous papers refer to a "BN filter", but always as an alternative terminology for the traditional BN decomposition, which does not impose a smoothing effect on the trend.

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 equation (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"-type 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 prior.

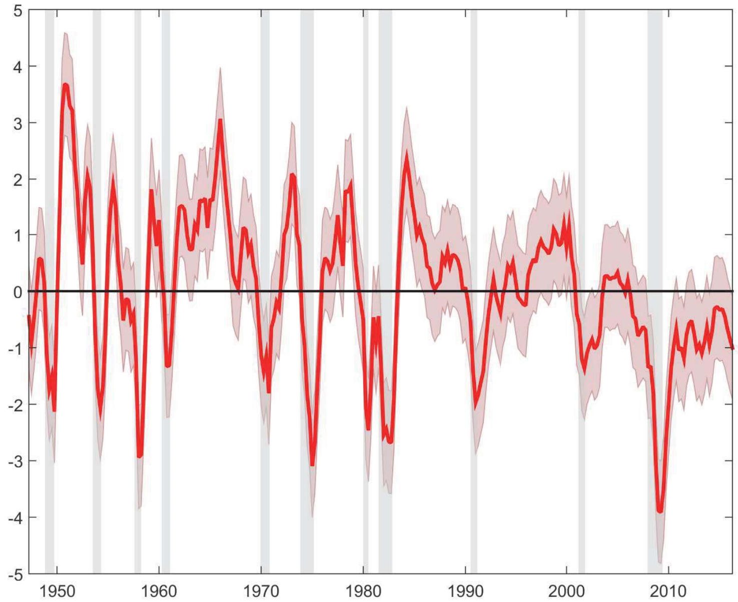
3. Given $\bar{\rho}$ and estimates of $\{\phi_j^*\}_{j=1}^{p-1}$, solve for restricted estimates of $\{\phi_j\}_{j=1}^{p}$ by inverting the Dickey-Fuller transformation and apply the BN decomposition as in equation (3).

Figure 3 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 estimated output gap that is consistent with widely-held beliefs about amplitude, persistence, and the sign of transitory movements in economic activity. In particular, the estimated output gap is large in amplitude, persistent, and moves procyclically with the NBER reference cycle. Referring back to Figure 1, the correlation with the CBO output gap is reasonably high at 0.75 rather than -0.30 for the estimate from a BN decomposition based on an AR(1) model. Meanwhile, 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.[12] The estimates appear reasonably precise and are significantly different from zero at many points throughout the sample, especially during recessions.[13]

[12] Details for how we calculate confidence bands for a BN cycle are provided in the online appendix.

[13] All of the methods considered in this paper 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 (2016) 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.

Figure 3: Estimated U.S. output gap from the BN filter with 95% confidence bands



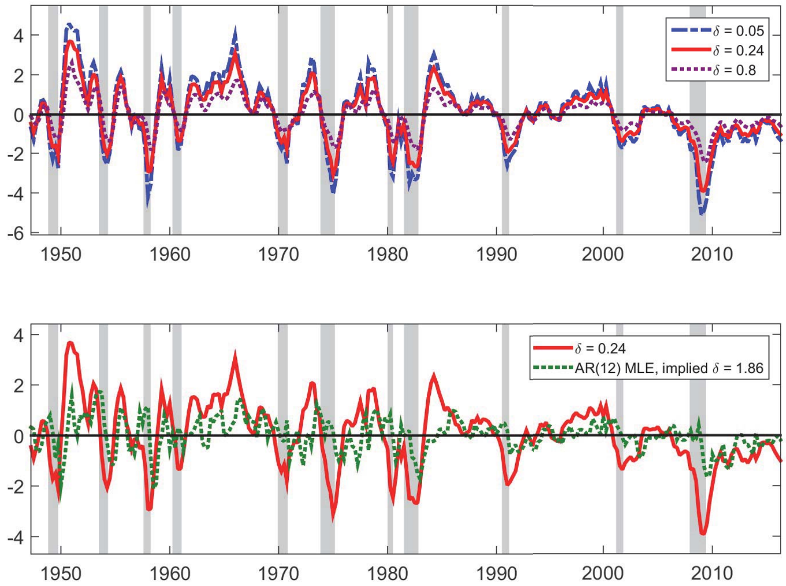
Notes: Units are 100 times natural log deviation from trend. "BN filter" refers to our proposed approach of estimating the output gap using the BN decomposition based on Bayesian estimation of an AR(12) model of U.S. quarterly real GDP growth with shrinkage priors on second-difference coefficients and 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.

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 4 plots the estimated output gap for $\delta \in \{0.05, 0.8\}$ compared against our approach where $\delta = 0.24$. The shape of the estimated output gap is little changed, with the persistence profile virtually unaltered. Indeed, the correlation between the different estimated output gaps varying the signal-to-noise ratio is well in excess of 0.95. Because the profile of fluctuations in the estimated output gap are similar even as we change the signal-to-noise ratio, the revision properties and real time forecast performance will be highly robust to the exact value of δ, at least as long as it is below one.[14]

Meanwhile, given this apparent robustness to different values of δ, we check whether our results are actually being driven by the AR(12) specification rather than imposing a low signal-

[14] All of the results for $\delta \in \{0.05, 0.8\}$ are reported in the online appendix.

Figure 4: 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 real GDP 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.

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 4 plots the two output gap estimates and makes it clear that imposing a low signal-to-noise ratio is crucial. For the freely estimated AR(12) model, we obtain $\hat{\delta} = 1.86$ and the correlation between the two estimates is only 0.39. Also, as shown below, the revision properties and forecast performance are not 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. 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. Therefore, to focus on the revision properties of the filtering method in particular, we consider "pseudo-real-time" analysis using the final vintage of data (from 2016Q2) rather than a full-blown real-time analysis with different vintages of data. However, our results are generally robust to consideration of data revisions.[15]

To evaluate the performance of the BN filter, we compare it to several other methods of trend-cycle decomposition. First, we consider BN decompositions based on various freely estimated ARMA forecasting models of $\{\Delta y_t\}$ and Kalman filtering for a UC model of $\{y_t\}$. In particular, we consider BN decompositions based on an AR(1) model, an AR(12) model, and an ARMA(2,2) model, all estimated via MLE. We also consider a multivariate BN decomposition based on a VAR(4) model of U.S. real GDP growth and the civilian unemployment rate, also estimated via MLE. For the UC model, we consider a similar specification to Harvey (1985) and Clark (1987) estimated via MLE. The AR(1) model is chosen based on SIC for the whole set of possible ARMA models. The AR(12) model allows us to understand the 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 Dungey et al., 2015; 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

[15] See the online appendix for the results in a full-blown real-time environment. Not only are the results generally robust, but the only major change is that the BN filter clearly outperforms the traditional BN decomposition based on an AR(1) model, 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 the BN cycle for it is a weighted average of 12 quarters of real GDP 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. More details are provided in the online appendix.

drift plus an AR(2) cycle. Meanwhile, for completeness, we also consider the Harvey-Clark UC model, similar to that considered by Orphanides and van Norden (2002). The Harvey-Clark UC model differs from the Watson (1986) model in that it also features a random walk drift in addition to a random walk trend. For this model, we retain the zero restrictions on correlations between movements in drift, trend, and cycle.

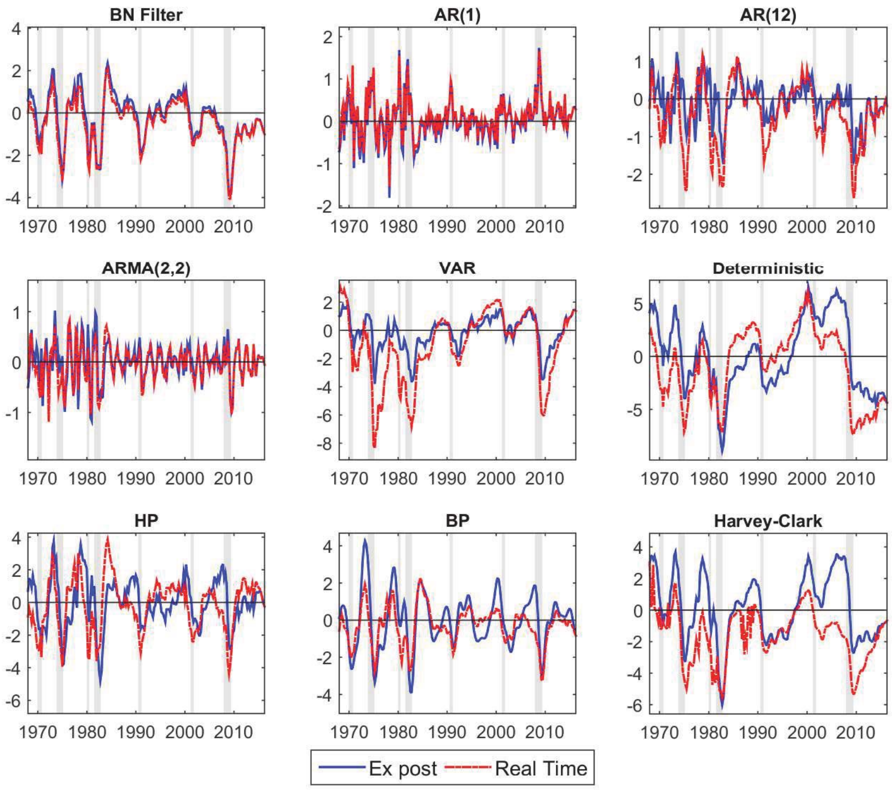
We also consider some popular methods of deterministic detrending and nonparametric filtering. In particular, we consider a deterministic quadratic trend, the Hodrick and Prescott (HP) (1997) filter, and the bandpass (BP) filter by Baxter and King (1999) and Christiano and Fitzgerald (2003). For the HP filter, we impose a smoothing parameter $\lambda = 1600$, as is standard for quarterly data. For the BP filter, we target frequencies 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). This reflects the difficulty of accurately forecasting future output growth.

Figure 5 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.[16] The first thing to notice is that all of the features of the output gap for the BN filter in Figure 3 carry over to the pseudo-real-time environment.[17] 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

[16]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 real GDP growth, the unemployment rate is backcast using its pseudo-real-time sample average to allow the estimation sample to always begin in 1947Q2.

[17]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.

Figure 5: 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 real GDP 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.

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 reasonably intuitive estimates based on the BN filter and the BN decompositions for the AR(12) model and the VAR model. Thus, being "intuitive" clearly cannot be 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 thing to notice in Figures 5 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 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.[18] This is the main reason we impose a shrinkage prior when considering the highly parameterized AR(12) model in our proposed approach. In particular, the shrinkage prior prevents overfitting, while the high lag order still allows for relatively rich dynamics. The revision properties of the BN filter estimates, which are more similar to those for the AR(1) model than for the AR(12) model based on MLE suggest that our proposed approach achieves a reasonable compromise between avoiding potential overfitting and allowing for richer dynamics. Finally, all of the estimates for the other popular methods in Figure 5 are heavily revised and clearly unreliable in real time.[19] In particular, the deterministic detrending is extremely

[18]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.

[19]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

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.

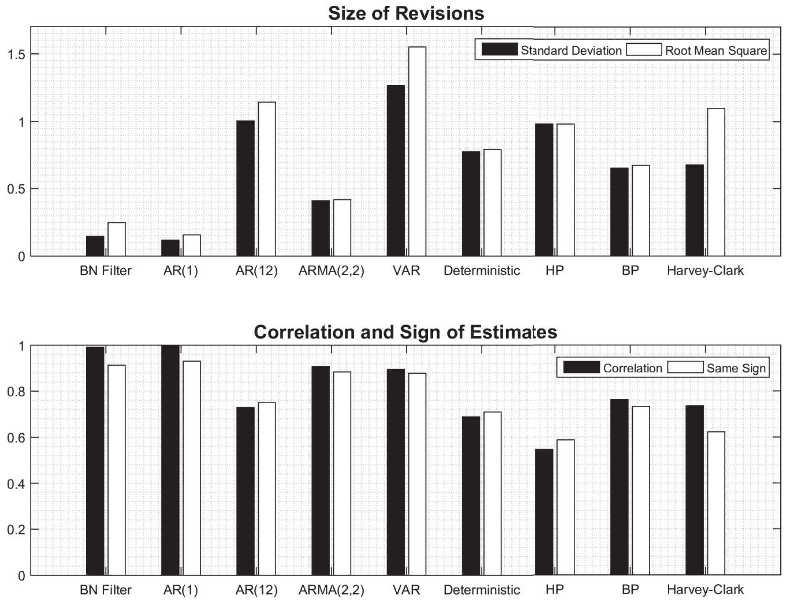
While eyeballing Figure 5 suggests the BN filter should be relatively appealing from a reliability perspective, we formally quantify these revision properties by calculating revision statistics, similar to the analysis in Orphanides and van Norden (2002), Edge and Rudd (2016), and Champagne et al. (2016). 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 as the different methods produce estimates with very different amplitudes.[20] Second, we calculate the correlation between the pseudo-real-time estimate of the output gap and the ex post estimate of the output gap. Third, we compute the frequency with which the pseudo-real-time estimate of the output gap has the same sign as the ex post estimate. We consider the evaluation sample of 1970Q1-2012Q4 to match with the starting point for the out-of-sample forecast comparison discussed in the next section and because the more recent estimates near the end of the full sample in 2016Q2 may end up becoming more heavily revised in the future.

Figure 6 presents the revision statistics. As was visually apparent in Figure 5, the BN filter does well in terms of size of revisions, with both the standard deviation and RMS statistics being less than one quarter of the 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 reasonably 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 quite poorly, with the standard deviation and RMS statistics about one standard de-

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.

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

Figure 6: Revision statistics for U.S. output gap estimates



Notes: See notes for Figure 5 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.

viation 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 relatively 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, within the class of BN decompositions, model parameter parsimony seems to be important for reliability of the estimated output gap. This is not much of a surprise given that models which are highly parameterized, such as the AR(12) model or the VAR(4) model, will tend to feature parameter estimates that can be more unstable with the consideration of future observations. 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 very 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 recap, we have proposed a BN filter that imposes a low signal-to-noise ratio when conducting the BN decomposition. When applied to U.S. log real GDP, the resulting output gap estimates are reliable in the Orphanides and van Norden (2002) sense of being subject to small revisions over time. The BN filter does better than other popular methods, except for the BN decomposition based on an AR(1) model estimated via MLE, which does marginally better in terms of revisions, but corresponds to a much higher signal-to-noise ratio and is economically indistinguishable from zero throughout the sample.[21]

To the extent that one is agnostic about the true signal-to-noise ratio, there is not much reason to deviate from using the BN decomposition based on an AR(1) model, especially if

[21] It is also statistically somewhat less distinguishable from zero in the sense that, as discussed in footnote 19, zero lies outside the 95% confidence band 30% of the time for the BN decomposition based on an AR(1) model versus 40% of the time for the BN filter.

standard model selection and revision properties are the main criteria for choosing an approach to estimating the output gap. In other words, one can really only justify using the BN filter if there is a compelling reason to believe that a low signal-to-noise ratio really represents the true state of the world. Whether a low or high signal-to-noise ratio represents the true state of the world remains unresolved in the empirical literature. While considerable empirical research has found support for the presence of a volatile stochastic trend in U.S. log real GDP (e.g., Nelson and Plosser, 1982; Morley et al., 2003), this view has not gone unchallenged (e.g., Cochrane, 1994; Perron and Wada, 2009).

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

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

For the first DGP, we parameterize the Watson (1986) UC model of $\{y_t\}$ with uncorrelated components as estimated for U.S. real GDP by Morley et al. (2003). We choose this DGP because it corresponds to a low signal-to-noise ratio, unlike the unrestricted UC model in Morley et al. (2003) that allows for correlation between permanent and transitory movements.[22] When considering model selection for possible ARMA specifications for $\{\Delta y_t\}$ given this DGP in finite samples, SIC will pick a low-order AR(p) model with reasonably high frequency, even though the true model has an ARMA(2,2) specification. Meanwhile, suppose there is some other observed variable $\{u_t\}$ that is related to the unobserved cycle $\{c_t\}$, but contains serially-

[22]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.

Table 1: Monte Carlo simulation for unobserved components process

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

Notes: We consider the following DGP:
$y_t = \tau_t + c_t$
$\tau_t = 0.81 + \tau_{t-1} + \eta_t,$
$c_t = 1.53c_{t-1} - 0.61c_{t-2} + \varepsilon_t,$
$u_t = -0.5c_t + v_t$
$v_t = 0.9v_{t-1} + \zeta_t,$
where $\eta_t \sim iidN(0, 0.69^2), \varepsilon_t \sim iidN(0, 0.62^2), \zeta_t \sim iidN(0,1)$, and η_t, ε_t, and ζ_t are mutually uncorrelated. $\delta = 0.50$ is the true value. "Correlation" refers to correlation between the true cycle and the estimated cycle. "Amplitude" is in terms of the standard deviation of percent deviation from trend. All estimated cycles are derived from BN decompositions for the respective models.

correlated "measurement error". Tests for Granger causality will often suggest that $\{u_t\}$ has predictive power for $\{\Delta y_t\}$ beyond a low-order univariate AR(p) process. Based on this, a researcher might consider a multivariate BN decomposition, as argued for by Cochrane (1994) in such a setting. For this DGP, we consider how well different cases of the BN decomposition would do in estimating the true cycle $\{c_t\}$.

Table 1 reports the results for the first DGP in a finite sample (T=250) and in population (T=500,000). The first thing to note 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 for this DGP, we find that SIC chooses a lag order for an AR(p) model of p=1 more than 95% of the time.

The next thing to note 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 too surprising given the inclusion of the true cycle in the forecasting model corresponds to a highly unrealistic scenario in which there exists an observed variable that perfectly captures economic slack. Indeed, if such a variable really did exist, there would be little reason to estimate the output gap in the first place rather than just monitoring the observed variable. Instead, a more realistic scenario is one in which there exists an observed variable that is related to economic slack but is also affected by persistent idiosyncratic factors (e.g., the unemployment rate). In order to capture such a scenario, we generate an artificial time series $\{u_t\}$ which is linked to the true cycle $\{c_t\}$ up to a 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.[23]

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.[24] Meanwhile,

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

[24]The BN decomposition based on a VAR(4) model of $\{\Delta y_t\}$ and the true cyclical component $\{c_t\}$ does not suffer from a downward bias because the cyclical component is observed and the true DGP has a (restricted) VAR(2) representation. 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 estimates – in this case their standard deviation – is different than the property of the true values. 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 optimal estimates to have different properties than the objects 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 have different properties in terms of their variability, as directly alluded to by their labels.

the finite sample results for the ARMA(2,2) model are much worse than the population results, with the correlation between the estimated cycle and the true cycle being close to zero. The relatively poor finite sample performance of the BN decomposition in this case likely reflects well-known difficulties with estimating ARMA parameters due to weak identification.

Turning to our proposed BN filter, we find that the estimated cycle shares the same relatively high correlation with the true cycle as a version of the BN decomposition that imposes the true signal-to-noise ratio (which, of course, is never known in practice) and the BN decomposition based on an AR(12) model. The AR(12) model does reasonably well given that it approximates the infinite-order AR representation of $\{\Delta y_t\}$. However, for this DGP, the BN decomposition based on the AR(12) model suffers from a larger downward bias in estimating the amplitude than our proposed approach. Indeed, the BN filter does even better in terms of amplitude than the BN decomposition imposing the true signal-to-noise ratio because it explicitly involves targeting $\bar{\delta}$ to maximize amplitude subject to a tradeoff with 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 a single-source-of-error process (see Anderson et al., 2006) that is parameterized to imply the same ARMA(2,2) process for $\{\Delta y_t\}$ as the first DGP. Thus, the same signal-to-noise ratio and all the same tendencies for SIC to pick a low-order AR model hold for this DGP. The only difference in a univariate context is a conceptual one about whether forecast errors represent true trend shocks (i.e., they are the "single source of error" in the process for $\{y_t\}$) or they are linear combinations of unobserved trend and cycle shocks, as was the case in the first DGP. See Morley (2011) for a full discussion of this conceptual distinction.

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

Table 2: Monte Carlo simulation for single source of error process

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

Notes: We consider the following DGP:
$y_t = \tau_t + c_t$
$\tau_t = 0.81 + \tau_{t-1} + \eta_t,$
$c_t = 1.53c_{t-1} - 0.61c_{t-2} + 0.42\eta_t - 0.18\eta_{t-1}$
$u_t = -0.5c_t + v_t$
$v_t = 0.9v_{t-1} + \zeta_t,$
$\eta_t \sim iidN(0, 0.69^2), \zeta_t \sim iidN(0,1)$, and η_t and ζ_t are mutually uncorrelated. "Correlation" refers to correlation between the true cycle and the estimated cycle. $\delta = 0.50$ is the true value. "Amplitude" is in terms of the standard deviation of percent deviation from trend. All estimated cycles are derived from BN decompositions for the respective models.

to-noise ratio for an AR model appears to be a more effective way at getting at the true cycle than estimating the true model when estimation involves weak identification issues. Second, the BN decomposition based on a VAR(4) model of $\{\Delta y_t\}$ and $\{u_t\}$ does worse than for the first DGP, suggesting that measurement error in an observed measure of economic slack offsets the benefits of having a forecast error represent the true trend shock. Again, imposing a low signal-to-noise ratio appears to be a more straightforward and effective way to get the estimated cycle closer to the true cycle than adding multivariate information, even if the multivariate information also decreases the signal-to-noise ratio, as discussed in Evans and Reichlin (1994). Determining the appropriate multivariate information is also a difficult econometric problem in itself, with finite-sample power issues and, at the same time, considerable danger of overfitting unless variable selection is handled carefully.[25] Meanwhile, even given the correct multivariate information, the practical issue of measurement error that effectively motivates the need to estimate the cyclical component in the first place means that a multivariate BN decomposition will suffer even in population. Third, although the BN decomposition based on an AR(12)

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

model estimated via MLE does relatively well, especially for this DGP, we know from the analysis in the previous section that, like the BN decomposition based on a VAR model, it suffers from larger revisions than our proposed approach.

The bottom line is that it is possible to think of a true state of the world in which standard model selection criteria and hypothesis testing would push a researcher towards an AR(1) model (based on parsimony), an ARMA(2,2) model (as estimation and testing eventually discovers the true model given enough data), or possibly a VAR model (based Granger causality tests), but the BN decomposition based on these models would do much worse at capturing the true cycle than our proposed approach. Although the ARMA(2,2) model is the correct specification, the BN decomposition for this model suffers in finite samples due to known estimation problems for such models. The BN decomposition for the AR(1) model performs poorly in large samples, as does the VAR(4) model when the multivariate information is measured with error. Meanwhile, even though the BN decomposition based on an AR(12) model does reasonably well, as would a VAR(4) model when the multivariate information is measured accurately, these versions of the BN decomposition still suffer from reliability issues. By contrast, the BN filter works well even in finite samples and is reliable 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 at the same time worry that imposing a low signal-to-noise ratio could produce a spurious cycle if the true state of the world for U.S. real GDP growth is more along the lines of an AR(1) model than the two DGPs considered above. In particular, if the BN filter produces a 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 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. Our forecast evaluation sample starts in 1970Q1. We use an expanding window for estimation. The first estimate of an output gap we have is for 1947Q2. We use the full extent of the data sample for forecast evaluation after adjusting for the maximum number of lags in the forecasting equation.

U.S. Output Growth Forecasts Nelson (2008) argues for using forecasts of future output growth as a way to evaluate competing estimates of the output gap. The underlying intuition is that if an estimated output gap suggests output is below trend, this should imply faster output growth at some point in the future when output adjusts towards the trend to close the gap. Conversely, if output is above trend, one should forecast slower output growth at some horizon for output to return back towards the trend. The point is that the true cycle will cross its unconditional mean of zero at some point, and a good estimate of the output gap should be able to forecast the effects of this reversion to mean. For an h-period-ahead output growth forecast, we consider a forecasting equation similar to Nelson (2008):

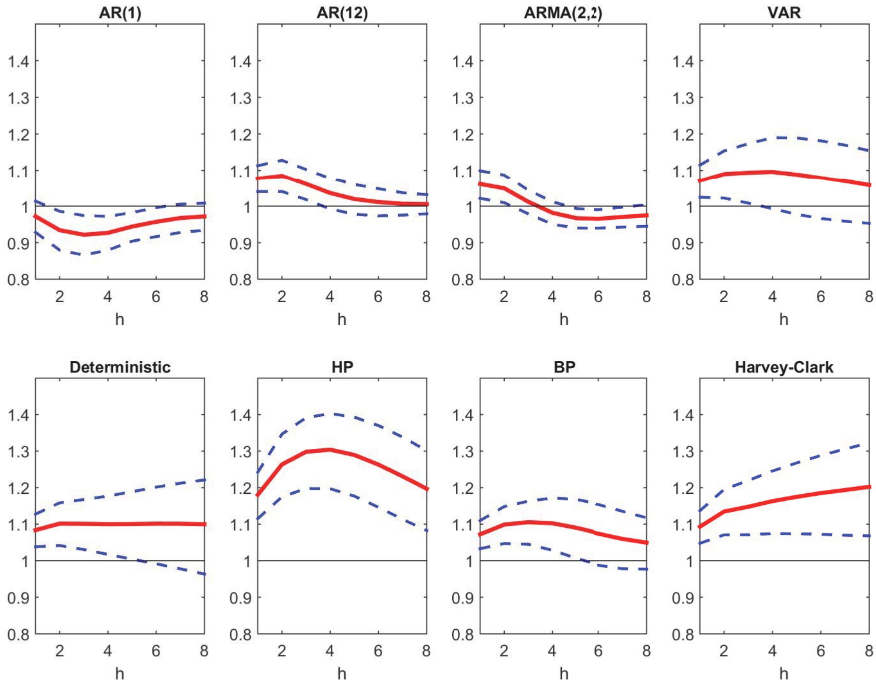
$$y_{t+h} - y_t = \alpha + \beta \hat{c}_t + \varepsilon_{t+h,t} \tag{7}$$

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

We make two observations about the forecasting results. First, the output gap estimates constructed using the BN decomposition do better at all horizons than those based on other

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



Notes: See notes for Figure 5 for descriptions of labels of methods. The graphs plot out-of-sample RRMSE compared to forecasts based on the BN filter estimated output gap. Out-of-sample evaluation begins in 1970Q1. The bands depict 90% confidence intervals from a two-sided Diebold and Mariano (1995) test of equal forecast accuracy.

methods, including the HP and BP filters. This further vindicates our choice to work with a BN decomposition. Second, within the class of BN decompositions, similar to with the revision statistics, parsimony or shrinkage priors seem to be key. In particular, the BN decomposition based on AR(12), ARMA(2,2), and VAR(4) models do worse than the BN decomposition based on an AR(1) model or the BN filter. Therefore, the results for forecasting output growth mimic many of the results we had for the revisions 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. Notably the RRMSE is close to one and is 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.

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

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

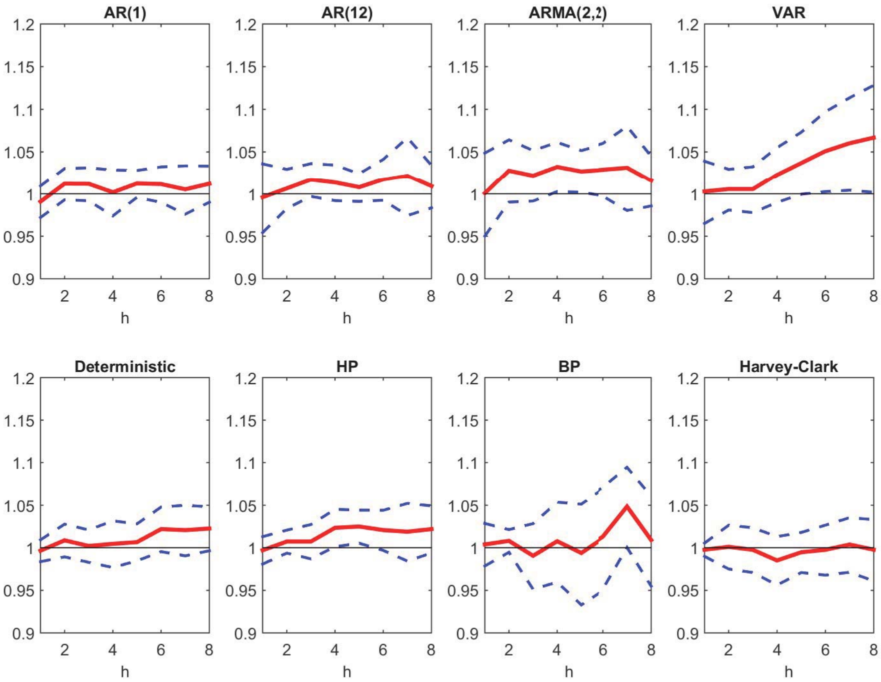
where π is U.S. CPI inflation.[26] We choose the lag orders of the forecasting equation, namely p and q above, using the SIC. As is commonly done (see, for example, Stock and Watson, 1999, 2008; Clark and McCracken, 2006), we apply the information criterion to the entire sample and run the pseudo-real-time 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 8 presents the out-of-sample forecasting results for the U.S inflation. Once again, as with the results for output growth, we compute 90% confidence intervals by inverting the Diebold and Mariano (1995) test.

As in the case of forecasting output growth, the BN filter estimate of the output gap does relatively well. In particular, imposing a low signal-to-noise ratio outperforms all other BN decompositions, although generally not significantly so. The BN filter also generally does better

[26]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.

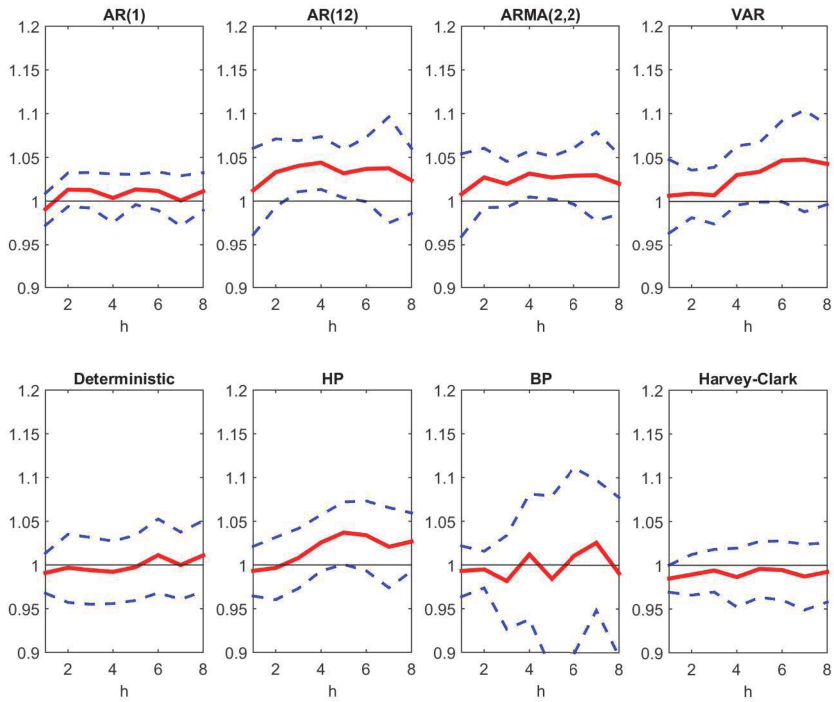
Figure 8: Out-of-sample U.S. inflation forecast comparison relative to the BN filter benchmark using pseudo-real-time output gap estimates



Notes: See notes for Figure 5 for descriptions of labels of methods. The graphs plot out-of-sample RRMSE compared to forecasts based on the BN filter estimated output gap. Out-of-sample evaluation begins in 1970Q1. The bands depict 90% confidence intervals from a two-sided Diebold and Mariano (1995) test of equal forecast accuracy.

than the HP filter, BP filter, and deterministic detrending, although again not significantly so. In contrast to the results for forecasting output growth, the differences in inflation forecast performance using the different output gap estimates are fairly small, with most RRMSEs within the 1.00 to 1.05 range, indicating the gains in changing the output gap estimates for forecasting inflation can be marginal at best and are generally not significant. To some extent, these results are not entirely surprising. Atkeson and Ohanian (2001) and Stock and Watson (2008) show that real-activity based Phillips Curve type forecasts may not be particularly useful for forecasting inflation. In some sense, then, our results are simply a manifestation of what is commonly found in the inflation forecasting literature. However, we note that our proposed approach is still competitive and may be slightly better than 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

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



Notes: See notes for Figure 5 for descriptions of labels of methods. The graphs plot pseudo-out-of-sample RRMSE compared to forecasts based on the BN filter estimated output gap. Forecast evaluation begins in 1970Q1, but is only a pseudo-out-of-sample evaluation given the use of revised output gap estimates. The bands depict 90% confidence intervals from a two-sided Diebold and Mariano (1995) test of equal forecast accuracy.

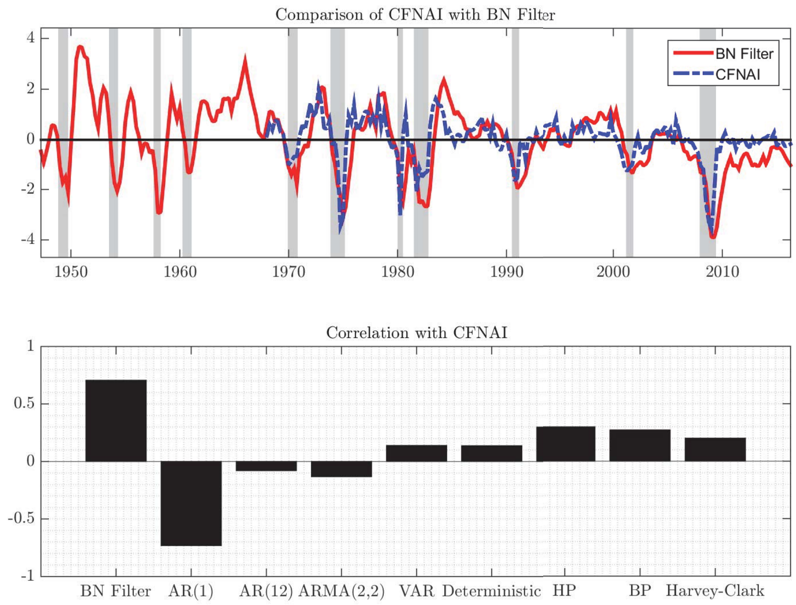
some horizons relative to approaches such as the HP filter and deterministic detrending. It is also noteworthy that none of the alternative methods outperform our proposed approach in a statistically significant way.

3.3 Are Large Revisions Useful for Understanding the Past?

In this subsection, we discuss whether heavily revised output gap estimates, although less useful for current analysis in real time, provide a better ex post understanding and interpretation of past economic activity. Without ever knowing the true output gap, it can be a challenge to evaluate the historical accuracy of output gap estimates. However, we attempt to address historical accuracy in three ways.

First, we consider the ability of revised estimates to predict future inflation. Figure 9

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



Notes: Units in top panel are 100 times natural log deviation from trend for the ex post estimated output gap for BN filter and reported index units for the Chicago Fed's National Activity Index (CFNAI). Shaded bars correspond to NBER recession dates. See notes for Figure 5 for descriptions of labels of methods. Ex post estimates are used in the calculation of correlations.

presents a pseudo-out-of-sample U.S. inflation forecast comparison using revised output gap estimates instead of real-time estimates. Notably, we find only small differences in the forecasting results, with the relative performance of heavily-revised approaches often deteriorating in comparison to using the pseudo-real-time estimates.[27] Thus, the large revisions for deterministic detrending, the HP filter, the BP filter, and the Harvey-Clark model are clearly not providing any additional insights into the historical values of the output gap that are relevant for inflation.

Second, we consider the relationship of the various revised estimates of the output gap with an alternative measure of U.S. economic activity that is constructed in a completely different way, namely the Chicago Fed's National Activity Index (CFNAI). The CFNAI is an index of activity based on 85 data series and Berge and Jordà (2011) find it provides particularly ac-

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