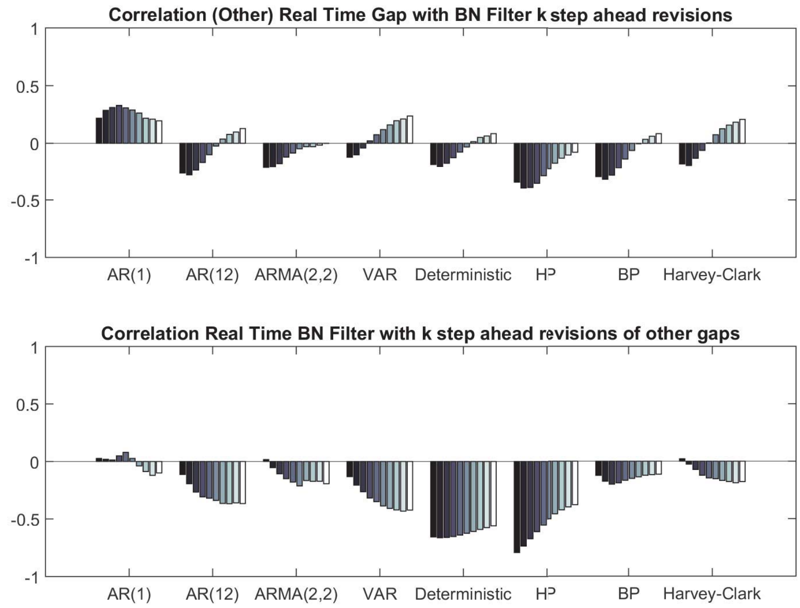
Figure 11: Relationship between pseudo-real-time estimates of the U.S. output gap and subsequent revisions across methods



Notes: See notes for Figure 5 for descriptions of labels of methods. The top panel presents the correlation of the *k*-step-ahead revision for the BN filter with the pseudo-real-time output gap estimate for each of the other methods. The bottom panel presents the correlation of the *k*-step-ahead revision for each of the other methods with the pseudo-real-time estimate for the BN filter. Each of the 10 bars indicates k = 1 to 10 step ahead reading off k = 1 from the leftmost to k=10 to the rightmost.

curate signals about the current state of the business cycle as determined by the retrospective evaluation of the NBER. The top panel of Figure 10 plots the revised CFNAI and the revised estimate for the BN filter. Despite our proposed approach being based only on the U.S. quarterly real GDP data series, it displays a remarkable similarity to the CFNAI, including in the correspondence of its movements with the NBER reference cycle. Meanwhile, the bottom panel presents correlations between the CFNAI and the various revised estimates of the output gap for different methods. The correlations confirm a strong positive relationship between the CF-NAI and the estimate based on the BN filter, with much weaker and sometimes even negative relationships for the other estimates.

Third, we consider whether future *k*-step ahead revisions of the output gap from alternative methods are correlated with the pseudo-real-time BN filter output gap and vice versa. The premise of this exercise is that, if the pseudo-real-time output gap from the BN filter is highly correlated with the true output gap, while other methods are less accurate but adjust towards the true output gap with their revisions, then the revisions should be correlated with the pseudo-real-time output gap from the BN filter. We also do the exercise in reverse as the same logic holds in the other direction. If revisions for the BN filter are correlated with the pseudo-real-time estimates for other methods, then this suggests the other estimates are more accurate. The results of this exercise are presented in Figure 11. We present correlations for $k = 1$ to 10 steps ahead. As the top panel reveals, the correlation of the k-step ahead revision of the BN filter never exceeds ± 0.4 with any of the other pseudo-real-time estimates, suggesting that the relatively small revisions for BN filter could not have been easily predicted by the pseudo-real-time estimates for the other methods. By contrast, in the bottom panel, the large negative correlations suggest that the other methods initially overestimate movements in the output gap relative to the BN filter, but adjust back towards the BN filter estimate with revisions.

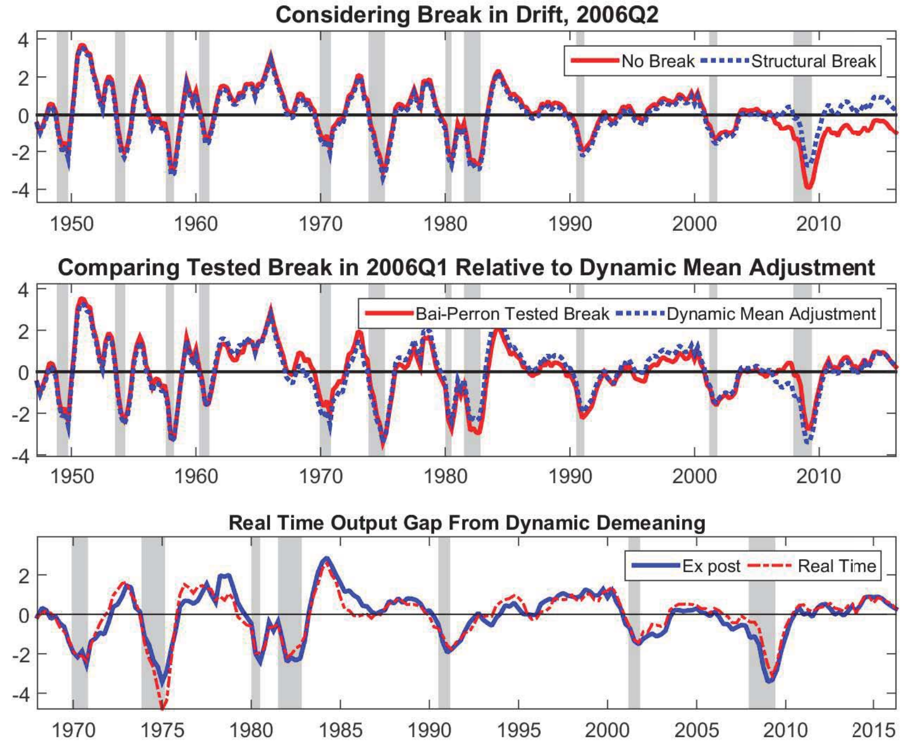
Overall, our analysis suggests that the large revisions for other trend-cycle decomposition methods are not capturing much about the true output gap that is not already captured by the BN filter. Also, our results suggest that our approach provides a convenient way to measure economic slack in that it appears to provide a shortcut to a large-scale multivariate approach that would also lead to a lower signal-to-noise ratio, while avoiding the overfitting and instability issues that inevitably arise with such multivariate analysis.

4 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 from the BN decomposition. More generally, Perron and Wada (2009, 2016) argue that estimates of

Figure 12: 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 3 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.

the output gap from different methods can be highly dependent on accounting for structural breaks. Therefore, we consider the effect of structural breaks on our estimates and propose a modification of the BN filter 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.[28] 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. The estimated

[28]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.

output gap allowing for a break in 2006Q1 is shown in the top panel of Figure 12.[29] Because we adjust for a slowdown in the long-run growth rate, the estimated output gap turns positive around 2010 and remains slightly positive from then on. An interesting 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 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.[30] 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 issues, we propose a slight modification to the BN filter 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.[31] That is, we construct deviations from mean as follows:

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

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 equation (3) for the dynamically demeaned data. This procedure loses some precision in the estimate of the mean compared to knowing the exact break date. But it is robust to multiple or gradual breaks and it

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

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

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

can be easily applied in real time.[32]

In the middle panel of Figure 12, 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 output gap estimate when allowing for a break detected by Bai and Perron (2003) procedures.[33] Meanwhile, in the bottom panel of Figure 12, 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.[34] Figure 13 plots the estimated gaps, reporting the signal-to-noise ratio that maximizes the amplitude-to-noise ratio for the output gap and the implied Okun's Law coefficient that estimates 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 visual comparison with the output gap.

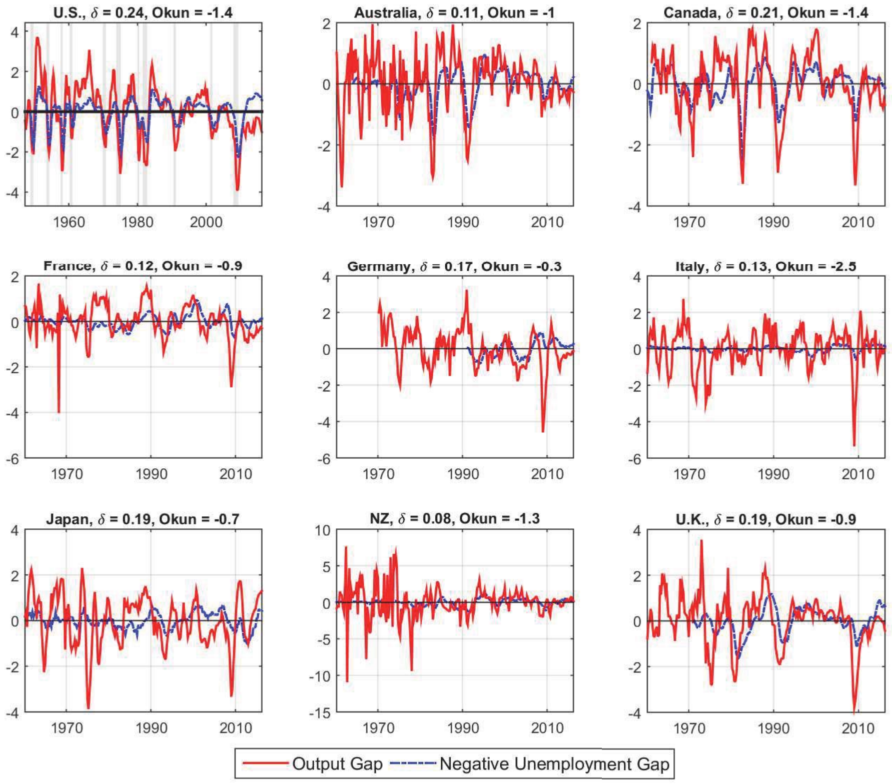
For the U.S. data, the estimated unemployment gap is highly (negatively) correlated with estimated output gap. When we regress the output gap against the unemployment gap, we obtain a coefficient of -1.4, which is slight lower than the consensus estimate of Okun's Law, but in agreement with a comparable analysis by Sinclair (2009), who estimates output and unemployment gaps in 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.

[32] For the first 40 quarters of the sample, we use the average growth rate for that 10 year period. However, by the start of our calculation of pseudo-real-time estimates in 1968Q1, the estimate of the mean is completely backward looking.

[33] 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.

[34] The data are sourced from the IMF Outlook.

Figure 13: 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 slope of Okun's law coefficient implying the percentage point change in the output gap per percentage point change in the unemployment gap. Shaded bars correspond to NBER recession dates.

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 in Figure 13, 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 not just for the United States, but for other countries as well.

5 Conclusion

In this paper, we have proposed a modification of the BN decomposition to directly impose a low signal-to-noise ratio. In particular, rather than focusing solely on model fit by freely estimating a time series forecasting model, we 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 to 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 apparent 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 relatively 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 motivate why it can be useful to impose a low signal-to-noise ratio. In particular, if the true state of the world is one in which there is an unobserved output gap that is large in amplitude and persistent, other methods 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 similarly to estimated output gap from the BN decomposition based on a freely estimated AR(1) model and better than for other methods, especially those that also impose a low signal-to-noise ratio. The revised estimate from the BN filter also appears to be more accurate than more heavily-revised output gap estimates in terms of its relationships with inflation and a well-known revised measure of U.S. economic activity, the Chicago Fed’s National Activity Index, that is constructed using a large number of economic variables. Furthermore, the real-time estimates from the BN filter are 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.

Finally, we show how to account for potential structural breaks in long-run growth rates when using the BN filter 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.

Berge TJ, Jordà O. 2011. Evaluating the classification of economic activity into recessions and expansions. *American Economic Journal: Macroeconomics* **3**: 246–277.

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

Edge RM, Rudd JB. 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. 2016. Is business cycle asymmetry intrinsic in industrialized economies? UNSW Business School Research Paper 2016-12, 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. 2008. Phillips curve inflation forecasts. NBER Working Papers 14322, National Bureau of Economic Research, Inc.

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

Online Appendix to *"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

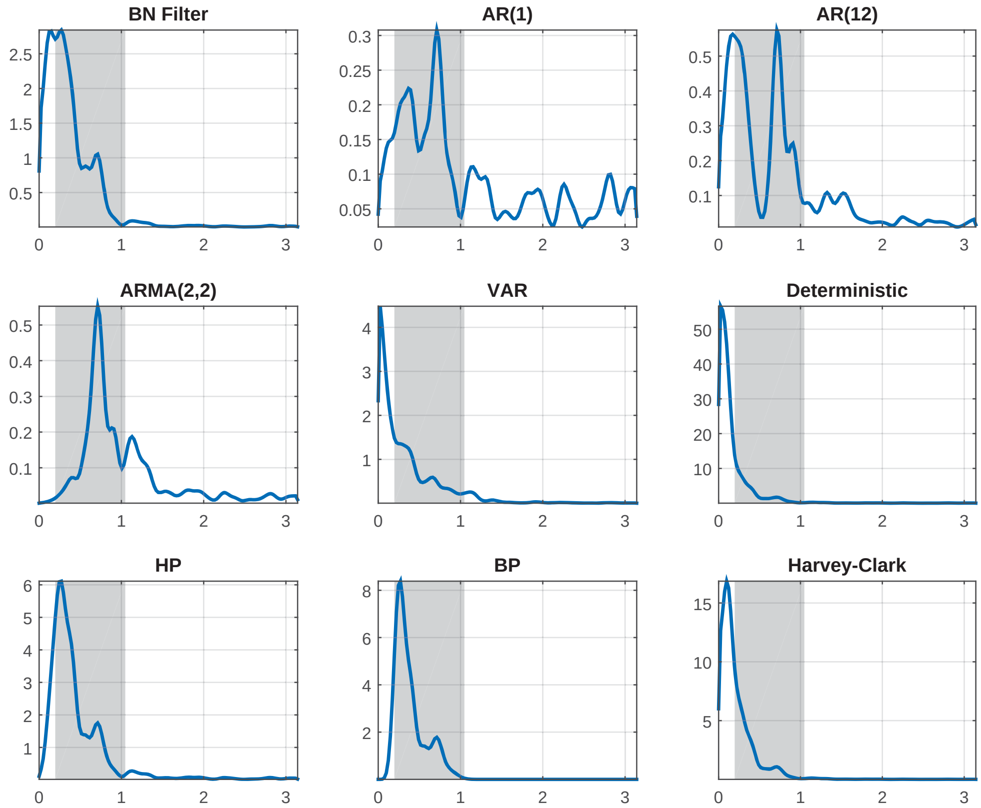
*The views expressed in this paper are those of the authors and do not necessarily represent the views of the Reserve Bank of New Zealand or the Bank for International Settlements. Kamber: gunes.kamber@bis.org Morley: james.morley@unsw.edu.au Wong: benjamin.wong@rbnz.govt.nz

A1 Sample Periodograms

Figure A1 plots the sample periodograms for the various ex post estimates of the U.S. output gap displayed in Figure 5 of the main text.[1] It is clear from the sample periodograms that the estimated output gap from the BN filter highlights frequencies that are similar to those highlighted by the HP and BP filters. The BN decompositions based on AR(1), AR(12), and ARMA(2,2) models appear to highlight higher frequencies (i.e., those with periods of less than 6 quarters), while the BN decomposition based on the VAR model, the deterministic detrending, and the Harvey-Clark UC model appear to highlight lower frequencies (i.e., those with periods greater than 32 quarters). Because the HP and BP filters are designed to highlight movements at business cycle frequencies, the similarity of the highlighted frequencies for the BN filter with those for the HP and BP filters suggest the BN filter is successful at capturing transitory movements in log real GDP at business cycle frequencies.

[1]For our calculations, we use the default periodogram function in Matlab.

Figure A1: Sample periodograms for estimated U.S. output gaps



Notes: y-axis units are spectral power and x-axis units are frequency. Shaded bars correspond to frequencies with periods between 6 and 32 quarters. "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. "AR(1)", "AR(12)", and "ARMA(2,2)" refer to BN decompositions based on the respective models estimated via MLE. "VAR" 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).

A2 Asymptotic Bias from Maximizing the Amplitude-to-Noise Ratio

We consider Monte Carlo simulation to investigate the asymptotic bias induced by selecting the signal-to-noise ratio based on maximizing the amplitude-to-noise ratio rather than freely estimating the signal-to-noise ratio. To do this, we consider the value of $\bar{\delta}$ chosen by our automatic selection procedure given T=500,000. By looking at such a large sample size, we are able to see whether our procedure really corresponds to a "dogmatic" prior in the sense of $\bar{\delta}$ being inconsistent for the true δ. We also isolate the impact of our procedure from any small sample biases that could occur in freely estimating the signal-to-noise ratio.

For our DGP, we use an AR(2) process:

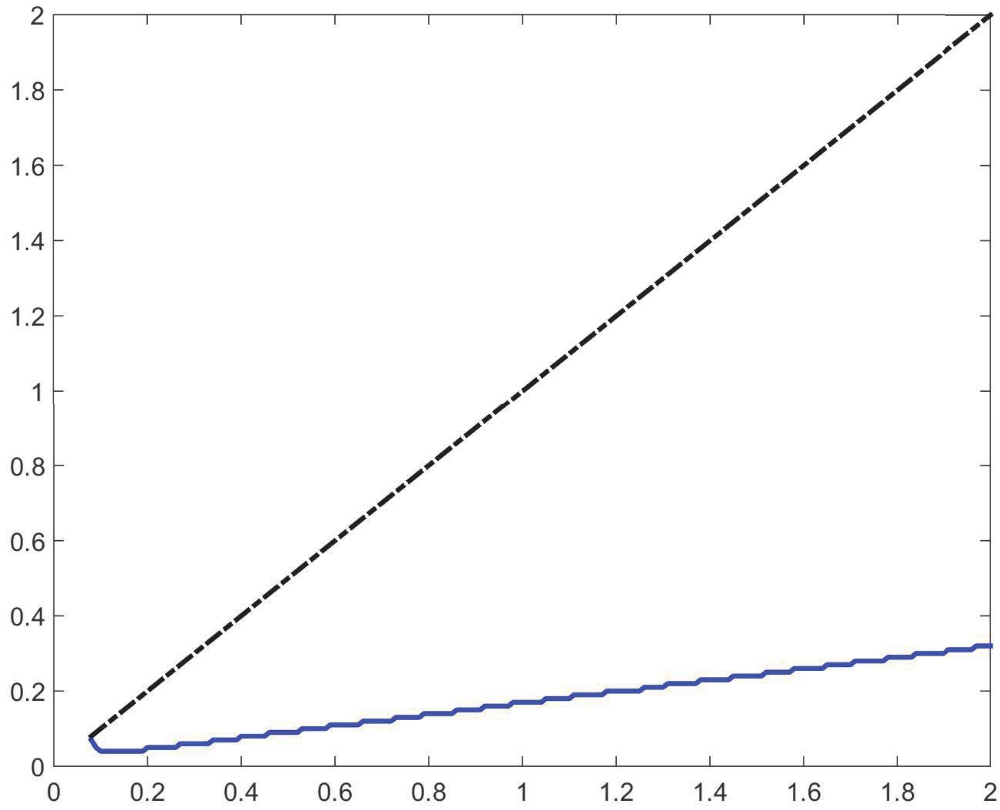
$$\Delta\tilde{y}_t = \rho\Delta\tilde{y}_{t-1} + \phi^*\Delta^2\tilde{y}_{t-1} + e_t, \tag{A.1}$$

where $e_t \sim iidN(0,1)$. An AR(2) process allows us to consider low values for the true signal-to-noise ratio δ. In particular, we fix $\phi^* = 0.8$ and adjust the value of ρ such that the true signal-to-noise ratio δ varies between 0.08 (the lowest value associated with a stationary process) and 2.00 in increments of 0.01. For each true δ, we simulate a time series with 500,000 observations and apply the BN filter to the simulated time series. This produces a corresponding $\bar{\delta}$ based on maximizing the amplitude-to-noise ratio.

Figure A2 plots the resulting $\bar{\delta}$ against the true δ for each simulation. The $\bar{\delta}$ chosen by our procedure is always below the 45 degree line, corresponding to a downward asymptotic bias in calculating the signal-to-noise ratio when selecting it based on maximizing the amplitude-to-noise ratio. The lack of consistency of our procedure in calculating the signal-to-noise ratio is what leads us to think of the BN filter as imposing a dogmatic prior. However, it is notable that this bias gets smaller as the true signal-to-noise ratio gets smaller and almost disappears when the true $\delta = 0.08$, which is very close to the boundary of the stationarity region for this AR(2) DGP with $\phi^* = 0.8$.[2] Thus, our approach to imposing a low signal-to-noise ratio can be reasonable when the true signal-to-noise ratio is low. Related, the Monte Carlo results in Section 3.1 of the main text suggest that, even when the AR(12) model used in the BN filter is misspecified, the downward asymptotic bias in calculating $\bar{\delta}$ is not too severe for a moderately low value of $\delta = 0.50$ and still leads to comparatively accurate inferences about the true cycle.

[2] We find that the result of almost no asymptotic bias when the true δ is very close to the the boundary of the stationarity region for the DGP is robust to different DGPs (e.g., it also holds for an AR(1) process, for which $\delta > 0.25$ is necessary for stationarity). A full investigation of this boundary effect would provide an interesting avenue for future research.

Figure A2: Monte Carlo simulation results for selecting the signal-to-noise ratio based on maximizing the amplitude-to-noise ratio



Notes: The solid line corresponds to the signal-to-noise ratio $\bar{\delta}$ on the y-axis that maximizes the amplitude-to-noise ratio when applying the BN filter for each true signal-to-noise ratio δ on the x-axis given an AR(2) DGP. The dashed line is the 45 degree line, along which the selected signal-to-noise ratio would be equal the true signal-to-noise ratio (i.e., $\bar{\delta} = \delta$). Values below the 45 degree line correspond to a downward asymptotic bias.

A3 Confidence Bands

Here we present the details for how we calculate confidence bands for a BN cycle. We then report the 95% confidence bands for all of the estimates of the U.S. output gap based on a BN decomposition considered in the main text. For comparison, we also report 95% confidence bands for estimates based on deterministic detrending using the standard error of the regression and for the Harvey-Clark UC model using the Kalman smoother. Because they are not model-based methods, we do not report confidence bands for estimates based on the HP and BP filters.

To develop our approach to calculating confidence bands for a BN cycle, first recall that, following Morley (2002), the BN cycle, c_t^{BN}, at time t for an AR(p) model is

$$c_t^{BN} = -[1 \quad 0 \quad \cdots \quad 0]F(I-F)^{-1}X_t, \tag{A.2}$$

where $X_t = (\Delta\tilde{y}_t, \Delta\tilde{y}_{t-1}, ..., \Delta\tilde{y}_{t-p+1})'$. Then it is straightforward to solve for the variance of the BN cycle, σ_c^2, as follows:

$$\sigma_c^2 = [1 \quad 0 \quad \cdots \quad 0]F(I-F)^{-1}\Sigma_X((I-F)')^{-1}F'[1 \quad 0 \quad \cdots \quad 0]', \tag{A.3}$$

where Σ_X is the variance of X_t and $vec(\Sigma_X) = (I - F \otimes F)^{-1}vec(Q)$, with Q being the variance-covariance matrix for the innovation vector of the companion form for the AR(p) model:

$$Q = \begin{bmatrix} \sigma_e^2 & 0 & \cdots \\ 0 & 0 & \ddots \\ \vdots & \ddots & \ddots \end{bmatrix}.$$

For the purpose of calculating confidence bands, we treat c_t^{BN} as an estimator of the true transitory component, c_t, of some unobserved components process $\{y_t\}$ with a trend that follows a random walk plus drift.[3] From the definition of the BN decomposition and assuming the reduced-form dynamics for $\{\Delta y_t\}$ can be fully captured by an AR(p) model, c_t^{BN} provides an unbiased estimator of the true c_t because τ_t^{BN} is an unbiased estimator of τ_t. Then, noting that the BN cycle in equation (A.2) inherits the assumed normality of $\Delta\tilde{y}_t$ and using the variance in equation (A.3), we can construct 95% confidence bands for the BN cycle by inverting a simple z-test that c_t is equal to a hypothesized fixed value c_t^0 as follows:

$$c_t^{BN} \pm 1.96\sigma_c. \tag{A.4}$$

[3]See Morley (2011) for a discussion of the distinction between interpreting the BN cycle as an estimator of the transitory component of an unobserved components process or as the transitory component of a single-source-of-error process. If it were the latter, then the mean squared error of the BN cycle as an estimator would be zero and there would be no uncertainty about the true transitory component (beyond any finite-sample model parameter uncertainty, which we have abstracted from for simplicity).

This is the approach we take in Figure 3 of the main text using implied estimates for F and Σ_X in equations (A.2) and (A.3). Meanwhile, when calculating confidence bands for the BN cycle based on other forecasting models, the approach is essentially the same, but the structure and/or elements of the state-space/companion form matrices will be different.

In our paper, we consider the BN decomposition to avoid having to take a stand on the particular underlying unobserved components process (or possibly single-source-of-error process) that leads to reduced-form AR(p) dynamics for $\{\Delta y_t\}$. It would be necessary to assume an exact specification of the underlying unobserved components process $\{y_t\}$ in order to construct confidence bands based on the mean squared error of the estimator of the transitory component c_t under the more standard assumption that it follows a stochastic process rather than treating it as a hypothesized fixed value c_t^0. Given an exact specification, one could, in principle, apply the Kalman filter to calculate the confidence bands based on the variance of the filtered estimator of the transitory component, which, following Morley et al. (2003), is the same as c_t^{BN} for the corresponding reduced-form forecasting model. However, a reduced-form AR(p) model for $\{\Delta y_t\}$ does not correspond to an identified unobserved components process with a transitory component that can be captured by a finite-order ARMA specification (see the discussion in Morley and Piger (2008)). Thus, this approach is not available in practice when dealing with an AR(p) forecasting model and we propose using the approach detailed here instead.[4]

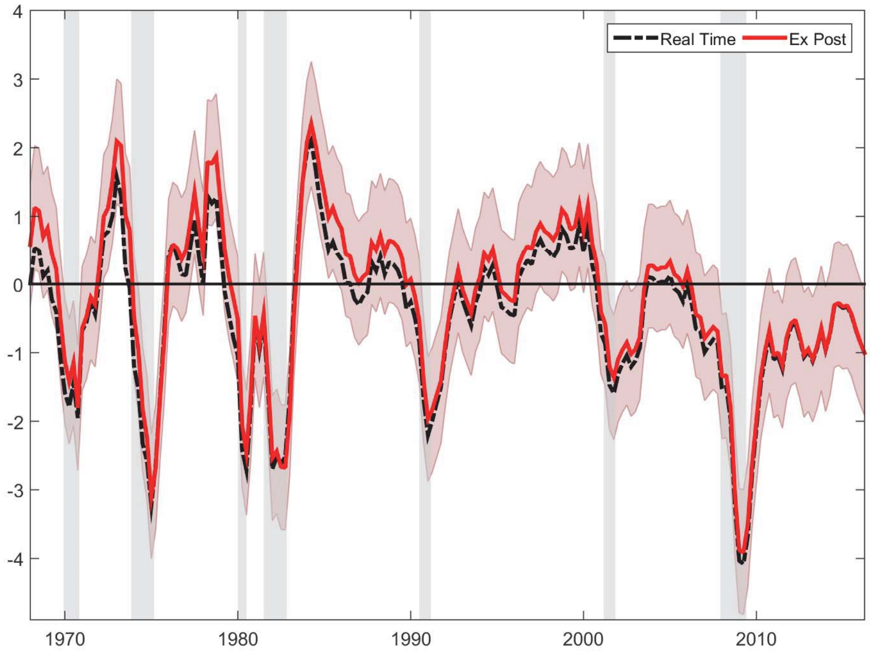
In terms of the U.S. output gap, Figure A3 plots the pseudo-real-time and ex post point estimates along with the ex post 95% confidence bands for the BN filter, the other BN decompositions, and deterministic detrending and the Harvey-Clark UC model. Figure A4 presents the proportion of output gap estimates that are statistically different from zero and the proportion of pseudo-real-time estimates that lie outside the ex post 95% confidence bands.[5] The BN filter is the most precise method in the sense of the ex post estimates being statistically different than zero more often than for the other methods, yet it is also the most accurate in the sense that its pseudo-real-time estimates always

[4]The Kalman filter approach is available when dealing with an ARMA(2,2) forecasting model. We find that the bands based on our proposed approach here are much closer to the bands from the Kalman filter for an unobserved components model of U.S. log real GDP with an AR(2) transitory component and correlated movements in trend and cycle that produces the same filtered estimates as the BN cycle for the ARMA(2,2) model, as in Morley et al. (2003), than to the implied perfect certainty about the true transitory component for a corresponding single-source-of-error process. In particular, our bands are about three-quarters the width of those from the Kalman filter. The relative narrowness of our bands compared to those from the Kalman filter potentially make them less accurate (definitely so if the truth were the unobserved components process in Morley et al. (2003)), but it also makes the inferences relatively more precise. However, even with our narrower bands, we almost never reject zero for the BN cycle based on an ARMA(2,2) model, while the pseudo-real-time estimates appear to be accurate in the sense of always remaining within the 95% bands on the ex post estimate.

[5]We only count the proportion of statistically significant estimates and pseudo-real-time estimates within the bands from 1970Q1-2012Q4 in order to match the evaluation sample for revision statistics in the main text.

Figure A3: Pseudo-real-time and ex post U.S. output gap estimates with 95% confidence bands

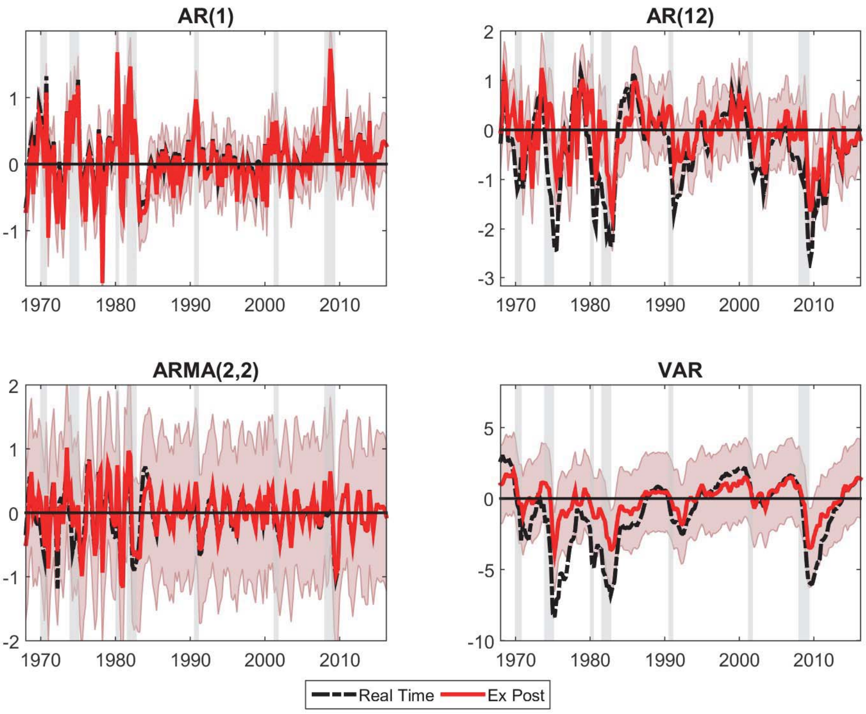
(a) BN filter



Notes: Units are 100 times natural log deviation from trend. Shaded bands around the ex post 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.

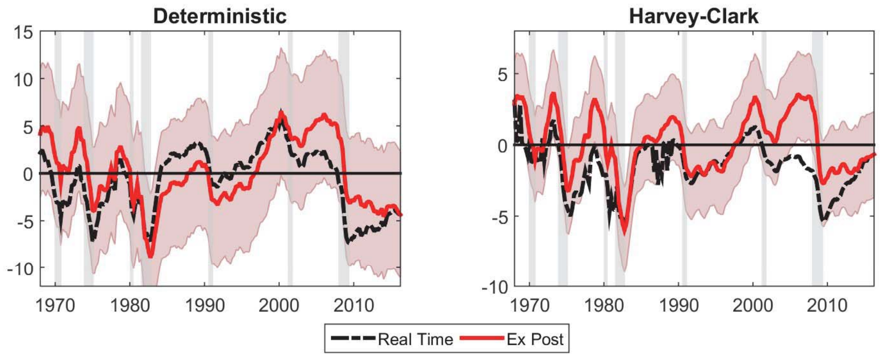
lie within ex post 95% bands. The BN decompositions based on the AR(12) and VAR models are particularly inaccurate, while still being less precise than the BN filter. The BN decomposition based on the AR(1) model is almost as precise as the BN filter and is accurate, while the BN decomposition based on the ARMA(2,2) model is not at all precise, although it is accurate. In terms of the other methods, the Harvey-Clark UC model produces inaccurate and relatively imprecise estimates, while deterministic detrending produces more accurate, but essentially uninformative estimates.

(b) BN decompositions for different forecasting models



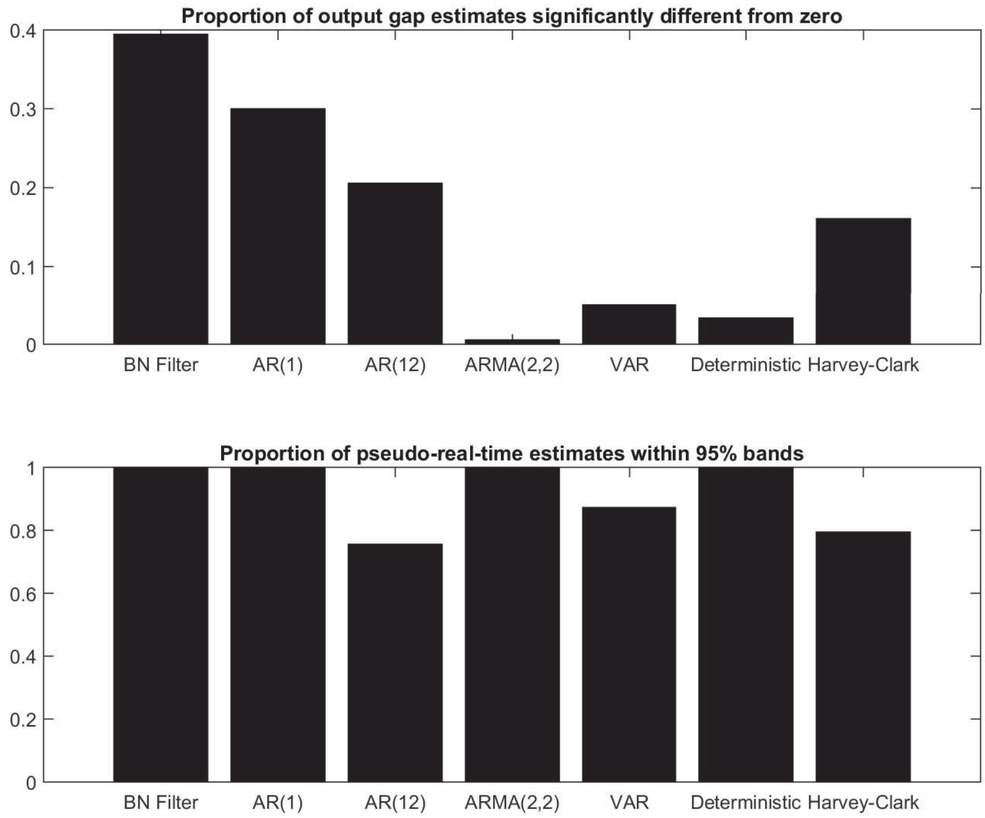
Notes: Units are 100 times natural log deviation from trend. See notes for Figure A1 for descriptions of labels of methods. Shaded bands around the ex post estimate correspond to a 95% confidence interval from inverting a z-test that the true output gap is equal to a hypothesized value using the standard deviation of the BN cycle. Shaded bars correspond to NBER recession dates.

(c) Some other popular methods



Notes: Units are 100 times natural log deviation from trend. See notes for Figure A1 for descriptions of labels of methods. Shaded bands around the ex post estimate correspond to a 95% confidence interval based on the standard error of the regression for deterministic detrending and the Kalman smoother for the Havey-Clark UC model. Shaded bars correspond to NBER recession dates.

Figure A4: Precision and accuracy of U.S. output gap estimates



Notes: See notes for Figure A1 for descriptions of labels of methods. Precision is in terms of frequency with which ex post estimates are significantly different than zero. Accuracy is in terms of frequency with which pseudo-real-time estimates lie within the ex post 95% confidence bands.

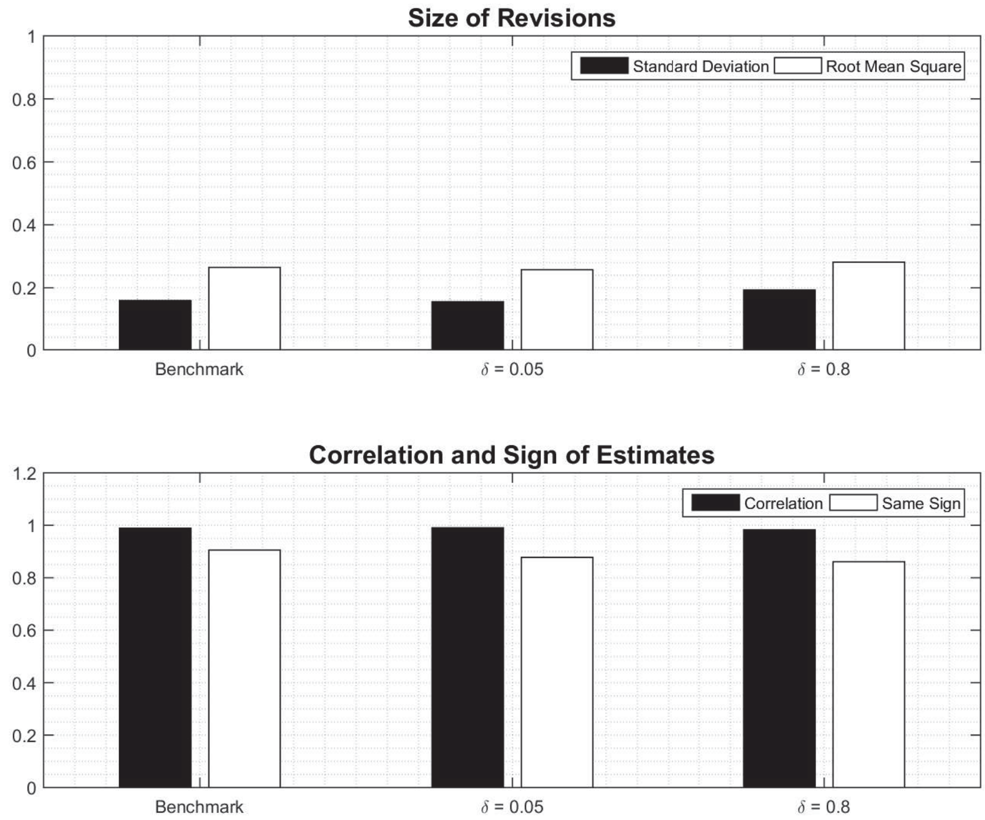
A4 Robustness to Other Low Values of δ

In the main text, we showed that changing δ to other low values affects the amplitude but not the shape of the estimated output gap for our proposed approach. In Figure A5, we verify that revision properties are robust to different low values of δ. In particular, we recalculate all of the revision statistics for $\delta \in \{0.05, 0.8\}$ and compare them to our benchmark case in which we chose δ in pseudo-real-time by maximizing the amplitude-to-noise ratio. Recall $\bar{\delta} = 0.24$ for the full sample and ranged between 0.21-0.26 in the pseudo-real-time analysis. The evaluation period is the same as in the main text. As can be seen in Figure A5, the revision statistics are very similar to the benchmark case.

We also redo the pseudo-out-of-sample forecast comparisons, once again for $\delta \in \{0.05, 0.8\}$. The sample period for forecast evaluation is kept the same as in the main text. Figures A6 and A7 present the output growth and inflation forecasting results in terms of relative root mean square errors (RRMSE), with the comparison being relative to the BN filter. As expected, because the shape of the estimated output gap is robust to different low values of δ, the forecast performance is very similar, with the relative root mean square error relative to the benchmark case extremely close to 1.

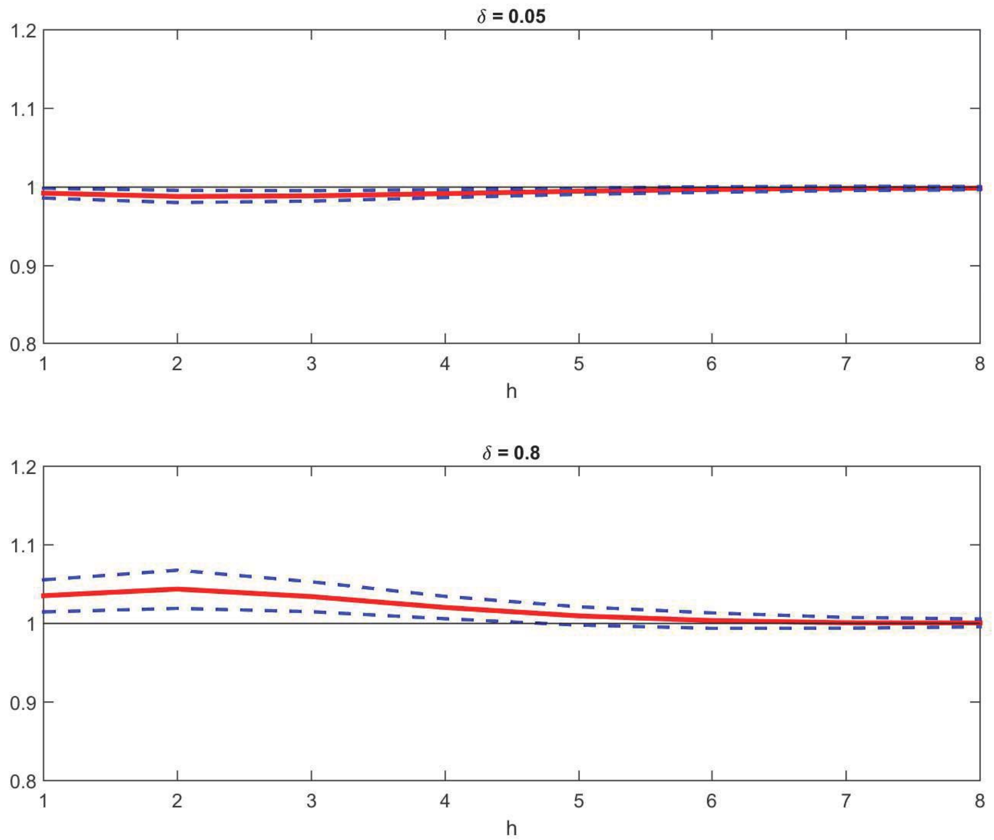
Based on these results, we conclude that our main findings are robust to different choices of a low value for δ. The only substantive effect of changing δ is an alteration in the amplitude of the estimated output gap, with revision properties and forecasting performance almost identical to our benchmark case.

Figure A5: Revision statistics for U.S. output gap estimates based on different signal-to-noise ratios



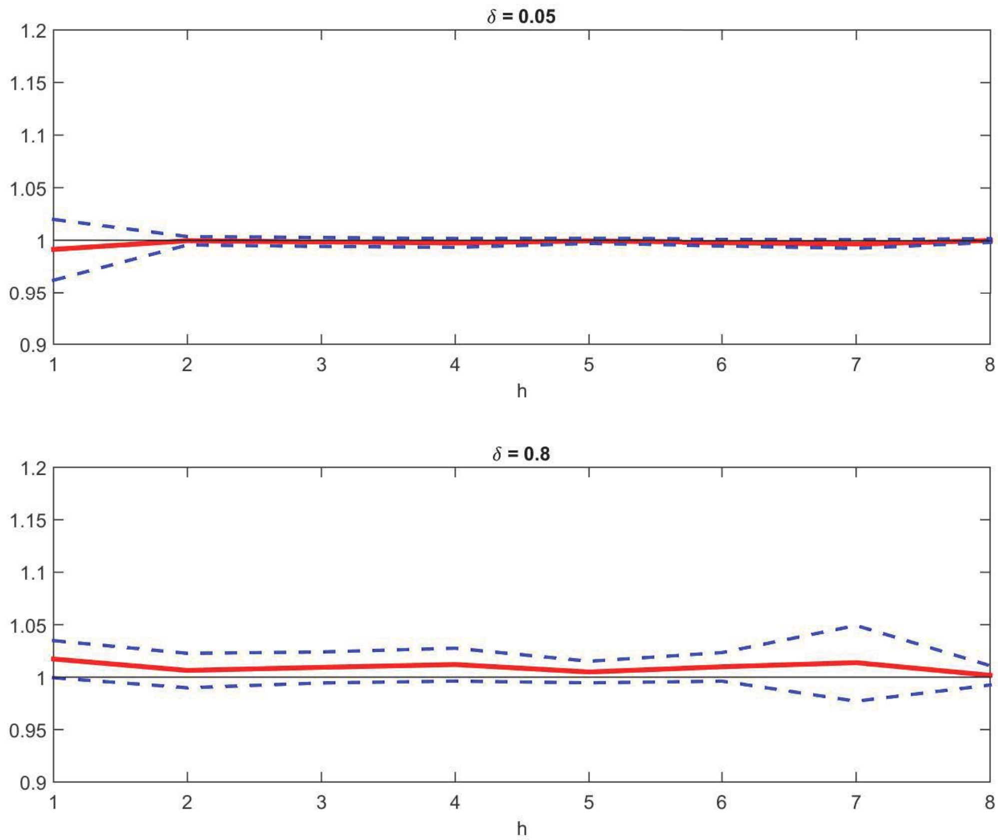
Notes: The benchmark selects $\bar{\delta}$ by maximizing the amplitude-to-noise ratio. For the full sample, $\bar{\delta} = 0.24$, while $\bar{\delta}$ varies between 0.21 - 0.26 in pseudo real time. Standard deviation and root mean square of revisions to the pseudo-real-time estimate of the output 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 A6: Out-of-sample U.S. output growth forecast comparisons for different signal-to-noise ratios relative to the BN filter benchmark using pseudo-real-time output gap estimates



Notes: The graphs plot out-of-sample RRMSE from a BN decomposition with $\delta \in \{0.05, 0.8\}$ compared to a BN filter based on maximizing the amplitude-to-noise ratio in pseudo-real-time. 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 A7: Out-of-sample U.S. inflation forecast comparisons for different signal-to-noise ratios relative to the BN filter benchmark using pseudo-real-time output gap estimates



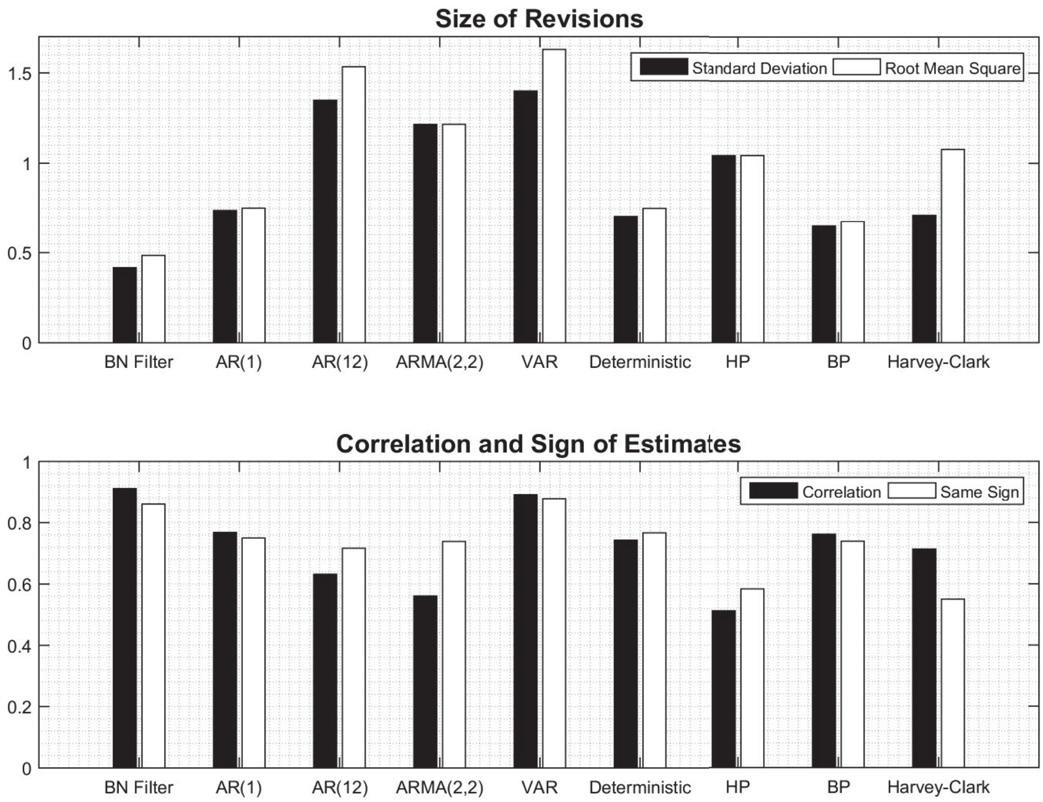
Notes: The graphs plot out-of-sample RRMSE from a BN decomposition with $\delta \in \{0.05, 0.8\}$ *compared to a BN filter based on maximizing the amplitude-to-noise ratio in pseudo-real-time. Out-of-sample evaluation begins in 1970Q1. The bands depict 90% confidence intervals from a two-sided Diebold and Mariano (1995) test of equal forecast accuracy.*

A5 Results Using Real Time Data

As Orphanides and van Norden (2002) demonstrate, most of the revisions to output gap estimates are due to the filtering method rather than the data revisions to real GDP. However, it is an open question whether our main findings still hold when working with real time data. Thus, here, we redo our analysis of revision statistics and out-of-sample forecasting performance using real time data. All real time vintages are from the Philadelphia Fed's real time datasets. Figure A8 reports the revision statistics and Figures A9 and A10 report the out-of-sample forecasting results. Once again, the sample period under evaluation matches that of the analysis in the main text.

With regards to revision statistics, the BN Filter does much better in comparison to other methods with real time data relative to the pseudo-real-time case. Most notably, the BN filter's revisions are now much smaller compared to the BN decomposition based on an AR(1) model. A key reason for this is the output gap estimate obtained from the BN decomposition based on an AR(1) model lacks persistence and depends on only the recent output growth, meaning large data revisions in recent output growth will have considerable influence on the real time estimate. Our approach continues to do better than other approaches. The results of the out-of-sample forecasting exercise are very similar to or slightly better than those obtained using pseudo-real-time estimates. We therefore conclude that our main findings are robust to the use of real time data.

Figure A8: Revision statistics for U.S. output gap estimates using real time data



Notes: See notes for Figure A1 for descriptions of labels of methods. Standard deviation and root mean square of revisions to the 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 real-time estimate and the ex post estimate of the output gap. "Same sign" refers to the proportion of 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.