Crawford School of Public Policy

CAMA

Centre for Applied Macroeconomic Analysis

Trend-Cycle Decomposition in the Presence of Large Shocks

CAMA Working Paper 24/2024
March 2024 (Revised in August 2024)

Güneş Kamber
International Monetary Fund

James Morley
The University of Sydney
Centre for Applied Macroeconomic Analysis, ANU

Benjamin Wong
Monash University
Centre for Applied Macroeconomic Analysis, ANU

Abstract

We introduce some refinements of the Beveridge-Nelson filter to produce more intuitive estimates of the output gap and address possible distortions from large shocks. We then compare how the Beveridge-Nelson filter and other popular univariate trend-cycle decomposition methods performed given the extreme outliers in the Covid-era data. Real-time estimates of the output gap based on the Hodrick-Prescott filter turn out to have been highly unreliable in the years just prior to the pandemic, although the estimated gap during the pandemic is similar to that of the more reliable Beveridge-Nelson filter. The Hamilton filter suffers from base effects that produce a mechanical spike in the estimated output gap exactly two years after the onset of the pandemic, in line with the filter horizon. Given projected data that includes a simulated Covid-like shock, both the Hodrick-Prescott and Hamilton filters overstate the true reduction in the output gap and fail to capture the implied movements in trend output. The Hodrick-Prescott filter generates a spurious transitory boom prior to the shock, while the Hamilton filter produces another mechanical spike two years after the shock and also an ongoing divergence in forecasted values of the output gap away from zero. Only the Beveridge-Nelson filter correctly forecasts trend and cycle movements when faced with a Covid-like shock.

Keywords

Beveridge-Nelson decomposition, output gap, real-time reliability

JEL Classification

C18, E17, E32

Address for correspondence:

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

ISSN 2206-0332

Trend-cycle decomposition in the presence of large shocks*†

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

[1]International Monetary Fund
[2]The University of Sydney
[3]Monash University, Australia
[4]CAMA, ANU

August 20, 2024

Abstract

We introduce some refinements of the Beveridge-Nelson filter to produce more intuitive estimates of the output gap and address possible distortions from large shocks. We then compare how the Beveridge-Nelson filter and other popular univariate trend-cycle decomposition methods performed given the extreme outliers in the Covid-era data. Real-time estimates of the output gap based on the Hodrick-Prescott filter turn out to have been highly unreliable in the years just prior to the pandemic, although the estimated gap during the pandemic is similar to that of the more reliable Beveridge-Nelson filter. The Hamilton filter suffers from base effects that produce a mechanical spike in the estimated output gap exactly two years after the onset of the pandemic, in line with the filter horizon. Given projected data that includes a simulated Covid-like shock, both the Hodrick-Prescott and Hamilton filters overstate the true reduction in the output gap and fail to capture the implied movements in trend output. The Hodrick-Prescott filter generates a spurious transitory boom prior to the shock, while the Hamilton filter produces another mechanical spike two years after the shock and also an ongoing divergence in forecasted values of the output gap away from zero. Only the Beveridge-Nelson filter correctly forecasts trend and cycle movements when faced with a Covid-like shock.

JEL Classification: C18, E17, E32
Keywords: Beveridge-Nelson decomposition, output gap, real-time reliability

---

*Email: Kamber: gkamber@imf.org Morley: james.morley@sydney.edu.au Wong: benjamin.wong@monash.edu

†We thank conference participants at the Africa Meeting of the Econometric Society in Abidjan and the Dolomiti Macro Meeting in Castelrotto for valuable comments. The BN filter described in this paper can be implemented for any time series at `https://bnfiltering.com` and we thank Cristian Corrado for his excellent work in developing the website. We also thank Abby Nguyen for helpful research assistance. Morley and Wong acknowledge funding support from the Australian Research Council (DP190100202, DP240100970 and DE200100693). The views expressed in this paper are those of the authors and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

1 Introduction

The Beveridge-Nelson (BN) filter approach to trend-cycle decomposition from Kamber, Morley, and Wong (2018) produces intuitive and reliable estimates of the output gap. This intuitiveness can be thought of in terms of what Rudd (2024) refers to as the BN filter's "respectable business cycle" that displays a reasonable coherence with the output gap implied by the Congressional Budget Office (CBO) measure of potential output and a strong positive association with the National Bureau of Economic Research (NBER) reference cycle, unlike what is found using the traditional BN decomposition based on freely-estimated univariate ARMA models following Beveridge and Nelson (1981). Reliability can be evaluated in different ways, with good revision properties providing a necessary condition for reliable estimates, as highlighted in Orphanides and van Norden (2002). Beyond the initial evidence presented in Kamber, Morley, and Wong (2018) that BN filter estimates i) have relatively small revisions, ii) provide a more accurate recovery of the true output gap for state-space processes than even a time series model that nests the true process, and iii) forecast output growth better out of sample than other approaches such as the Hodrick and Prescott (1997) (HP) filter and the Christiano and Fitzgerald (2003) bandpass filter, there have been a number of subsequent studies demonstrating the comparative reliability of the BN filter, including Barbarino, Berge, and Stella (2024), Jönsson (2024), and Kuang, Mitra, and Tang (2024).[1]

However, when faced with large shocks in macroeconomic data such as occurred during the COVID-19 pandemic, there is a question whether the BN filter retains its favorable real-time reliability documented in the previous studies. These extreme outliers have created a need to modify many standard time series techniques, such as how to estimate vector autoregressions with Covid-era data in Lenza and Primiceri (2022) or conduct multivariate

[1] Barbarino, Berge, and Stella (2024) conduct a comprehensive analysis of real-time stability for a wide range of approaches to estimating the output gap and find that the BN filter outperforms the HP filter, the Hamilton (2018) filter, CBO estimates, and the Federal Reserve's Tealbook estimates in terms of producing small revisions in the form of low "noise-to-signal" ratios, as well as more accurate forecasts of inflation, output growth, and the unemployment rate for multiple forecasting horizons and subsamples. Rather than focus on a particular sample of data, Jönsson (2024) considers the frequentist revision properties given the same state-space processes in Kamber, Morley, and Wong (2018) and finds the BN filter performs best of all methods considered, including the HP filter and the bandpass filter. Kuang, Mitra, and Tang (2024) find that using the BN filter would help central banks minimize policy mistakes in a theory-based New Keynesian setting with adaptive learning. Specifically, they show that the BN filter provides the most reliable estimates in terms of revision properties and is almost as good as the bandpass filter (but even better when considering a model with an inflation target of 4% instead of 2%) and better than other methods including the HP and Hamilton filters in terms of stability over a wide range of policy parameters and best or equal in terms of minimizing welfare loss with the original Taylor rule coefficients or optimally-chosen coefficients.

trend-cycle decomposition in Holston, Laubach, and Williams (2023) and Morley, Rodriguez-Palenzuela, Sun, and Wong (2023). In this paper, we introduce some modifications or "refinements" of the BN filter that are designed to be useful for practitioners and to address possible distortions from large shocks. The four refinements are i) a new trend-smoothness loss function to select a signal-to-noise ratio when detrending, ii) an iterative dynamic demeaning approach when estimating time-varying drift, iii) error bands that allow for time-varying volatility, and iv) backcasting based on time reversibility to produce an estimated cycle for the full sample of the variable being detrended. We then compare how the Beveridge-Nelson filter and other popular univariate trend-cycle decomposition methods have performed given the Covid-era data.

Our first main finding is that, perhaps surprisingly, both the original and refined versions of the BN filter had good revision properties during the pandemic, although they produce some different implications about the output gap. The refined BN filter output gap is closer in amplitude to the HP filter, with a smoother estimated trend than for the original BN filter around the COVID-19 recession and time-varying error bands reflecting heightened uncertainty about the output gap since the pandemic. Using our new trend-smoothness loss function, the selected signal-to-noise ratio corresponds to trend shocks accounting for only 2% of the forecast error variance instead the 24% found in Kamber, Morley, and Wong (2018) when maximizing the amplitude-to-noise ratio. Notably, this is closer to the 5% fixed level considered in Rudd (2024) and Kuang, Mitra, and Tang (2024) and is the primary reason why our estimates end up being even more intuitive than with the original BN filter. The iterative dynamic demeaning procedure for the refined BN filter produces more positive estimates of the output gap from the mid-2010s up to the pandemic that are more in line with the final-vintage HP filter estimates.

Our second main finding is that the real-time HP filter estimates of the output gap just prior to the pandemic turn out to be particularly unreliable. Specifically, the real-time HP filter estimates in 2018-2019 suggested that the economy was very close to trend, while the estimates for the same time period have been subsequently revised upwards closer to the more reliable estimates from the refined BN filter. Thus, the refined BN filter would have provided a better prediction of the final-vintage HP filter estimates than a one-sided HP filter. At the same time, the upward revisions for the HP filter estimates just prior to the pandemic may have been at least partly mechanical, as we demonstrate with projections of future output growth that include a large one-time simulated Covid-like shock and result in the HP filter estimating a spurious transitory boom prior to the shock.

As hinted at by the criticism of the revision properties of the HP filter in Hamilton (2018), the Hamilton filter is more reliable than the HP filter and this holds during the pandemic. However, our third main finding is that the Hamilton filter suffers from base effects that produce a mechanical spike in the estimated output gap exactly two years after the onset of the pandemic, in line with the filter horizon. This sudden jump in the estimated output gap in 2022 is unintuitive and strongly at odds with the estimates for the BN and HP filters. A modified Hamilton filter due to Quast and Wolters (2022) is, as advertised, even more reliable than the original and avoids such a large discrete jump in the estimated output gap in 2022 given that it averages over different horizons. However, both the original and modified Hamilton filters generate distorted output gap forecasts given the projections of future output growth that include a large one-time simulated Covid-like shock, including a divergence in forecasted values away from zero given projected values of future output growth at a constant level.

In addition to being the most reliable in real time, including with the Covid-era data, our fourth main finding is that the BN filter correctly forecasts trend and cycle movements in the projected data with a Covid-like shock. Given all of the results, we argue for the use of the refined BN filter in particular when seeking both intuitive and reliable estimates of the output gap in the presence of large shocks and also when considering data augmented with long-term projections, as is often done in policy settings.

The rest of this paper is organised as follows: First, we recap the details of the BN filter and introduce some refinements. Second, we consider the real-time reliability of the original and refined BN filters when faced with Covid-era data and compare revision properties to those of the HP filter and the original and modified Hamilton filters. Third, we look at what the different methods imply about forecasts of the output gap given a projection of future real GDP that reflects a plausible assumption about the long-run growth rate of the economy and also allows for a Covid-like shock. We end with some brief conclusions.

2 The BN filter with refinements

We begin with a summary of the original BN filter developed in Kamber, Morley, and Wong (2018) to motivate our proposed refinements.

2.1 The Kamber, Morley, and Wong (2018) BN Filter

In Kamber, Morley, and Wong (2018), we employed the definition of trend from Beveridge and Nelson (1981) as the long-horizon conditional expectation of a time series minus any *a priori* known (i.e., deterministic) future movements in the time series. Denoting $\{y_t\}$ as a time series with a trend component that follows a random walk with deterministic drift, the BN trend at time t, τ_t^{BN}, is

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

The basic intuition behind the BN decomposition is that the long-horizon conditional expectation of a time series is the same as the long-horizon conditional expectation of the trend component under the assumption that the conditional expectation of the remaining cyclical component goes to zero at long horizons. By removing the deterministic drift, $\mathbb{E}[\Delta y_t]$, the conditional expectation in (1) remains finite and becomes an optimal (minimum mean squared error) estimate of the current trend component (see Watson, 1986; Morley, Nelson, and Zivot, 2003).

The BN filter in Kamber, Morley, and Wong (2018) implements the BN decomposition by applying the definition in (1) given a restricted AR(p) forecasting model for $\{\Delta y_t\}$ to calculate conditional expectations, where the restriction allows the imposition of a "signal-to-noise" ratio to imply a smooth trend, similar to the HP filter. More precisely, we specify an AR(p) model:

$$\Delta y_t = \mu + \sum_{j=1}^{p} \phi_j (\Delta y_{t-j} - \mu) + e_t, \;\; e_t \sim N(0, \sigma_e^2), \tag{2}$$

where μ is equal to $\mathbb{E}[\Delta y_t]$ and, therefore, denotes the deterministic drift in $\{y_t\}$.[2] Modelling a time series as a finite-order AR process in first differences corresponds to a stochastic trend in the level because a forecast error for the first differences will have an implied *permanent*

[2]Our choice of exposition is in terms of the demeaned form of the AR(p) model rather than specifying an intercept form in order to make it clear where the drift entering into the calculation of the BN decomposition comes from. If one were to estimate an intercept, denoting c as the intercept, the drift would simply correspond to $c/(1 - \sum_i^p \phi_i)$.

effect on the long-horizon conditional expectation of $\{y_t\}$. Using the state-space approach to calculating the BN decomposition from Morley (2002), the BN cycle at time t, c_t^{BN}, for this model is

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

where $X_t = (\Delta y_t - \mu, \Delta y_{t-1} - \mu, ..., \Delta y_{t-p+1} - \mu)'$, and F is the companion matrix for the AR(p) model:

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

Defining the signal-to-noise ratio $\delta \equiv \sigma^2_{\Delta\tau}/\sigma^2_e$, where $\sigma^2_{\Delta\tau} \equiv var(\Delta\tau_t)$ is the implied variance of the change in trend and σ^2_e is the forecast error variance from (2), Kamber, Morley, and Wong (2018) note that δ maps directly to the coefficients of the AR(p) model, so that one can impose a signal-to-noise ratio for the model in (2) simply by setting a restriction on the AR coefficients. Specifically, we can transform the AR(p) model in (2) into its Dickey-Fuller representation:

$$\Delta y_t = \mu + \rho(\Delta y_{t-1} - \mu) + \sum_{j=1}^{p-1} \phi_j^*(\Delta^2 y_{t-j} - \mu) + e_t, \tag{4}$$

where $\rho \equiv \phi_1 + \phi_2 + \ldots + \phi_p$ and $\phi_j^* \equiv -(\phi_{j+1} + \ldots + \phi_p)$. Then, there is a direct mapping from δ to the sum of the AR coefficients ρ in the Dickey-Fuller representation, with $\delta = (1-\rho)^{-2}$. This mapping underpins a key motivation of Kamber, Morley, and Wong (2018) given that freely estimating an AR(p) model for real GDP growth often yields an implied signal-to-noise ratio in excess of 1, whereas common methods of removing the trend in macroeconomic time series like the HP filter explicitly or implicitly impose a much lower signal-to-noise ratio. The procedure developed in Kamber, Morley, and Wong (2018) thus allows one to impose a low signal-to-noise ratio by setting $\bar{\delta}$. Specifically, an AR(p) forecasting model can be estimated imposing a particular signal-to-noise ratio $\bar{\delta}$ by first fixing the value of ρ in (4) as follows:

$$\bar{\rho} = 1 - 1/\sqrt{\bar{\delta}}. \tag{5}$$

The $\{\phi_j^*\}_{j=1}^{p-1}$ coefficients in (4) can then be estimated using Bayesian methods with Minnesota-like priors to mitigate possible overfitting when specifying a high-order AR(p) model that can allow for a relatively complicated autocovariance structure for $\{\Delta y_t\}$. Finally, given $\bar{\rho}$ and the estimates of $\{\phi_j^*\}_{j=1}^{p-1}$, the BN cycle for a fixed $\bar{\delta}$ can be calculated by inverting the Dickey-Fuller transformation to solve for the restricted estimates of $\{\phi_j\}_{j=1}^{p}$ and then following (3) given the companion matrix F.

Similar to setting the smoothing parameter for the HP filter to 1600, implementation of the BN filter procedure could be as straightforward as fixing the signal-to-noise ratio at a particular value like $\bar{\delta} = 0.05$, which would correspond to the numerically-specific assumption that only 5% of the quarterly forecast-error variance for output growth is due to trend shocks and was the approach taken, for example, in Rudd (2024) and Kuang, Mitra, and Tang (2024) However, because any particular choice for $\bar{\delta}$ might appear somewhat arbitrary in practice, we proposed an automatic selection of $\bar{\delta}$ in Kamber, Morley, and Wong (2018) based on maximizing the amplitude-to-noise ratio. The intuition of this approach is that, given practitioners often have a less numerically-specific, but generally low signal-to-noise ratio (and thus high amplitude cycle) in mind, this criterion allows for a trade-off between a larger amplitude cycle against the poorer fit of a corresponding restricted forecasting model.

Time-varying drift While the key innovation with of our original BN filter was to impose a low signal-to-noise ratio in an AR(p) model when performing the BN decomposition, we also explicitly allowed for the possibility of time-varying drift. This accommodation was important because empirical evidence suggests the possibility of breaks in drift for many time series, including US real GDP, and so allowing for time-varying drift broadens the utility of the BN filter for practitioners. Generally, the drift term (i.e. μ in (2) and (4)) needs to be estimated. We originally considered two possibilities for how to deal with time-varying drift in applied work. For the first approach, we proposed testing for structural breaks in μ using Bai and Perron (2003) procedures, and, if breaks in μ appear to have occurred, adjusting for these breaks by demeaning $\{\Delta y_t\}$ using subsample averages given estimated breakdates and then proceeding with the BN filter using the demeaned series to estimate the $\{\phi_j^*\}_{j=1}^{p-1}$ coefficients. In an output gap application, this amounts to finding breaks in the mean growth rate of real GDP and demeaning the growth rate based on the estimated breakdates. However, while such an approach is useful for *ex-post* analysis, it is less useful for real-time analysis. This is because breaks are, by their nature, two-sided, and one needs to know the future growth rate before confidently dating the timing of a break *ex post*, a luxury that practitioners who need to estimate the output gap in real time do not enjoy. We therefore also proposed a second approach of using a 40-quarter rolling window to estimate a time-varying drift.[3] That is, we set $\hat{\mu}_t = \frac{1}{40}\sum_{i=1}^{40} \Delta y_{t-i+1}$ and thereafter dynamically demean the first differences of the time series being detrended.

The dynamic demeaning procedure has a few advantages over other real-time approaches. First, in the case of a structural break in drift, it immediately starts to adjust the estimated drift following a break and still provide an unbiased estimate for all but the 39 quarters from the date of the break, while a real-time test for a structural break might not allow a break (e.g., if using a 15% of the sample trimming rule, which is more than 11 years for our sample) or find a significant break for as many or more quarters, thus leading to a comparatively larger bias until the break is detected. Second, in the case of more gradual structural change, it leads to less bias in estimating drift than an expanding window or approximating the structural change with a discrete break that would again be hard to detect in real time. A formal model of gradual structural change such as based on a random walk with small shocks might lead to more precise inferences given structural change that follows the assumed pattern, but dynamic demeaning is easier to implement and is robust to pile-up problems when estimating the variance of an unobserved random walk with small shocks. Meanwhile, if the drift did not actually change or was largely unchanged, dynamic demeaning makes little practical difference to the estimated cyclical component as long as $\{\Delta y_t\}$ is not very persistent such that the estimated mean will be relatively precise given 40 observations.

2.2 Refinements

Next, we introduce four refinements of the BN filter. First, we propose a new trend-smoothness loss function to select the signal-to-noise ratio when detrending. Second, we provide an iterative

[3] We can allow for different possible windows for estimating a time-varying drift, but we use a 40-quarter rolling window when considering estimation of the output gap given that the window is long enough such that temporary effects of business cycles on growth rates should average out and the estimated drift will also be reasonably precise, while it is short enough such that it allows for substantial low frequency changes in long-run growth. In principle, one could start with a longer window and shorten it until results for implied time-varying drift become either relatively unchanged or very imprecise with a further shortening of the window. We find 40 quarters works well in practice for macroeconomic data.

dynamic demeaning approach to address outliers that can affect the average change in the cycle over the window used to estimate time-varying drift. Third, we construct error bands that allow for time-varying volatility. Fourth, we conduct backcasting based on time reversibility, producing an estimated cycle for the full sample of the variable being detrended. The details of these refinements are presented in turn.

2.2.1 A trend-smoothness loss function when selecting $\bar{\delta}$

As an alternative to maximizing the amplitude-to-noise ratio considered in Kamber, Morley, and Wong (2018), we propose setting $\bar{\delta}$ to minimize the variance of the change in trend subject to a positive signal-to-noise ratio. Specifically, $\bar{\delta}$ is chosen as follows:

$$\bar{\delta} = \arg\min_{\delta>0} \quad \sigma^2_{\Delta\tau}(\delta) \tag{6}$$

Obviously, imposing the limiting case of $\delta = 0$ would truly minimize the variance of trend shocks by implying no stochastic trend at all – i.e., $\sigma^2_{\Delta\tau}(0) = 0$. However, that limiting case corresponds to a unit MA root for $\{\Delta y_t\}$ and is, therefore, incompatible with our use of a finite-order AR model for which the implied δ is bounded away from zero, as discussed in Kamber, Morley, and Wong (2018). In applying the BN filter with a finite-order AR model, we take it as a maintained assumption, possibly motivated by unit root and/or stationarity tests, that the time series $\{y_t\}$ being detrended has a non-zero stochastic trend component. Importantly, there is an interior solution and near-zero values of $\delta > 0$ will not be close to the optimum if the fit of the model deteriorates as $\delta \to 0$ (i.e., $\lim_{\delta\to 0}\sigma^2_e(\delta)' < 0$), which is what we find in practice. Specifically, noting that the variance of trend shocks $\sigma^2_{\Delta\tau}(\delta) = \delta\sigma^2_e(\delta)$ and, therefore, $\sigma^2_{\Delta\tau}(\delta)' = \delta\sigma^2_e(\delta)' + \sigma^2_e(\delta)$ and $\sigma^2_{\Delta\tau}(\delta)'' = \delta\sigma^2_e(\delta)'' + 2\sigma^2_e(\delta)'$, it is easy to show that $\lim_{\delta\to 0}\sigma^2_{\Delta\tau}(\delta)' > 0$ and $\lim_{\delta\to 0}\sigma^2_{\Delta\tau}(\delta)'' < 0$, meaning that, as $\delta \to 0$, there will be a trade-off in determining $\sigma^2_{\Delta\tau}(\delta) = \delta\sigma^2_e(\delta)$ between the direct reduction in the signal-to-noise term δ and an implied worsening of the fit term $\sigma^2_e(\delta)$, with an interior local minimum resulting as it requires $\sigma^2_{\Delta\tau}(\bar{\delta})' = 0$ and $\sigma^2_{\Delta\tau}(\bar{\delta})'' > 0$.[4]

[4]If the model fit also deteriorates for large enough δ (i.e., $\lim_{\delta\to\infty}\sigma^2_e(\delta)' > 0$), this implies $\sigma^2_e(\delta^*)' = 0$ and $\sigma^2_e(\delta^*)'' > 0$ for some $\delta^* > 0$, which, in turn, implies $\sigma^2_{\Delta\tau}(\delta^*)' > 0$ and $\sigma^2_{\Delta\tau}(\delta^*)'' > 0$. This further implies $\sigma^2_{\Delta\tau}(\underline{\delta})'' = 0$ for some smaller value of $0 < \underline{\delta} < \delta^*$ for which $\sigma^2_e(\underline{\delta})' < 0$ and $\sigma^2_e(\underline{\delta})'' > 0$. Then solving for an optimum by finding $\bar{\delta} > \underline{\delta}$ such that $\sigma^2_{\Delta\tau}(\bar{\delta})' = 0$ will correspond to a local interior minimum because $\sigma^2_{\Delta\tau}(\delta)'' > 0$ for $\delta > \underline{\delta}$. An exact analytical solution depends on the noise function $\sigma^2_e(\delta)$. In practice, we

We see this trend-smoothness loss function as being in the spirit of trend-cycle decomposition methods that impose explicit smoothness priors on the trend, such as Harvey, Trimbur, and Van Dijk (2007), or impose a low signal-to-noise, such as the HP filter. The trend-smoothness loss function should be well suited to events such as the COVID-19 pandemic that involve extreme movements in the data without an obvious *ex ante* notion of whether the trend moves much as a result. At the same time, smoothing the trend in this way has the potential to alter the reliability properties of the BN filter given that other methods which smooth the trend, such as the HP filter, are notably less reliable and can generate spurious cycles (Cogley and Nason, 1995). We also note that this sort of trend smoothness loss function can also been useful beyond the univariate setting. In particular, in related work estimating the euro area output gap using a multivariate BN decomposition, Morley, Rodriguez-Palenzuela, Sun, and Wong (2023) borrow from the approach proposed here by minimizing the variance of the change in trend to select the key shrinkage hyperparameter for Bayesian estimation of a large VAR and find that doing so is particularly important in terms of estimating the depth of the euro area output gap during the COVID-19 recession.

2.2.2 Iterative dynamic demeaning

To understand our proposed iterative dynamic-demeaning approach, it is helpful to revisit the basic assumption that the trend of $\{y_t\}$ is a random walk with drift. That is, the non-zero unconditional expectation $\mathbb{E}[\Delta y_t] = \mu \neq 0$ is associated with the trend, not the cycle. Recall the basic trend-cycle identity:

$$y_t = \tau_t + c_t. \tag{7}$$

Then, it directly follows that

$$\Delta y_t = \Delta \tau_t + \Delta c_t \tag{8}$$

and

$$\mathbb{E}[\Delta y_t] = \mathbb{E}[\Delta \tau_t] + \mathbb{E}[\Delta c_t]. \tag{9}$$

consider a numerical grid to find the $\bar{\delta}$ that minimizes the $\sigma^2_{\Delta\tau}(\delta)$. It is straightforward with the grid to confirm existence and uniqueness of the local minimum in any given setting. We always find a unique interior minimum in practice.

The assumption that drift is associated with the trend is equivalent to assuming $\mathbb{E}[\Delta\tau_t] = \mu$ and $\mathbb{E}[\Delta c_t] = 0$, which directly follows from the maintained assumption $\mathbb{E}[c_t] = 0$ when applying the BN decomposition.

When we use the sample mean of $\{\Delta y_t\}$ to estimate μ, we are implicitly assuming that the sample mean of $\{\Delta c_t\}$ is close to zero, consistent with its unconditional expectation. That is, the trend-cycle identity in (7) also implies

$$\frac{1}{T}\sum_{t=1}^{T}\Delta y_t = \frac{1}{T}\sum_{i=t}^{T}\Delta\tau_t + \frac{1}{T}\sum_{t=1}^{T}\Delta c_t \tag{10}$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\Delta c_t \approx 0 \Rightarrow \frac{1}{T}\sum_{i=t}^{T}\Delta\tau_t \approx \frac{1}{T}\sum_{t=1}^{T}\Delta y_t. \tag{11}$$

Thus, the sample mean of $\{\Delta y_t\}$ approximates the sample mean of $\{\Delta\tau_t\}$, and thus provides an easy way to estimate the drift associated with the trend.

However, when we consider dynamic demeaning to account for time-varying drift, the sample for our rolling-window estimation can become small enough that the average change in the cycle over the rolling window might not be close to zero. Of course, when estimating the cycle, we obviously do not observe it *a priori*, so we cannot simply remove the sample mean of the change in the cycle to estimate the drift associated with trend growth. But if our estimated cycle using $\{\Delta y_t\}$ to estimate time-varying drift implies a non-zero sample mean of the change in the cycle, then we can consider iterative estimation of trend growth until the estimated cycle is consistent with the sample mean of the estimated change in the cycle. Specifically, initially setting $c_t^{\{0\}} = 0$ and $j = 1$, we repeatedly estimate the cycle using dynamic demeaning according to $\mu_t^{\{j\}} = \frac{1}{40}\sum_{i=1}^{40}\Delta y_{t-i+1} - \frac{1}{40}\sum_{i=1}^{40}\Delta c_{t-i+1}^{\{j-1\}}$ for j iterations until $c_t^{\{j\}} \approx c_t^{\{j-1\}}$ up to some arbitrary level of precision. Note that $c_t^{\{1\}}$ is just our original dynamic-demeaning estimate using $\{\Delta y_t\}$ to estimate time-varying drift. The iterative approach helps address unusual movements in a cycle that can result in the average change in the cycle over the rolling window being significantly different than zero. We are motivated by the extreme outliers associated with the pandemic to consider this approach, although we find relatively little difference in estimates during the pandemic using the original or iterative approaches given that the change in the cycle, while large, was not persistent *ex post*. We find larger changes in estimates at other points of the sample period during more persistent downturns such as the Great Recession or booms such as in the mid-1980s.

2.2.3 Time-varying error bands

In the appendix of Kamber, Morley, and Wong (2018), we presented a method to assess the uncertainty associated with the BN filter estimates. In particular, based on equations (3) and (4), we solved for the variance of the BN cycle, σ_c^2, as follows:

$$\sigma_c^2 = [1 \quad 0 \quad \cdots \quad 0] F(I-F)^{-1}\Sigma_X((I-F)')^{-1}F'[1 \quad 0 \quad \cdots \quad 0]', \tag{12}$$

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:

$$Q = \begin{bmatrix} \sigma_e^2 & 0 & \cdots \\ 0 & 0 & \ddots \\ \vdots & \ddots & \ddots \end{bmatrix}. \tag{13}$$

Because the estimated trend provides an unbiased estimator of the true trend, we proposed constructing a 95% confidence interval for the BN cycle by inverting a simple z-test for different values of c_t as follows:

$$c_t^{BN} \pm 1.96\sigma_c. \tag{14}$$

While useful, error bands based on this approach have constant width. However, if the conditional volatility of $\{\Delta y_t\}$ were actually time varying, the originally proposed approach might incorrectly estimate the degree of estimation uncertainty at different points of the sample, which seems particularly relevant when faced with events such as the COVID-19 pandemic when there was unprecedented volatility in real GDP growth.[5] Thus, the refinement we propose for

[5] We note that accounting for changes in volatility is less relevant for the estimated cycle based on (3) because modelling heteroskedasticity has minimal impact on estimates of the autoregressive coefficients in the companion matrix F. Morley and Piger (2012) found little effect of accounting for heteroskedasticity when conducting trend-cycle decomposition based on linear and nonlinear AR models estimated via MLE and it would be even less so for the AR coefficients in the BN filter given the restriction on their sum implied by $\bar{\delta}$ and the use of Minnesota-like priors to estimate the $\{\phi_j^*\}_{j=1}^{p-1}$ coefficients. Thus, it is not such an issue when conducting the BN decomposition to ignore potential heteroskedasticity. It is only when constructing confidence intervals that this heteroskedasticity matters more. Related, accounting for parameter uncertainty would have a relatively small effect on the error bands compared to the sampling uncertainty about c_t.

the construction of error bands is to allow for the possibility of time-varying volatility. In particular, our simple proposal is to estimate time-varying conditional volatility of $\{\Delta y_t\}$ using a rolling window, although one could certainly consider a more complicated approach such as a formal model of conditional volatility. Specifically, given window size k, we replace the estimate of conditional volatility σ_e^2 in (13) based on least squares residuals from (2) with estimates of $\sigma_{e,t}^2$ using a rolling window from $t-k+1$ to t. We consider $k = 40$ quarters as the choice for the rolling window to align with our choice for dynamic demeaning.[6]

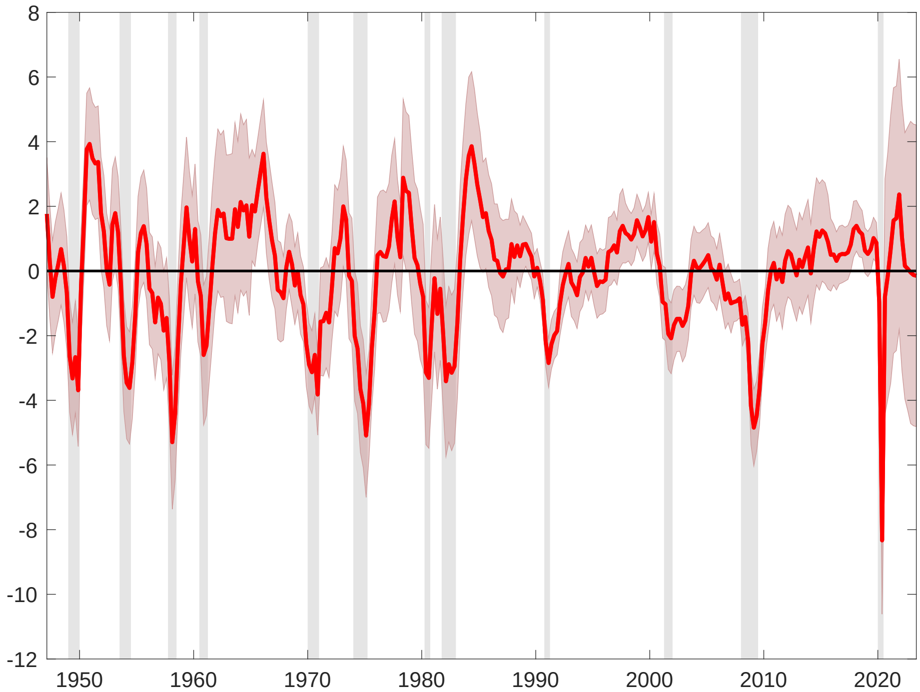
2.2.4 Iterative backcasting to generate trend-cycle estimates for y_1

Our last refinement involves iterative backcasting of initial observations for estimation by utilizing the time reversibility of any linear time series process that can be described in terms of ARMA dynamics (see, for example, Ramsey and Rothman, 1996). Our motivation for this refinement is to be comparable to trend-cycle decomposition methods such as the HP filter that produce estimates of trend and cycle for the first observation of the level of the time series being decomposed, y_1. Given the BN decomposition makes use of a forecasting model in first differences, the original BN filter produces estimates for the first available observation of the first differences (i.e., $\Delta y_2 = y_2 - y_1$), corresponding to the second observation in levels, y_2. In order to obtain reasonable trend and cycle estimates from the BN filter for the first observation in levels, we would need a better estimate of the first difference associated with the initial observation (i.e., $\Delta \hat{y}_1$) than just the estimated drift, $\hat{\mu}$, which is what we used for backcasting initial observations when estimating the restricted AR(p) model for the original BN filter in Kamber, Morley, and Wong (2018). With our proposed iterative backcasting, we find estimates of initial observations by using the restricted AR(p) model to first forecast $p+1$ observations in first differences out of sample and then reversing the time series in differences including these forecasted observations in order to backcast $p+1$ initial observations in first differences. This forecasting, reversing, and backcasting is done iteratively until the estimated initial observations and out-of-sample forecasts converge to an arbitrarily small tolerance. In practice, we find this convergence is almost immediate. Then, with $p+1$ backcast initial

[6] As with dynamic demeaning, the estimates for the first 40 quarters are constant and based on the first 40 quarters. The 40-quarter rolling window is again motivated by being long enough to provide relatively precise variance estimates, although it could overstate the persistence of changes in conditional volatility. Tailoring a formal model of conditional volatility to a given time series could be worthwhile in settings where there are concerns about higher-frequency changes in conditional volatility.

observations, $\Delta\hat{y}_1, \Delta\hat{y}_0, ..., \Delta\hat{y}_{-p+1}$, we can use (3) to calculate the BN cycle (and, therefore, BN trend) for y_1.

Figure 1: Estimates of the US output gap using the refined BN filter



Notes: Units are 100 times natural log deviation from trend. Sample period is 1947Q1 to 2023Q2. Shaded bands around the estimate correspond to 95% confidence intervals based on inverting a z-test that the true output gap is equal to a hypothesized value using the standard deviation of the BN cycle estimated with rolling-window estimates of time-varying conditional volatility for output growth. Shaded bars correspond to NBER recession dates.

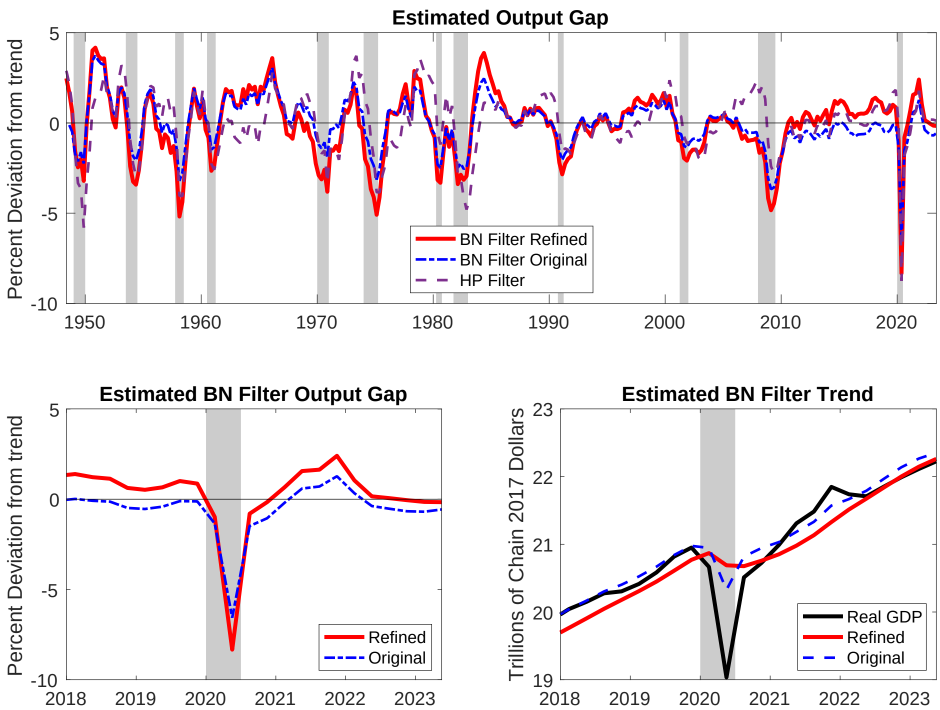
2.3 Estimates of the output gap using the refined BN filter

We first present estimates of the output gap using US real GDP data from 1947Q1 to 2023Q2 when considering the BN filter with all of the refinements. Figure 1 plots these estimates, along with 95% confidence intervals. Similar to what we found in Kamber, Morley, and Wong (2018), the refined BN filter produces intuitive estimates of the output gap, with movements in it being well aligned with the NBER reference cycle. We note there is a larger degree of estimation uncertainty in the earlier part of the sample than during the Great Moderation, but a higher level of uncertainty has returned since the pandemic given the inclusion of outlier growth rates associated with the pandemic in the rolling window when estimating time-varying conditional volatility of output growth.[7] The $\bar{\delta}$ selected given the new loss function is 0.02,

[7]Shortening the rolling window produces even wider error bands right around the pandemic, but does not have as much effect on the bands at other points of the sample. Given how short lived the high volatility in output growth was in 2020, a formal model of time-varying volatility that adapts to persistence of different

implying that only 2% of the quarterly forecast-error variance for output growth corresponds to trend shocks. This contrasts with the $\bar{\delta} = 0.24$ found in Kamber, Morley, and Wong (2018), which while yielding intuitive and reliable estimates of the output gap, may have been a larger signal-to-noise ratio than believed by many practitioners seeking to estimate a smooth trend.

Figure 2: Comparison of the original and refined BN filter



Notes: Units are 100 times natural log deviation from trend for the first two panels and trillions of 2017 dollars for the third panel. Sample period is 1947Q1 to 2023Q2 for the top panel and 2018Q1 to 2023Q2 for the bottom panels. Shaded bars correspond to NBER recession dates.

To help understand the role of the refinements, Figure 2 compares the results for the refined BN filter with those for the updated sample based on the original BN filter proposed in Kamber, Morley, and Wong (2018). The top panel of Figure 2 plots the estimated output gap for the refined BN filter together with estimated output gap for the original BN filter, while also including the output gap based on the HP filter using the typical smoothing parameter of 1600 as a point of reference. In general, the refined BN filter produces an estimated output gap with fluctuations that are very similar to those for the original BN filter (the correlation is 92%). However, there are some key differences. The refined BN filter estimates display a more similar amplitude to that of the HP filter estimates than the original BN filter estimates, with the refined BN filter and HP filter estimates being reasonably similar to each other around the

episodes would presumably imply a quicker return to the tighter bands from before the pandemic than our rolling window approach.

pandemic (fully acknowledging that the estimates near the end of the sample could be revised in the future). The difference in amplitude between the original and refined BN filter estimates are particularly notable during the 1970s recessions, the mid-1980s, the Great Recession, and the COVID-19 pandemic.[8] Specifically, because the new loss function is explicitly designed to minimize the change in trend, it mechanically attributes a larger proportion of the fluctuations in real GDP to the cycle.[9] All in all, the comparison with the HP filter suggests that the new loss function for the refined BN filter comes closer to the implicit loss function of practitioners who continue to use the HP filter despite repeated warnings not to do so (e.g., Cogley and Nason, 1995; Hamilton, 2018), perhaps because they believe the two-sided *ex-post* estimates from the HP filter approximate the true cycle when defined in terms of an ideal high-pass filter (see the discussions in Pedersen, 2001; Cogley, 2001).

To provide some sense of how our updated approach deals with the pandemic, the bottom panels of Figure 2 compare the refined BN filter with the original BN filter, zooming in around the COVID-19 period in particular. For the comparison, we present both the output gap and trend output estimates. The refined BN filter attributes more of the fall of real GDP during 2020Q2 to being transitory rather than permanent. The corresponding trend estimates show that the refined BN filter implies a relatively smooth reversal in trend output, consistent with the new loss function. The original BN filter, on the other hand, implies a sharp drop in trend in 2020Q2 followed by a sharp increase in 2020Q3. From an *ex-post* perspective, the refined BN filter produces estimates of the output gap and trend output that are arguably more consistent with the economic narrative of the time, at least if one associates the output gap with transitory movements in real GDP rather than necessarily making distinctions between sources of these movements in terms of aggregate demand and supply. In particular, the outlier fall in real GDP growth in 2020Q2 was clearly associated with temporary and massive restrictions on economic activity such as lockdowns that were largely reversed in 2020Q3, which suggests an

[8]In addition to being more similar to the HP filter estimates, the larger amplitude of the output gap for the refined BN filter can be thought of as being more intuitive because the estimates are generally in line with the magnitude of the CBO output gap at these times as well. For example, the CBO output gap was also close to -5% at the troughs of the recessions in 1975 and 2009.

[9]We note that the original procedure selects a $\bar{\delta}$ of 0.15 for the updated sample, which is lower than the selected value of 0.24 for the original sample from 1947Q1 to 2016Q2 used in Kamber, Morley, and Wong (2018). Nonetheless, this is still much higher than the $\bar{\delta}$ of 0.02 selected with the proposed new loss function and would increase the loss in (6) by 39% compared to the new procedure, with the standard deviation of trend shocks increasing from 0.45% to 0.53%. The economic significance of this difference is illustrated in the last panel of Figure 2 when comparing trend estimates during the pandemic, as discussed below.

economic narrative that most of the decline in activity due to the pandemic was transitory.[10] The differences in estimates around the pandemic are largely about the new loss function, with some differences also reflecting iterative dynamic demeaning given lower estimated trend growth throughout the 2010s corresponding to a generally higher level of the output gap, especially just prior to the pandemic.

3 Real-time reliability when including Covid-era data

We now compare the real-time reliability of the BN filter against some widely-used trend-cycle methods by examining their revision properties when including Covid-era data in the analysis. This evaluation is motivated by the well-documented general unreliability of output gap estimates in real time, as most prominently demonstrated by Orphanides and van Norden (2002). A key attraction of the original BN filter is that it appears to have revision properties which largely circumvent the Orphanides and van Norden (2002) critique. It is thus of interest to understand whether both the original and refined BN filters retain such good revision properties when confronted with Covid-era data and how they compare to each other and other widely-used methods.

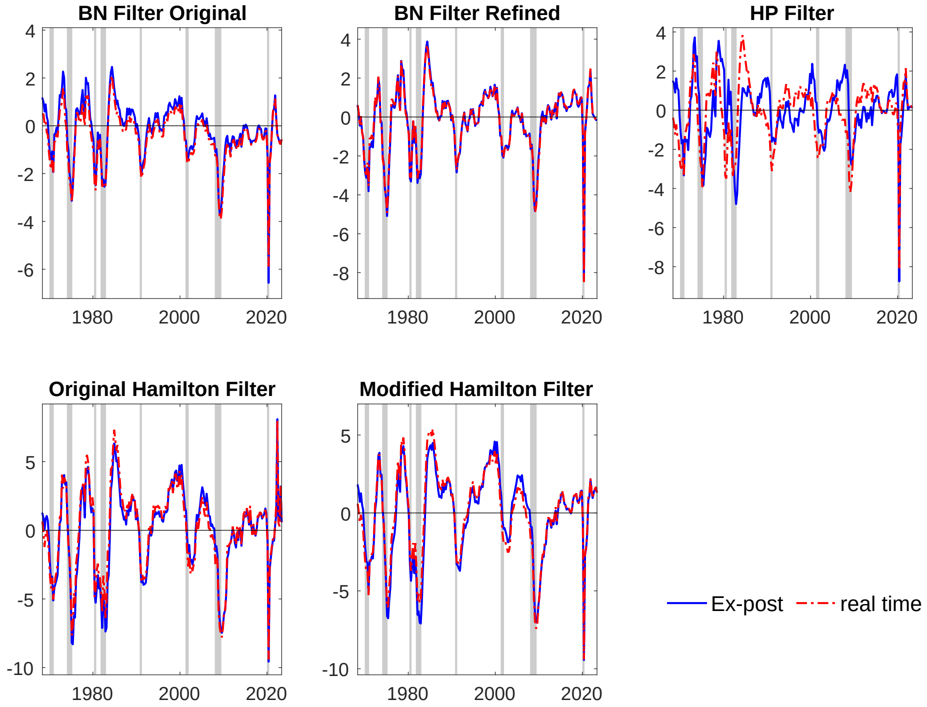
In terms of the other methods, we consider the HP and Hamilton filters detailed in Hodrick and Prescott (1997) and Hamilton (2018), respectively. The HP filter is a natural choice given its near ubiquitous use in academia and policy environments, and we consider the standard implementation of the HP filter using a smoothing parameter of 1600, as in Figure 2.[11] The Hamilton filter is also an obvious choice given its recent emergence due to its ease of use and also its purported ability to address various shortcoming of the HP filter. Furthermore, since the publication of Kamber, Morley, and Wong (2018), Quast and Wolters (2022) have shown that the Hamilton filter possesses similarly good revision properties when comparing to the BN filter. The Hamilton filter output gap corresponds to the residuals from a least squares regression of h-step-ahead log real GDP on its contemporaneous value and three lags as well as

[10] It is possible that the disruption associated with the pandemic also had some more persistent and possibly permanent negative implications for economic activity, consistent with our finding of a small reduction in trend when using the refined BN filter or even the HP filter.

[11] As noted in Kamber, Morley, and Wong (2018), the estimates and revision properties remain very similar when padding the data with, say, AR(4) forecasts to try to address end-point problems, as done, for example, in Edge and Rudd (2016).

a constant. We consider two versions of the Hamilton filter: the original version in Hamilton (2018), which sets $h = 8$, and the modified version in Quast and Wolters (2022), which averages over gaps obtained from $h = 4$ to $h = 12$.

Figure 3: Ex-post versus pseudo-real-time output gap estimates



Units are 100 times natural log deviation from trend. Sample period is 1947Q1 to 2023Q2. Shaded bars correspond to NBER recession dates.

Figure 3 plots *ex-post* versus pseudo-real-time estimates of the US output gap for the different methods. By *"ex post"*, we are referring to the output gap estimated using the full 1947Q1-2023Q2 final-vintage sample of data. For the pseudo-real-time estimates, we sequentially estimate the output gap starting with a shorter sample from 1947Q1 to 1970Q1 and then, adding one observation of real GDP at a time, retain the last estimated value of the output gap for each pseudo-real-time sample. This is analogous to the filtered estimate of the output gap if one were estimating the gap in real time using the Kalman filter, albeit using the final vintage of data.[12]

Consistent with Orphanides and van Norden (2002) and Kamber, Morley, and Wong (2018),

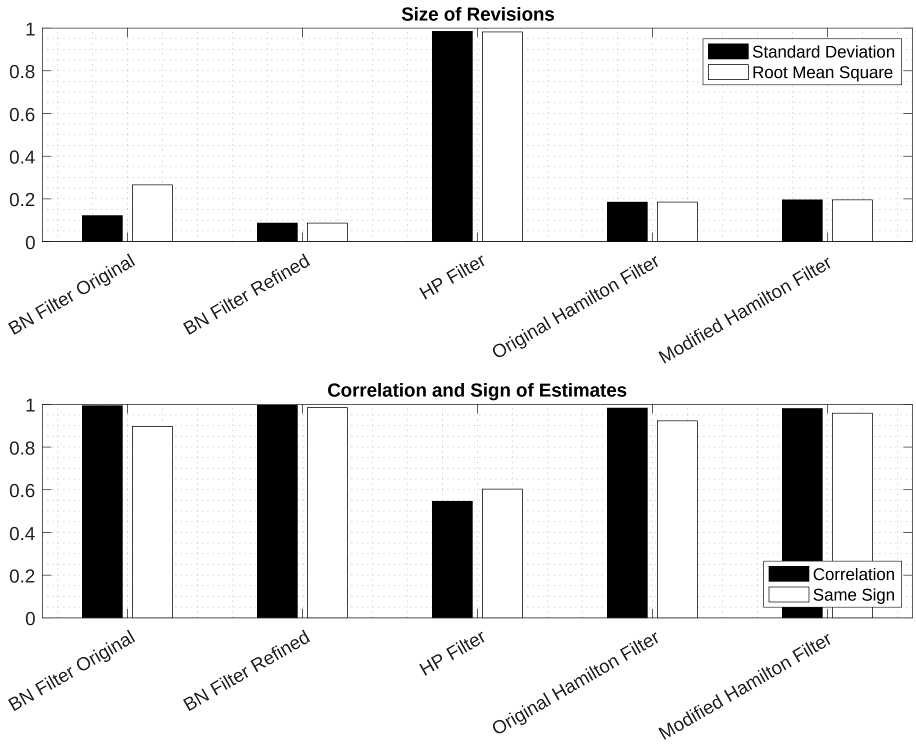
[12]We consider just a single vintage of data in order to isolate the role of the methods rather than data revisions for three reasons. First, Orphanides and van Norden (2002) show that most of the revisions for output gap estimates are usually due to the method rather than data revisions. Second, none of the methods are explicitly designed to address data revisions, so using a single vintage of data isolates the analysis to how well the various methods are able to deal with the so-called "end-point" problem. Third, because none of these methods are designed to address data revisions, we show in Kamber, Morley, and Wong (2018) that it makes relatively little difference whether one uses real-time data or a single vintage if one were just seeking to understand the comparative revision properties of the methods under consideration.

Figure 3 shows that the output gap based on the HP filter is often heavily revised, with the pseudo-real-time and *ex-post* estimates diverging considerably. Notably, despite being relatively close to the end of the sample, the one-sided real-time HP filter estimates of the output gap just prior to the pandemic are highly unreliable, with the *ex-post* estimates revised upwards substantially for 2018-2019 in particular from being very close to zero to being closer to the real-time estimates for the refined BN filter and both versions of the Hamilton filters.

Despite the inclusion of Covid-era data, it is clear from Figure 3 that the original BN filter retains the good revision properties reported in Kamber, Morley, and Wong (2018) and also confirmed in Quast and Wolters (2022), Barigozzi and Luciani (2023), and Barbarino, Berge, and Stella (2024). Meanwhile, the refined BN filter displays even slightly better revision properties than the original BN filter. Note that because we estimate the output gaps in pseudo real time, we are also allowing $\bar{\delta}$ to change in the real-time setting. While the pandemic did change the selected $\bar{\delta}$, most prominently for the original BN filter from the originally published 0.24 to 0.15 at the final data point, this makes little difference to the *ex-post* versus pseudo-real-time estimates. In fact, it is precisely because of the BN filter's reliability that $\bar{\delta}$ falls, although it falls by less than 0.01 for the refined BN filter. Given that the historical and *ex-post* estimates do not diverge after the pandemic, the only way that the BN filter can retain its reliability when considering either the amplitude-to-noise ratio or the variance of the change in trend in the presence of extreme outliers that boost the sample variance of the forecast errors is through a corresponding decrease in the signal-to-noise ratio δ given "noise" is being measured by σ_e^2. Meanwhile, the slight downward shift in the *ex-post* estimates for the original BN filter is due to slower average output growth in recent years, which the refined BN filter accounts for with iterated dynamic demeaning.

Turning to the Hamilton filter, estimates for both the original and modified versions are also little revised in Figure 3, consistent with what was documented by Quast and Wolters (2022). An interesting finding, though, is that, while the estimated output gaps between the original and modified versions were very similar before the pandemic, they are quite different afterwards. In particular, the original Hamilton filter output gap displays a large mechanical spike in the estimated output gap in 2022Q2, exactly two years after the onset of the pandemic. The spike is mechanical because it was perfectly predictable prior to 2022Q2 given base effects for the 8-quarter-ahead projection of the level of log real GDP for 2022Q2 due to the low value of log real GDP in 2020Q2 and then for 2022Q3 due to the largely recovered value of log real GDP in 2020Q3.[13] We will return to this issue of how and why the original and modified versions of the Hamilton filter are so different after the pandemic in Section 4.

Figure 4: Revision properties of output gap estimates



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 to 2018Q2.

To round out the real-time reliability analysis, Figure 4 presents the formal revision statistics for the various methods, albeit dropping the last five years of the sample that include the pandemic when calculating the statistics because the more recent *ex-post* estimates near the end of the sample may end being heavily revised in the future. These statistics were originally

[13] The predictable base effects may be easiest to think about by considering the simplified version of the Hamilton filter also discussed in Hamilton (2018) that constructs the output gap as the eight-quarter difference in log real GDP – i.e., $y_t - y_{t-8} = \Delta y_t + \Delta y_{t-1} + ... + \Delta y_{t-7}$. From 2020Q3 to 2022Q1, the eight-quarter difference will include the 2020Q2 and 2020Q3 quarterly growth rates, which were largely offsetting. So, if all other quarterly growth rates were much closer to the drift μ, the eight-quarter difference from 2020Q3 to 2022Q1 would be close to $6 \times \mu$. Then, predictably in 2022Q2, the output gap will jump up by about $-\Delta y_{2020Q2} + \mu$ as the very negative quarterly growth for 2020Q2 is dropped from the eight-quarter growth rate and the 2022Q2 growth rate, which is assumed to be closer to the drift μ, is added. Furthermore, predictably in 2022Q3, the output gap will fall back down by $-\Delta y_{2020Q3} + \mu$ to close to $8 \times \mu$ as the very positive growth rate for 2020Q3 is dropped and the 2022Q3 growth rate, which is again assumed to be closer to the drift μ, is added. In this sense, it is predictable changes in the "base" for the eight-quarter difference (i.e., y_{t-8}) that mechanically explain the positive spike exactly 8 quarters after a large negative spike.

proposed by Orphanides and van Norden (2002). The top panel plots the size of the revisions between the pseudo-real-time and *ex-post* estimates in terms of standard deviation and root mean square, both normalized by the standard deviation of the *ex-post* estimated gap, as suggested by Orphanides and van Norden (2002). The bottom panel plots the correlation between the pseudo-real-time and *ex-post* estimates and proportion of these estimates that share the same sign. As one might suspect from Figure 3, all variants of the BN filter and the Hamilton filter have good revision statistics, whereas the HP filter does poorly. On some level, this is perhaps not entirely surprising. As first pointed out by Kamber, Morley, and Wong (2018), the BN filter does well because it is a one-sided filter which does not rely on future information, unlike the HP filter and the various other filters considered by Orphanides and van Norden (2002). The Hamilton filter also falls in the category of a one-sided filter, so it naturally also does well on these metrics as long as the estimated regression parameters remain little changed with the addition of new data. Finally, we note that, even though the original Hamilton and BN filters do well on these metrics in an absolute sense, their updated versions seem to do better on at least some of the metrics, especially for the refined BN filter, which suggest that the updated versions can provide some valuable enhancements without sacrificing real-time reliability. Again, this is a particularly notable finding for the refined BN filter given that the trend-smoothing criterion of the HP filter seems to be a source of its unreliability, while the trend-smoothness loss function for the BN filter perhaps surprisingly does not seem to lead to any deterioration in its reliability.[14]

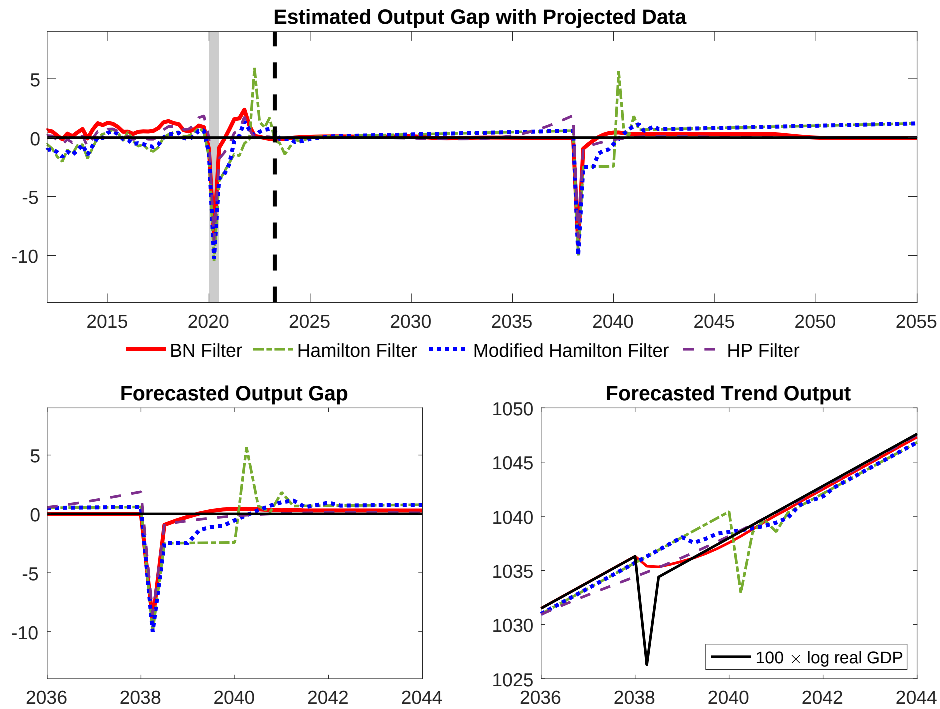
4 Trend-cycle decomposition with projected data

Next, we augment the US real GDP data from 1947Q1 to 2023Q2 with projections of future output growth from 2023Q3 to 2055Q4 that include a simulated Covid-like shock occurring 15 years in the future in 2038Q3 to see how the different methods would forecast the output gap and process an outlier shock for which we know by assumption the true impact on trend and cycle. In particular, we project that real GDP generally grows from the end of our sample

[14]Similar to how the refined BN filter provides a better prediction of the final-vintage HP filter output gap for 2018-2019 than the one-sided real-time HP filter, we find that the correlation with the final-vintage HP filter output gap is actually a bit higher for the real-time refined BN filter than for the one-sided real-time HP filter at 58% versus 55% over the evaluation sample considered in Figure 4 and 65% versus 60% when extending the evaluation sample to 2023Q2 to include the pandemic.

by 0.6% in every quarter, coinciding with a drift of about 2.5% real GDP growth per annum, close to the historical norm. However, in 2038Q3, we also project a (i) a one-time permanent 2.5% reduction in the level of real GDP and (ii) an additional 7.5% transitory reduction in the level of real GDP that lasts only one quarter. Our scenario thus assumes an overall 10% reduction in real GDP in 2038Q3, followed by a partial recovery in 2038Q4. That is, $1/4^{\text{th}}$ of the overall Covid-like shock corresponds to a reduction in trend and the remaining $3/4^{\text{ths}}$ of the shock corresponds to a decline in the output gap. We then apply the different methods to the augmented data to forecast the output gap under this scenario. Note that because only $3/4^{\text{ths}}$ of the Covid-like shock is transitory, some of the projected reduction in real GDP should be attributed to the trend, not the cycle.

Figure 5: Estimates with projected data



Units are 100 times natural log deviation from trend for the first two panels and 100 times natural logs for the third panel. Estimates based on data from 1947Q1 to 2023Q2 augmented with projected data from 2023Q3 to 2055Q4 are reported for 2012Q1 to 2055Q4 in the top panel and 2036Q1 to 2044Q4 in the bottom panels. Sample period is 1947Q1 to 2023Q2 for the top panel and 2018Q1 to 2023Q2 for the bottom panels. The vertical dashed black line denotes 2023Q3, which is the beginning of the forecast period. The shaded bar corresponds to NBER recession dates.

The top panel of Figure 5 plots the estimated output gaps using the augmented data for the refined BN filter, both versions of the Hamilton filter, and the HP filter, with the vertical dashed black line marking the beginning of the forecasted output gaps based on projected real GDP. Before the start of the projection, it can be seen that the output gap estimates are reasonably consistent with the estimated output gaps reported in Figures 2 and 3 even though estimation now considers the augmented data up to 2055. Specifically, all of the estimated output gaps suggest that the COVID-19 recession mostly represented a transitory decline in output, consistent with the economic narrative that economic activity that was constrained by mitigation measures in 2020Q2 was largely restored in 2020Q3. Furthermore, as discussed previously, even though both versions of the Hamilton filter produced output gap estimates that were very similar prior to the pandemic, this does not appear to be the case after the pandemic. The differences for the original and modified Hamilton filters are thus entirely driven by the large shock at the onset of the pandemic. To be sure, in the absence of outliers, the original and modified Hamilton filter would produce very similar estimates, as also shown by Quast and Wolters (2022). However, when extreme outliers occur, the two versions of the Hamilton filter will produce vastly different estimates of the output gap. In particular, because the modified version of the Hamilton filter averages the gap obtained from $h = 4$ to $h = 12$, the base effects of the outlier quarterly growth rates affect the modified filter starting at four quarters after the large shock (2021Q2), which is when we can see the estimated output gaps for the original and modified Hamilton filters start diverging. Of course, because the base effects of the extreme outliers are averaged in the modified version, the modified version does not produce the same large spike after eight quarters (2022Q2) as found with the original Hamilton filter, so the modified Hamilton filter might be seen as somewhat better than the original Hamilton filter in the presence of large shocks.

Turning to the forecasts in Figure 5, because the output gap estimates are all close to zero in 2023Q2 and output growth is set to trend growth for the 15 years before the outlier shock in 2038Q3, we would expect the forecasted output gap to converge to zero as the effects of previous shocks dissipate. This is the case for the BN and HP filters. However, for both versions of the Hamilton filter, the forecasted output gap diverges slowly away from zero over the forecast horizon. This is due to the fact that the actual growth in the projection of 0.6% per quarter is not identical to what would be predicted based on the regression model, which will largely reflect the average growth rate over the full sample (including the augmented data). In principle, the longer the projection sample, the closer the predicted growth should get to the assumed growth. But, in practice, because there is no stochastic variation in projected real GDP for most of the projection (other than the Covid-like shock in 2038, which we discuss below), the regression for the Hamilton filter would become more and more akin to a regression of a linear time trend on lagged linear time trends and a constant, thus moving towards a singularity in the limit and not actually identifying coefficients that sum to one with the constant converging to the underlying drift. It is possible to find a precise growth rate for which the Hamilton filter would imply a forecasted output gap that is a flat line close to zero. But this is a knife-edge case and there is no reason the corresponding growth rate would be the same projection of output growth that a policymaker would want to consider. Meanwhile, consistent with nearly identical estimates in the absence of outliers, this issue with the drifting forecast is not addressed by the modified Hamilton filter, which produces an indistinguishable forecast of the output gap that drifts upwards on the same path as for the original Hamilton filter.[15]

In terms of the simulated Covid-like shock in 2038, we can see it has very similar effects on the forecasted output gaps to what happened with the estimated output gaps around the pandemic. Zooming in on the period just before and after the simulated shock in the bottom left panel of Figure 5, the original Hamilton filter again implies a big mechanical spike exactly two years after the shock. The modified Hamilton filter smoothes this effect out to an extent, but still implies a persistently large positive output gap for a few years before continuing back on the same upward trajectory as the original Hamilton filter once the effects of the Covid-like shock on the estimates have died out. Both versions of the Hamilton filter also overstate the decline in the output gap in 2038Q3, falling by about 10.6 percentage points even though the actual transitory decrease in output was 7.5% in the simulation. Therefore, given output growth was -10% in 2038Q3, the Hamilton filter implies a small increase in trend output even though the true trend actually fell by 2.5%, with the forecasted trend remaining well above the true trend for a number of quarters afterwards, as can be seen in the bottom right panel of Figure 5.[16] Both the HP and BN filters capture an output gap of -7.5% in 2038Q3, but this results from very different dynamics because the HP filter gap was at a spuriously positive level prior to the shock that looks similar to what was found with the final-vintage estimates based on the HP filter prior to the pandemic. Thus, the HP filter output gap actually falls by close to 10 percentage points with the simulated shock, similar to the Hamilton filter, and there is

[15] It is not clear how to further modify the Hamilton filter to address this divergence or to account for structural change in drift such as we do with iterated dynamic demeaning for the BN filter.

[16] For the projected data, real GDP follows the true trend, except for in 2038Q3, when it is also affected by a 7.5% transitory reduction in its level.

an implied offsetting increase in trend output. Corresponding to the spurious positive output gap, the implied level of trend for the HP filter is well below the true trend prior to the shock, as can be seen in the bottom right panel of Figure 5. Both the HP and BN filter output gaps eventually return to zero in the long run, with the HP filter adjusting somewhat faster than the BN filter.[17] Overall, though, only the BN filter captures the general movements in trend and cycle around the Covid-like shock correctly, with economically very different implications about trend output in particular compared to the HP and Hamilton filters.

5 Conclusions

We have introduced some refinements of the BN filter in Kamber, Morley, and Wong (2018) and investigated whether the resulting output gap estimates remains intuitive and reliable in the face of large shocks, such as occurred with the COVID-19 pandemic. We find that the BN filter is still reliable in terms of its revision properties, with the refinements enhancing the intuitiveness of the resulting output gap estimates. Comparing against other popular methods, the HP filter also produces intuitive output gap estimates, but its well-known lack of real-time reliability remains a problem. The Hamilton filter, on the other hand, has good revision properties, but appears less suited to deal with extreme outliers such as occurred during the pandemic because it mechanically produces a future spike in the estimated output gap following an outlier shock exactly in line with the filter horizon. It also produces unintuitive forecasts of the output gap that diverge away from zero given plausible projected values of future output growth. From this perspective, especially when considering large shocks such as experienced during the pandemic and also data augmented with long-term projections as often considered in policy settings, the BN filter appears to be the only of the three univariate trend-cycle decomposition methods under consideration that produces both intuitive and reliable estimates of the output gap.

References

Bai, Jushan and Pierre Perron (2003): "Computation and Analysis of Multiple Structural Change Models," *Journal of Applied Econometrics*, 18 (1), 1–22.

[17]A persistent small positive output gap for the BN filter before returning to zero just over 10 years after the shock reflects some noise from the iterative dynamic demeaning procedure given our assumption of no change in the trend growth in the projected data, while allowing for dynamic demeaning would clearly be more important for projected data that includes assumed changes in long-run trend growth.

Barbarino, Alessandro, Travis J. Berge, and Andrea Stella (2024): "The Stability and Economic Relevance of Output Gap Estimates," *Journal of Applied Econometrics*, forthcoming.

Barigozzi, Matteo and Matteo Luciani (2023): "Measuring US Aggregate Output and Output Gap Using Large Datasets," *Review of Economics and Statistics*, 105 (6), 1500–1514.

Beveridge, Stephen and Charles R. Nelson (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 (2), 151–174.

Christiano, Lawrence J. and Terry J. Fitzgerald (2003): "The Band Pass Filter," *International Economic Review*, 44 (2), 435–465.

Cogley, Timothy (2001): "Alternative Definitions of the Business Cycle and Their Implications for Business Cycle Models: A Reply to Torben Mark Pederson," *Journal of Economic Dynamics and Control*, 25 (8), 1103–1107.

Cogley, Timothy and James M. Nason (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 (1–2), 253–278.

Edge, Rochelle M. and Jeremy B. Rudd (2016): "Real-Time Properties of the Federal Reserve's Output Gap," *The Review of Economics and Statistics*, 98 (4), 785–791.

Hamilton, James D. (2018): "Why You Should Never Use the Hodrick-Prescott Filter," *The Review of Economics and Statistics*, 100 (5), 831–843.

Harvey, Andrew C., Thomas M. Trimbur, and Herman K. Van Dijk (2007): "Trends and Cycles in Economic Time Series: A Bayesian Approach," *Journal of Econometrics*, 140 (2), 618–649.

Hodrick, Robert J and Edward C Prescott (1997): "Postwar U.S. Business Cycles: An Empirical Investigation," *Journal of Money, Credit and Banking*, 29 (1), 1–16.

Holston, Kathryn, Thomas Laubach, and John C. Williams (2023): "Measuring the Natural Rate of Interest after COVID-19," Federal Reserve Bank of New York Staff Report 1063.

Jönsson, Kristian (2024): "Simulation-Based Analysis of Real-Time Reliability for Trend/Cycle Decompositions," *Journal of Business Cycle Research*, forthcoming.

Kamber, Güneş, James Morley, and Benjamin Wong (2018): "Intuitive and Reliable Estimates of the Output Gap from a Beveridge-Nelson Filter," *The Review of Economics and Statistics*, 100 (3), 550–566.

Kuang, Pei, Kaushik Mitra, and Li Tang (2024): "Output Gap Estimation and Monetary Policy with Imperfect Knowledge," SSRN working paper.

Lenza, Michele and Giorgio E. Primiceri (2022): "How to Estimate a Vector Autoregression after March 2020," *Journal of Applied Econometrics*, 37 (4), 688–699.

Morley, James and Jeremy Piger (2012): "The Asymmetric Business Cycle," *The Review of Economics and Statistics*, 94 (1), 208–221.

Morley, James, Diego Rodriguez-Palenzuela, Yiqiao Sun, and Benjamin Wong (2023): "Estimating the Euro Area Output Gap Using Multivariate Information and Addressing the COVID-19 Pandemic," *European Economic Review*, 153, 104385.

Morley, James C. (2002): "A State-Space Approach to Calculating the Beveridge-Nelson Decomposition," *Economics Letters*, 75 (1), 123–127.

Morley, James C., Charles R. Nelson, and Eric Zivot (2003): "Why Are the Beveridge-Nelson and Unobserved-Components Decompositions of GDP So Different?" *The Review of Economics and Statistics*, 85 (2), 235–243.

Orphanides, Athanasios and Simon van Norden (2002): "The Unreliability of Output-Gap Estimates in Real Time," *The Review of Economics and Statistics*, 84 (4), 569–583.

Pedersen, Torben Mark (2001): "The Hodrick-Prescott Filter, the Slutzky Effect, and the Distortionary Effect of Filters," *Journal of Economic Dynamics and Control*, 25 (8), 1081–1101.

Quast, Josefine and Maik H. Wolters (2022): "Reliable Real-Time Output Gap Estimates Based on a Modified Hamilton Filter," *Journal of Business & Economic Statistics*, 40 (1), 152–168.

Ramsey, James B and Philip Rothman (1996): "Time Irreversibility and Business Cycle Asymmetry," *Journal of Money, Credit and Banking*, 28 (1), 1–21.

Rudd, Jeremy B. (2024): *A Practical Guide to Macroeconomics*, no. 9781009465793 in Cambridge Books, Cambridge University Press.

Watson, Mark W. (1986): "Univariate Detrending Methods with Stochastic Trends," *Journal of Monetary Economics*, 18 (1), 49–75.