Economics Working Paper Series

2018 - 4

Estimating and Accounting for the Output Gap with Large Bayesian Vector Autoregressions

James Morley and Benjamin Wong

September 2018

Estimating and Accounting for the Output Gap with Large Bayesian Vector Autoregressions *†

James Morley[1] and Benjamin Wong[2]

[1]University of Sydney
[2]Monash University, Australia

September 10, 2018

Abstract

We consider how to estimate the trend and cycle of a time series, such as real GDP, given a large information set. Our approach makes use of the Beveridge-Nelson decomposition based on a vector autoregression, but with two practical considerations. First, we show how to determine which conditioning variables contain relevant information by directly accounting for the Beveridge-Nelson trend and cycle in terms of contributions from different forecast errors. Second, we employ Bayesian shrinkage to avoid overfitting in finite samples when estimating models that are large enough to include many possible sources of information. An empirical application with up to 138 variables covering various aspects of the U.S. economy reveals that the unemployment rate, inflation, and, to a lesser extent, housing starts, aggregate consumption, stock prices, real money balances, and the federal funds rate all contain relevant information beyond that in output growth for estimating the output gap, with estimates largely robust to incorporating additional variables.

JEL Classification: C18, E17, E32

Keywords: Beveridge-Nelson decomposition, output gap, Bayesian estimation, multivariate information

*Email: Morley: james.morley@sydney.edu.au Wong: benjamin.wong@monash.edu

†We thank the Editor, Barbara Rossi, three anonymous referees, Hilde Bjørnland, Kalvinder Shields, participants at the 2018 AEA Meetings, the 25^{th} Symposium of the Society for Nonlinear Dynamics and Econometrics, the 14^{th} International Symposium on Econometric Theory and Applications, the 11^{th} RCEA Bayes Workshop, and seminar and workshop audiences at Reserve Bank of New Zealand, University of Melbourne, Australian National University, Monash University, University of Tasmania, Hitotsubashi University, Reserve Bank of Australia, and Victoria University of Wellington for helpful comments and suggestions. Any errors are our own.

1 Introduction

Interpretation of macroeconomic data often involves decomposing a time series into trend and cycle, especially as related concepts such as the neutral rate of interest, the output gap, and trend inflation are crucial inputs into macroeconomic policy decision-making. The macroeconomic literature is replete with statistical methods for conducting such decompositions (e.g., Hodrick and Prescott, 1997; Christiano and Fitzgerald, 2003). These methods are typically univariate in nature and so only rely on the single variable being detrended for implementation. A challenge with such a univariate approach is that the interpretation of the estimated trend and cycle from a statistical filter often needs to be corroborated "off-model" with other sources of information. It is possible to directly allow for multivariate information to help conduct and interpret trend-cycle decompositions (e.g., Kozicki, 1999; Garratt et al., 2006; Sinclair, 2009; Garratt et al., 2016; Chan and Grant, 2017; Barigozzi and Luciani, 2017), but practical challenges remain in terms of determining exactly which variables should be included in the information set or even with how large the information set can be while still keeping estimation tractable.

We address these practical challenges to processing multivariate information within the context of a particular approach to estimating trend and cycle, namely the Beveridge and Nelson (BN) (1981) decomposition based on a vector autoregression (VAR), as considered in Evans and Reichlin (1994), amongst others. First, we show how to determine which conditioning variables contain relevant information by directly accounting for the BN trend and cycle in terms of contributions from different forecast errors in the VAR. This accounting can be used to define the relevant information set and provides interpretability in terms of which sources of information are most important for estimating trend and cycle for a target variable. Furthermore, given an identification scheme that maps forecast errors to structural shocks, it can also be used for a structural decomposition of movements in trend and cycle. Second, we employ Bayesian shrinkage to avoid overfitting in finite samples when estimating models that are large enough to include many possible sources of information. Evans and Reichlin (1994) show that the multivariate BN decomposition based on a VAR estimated by least squares can be quite sensitive to the number of conditioning variables due to sampling error, perhaps explaining its relative lack of use compared to univariate methods in practice. By considering a standard Minnesota prior with a key hyperparameter calibrated to minimize the pseudo-out-of-sample forecast error variance for the target variable, we find that the degree of shrinkage for our Bayesian VARs (BVARs) increases with the size of the model, mitigating the effects of increasing sampling error for larger systems, while still allowing the likelihood to dominate the prior for coefficients on relevant conditioning variables.

An application with up to 138 variables covering various aspects of the U.S. economy reveals that the unemployment rate, inflation, and, to a lesser extent, housing starts, aggregate consumption, stock prices, real money balances, and the federal funds rate all contain relevant information beyond that in output growth for estimating the output gap. Estimates are largely robust to incorporating additional variables, but highly sensitive to dropping a key informational variable such as the unemployment rate. Notably, our benchmark BVAR forecasts output growth better than an AR(1) model and the estimated output gap performs favorably against other measures used in policy environments in terms of having a strong negative correlation with future output growth and positive correlation with future inflation. Our results are robust to allowing for possible structural change in long-run growth and to consideration of real GDI instead of real GDP as the measure of aggregate output, while Monte Carlo analysis using empirically-motivated data generating processes shows that Bayesian shrinkage makes finite-sample estimates of the BN cycle more robust to misspecification of the size of a system compared to those based on least squares estimation.

The remainder of this paper proceeds as follows: Section 2 discusses the BN decomposition in a multivariate setting and how to determine the relevant sources of information by accounting for the BN trend and cycle in terms of contributions from different forecast errors. Section 3 describes how we employ Bayesian shrinkage to avoid overfitting in finite samples when estimating models that are large enough to include many possible sources of information. Section 4 reports the results for our empirical application estimating the U.S. output gap using a dataset with 138 variables. Section 5 considers robustness to accounting for structural change in long-run growth, consideration of real GDI instead of real GDP, and misspecifying the size of a system in finite samples. We conclude by briefly summarizing our contribution and suggesting possible extensions.

2 The BN Decomposition in a Multivariate Setting

Beveridge and Nelson (1981) define the trend of a time series as its long-horizon conditional expectation minus any future deterministic drift. In particular, letting $\{y_t\}$ be a time series process with a trend component that follows a random walk with a constant drift μ, the BN trend, τ_t, at time t is

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

The BN cycle, c_t, can then be calculated simply as the difference between the observed time series and the BN trend:

$$c_t = y_t - \tau_t. \tag{2}$$

The intuition behind the BN decomposition is that the long-horizon conditional expectation of the cyclical component of a time series process is zero, meaning that the long-horizon conditional expectation of the time series will just reflect its trend. Therefore, one only needs to specify a forecasting model for the time series to estimate the trend based on the implied long-horizon conditional expectation. In a univariate setting, ARIMA models have often been considered (e.g., Beveridge and Nelson, 1981; Morley et al., 2003). In a multivariate setting, linear VARs have been considered (e.g., Evans and Reichlin, 1994).

For the multivariate setting, let $\mathbf{\Delta x_t}$ represent a vector of n stationary variables that includes the first difference of the target variable y_t.[1] We assume that $\mathbf{\Delta x_t}$ has a finite-order VAR(p) representation with the following companion form:

$$(\mathbf{\Delta X_t} - \boldsymbol{\mu}) = \mathbf{F}(\mathbf{\Delta X_{t-1}} - \boldsymbol{\mu}) + \mathbf{H}\boldsymbol{e_t}, \tag{3}$$

where $\mathbf{\Delta X_t} = \{\mathbf{\Delta x'_t}, \mathbf{\Delta x'_{t-1}}, ..., \mathbf{\Delta x'_{t-p+1}}\}'$, $\mathbf{F}$ is the companion matrix, $\boldsymbol{\mu}$ is a vector of unconditional means, $\mathbf{H}$ maps the VAR forecast errors to the companion form, and $\boldsymbol{e_t}$ is a vector of serially uncorrelated forecast errors with covariance matrix $\mathbf{\Sigma}$.[2] Given stationarity, $(\mathbf{I} - \mathbf{F})^{-\mathbf{1}}$ exists and, from equation (3), the cumulative sum at time t of expected future deviations of

[1] By framing the stationary variables as being in differences, we can apply the BN decomposition to the integrated levels, $\mathbf{x_t}$, of these variables, which importantly includes the target variable y_t, although variables that are stationary in their levels could also be included in $\mathbf{\Delta x_t}$ and the BN decomposition would implicitly be applied to the accumulation of their levels.

[2] It should be noted that equation (3) can accommodate cointegration according to a vector error correction representation by including the long-run equilibrium errors in $\mathbf{\Delta X_t}$, with particular restrictions on $\mathbf{F}$ and $\mathbf{H}$. See Morley (2002) for an example.

the vector process from its unconditional mean can be written as

$$\mathbb{E}_t \sum_{j=1}^{\infty} (\mathbf{\Delta X_{t+j}} - \boldsymbol{\mu}) = \mathbf{F}(\mathbf{I} - \mathbf{F})^{\mathbf{-1}}(\mathbf{\Delta X_t} - \boldsymbol{\mu}). \tag{4}$$

Then, denoting $\boldsymbol{\tau_t}$ and $\mathbf{c_t}$ as vectors of BN trends and BN cycles, respectively, these can be solved following Morley (2002) as

$$\begin{aligned} \boldsymbol{\tau_t} &= \mathbf{X_t} + \mathbf{F}(\mathbf{I} - \mathbf{F})^{\mathbf{-1}}(\mathbf{\Delta X_t} - \boldsymbol{\mu}) && (5) \\ \mathbf{c_t} &= -\mathbf{F}(\mathbf{I} - \mathbf{F})^{\mathbf{-1}}(\mathbf{\Delta X_t} - \boldsymbol{\mu}). && (6) \end{aligned}$$

The crucial question in the multivariate setting is which conditioning variables to include in the model. In principle, the simple answer based on the definition of the BN decomposition would be to include all variables that contain relevant information for forecasting the target variable y_t based on the true data generating process (DGP). In particular, denote $\mathbf{w_t}$ as a strictly multivariate set of variables such that $\mathbf{w_t}$ is a vector with $n^* > 1$ rows that includes Δy_t and follows a finite-order VAR(p^*) process. Then, given population values for $\mathbf{F}$ and $\boldsymbol{\mu}$ and setting $\mathbf{\Delta X_t} = \mathbf{W_t}$, where $\mathbf{W_t} = \{\mathbf{w'_t}, \mathbf{w'_{t-1}}, ..., \mathbf{w'_{t-p^*+1}}\}'$, equations (5) and (6) would recover the true trends and cycles for $\mathbf{x_t}$, including for y_t.

We consider the case where a large available dataset includes $\mathbf{w_t}$, as well as some additional extraneous variables that are not relevant for forecasting future values of the target variable. Then, letting $\mathbf{V_t} \subset \mathbf{W_t} \subset \mathbf{Z_t}$, we can make the following two observations: First, Proposition 1 in Evans and Reichlin (1994) directly implies that the BN cycle will be different when conditioning on a smaller information set $\mathbf{V_t}$ instead of $\mathbf{W_t}$, with a strictly smaller variance of the cycle in the case of omitted variables. Second, it is straightforward to reason from the proposition that the BN cycle will be the same when conditioning on a larger information set $\mathbf{Z_t}$ instead of $\mathbf{W_t}$.[3]

Based on these two observations, we propose a practical way to determine which conditioning variables contain relevant information. To understand our approach, let $\boldsymbol{\Gamma_i} \equiv \mathbf{F^i}(\mathbf{I} - \mathbf{F})^{\mathbf{-1}}$ for notational convenience and repeatedly lag and substitute equation (3) into equation (6) to get

[3]This equivalence is given population values for $\mathbf{F}$ and $\boldsymbol{\mu}$. The BN cycles would differ when allowing for sampling error, with a strictly larger variance in the case of extraneous variables and estimation based on least squares or maximum likelihood. We address parameter estimation and how to avoid overfitting in finite samples given large models in the next section.

the following expression for $\mathbf{c_t}$ as a function of historical forecast errors:

$$\begin{aligned}
\mathbf{c_t} &= -\mathbf{\Gamma_1}(\mathbf{\Delta X_t} - \boldsymbol{\mu}) \\
&= -\mathbf{\Gamma_1}\left\{\mathbf{F}(\mathbf{\Delta X_{t-1}} - \boldsymbol{\mu}) + \mathbf{H}\boldsymbol{e}_t\right\} \\
&= -\mathbf{\Gamma_1}\mathbf{H}\boldsymbol{e}_t - \left\{\mathbf{F}\mathbf{\Gamma_1}(\mathbf{\Delta X_{t-1}} - \boldsymbol{\mu})\right\} \\
&= -\mathbf{\Gamma_1}\mathbf{H}\boldsymbol{e}_t - \left\{\mathbf{\Gamma_2}(\mathbf{\Delta X_{t-1}} - \boldsymbol{\mu})\right\} \\
&= -\sum_{i=0}^{t-1}\mathbf{\Gamma}_{\boldsymbol{i+1}}\mathbf{H}\boldsymbol{e}_{\boldsymbol{t-i}} - \mathbf{\Gamma}_{\boldsymbol{t+1}}(\mathbf{\Delta X_0} - \boldsymbol{\mu}) \\
&\approx -\sum_{i=0}^{t-1}\mathbf{\Gamma}_{\boldsymbol{i+1}}\mathbf{H}\boldsymbol{e}_{\boldsymbol{t-i}}, \qquad (7)
\end{aligned}$$

where the approximation in the last line should be highly accurate for all but the first few time periods in a sample given that $\mathbf{\Gamma}_{\boldsymbol{t+1}}$ exponentially decays to zero as t increases for a stationary vector process.[4] Then, defining a selection vector $\mathbf{s_{r,q}}$ as a column of zeros in all r rows except for a 1 in the q^{th} row and assuming n variables and p lags in the VAR, we can account for the contribution of the forecast error for the k^{th} variable to the BN cycle of the l^{th}-ordered target variable as

$$c_{k,t} = -\sum_{i=0}^{t-1}\mathbf{s_{np,l}}'\mathbf{\Gamma}_{\boldsymbol{i+1}}\mathbf{H}\mathbf{s_{n,k}}\mathbf{s_{n,k}}'\boldsymbol{e}_{\boldsymbol{t-i}}. \qquad (8)$$

Given this calculation for the contribution of a particular variable, our proposed approach is to start with a VAR based on the entire available dataset and then drop variables with forecast errors that only contribute a negligible amount of variation to the BN cycle of the target variable until doing so leads a meaningful change in the cycle. In particular, Proposition 1 in Evans and Reichlin (1994) means that the BN cycle will not change as we go from $\mathbf{\Delta X_t} = \mathbf{Z_t}$ to $\mathbf{\Delta X_t} = \mathbf{W_t}$, but its amplitude will shrink as we go from $\mathbf{\Delta X_t} = \mathbf{W_t}$ to $\mathbf{\Delta X_t} = \mathbf{V_t}$. Thus, this procedure should result in determining which conditioning variables contain relevant information and belong in $\mathbf{w_t}$.

We note that this approach is not equivalent to simply including any variable that Granger causes the target variable in $\mathbf{\Delta x_t}$, as might be inferred from Remark 1 to Proposition 1 in Evans and Reichlin (1994). The issue is that Granger causation is a sufficient, but not necessary condition for belonging to $\mathbf{w_t}$. In particular, because the BN decomposition is based on an

[4] There is no approximation if the initial condition (i.e., $\mathbf{\Delta X_0} - \boldsymbol{\mu}$) is set to zero when backcasting based on to its unconditional expectation, which is what we do in our empirical application.

infinite-horizon forecast, a variable might not Granger cause the target variable, but still be relevant if it Granger causes another variable that Granger causes the target variable.[5] The key is that any variable for which removal from $\mathbf{\Delta x_t}$ alters the BN cycle for y_t belongs in $\mathbf{w_t}$.

As part of our approach, we can determine the relative importance of different sources of information in calculating the BN cycle by looking at the variance of the contribution $c_{k,t}$ for each remaining variable in the VAR. However, this is clearly not a strict variance decomposition as any correlation between forecast errors across equations will mean there are also non-zero covariance terms that affect the total variance of c_t. Likewise, we can obtain a historical decomposition in terms of which forecast errors explain the observed BN cycle by looking directly at $c_{k,t}$, with $c_t = c_{1,t} + c_{2,t} + ... + c_{n,t}$, but again this will not have a structural interpretation given correlation between forecast errors across equations. However, if we have an identification scheme that maps orthogonal structural shocks to forecast errors according to $\mathbf{A}\boldsymbol{\varepsilon_t} = \boldsymbol{e_t}$, where $\boldsymbol{\varepsilon_t}$ is a vector of structural shocks with covariance matrix $\mathbf{I}$, implying $\mathbf{\Sigma} = \mathbf{AA'}$, we can substitute in $\mathbf{A}\boldsymbol{\varepsilon_t}$ for the forecast errors in equation (8) to examine the role of the identified orthogonal structural shocks in driving the cycle.[6]

For completeness, we note that to solve for the BN trend growth as a function of forecast errors, we can difference equation (5) to get

$$\begin{aligned} \boldsymbol{\Delta\tau_t} &= \mathbf{X_t} + \mathbf{\Gamma_1}(\mathbf{\Delta X_t} - \boldsymbol{\mu}) - \{\mathbf{X_{t-1}} + \mathbf{\Gamma_1}(\mathbf{\Delta X_{t-1}} - \boldsymbol{\mu})\} \\ &= \boldsymbol{\mu} + \mathbf{\Gamma_0 H}\boldsymbol{e_t}. \end{aligned} \tag{9}$$

Again, we can account for the contribution of the forecast error for the k^{th} variable to the BN trend growth of the l^{th}-ordered target variable as

$$\Delta\tau_{k,t} = \mathbf{s_{np,l}}'\boldsymbol{\mu} + \mathbf{s_{np,l}}'\mathbf{\Gamma_0 H s_{n,k} s_{n,k}}'\boldsymbol{e_t}. \tag{10}$$

Similar to the case with the BN cycle, this calculation allows us to determine the importance

[5]This broader concept of forecast relevancy has previously been studied and described as "Granger Causal Priority" (e.g., Jarociński and Maćkowiak, 2017) or "long-run Granger causality" (e.g., Dufour and Renault, 1998).

[6]We present an example in the online appendix based on standard identification schemes for oil price shocks and monetary policy shocks. Importantly for large models such as in our empirical application, orthogonality makes it possible to examine the causal effects of a subset of structural shocks without necessarily identifying all of the structural shocks in a system.

of different sources of information in calculating the BN trend and possibly examine the role of the identified orthogonal structural shocks in driving the trend by substituting in $\mathbf{A}\boldsymbol{\varepsilon}_t$ for the forecast errors in equation (10).[7]

3 Estimation in Finite Samples

Evans and Reichlin (1994) show that estimates of the U.S. output gap based on multivariate BN decompositions are fundamentally different than those based on a univariate BN decomposition. In particular, they find that, in contrast to univariate estimates for a low-order AR model of output growth, multivariate estimates using information from the unemployment rate, aggregate consumption, and indices of leading and coincident indicators are large in amplitude and positively associated with NBER reference cycles. They interpret this finding as reflecting the relevance of the multivariate information for forecasting output growth, specifically in terms of capturing negative serial correlation at long horizons that is not captured by a typical univariate time series model.[8]

However, Evans and Reichlin (1994) also acknowledge that estimates of trend and cycle are sensitive to sampling error in finite samples. In particular, they note that the variance of the BN trend should be invariant to the information set in theory, but is not in practice due to large sampling error when estimating highly-parameterized models. Meanwhile, even given the same information set, they find that the amplitude of the BN cycle is sensitive to model specification, especially lag length. The general issue is that including more information in a forecasting model estimated by least squares or maximum likelihood in a finite sample will mechanically increase the implied predictability of output growth and, therefore, the amplitude of the BN cycle according to Proposition 1 in Evans and Reichlin (1994). This problem explains why Evans and Reichlin (1994) and others who consider the multivariate BN decomposition typically keep the number of variables in the information set relatively small and rely heavily on Granger causality tests to justify inclusion of variables even though the tests can suffer from

[7]In related work, Kamber and Wong (2018) use this approach to determine the roles of foreign and domestic shocks in driving both trend and cycle of inflation in small open economies using a block exogeneity assumption for the foreign variables.

[8]This interpretation is directly supported by the findings in Kamber et al. (2018) that imposing a low signal-to-noise ratio in terms of the variance of trend shocks as a fraction of the overall forecast error variance restricts a univariate model to imply negative serial correlation and produces a BN cycle estimate of the U.S. output gap with large amplitude and positive association with NBER reference cycles.

low power in finite samples and variables could contain relevant information even if they do not directly Granger cause output growth, as discussed in the previous section.

Here, we propose how to address practical concerns about overfitting in finite samples when estimating large models to ensure inclusion of all relevant information. In particular, we are motivated by the literature on using shrinkage priors in Bayesian VARs (BVARs) for forecasting going back to Litterman (1986) and Robertson and Tallman (1999), but especially more recent studies on estimating large systems corresponding to entire datasets, such as Banbura et al. (2010). Building on these studies, we employ a Minnesota-type shrinkage prior with a key hyperparameter that we calibrate to minimize the pseudo-out-of-sample forecast error variance for our target variable y_t.

The implementation of our shrinkage prior is best illustrated by directly considering the VAR(p) for the demeaned vector of variables, $\mathbf{\Delta\tilde{x}_t} \equiv \mathbf{\Delta x_t} - \boldsymbol{\mu}$:

$$\begin{aligned}
\mathbf{\Delta\tilde{x}_t} &= \boldsymbol{\Phi_1}\mathbf{\Delta\tilde{x}_{t-1}} + \ldots + \boldsymbol{\Phi_p}\mathbf{\Delta\tilde{x}_{t-p}} + \boldsymbol{e_t} \\
&= \begin{bmatrix} \boldsymbol{\Phi_1} & \boldsymbol{\Phi_2}\ldots\boldsymbol{\Phi_p} \end{bmatrix} \begin{bmatrix} \mathbf{\Delta\tilde{x}_{t-1}} \\ \mathbf{\Delta\tilde{x}_{t-2}} \\ \vdots \\ \mathbf{\Delta\tilde{x}_{t-p}} \end{bmatrix} + \boldsymbol{e_t} \\
&= \begin{bmatrix} \phi_1^{11} & \ldots & \phi_1^{1n} & \phi_2^{11} & \ldots & \phi_2^{1n} & \ldots & \ldots & \phi_p^{1n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots \\ \phi_1^{n1} & \ldots & \phi_1^{nn} & \phi_2^{n1} & \ldots & \phi_2^{nn} & \ldots & \ldots & \phi_p^{nn} \end{bmatrix} \begin{bmatrix} \mathbf{\Delta\tilde{x}_{t-1}} \\ \mathbf{\Delta\tilde{x}_{t-2}} \\ \vdots \\ \mathbf{\Delta\tilde{x}_{t-p}} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ \vdots \\ e_{n,t} \end{bmatrix}, \qquad (11)
\end{aligned}$$

where $\mathbb{E}(\boldsymbol{e_t'e_t}) = \boldsymbol{\Sigma}$ and $\mathbb{E}(\boldsymbol{e_t'e_{t-i}}) = \mathbf{0}$ $\forall i > 0$. The use of demeaned observations $\mathbf{\Delta\tilde{x}_t}$ is equivalent to setting a flat prior on the unconditional means. Shrinkage is then applied to the slope coefficients using a Minnesota-type prior specification. In particular, letting ϕ_i^{jk} denote the slope coefficient of the i^{th} lag of variable k in the j^{th} equation of the VAR in equation (11), we set the prior means and variances of the slope coefficients as follows:

$$\mathbb{E}[\phi_i^{jk}] = 0 \tag{12}$$

$$Var[\phi_i^{jk}] = \begin{cases} \frac{\lambda^2}{i^2}, & j = k \\ \frac{\lambda^2}{i^2}\frac{\sigma_j^2}{\sigma_k^2}, & \text{otherwise}, \end{cases} \tag{13}$$

where the degree of shrinkage is governed by the hyperparameter λ, with $\lambda \to 0$ shrinking to the assumption that the variables in the VAR are independent white noise processes or, equivalently for all of the differenced variables in the VAR, independent random walk processes in levels. However, given $\lambda > 0$, the prior is not dogmatic and, following standard results for Bayesian estimation of regression models (e.g., see Koop, 2003), the posterior will converge asymptotically to the population parameters assumed in the analysis in the previous section.[9] Following the standard Minnesota prior structure, the factor $1/i^2$ shrinks coefficients at longer lags closer to zero. The variances σ_j^2 and σ_k^2 are set to the variances of residuals from AR(4) models estimated using least squares for the corresponding variables as per the usual practice (e.g., Banbura et al., 2010; Koop, 2013).

In terms of setting the shrinkage hyperparameter λ, we are motivated by methods in previous studies on forecasting with large BVARs that generally imply more shrinkage – i.e., setting λ closer to zero – as the number of variables in the model increases. The basic idea is to avoid overfitting in finite samples as the number of parameters in the VAR proliferates with the number of variables. Banbura et al. (2010) suggest choosing λ so as to match the in-sample fit of a large model with that of a smaller 3-variable system based on least squares, Giannone et al. (2015) suggest placing a hierarchical prior on λ and directly estimating it, while Carriero et al. (2015) suggest choosing λ to maximize the marginal data density of the model. The first approach is very application-specific in terms of the choice of the smaller 3-variable system, although we draw from this general idea by choosing λ to maximize the fit of a particular target

[9]One possible concern is that the shrinkage to random walk processes is so strong that there is a significant downward bias in the variance of the estimated cycle for the target variable. To address this, we consider a modification of the Minnesota prior in the online appendix that shrinks to a small signal-to-noise ratio in terms of the variance of trend shocks as a fraction of the overall forecast error variance that is consistent with what was found in the univariate analysis of Kamber et al. (2018) rather than the larger value of 1 implied by a random walk. We find that this modification has virtually no impact on the estimated cycle in our empirical application, including when considering as many as 138 variables, suggesting that, even with small λ, the standard Minnesota prior is diffuse enough that it is dominated by the likelihood for the implied signal-to-noise ratio. See the online appendix for full details.

variable rather than the fit of the entire system, which is the implicit goal of the second and third approaches. The problem with focusing on the fit of the entire system is that it leads the BN cycle to always be sensitive to the size of the model. Intuitively, as the number of variables increases, relatively less weight is put on the fit of the target variable, causing the BN cycle to change even as extraneous information is added to the system. Meanwhile, we focus on fit in terms of point forecasts, while the second and third approaches are technically related to density forecasts, although point and density forecast accuracy are often closely related in practice. Our focus on point forecast accuracy is consistent with the definition of the BN decomposition in terms of a long-horizon point forecast.

The specifics of our approach are that we conduct numerical optimization to find the λ that minimizes the one-step-ahead root mean squared forecast error (RMSFE) for the target variable y_t over an evaluation sample using pseudo-real-time estimation based on an expanding window starting with a particular initial fraction of the full sample.[10] In addition to generally providing more shrinkage as the number of variables included in the model increases, our focus on pseudo-out-of-sample RMSFE serves another purpose within the context of performing a trend-cycle decomposition with the BN decomposition. In particular, Nelson (2008) argues that an AR(1) model of output growth is a sensible choice for performing a BN decomposition for aggregate output because, while extremely parsimonious, it produces comparatively good out-of-sample forecasts. Thus, we view a competitive out-of-sample forecast for output growth relative to a univariate AR(1) model as crucial to address Nelson's critique of standard approaches to trend-cycle decomposition such as the HP filter that have implicit out-of-sample forecasts of output growth that perform much worse (e.g., see Kamber et al., 2018).

Conditional on λ and the assumption of Normality for the variables, the calculation of posterior moments for the slope coefficients is straightforward, as the natural conjugacy of the prior implies that we can implement estimation using least squares with dummies observations (e.g., Del Negro and Schorfheide, 2011; Woźniak, 2016). For brevity, we relegate these details to the online appendix.

[10]For our application, we start our recursive estimation with the initial 20 years of data (roughly one-third of the 230 quarterly observations) and use the remaining almost 40 years of data for evaluation of the RMSFE, although we note that the results for our benchmark model are quite robust to an initial sample as short as the first 10 years of data and the remaining almost 50 years of data used for evaluation. We also note that λ can be determined by a simple grid search, which we do in our application in order to report how the RMSFE changes for different values of the hyperparameter. However, we generally recommend numerical optimization, as it is much faster.

4 Application to the U.S. Output Gap

For our empirical application, we consider a dataset with 138 variables covering various aspects of the U.S. economy. In our benchmark model, we focus on the following 23 variables: the oil price, real GDP, the CPI, the unemployment rate, hourly earnings, the fed funds rate, stock prices, the slope of the yield curve, GDP deflator, employment, personal income, real PCE, industrial production, capacity utilization, housing starts, producer price index for all commodities, PCE deflator, hours worked, nonfarm real output per hour, total reserves, non-borrowed reserves, real M1, and real M2. This choice of variables is informed by the 20-variable model in Banbura et al. (2010), which in turn was informed by an influential monetary VAR study by Christiano et al. (1999). In particular, Banbura et al. (2010) suggest that their medium-sized 20-variable model covers a sufficiently broad set of information for macroeconomic forecasting purposes and so we believe it serves as a reasonable starting point for a model which should contain most, if not all, of the relevant information for estimating the output gap via a multivariate BN decomposition. However, for completeness and comparability to the larger BVAR in Banbura et al. (2010), as well as FAVARs (e.g., Bernanke et al., 2005), we also consider a model with the full 138 variables in the dataset, where many of the additional variables are subcomponents of the 23 variables in our benchmark model. All of the raw data are sourced from IFS and FRED and considered at a quarterly frequency for a sample period covering 1959-2016. Definitions and details of the data are available in the online appendix.

We take natural logarithms of the data when appropriate and then differences if either a unit root test cannot reject at a 5% level of significance or a t-test can reject a constant mean across the first and second halves of the sample at a 10% level of significance.[11] We transform the data in this way because stationarity is necessary to construct BN trends and cycles following the method described in Section 2.[12] All series, once transformed to be stationary, are backcast using a sample average so as to keep the initial observations as part of the estimation sample from 1959Q3 to 2016Q4. Unless otherwise noted, the lag length is always set to $p = 4$, as is

[11]This relatively simple approach to addressing possible structural change is for convenience given the large number of variables under consideration. However, it would certainly be possible to consider more formal tests for structural breaks at unknown breakpoints and adjust series accordingly. Meanwhile, we consider a possible break in mean of real GDP growth and other variables as a robustness check in Section 5 below.

[12]Preliminary analysis showed that, despite shrinkage to white noise processes, incorporating very persistent variables in the BVAR, including those with apparent large shifts in their level, results in BN cycles that appear to drift up or down over rather than reverting to a mean of zero.

often considered for quarterly data.

Before estimating our 23-variable benchmark model using Bayesian shrinkage, we motivate our approach by considering the estimated output gap based on the BN decomposition using least squares estimation of smaller models.[13] In particular, Figure 1 plots the estimated output gap for the cases of a univariate AR model of output growth, a bivariate VAR model of output growth and the unemployment rate, and a four-variable VAR model of output growth, the unemployment rate, CPI inflation, and the growth rate of industrial production. Consistent with the findings in Evans and Reichlin (1994), the inclusion of multivariate information serves to capture negative serial correlation in output growth at longer horizons and leads to more intuitive estimates of the output gap that are positively associated with NBER reference cycles. However, it is notable just how sensitive the estimates are to the information included in the forecasting model, with the amplitude of the estimated output gap increasing substantially as more information is added. This result could be because, as we will see with our analysis below, the unemployment rate and CPI inflation are particularly important conditioning variables for estimating the output gap. But Proposition 1 in Evans and Reichlin (1994) also means that the amplitude of the BN cycle for the target variable will necessarily increase in a purely mechanical way when more variables are added to the forecasting model given least squares estimation in finite samples. Thus, some of the increase in amplitude that can be seen for additional variables in Figure 1 is likely to be overstated due to sampling error. At the same time, it is possible that other variables are relevant too, but adding them would mechanically increase the amplitude further due to sampling error.

Our proposed approach attempts to overcome the shortcomings with conducting a multivariate BN decomposition using least squares. In particular, we determine which additional variables should be included by looking at contributions of forecast errors to the BN cycle and we use Bayesian shrinkage to avoid overfitting when estimating models large enough to include many possible sources of information. To begin, Figure 2 plots the estimated output gap for our benchmark model, along with 90% confidence bands calculated using the approach for the BN decomposition in Kamber et al. (2018). The output gap is positively associated with NBER reference cycles and tends to be most significantly different from zero during recessions and the

[13]Least squares estimation is equivalent to maximum likelihood under an assumption of Normality of the variables that is also maintained when conducting the Bayesian estimation.

ends of expansions.[14] Despite considering 23 variables in our benchmark BVAR, the amplitude of the estimated output gap is reasonably similar to that of the four-variable case in Figure 1 based on least squares rather than being much larger, as would be the case using least squares instead of Bayesian shrinkage.

An immediate question arises as to whether all 23 variables in the benchmark model contain relevant information. To answer this question, we consider the contributions of different forecast errors to the estimated output gap based on the formula given in equation (8). Figure 3 reports the standard deviations of these contributions in order to measure the relevance of different sources of information in the forecasting model. The first thing to notice is that a number of variables appear to contain more information about the output gap than output growth itself. This directly explains why the estimated output gap in Figure 2 is so different than the univariate case in Figure 1. The most important conditioning variables are the unemployment rate and CPI inflation that were also considered in the multivariate models in Figure 1. However, other variables appear relevant too, especially real PCE, housing starts, the federal funds rate, real M1, and stock prices.

To confirm the relevance of different sources of information identified by our approach, we also consider an eight-variable BVAR that includes the seven most informationally-relevant conditioning variables implied by Figure 3, in addition to output growth itself. For completeness, we also consider a 138-variable BVAR that includes our whole dataset to see if we have omitted any important information in our benchmark model. Finally, we consider what happens if we drop the most relevant conditioning variable, the unemployment rate, from our benchmark model.

Figure 4 plots the estimated output gap for BVARs with different conditioning variables. The top panel compares the eight-variable, 23-variable, and 138-variable cases. The estimates in all three cases are very similar, except for some differences for the 138-variable case post 2008 that we will examine in our robustness analysis below in Section 5. Notably, by dropping the less informationally-relevant variables from the benchmark model, the estimated output gap is barely affected. This suggests that our approach to selecting conditioning variables based on informational relevance works well in practice. Meanwhile, it is not simply the case

[14] We note that the degree of uncertainty about the exact level of the output gap is consistent with findings for other measures of the output gap in Garratt et al. (2014).

that the estimated output gap is so similar across the different information sets due to Bayesian shrinkage. In particular, the bottom panel of Figure 4 compares the benchmark 23-variable case with the same model except without the unemployment rate and the results make it clear that the estimated output gap can be quite sensitive to the omission of even just one conditioning variable if it contains highly relevant information.[15]

Figure 5 plots the pseudo-out-of-sample RMSFE for output growth as a function of the shrinkage hyperparameter λ for the eight-variable, 23-variable, and 138-variable cases. As expected, our approach produces more shrinkage as the system gets larger, with the λ that minimizes RMSFE decreasing as more variables are added. Interestingly, the minimum RMSFE itself does not change much across models and is virtually identical in the 23-variable and 138-variable cases. But, notably, the RMSFE is lower than in the random walk case, which corresponds to $\lambda = 0$ based on a dogmatic Minnesota prior, and the case of an AR(1) model estimated using least squares for which the RMSFE is given by the horizontal line in Figure 5. An AR(1) model is notoriously challenging to beat in out-of-sample forecasts of output growth, as noted in the critique by Nelson (2008) mentioned in the previous section. The fact, then, that the estimated output gap in Figure 2 is based on a model that performs better than an AR(1) model in out-of-sample forecasts provides some assurance that it is not spurious.

A related question is whether our estimated output gap is relevant for policy, as well as how it compares to other measures typically used in a policy environment. To examine this, we consider the correlations between different measures of the output gap and future values of important macroeconomic variables. In particular, a relevant output gap should have a negative relationship with future output growth and a positive relationship with future inflation. We find that our estimated output gap has a correlation of -0.44 with output growth over the subsequent four quarters (i.e., $corr(c_t, y_{t+4} - y_t)$, where c_t is the BN cycle for our benchmark model and y_t is the natural logarithm of quarterly real GDP). The corresponding correlations for the CBO output gap, a one-sided HP filter, and the Hamilton (forthcoming) filter are -0.18, 0.18, and 0.00, respectively. Thus, our output gap performs best, despite the CBO output gap in particular being a revised measure, with revisions likely influenced by realized values of future output growth.[16] Consistent with the critique of the HP filter by Hamilton (forthcoming), the

[15] This finding suggests that it is the inclusion of key informational variables for the BN cycle of the target variable that matters, not the size of the information set per se.

[16] To understand the problem with using a revised measure for this evaluation, note that the standard two-

output gap based on his approach performs better than the one-sided HP filter measure of the output gap. However, it does not provide any guidance on the future direction of real activity for the economy that one would expect an accurate measure of the output gap to do. At the same time, our measure is explicitly constructed based on optimizing (one-step-ahead pseudo-out-of-sample) forecast accuracy for output growth, so the strong negative correlation with future output growth in our case is perhaps not so surprising. However, our estimated output gap also has a correlation of 0.24 with inflation over the subsequent four quarters (i.e., $corr(c_t, p_{t+4} - p_t)$, where p_t is the natural logarithm of the CPI) compared to corresponding correlations for the CBO output gap, a one-sided HP filter, and the Hamilton filter of 0.28, -0.03, and 0.07, respectively. That is, our approach does almost as well as the CBO output gap that is, again, a revised measure and much better than the one-sided HP filter and the Hamilton filter.[17] Taken together, these correlations suggest the output gap estimated using a BVAR provides a useful measure for monitoring the stance of the economy in terms of how it relates to future output growth and inflation.

5 Robustness

In this section, we consider the robustness of our empirical results to accounting for possible structural change in the long-run growth rate of the U.S. economy and to an alternative measure of aggregate economic activity, with a particular focus on understanding the behavior of the output gap in the Great Recession and its immediate aftermath. We also consider the robustness of our approach to possible misspecification of the size of a system according to empirically-motivated Monte Carlo simulations.

First, recall that Figure 4 in the previous section suggests that the estimated output gap

sided HP filter also produces a measure of the output gap that is mechanically influenced by realized values of future output growth. It has a correlation of -0.52 with future output growth, but this is clearly spurious in the sense that it must reflect a 'look-ahead bias' from being influenced by realized values of future output growth in its calculation given that the one-sided HP filter output gap has a correlation of 0.18 with future output growth.

[17] As before, the influence of realized values of future output growth in driving a spurious correlation can be seen by comparing the correlation for the standard two-sided HP filter, which is 0.26, with the correlation of -0.03 for the one-sided HP filter. The bandpass filter is also two-sided and the correlations of its measure of the output gap with future output growth and future inflation are very similar to those for the two-sided HP filter at -0.54 and 0.32, respectively. A full analysis of real-time out-of-sample forecasting performance of different measures of the output gap such as in Guérin et al. (2015) and Kamber et al. (2018) would be worth exploring, but is left to future research given challenges in accounting for data revisions for such a large dataset.

is not completely robust to the size of the system during the Great Recession, although it is, perhaps surprisingly, always smaller in magnitude than in the 1981-82 recession. In a univariate setting, accounting for a break in long-run output growth around the Great Recession has been shown to significantly alter inference about the output gap based on long-run forecasts (e.g., Eo and Morley, 2017; Kamber et al., 2018). Thus, we check if this is also the case in a multivariate setting, with the estimates across different-sized systems possibly influenced to differing degrees by a failure to account for structural change.

For our estimation sample, Bai and Perron (2003) procedures suggest a significant break in the unconditional mean of output growth estimated in 2006Q2. To account for this break, we allow for a one-time break in the mean of all variables in each BVAR in 2006Q2. To the extent that other variables share a common break with output growth, this will be taken into account, while variables without such a break will be little affected by allowing for a redundant break in mean in estimation.[18]

Figure 6 plots the estimated output gap when allowing for a structural break in 2006Q2. The top panel compares the estimated output gap for a 23-variable BVAR allowing for a break to the benchmark case without a break presented in the previous section. The estimates are reasonably robust, although the amplitude is a bit larger when allowing for a break, especially in the latter half of the sample. Meanwhile, the bottom panel compares the estimated output gaps for the eight-variable, 23-variable, and 138-variable cases allowing for a structural break in each case. Notably, the estimated output gap is more robust across all three cases post 2008 than in in Figure 4, suggesting that a failure to account for structural change could help explain the differences for the 138-variable case reported in the previous section.

Notably, however, the estimated output gap remains smaller in magnitude in the Great Recession than in the 1981-82 recession even when accounting for a structural change in long-run growth. One possibility is that this result could reflect measurement issues with real GDP that may have been particularly severe around the Great Recession and may be mitigated by considering a measure of real GDI instead (see, for example, Nalewaik, 2010). To check this, we construct real GDI using nominal GDI and the GDP deflator from FRED for the same sample period as the other series and apply our approach to estimate the output gap using real GDI

[18] We leave a more in-depth analysis of structural breaks in a multivariate setting to future research.

instead of real GDP.[19]

Figure 7 plots the estimated output gap using real GDP versus real GDI for the benchmark BVAR specification. In terms of the shape of the output gap, the estimates are reasonably robust across the two measures. However, the amplitude of the estimated output gap is a bit larger in the case of real GDI, suggesting via Proposition 1 in Evans and Reichlin (1994) that it is somewhat more closely related to the conditioning variables in our benchmark model, which is consistent with the general findings in Nalewaik (2010) about GDI. Also, the estimated output gap is larger in magnitude in the Great Recession when using real GDI. However, we note that it is still not as large in magnitude as in the 1981-82 recession.

To understand the possibly surprising, but seemingly robust result about the relative magnitudes of the output gap in 1981-82 versus the Great Recession, we consider a historical decomposition of trend growth based on the formula given in equation (10). The smaller magnitude of the estimated output gap in 2007-09 despite the larger drop in log output compared to the 1981-82 recession directly corresponds to the idea that the trend fell by more in the Great Recession, consistent with the assessment by the President of the Federal Reserve Bank of St. Louis, James Bullard, that the Great Recession resulted in large permanent decreases in output that cannot be expected to be reversed (see Bullard, 2012). Meanwhile, according to the historical decomposition, the large decline in trend in the Great Recession can be substantially accounted for by large negative forecast errors for real PCE growth that were less prominent in the 1981-82 recession.[20]

Figure 8 presents the results on the role of consumption in explaining the differences in trend growth between the early 1980s and the post-2008 period. The top panel reports the change in trend output (cumulated over time) that is accounted for by the forecast errors for real PCE growth, comparing two periods, 1980Q1-1983Q1 and 2008Q1-2013Q4.[21] Evidently, the forecast errors for consumption only lower the estimate of trend output in the early 1980s by less than -0.5%. By contrast, the forecast errors for consumption lower the estimate of

[19]The FRED mnemonic for real GDI is A261RX1Q020SBEA. The ADF and structural break test results are the same for real GDI as for real GDP.

[20]As discussed in Section 2, this is not a structural decomposition because forecast errors can be correlated across variables. However, to the extent that a basic version of the permanent income hypothesis holds, aggregate consumption will follow a random walk and shocks to it would correspond to shocks to permanent income. Thus, we can interpret the differences in trend growth across recessions as possibly being due to differences in shocks to permanent income.

[21]The estimates are based on the benchmark model for real GDP in the previous section, but they are largely robust to the other models, accounting for a structural break in 2006, or using real GDI.

trend output by about 2% in the post-2008 period. To help see what is going on, the bottom panel plots consumption growth during the two time periods. Consumption growth remained sluggish throughout the Great Recession compared to the early 1980s, when a sharp fall in the consumption growth rate in 1980Q2 was immediately reversed in the following quarter and consumption growth was actually positive for most of the 1981-82 recession.

The final robustness issue that we consider is what happens if the size of the system is misspecified. In particular, we conduct empirically-motived Monte Carlo simulations based on three DGPs corresponding to i) an estimated univariate AR model of U.S. output growth using Bayesian shrinkage, ii) our estimated eight-variable BVAR, and iii) our estimated benchmark 23-variable BVAR. We generate 1000 artificial datasets for each of the three DGPs of length $T = 230$, consistent with the length of the sample in our empirical analysis. The true output gap is defined as the BN decomposition of output based on the population parameters for a given DGP. For each artificial dataset, we estimate univariate and multivariate models using least squares and Bayesian estimation. In each case, we estimate the output gap via the BN decomposition based on the estimated parameters. For each draw, we calculate the root mean squared error (RMSE) of the estimated output gap relative to the true cycle. For each DGP, we therefore consider one correct specification in terms of the number and set of variables and two other specifications that do not coincide with the number or set of variables in the underlying DGP.[22]

Table 1 reports the results for our Monte Carlo analysis. The top panel (a) presents the mean and standard deviation of the RMSEs across draws in each case of estimated model size for a given DGP. The bottom panel (b) presents the proportion of Monte Carlo draws where our BVAR approach has a lower RMSE compared to estimates based on least squares. In every case, Bayesian estimation is more accurate than least squares, although the differences increase with the size of the system and the size of the estimated model. In cases where the size of the model is correctly specified (i.e., the diagonals in both panels (a) and (b)), it is straightforward to see that the differences in mean and standard deviation of the RMSE increase with the

[22]For variables that do not feature in the underlying DGP, we use historical realized values from our dataset in our empirical analysis as the additional data considered when estimating larger models for the Monte Carlo analysis. For example, for our eight-variable DGP, we use the eight simulated series for the variables in the DGP and the 15 additional historical data series for the variables that are not in the DGP when estimating a 23-variable model. By construction, the additional variables are extraneous and will be irrelevant in population, but will lead to sampling error in finite-sample estimation.

number of variables/parameters, with the mean RMSE twice as large for least squares than Bayesian estimation in the 23-variable case and the BVAR producing a more accurate estimate of the output gap more than 99% of the time. Notably though, Bayesian shrinkage is not only useful when the size of the model is correctly specified, but also when it is not (i.e., the off-diagonals in both panels (a) and (b)). When the model size is misspecified, our Monte Carlo results suggest that shrinkage still helps keep the estimate of the output gap comparatively accurate. In particular, we find that, under misspecification, our approach produces a more accurate estimate of the output gap between 98-100% of the time when the estimated model is too large and between 60-81% of the time when the estimated model is too small. Thus, there is possibly some justification for estimating a larger BVAR than might be necessary, as we may have done with our 23-variable benchmark model in our empirical application. In particular, the mean RMSEs for the BVARs in our Monte Carlo analysis are very similar for eight- and 23-variable models given one- and eight-variable DGPs, while it is, of course, lower for the 23-variable DGP.

6 Conclusion

In this paper, we have shown how to apply the Beveridge-Nelson decomposition to obtain estimates of trend and cycle using large vector autoregressions estimated with Bayesian shrinkage. We have also shown how to account for and interpret the various sources of multivariate information contributing to the estimates of trend and cycle. In our empirical application, we present estimates of the U.S. output gap based on information sets containing as many as 138 variables. We find that the unemployment rate, inflation, and, to a lesser extent, housing starts, aggregate consumption, stock prices, real money balances, and the federal funds rate contain relevant information for estimating the U.S. output gap. Our findings are robust to consideration of structural change and using a real GDI measure of aggregate output. Monte Carlo analysis suggests that the Bayesian approach produces estimates of the cycle that are closer to population values in finite samples than using least squares and that are more robust to misspecification of the relevant information set.

We view two advantages of the approach proposed in this paper that motivate future extensions and applications. The first advantage is that casting the detrending problem within a regression framework allows us to utilize complicated datasets for estimating trend and cycle. In particular, many time series problems can naturally be cast into the vector autoregressions considered in this paper. For example, policy institutions often construct an output gap measure by monitoring a very broad set of data at differing frequencies. One could cast the problem into a vector autoregression with mixed frequencies and thus allow information from monthly data to directly enter the problem of nowcasting the output gap, even though real GDP is often only available at a quarterly frequency. Likewise, large vector autoregressions with time-varying parameters (e.g., Koop and Korobilis, 2013) or combinations of output gaps for different models (e.g., Morley and Piger, 2012; Garratt et al., 2014; Guérin et al., 2015) could be considered. Another potential extension is joint detrending. Although we only target a single variable in order to estimate the output gap, it would be reasonably straightforward to modify our approach to target multiple variables at once in order to obtain estimates of, say, the natural rate of interest, trend inflation, and the natural rate of unemployment, in addition to the output gap, within a unified and consistent multivariate framework. A second advantage of our approach is the ability to interpret trend and cycle by appealing further to tools from the well-developed literature on structural vector autoregressions. This would allow us to meaningfully discuss shocks driving the trend and cycle and to attribute causality. The standard frameworks of trend-cycle decomposition using time series methods like unobserved components models can struggle to attribute causality, in addition to being more difficult to estimate than the models we propose, especially given large information sets. For example, Kamber and Wong (2018) employ the methods introduced in this paper to estimate the role of foreign shocks in driving trend inflation and the inflation gap for a number of small open economies. One could similarly use the tools we introduce in this paper to answer relevant policy questions such as what drives low neutral interest rates or what drives financial cycles. However, we leave this analysis to future research.

References

Bai J, Perron P. 2003. Computation and Analysis of Multiple Structural Change Models. *Journal of Applied Econometrics* **18**: 1–22.

Banbura M, Giannone D, Reichlin L. 2010. Large Bayesian Vector Auto Regressions. *Journal of Applied*

Econometrics **25**: 71–92.

Barigozzi M, Luciani M. 2017. Common Factors, Trends, and Cycles in Large Datasets. Finance and economics discussion series, Board of Governors of the Federal Reserve System (U.S.).

Bernanke BS, Boivin J, Eliasz P. 2005. Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach. *The Quarterly Journal of Economics* **120**: 387–422.

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.

Bullard J. 2012. Inflation Targeting in the USA. Speech at the Union League Club of Chicago, Breakfast@65West Chicago, Feb 6, 2012.

Carriero A, Clark TE, Marcellino M. 2015. Bayesian VARs: Specification Choices and Forecast Accuracy. *Journal of Applied Econometrics* **30**: 46–73.

Chan JCC, Grant AL. 2017. Measuring the Output Gap using Stochastic Model Specification Search. CAMA Working Papers 2017-02, Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University.

Christiano LJ, Eichenbaum M, Evans CL. 1999. Monetary Policy Shocks: What Have We Learned and to What End? In Taylor JB, Woodford M (eds.) *Handbook of Macroeconomics*, volume 1 of *Handbook of Macroeconomics*, chapter 2. Elsevier, 65–148.

Christiano LJ, Fitzgerald TJ. 2003. The Band Pass Filter. *International Economic Review* **44**: 435–465.

Del Negro M, Schorfheide F. 2011. Bayesian Macroeconometrics. In Geweke J, Koop G, Van Dijk H (eds.) *The Oxford Handbook of Bayesian Econometrics*. Oxford University Press, 293–389.

Dufour JM, Renault E. 1998. Short Run and Long Run Causality in Time Series: Theory. *Econometrica* **66**: 1099–1126.

Eo Y, Morley J. 2017. Why Has the U.S. Economy Stagnated since the Great Recession? Working Paper 2017-14, University of Sydney, School of Economics.

Evans G, Reichlin L. 1994. Information, Forecasts, and Measurement of the Business Cycle. *Journal of Monetary Economics* **33**: 233–254.

Garratt A, Lee K, Shields K. 2016. Information Rigidities and the News-Adjusted Output Gap. *Journal of Economic Dynamics and Control* **70**: 1–17.

Garratt A, Mitchell J, Vahey SP. 2014. Measuring Output Gap Nowcast Uncertainty. *International Journal of Forecasting* **30**: 268–279.

Garratt A, Robertson D, Wright S. 2006. Permanent vs Transitory Components and Economic Fundamentals. *Journal of Applied Econometrics* **21**: 521–542.

Giannone D, Lenza M, Primiceri GE. 2015. Prior Selection for Vector Autoregressions. *The Review of Economics and Statistics* **97**: 436–451.

Guérin P, Maurin L, Mohr M. 2015. Trend-Cycle Decomposition of Output and Euro Area Inflation Forecasts: A Real-Time Approach Based on Model Combination. *Macroeconomic Dynamics* **19**: 363–393.

Hamilton JD. forthcoming. Why You Should Never Use the Hodrick-Prescott Filter. *The Review of Economics and Statistics* doi: 10.1162/REST_a_00706.

Hodrick RJ, Prescott EC. 1997. Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit and Banking* **29**: 1–16.

Jarociński M, Maćkowiak B. 2017. Granger Causal Priority and Choice of Variables in Vector Autoregressions. *The Review of Economics and Statistics* **99**: 319–329.

Kamber G, Morley J, Wong B. 2018. Intuitive and Reliable Estimates of the Output Gap from a Beveridge-Nelson Filter. *The Review of Economics and Statistics* **100**: 550–566.

Kamber G, Wong B. 2018. Global Inflation and Trend Inflation. BIS Working Papers 688, Bank for International Settlements, Basel.

Koop G. 2003. *Bayesian Econometrics.* John Wiley & Sons Inc.

Koop G, Korobilis D. 2013. Large Time-Varying Parameter VARs. *Journal of Econometrics* **177**: 185 – 198.

Koop GM. 2013. Forecasting with Medium and Large Bayesian VARs. *Journal of Applied Econometrics* **28**: 177–203.

Kozicki S. 1999. Multivariate Detrending under Common Trend Restrictions: Implications for Business Cycle Research. *Journal of Economic Dynamics and Control* **23**: 997–1028.

Litterman RB. 1986. Forecasting with Bayesian Vector Autoregressions: Five Years of Experience. *Journal of Business & Economic Statistics* **4**: 25–38.

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, 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.

Nalewaik JJ. 2010. The Income- and Expenditure-Side Estimates of U.S. Output Growth. *Brookings Papers on Economic Activity* **41**: 71–127.

Nelson CR. 2008. The Beveridge-Nelson Decomposition in Retrospect and Prospect. *Journal of Econometrics* **146**: 202–206.

Robertson JC, Tallman EW. 1999. Vector Autoregressions: Forecasting and Reality. *Economic Review* : 4–18.

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.

Woźniak T. 2016. Bayesian Vector Autoregressions. *Australian Economic Review* **49**: 365–380.

Table 1: Monte Carlo Results

(a) Mean and standard deviation of RMSEs

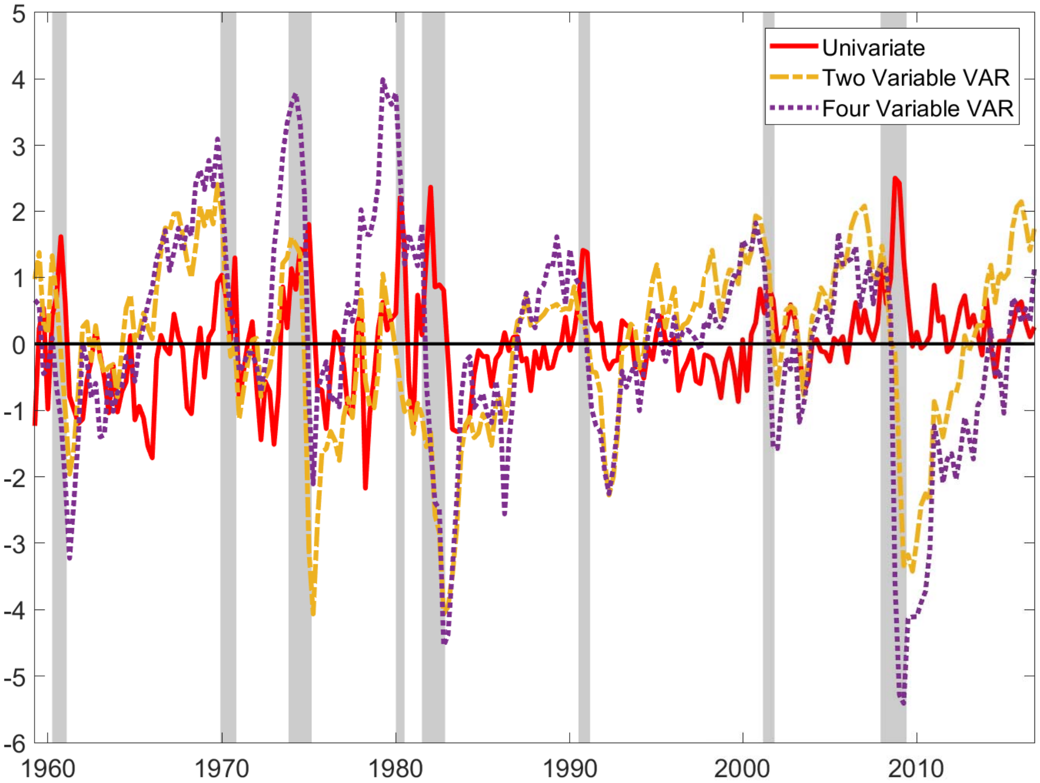
			Estimated Model Size:		
			1 variable	**8 variables**	**23 variables**
	1 variable	**Shrinkage**	0.30	1.18	1.03
			(0.171)	(0.682)	(0.518)
		OLS	0.33	2.26	4.54
			(0.231)	(1.011)	(2.221)
	8 variables	**Shrinkage**	1.45	0.81	0.82
DGP:			(0.198)	(0.409)	(0.355)
		OLS	1.48	1.03	1.66
			(0.231)	(0.553)	(0.684)
	23 variables	**Shrinkage**	1.34	0.73	0.64
			(0.194)	(0.379)	(0.366)
		OLS	1.38	0.95	1.28
			(0.234)	(0.538)	(0.569)

(b) Proportion of Monte Carlo Draws $\mathrm{RMSE}_{\mathrm{Shrinkage}} < \mathrm{RMSE}_{\mathrm{MLE}}$

		Estimated Model Size:		
		1 variable	**8 variables**	**23 variables**
	1 variable	0.53	0.98	1.00
DGP:	**8 variables**	0.60	0.74	0.97
	23 variables	0.67	0.81	0.99

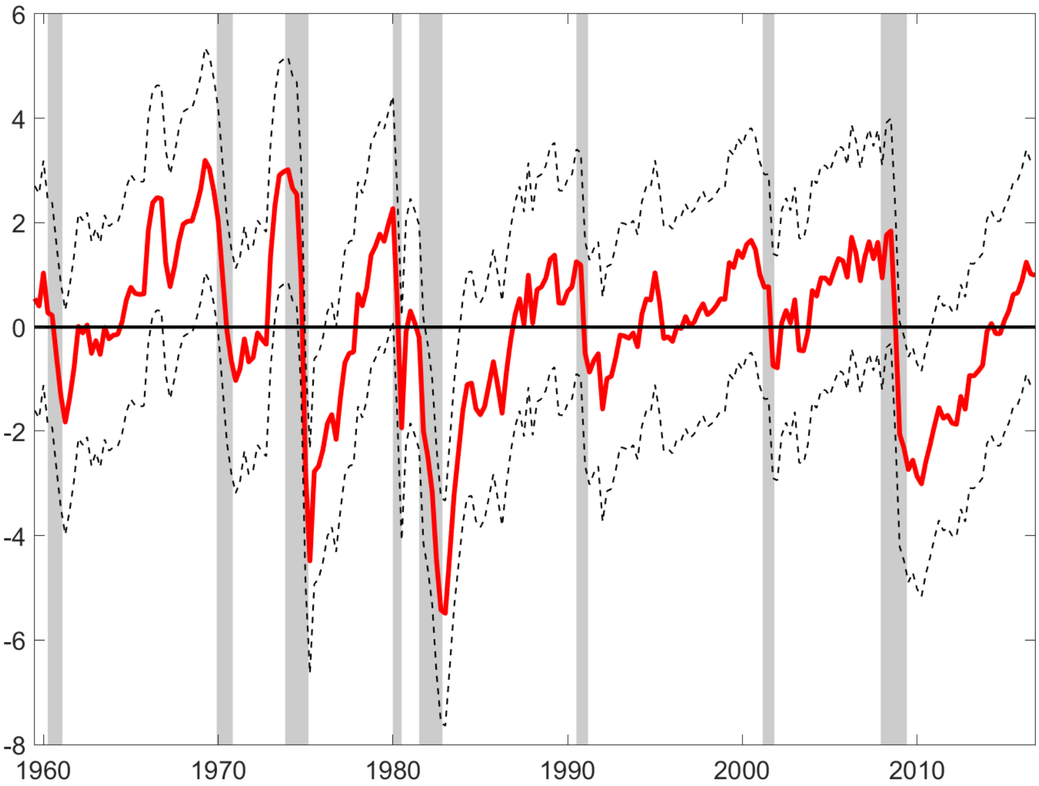
Notes: Root mean squared errors (RMSEs) are calculated for estimates in comparison to the true output gap for a given simulation under the respective DGP. Estimated model size indicates the number of variables in the information set used to estimate the VAR. The three DGPs contain one, eight, and 23 variables, respectively. Panel (a) presents the mean and standard deviation (in parentheses) of the RMSEs across 1000 Monte Carlo draws. Shrinkage and OLS refer respectively to using our procedure or least squares to estimate the VAR. Panel (b) counts the proportion of Monte Carlo draws where shrinkage produces an output gap that has a lower RMSE relative to the true output gap compared to estimating the VAR using least squares.

Figure 1: Estimated U.S. Output Gap from Univariate and Multivariate BN Decompositions



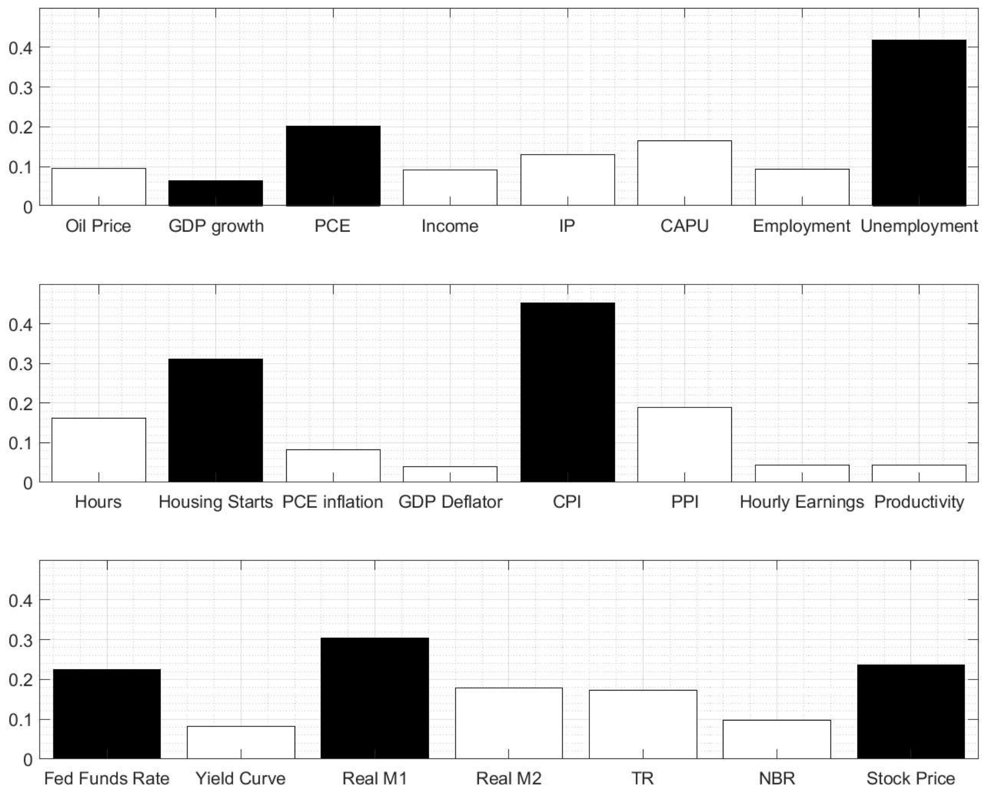
Notes: Units are 100 times natural log deviation from trend. Shaded bars correspond to NBER recession dates. The output gaps are based on a BN decomposition for various forecasting models estimated using least squares. 'Univariate' refers to the estimated output gap for an AR(4) of output growth. 'Two Variable VAR' refers to the estimated output gap for a bivariate VAR(4) with output growth and the unemployment rate. 'Four Variable VAR' refers to the estimated output gap for a four-variable VAR(4) with output growth, the unemployment rate, the growth rate of industrial production, and quarterly CPI inflation.

Figure 2: Estimated U.S. Output Gap for Benchmark BVAR



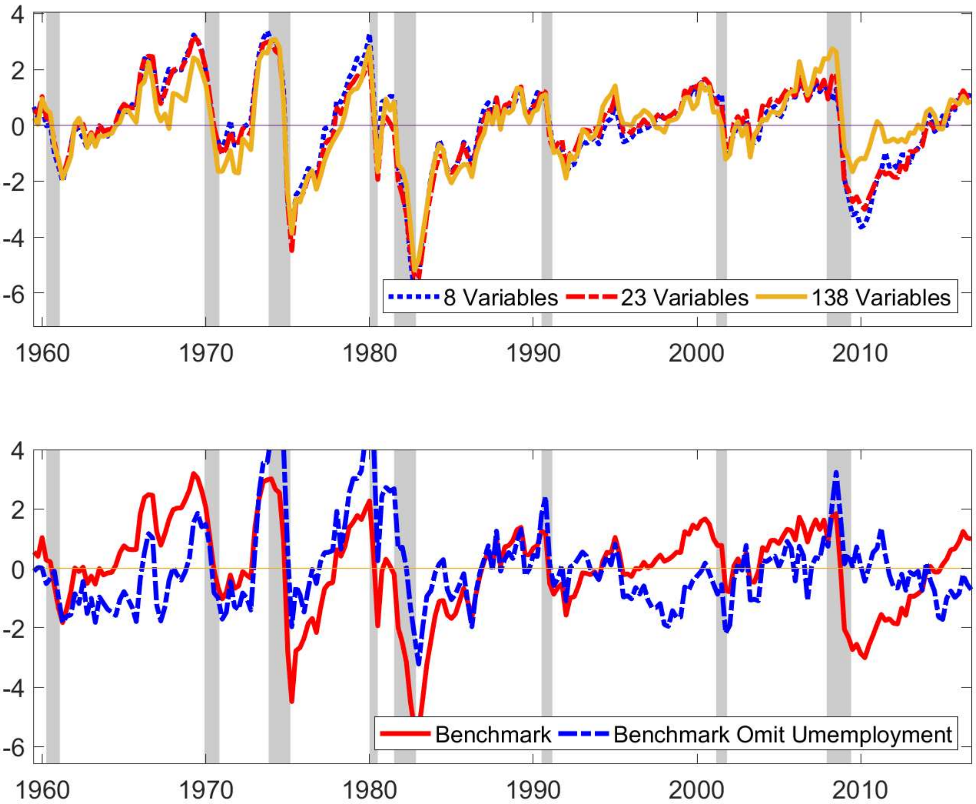
Notes: Units are 100 times natural log deviation from trend. The dotted lines indicated the bounds of the 90% credible set, calculated using the approach detailed by Kamber et al. (2018). Shaded bars correspond to NBER recession dates.

Figure 3: Standard Deviations of Informational Contributions



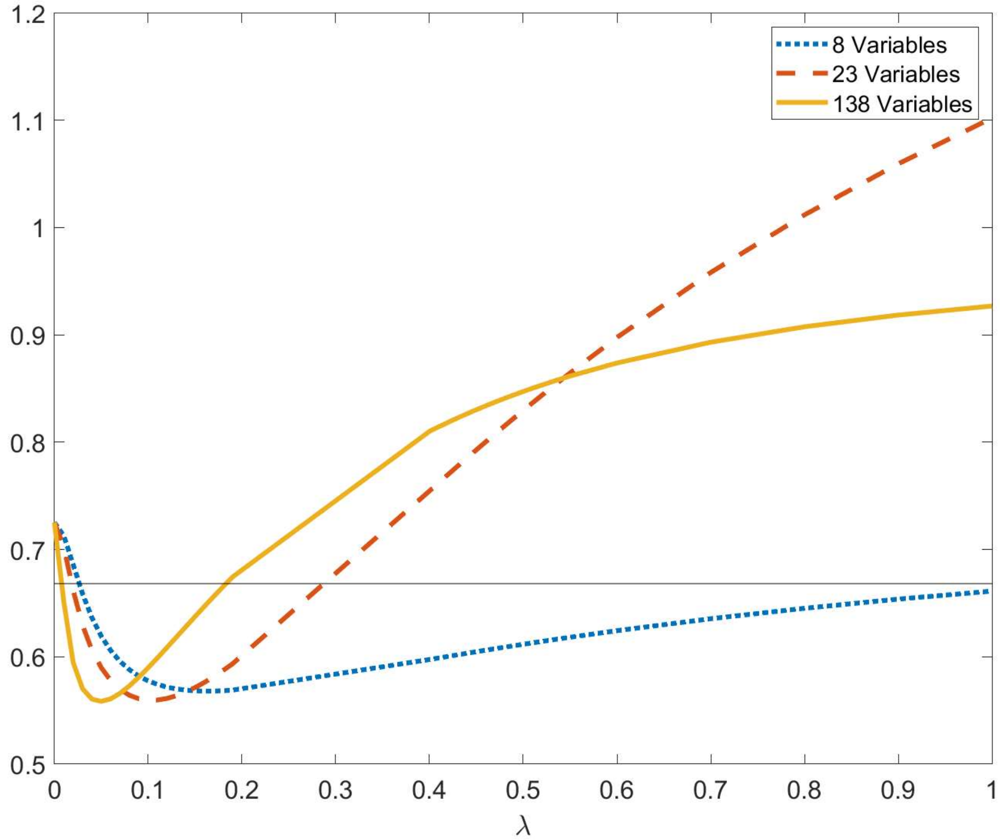
Notes: Units are standard deviations. Contributions to the estimated output gap for each variable are calculated using equation (8). *Output growth and the variables with the seven highest shares are highlighted in black.*

Figure 4: Estimated U.S. Output Gap for Various-Sized BVARs



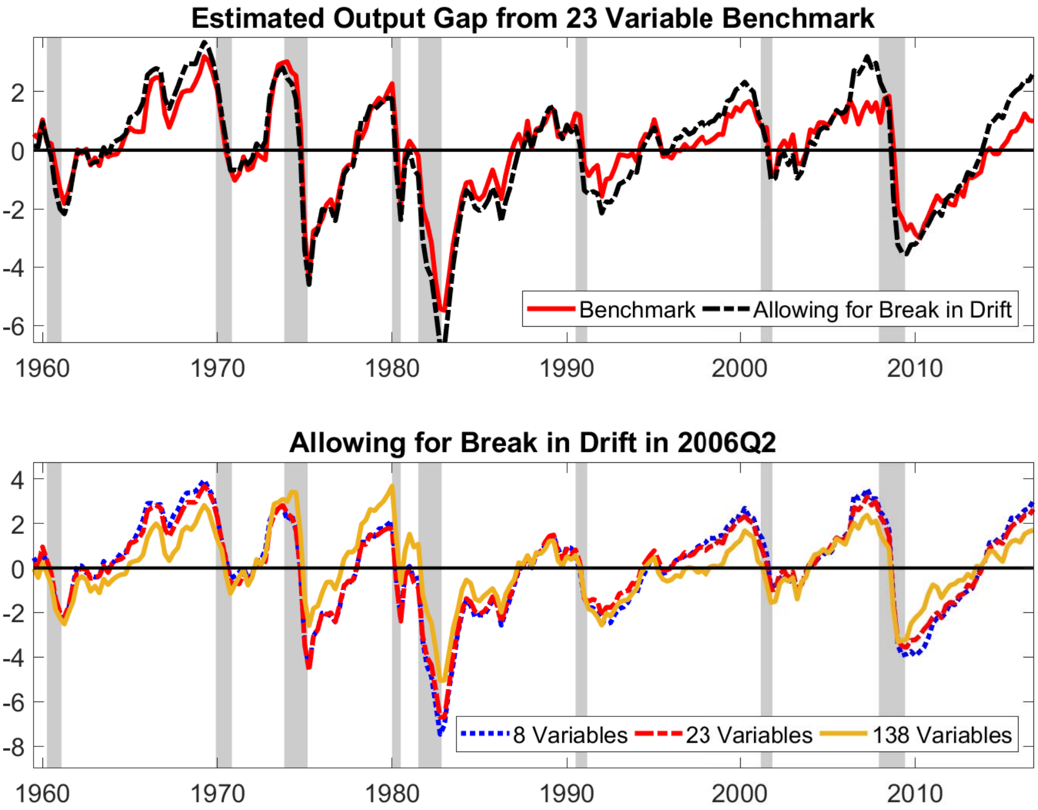
Notes: Units are 100 times natural log deviation from trend. Shaded bars correspond to NBER recession dates.

Figure 5: One-Step-Ahead Pseudo-Out-of-Sample Root Mean Square Forecast Error for Various-Sized BVARs



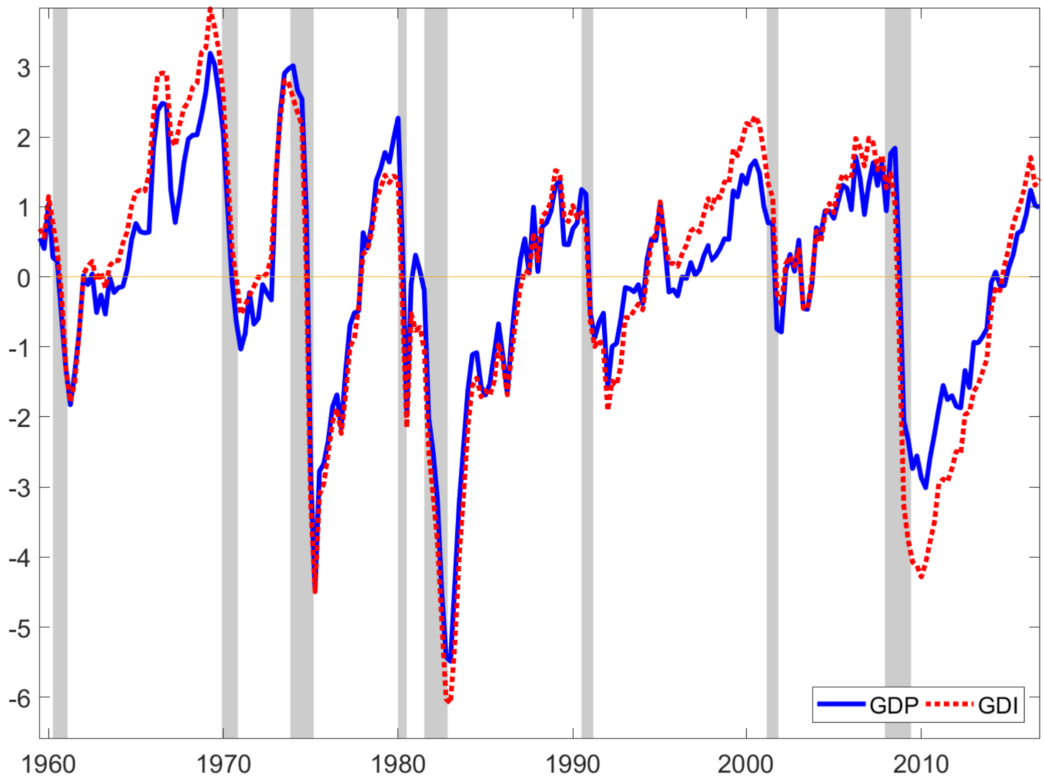
Notes: The horizontal axis represents the tightness of the prior on the hyperparameter λ*, with lower values corresponding to a tighter prior. The vertical axis represents the one-step-ahead pseudo-out-of-sample root mean square forecast error. '8 variables', '23 variables', and '138 variables' refer to the size of the various BVARs. The horizontal line is the one-step-ahead pseudo-out-of-sample root mean square forecast error based on an AR(1) model of output growth estimated by least squares.*

Figure 6: Estimated U.S. Output Gap Allowing for Break in Long-Run Growth



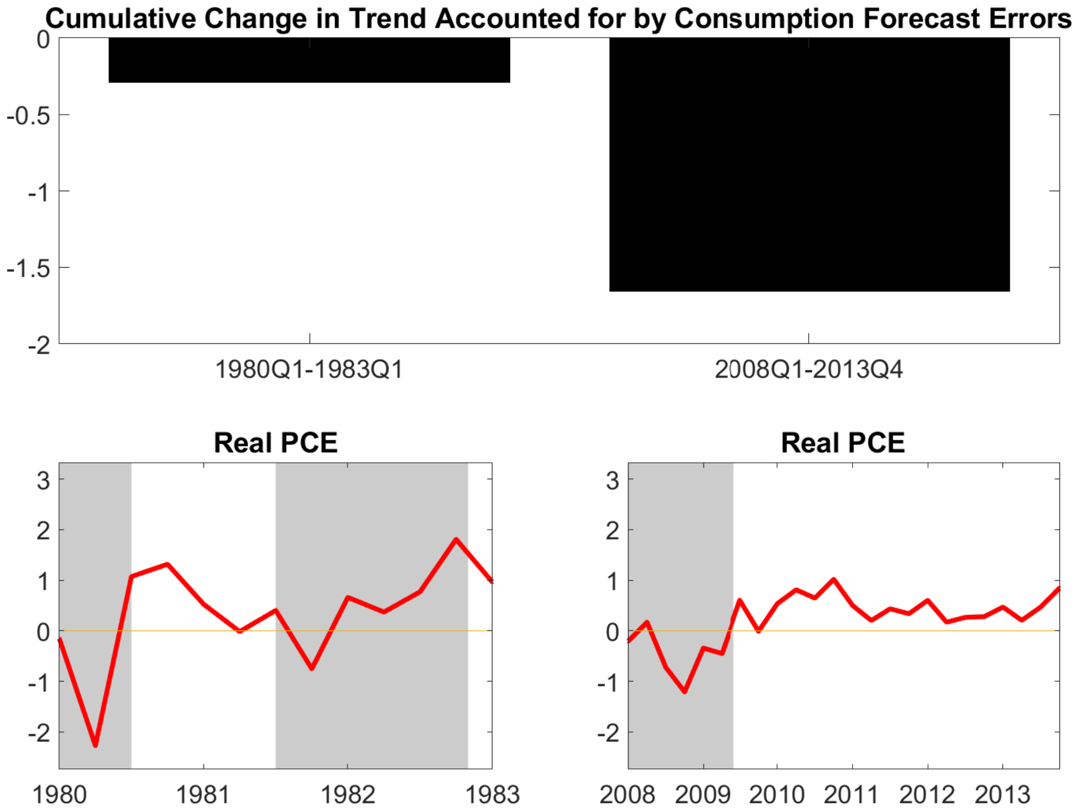
Notes: Units are 100 times natural log deviation from trend. Shaded bars correspond to NBER recession dates. The timing of a break in 2006Q2 is based on the estimated date according to Bai and Perron (2003) procedures.

Figure 7: Estimated U.S. Output Gap Using Real GDP versus Using Real GDI



Notes: Units are 100 times natural log deviation from trend. Shaded bars correspond to NBER recession dates. Real GDI is calculated using nominal GDI and the GDP deflator.

Figure 8: Role of Aggregate Consumption in Accounting for the Estimated U.S. Output Gap



Notes: Cumulative change in trend output accounted for by aggregate consumption is in terms of 100 times natural logs. Shaded bars correspond to NBER recession dates. Real personal consumption expenditure is plotted as a quarterly percent change.

Online Appendix to

Estimating and Accounting for the Output Gap with Large Bayesian Vector Autoregressions *

James Morley[1] and Benjamin Wong[2]

[1]University of Sydney

[2]Monash University, Australia

September 9, 2018

*Email: Morley: james.morley@sydney.edu.au Wong: benjamin.wong@monash.edu

A1 Data

IFS in the mnemonic column refers to a series being sourced from the International Financial Statistics. Otherwise, the data are sourced from the Federal Reserve Economic Data (FRED) and the FRED mnemonic is provided. The "Adjust" column refers to any data transformations: 'ln' indicates natural logarithms have been taken and 'Δ^i' indicates the variable has been differenced i times. Differencing is conducted if a Chow test for a change in mean from the first half to the second half of the sample is significant at the 10% level and/or an augmented Dicky-Fuller test rejects a unit root at the 5% level. An '**x**' in the 'BM' column indicates that a variable is included in the 23-variable benchmark BVAR.

Series	**Mnemonic**	**Adjust**	**BM**
U.S.: Commodity Price: W Texas Interm Spot Price (US$/Barrel)	IFS	ln, Δ	**x**
Real Gross Domestic Product, 3 Decimal	GDPC96	ln, Δ	**x**
Real Personal Consumption Expenditures	PCECC96	ln, Δ	**x**
Personal Consumption Expenditures: Durable Goods	PCDGx	ln, Δ	
Personal Consumption Expenditures: Services	PCESVx	ln, Δ	
Personal Consumption Expenditures: Nondurable Goods	PCNDx	ln, Δ	
Real Gross Private Domestic Investment, 3 decimal	GPDIC96	ln, Δ	
Fixed Private Investment	FPIx	ln, Δ	
Gross Private Domestic Investment: Fixed Investment: Nonresidential: Equipment	Y033RC1Q027SBEAx	ln, Δ	
Private Nonresidential Fixed Investment	PNFIx	ln, Δ	
Private Residential Fixed Investment	PRFIx	ln, Δ	
Shares of gross domestic product: Gross private domestic investment: Change in private inventories	A014RE1Q156NBEA	Δ	
Real Government Consumption Expenditures and Gross Investment	GCEC96	ln, Δ	
Real Government Consumption Expenditures and Gross Investment: Federal	A823RL1Q225SBEA	ln, Δ	
Federal Government Current Receipts	FGRECPTx	ln, Δ	
State and Local Consumption Expenditures & Gross Investment	SLCEx	ln, Δ^2	
Real Exports of Goods and Services, 3 Decimal	EXPGSC96	ln, Δ	
Real Imports of Goods and Services, 3 Decimal	IMPGSC96	ln, Δ	
Real Disposable Personal Income	DPIC96	ln, Δ	**x**
Nonfarm Business Sector: Real Output	OUTNFB	ln, Δ	
Business Sector: Real Output	OUTBS	ln, Δ	
Industrial Production Index	INDPRO	ln, Δ	**x**
Industrial Production: Final Products (Market Group)	IPFINAL	ln, Δ	
Industrial Production: Consumer Goods	IPCONGD	ln, Δ	

Industrial Production: Materials	IPMAT	ln, Δ	
Industrial Production: Durable Materials	IPDMAT	ln, Δ	
Industrial Production: Nondurable Materials	IPNMAT	ln, Δ	
Industrial Production: Durable Consumer Goods	IPDCONGD	ln, Δ	
Industrial Production: Durable Goods: Automotive products	IPB51110SQ	ln, Δ	
Industrial Production: Nondurable Consumer Goods	IPNCONGD	ln, Δ	
Industrial Production: Business Equipment	IPBUSEQ	ln, Δ	
Industrial Production: Consumer energy products	IPB51220SQ	ln, Δ	
Capacity Utilization: Manufacturing (SIC)	CUMFNS	Δ	x
All Employees: Total Nonfarm Payrolls	PAYEMS	ln, Δ	
All Employees: Total Private Industries	USPRIV	ln, Δ	
Civilian Employment Level	CE16OV	ln, Δ	x
Civilian Labor Force Participation Rate	CIVPART	Δ	
Civilian Unemployment Rate	UNRATE		x
Unemployment Rate: 16 to 19 years	LNS14000012	Δ	
Unemployment Rate: 20 years and over, Men	LNS14000025	Δ	
Unemployment Rate: 20 years and over, Women	LNS14000026	Δ	
Number of Civilians Unemployed for Less Than 5 Weeks	UEMPLT5	ln, Δ	
Number of Civilians Unemployed for 5 to 14 Weeks	UEMP5TO14	ln, Δ	
Number of Civilians Unemployed for 15 to 26 Weeks	UEMP15T26	ln, Δ	
Number of Civilians Unemployed for 27 Weeks and Over	UEMP27OV	ln, Δ	
Employment Level: Part-Time for Economic Reasons, All Industries	LNS12032194	ln, Δ	
Business Sector: Hours of All Persons	HOABS	ln, Δ	
Nonfarm Business Sector: Hours of All Persons	HOANBS	ln, Δ	x
Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing	AWHMAN	ln, Δ	
Average Weekly Overtime Hours of Production and Nonsupervisory Employees: Manufacturing	AWOTMAN	ln, Δ	
Housing Starts: Total: New Privately Owned Housing Units Started	HOUST	ln, Δ	x
Privately Owned Housing Starts: 5-Unit Structures or More	HOUST5F	ln, Δ	
Housing Starts in Midwest Census Region	HOUSTMW	ln, Δ	
Housing Starts in Northeast Census Region	HOUSTNE	ln, Δ	
Housing Starts in South Census Region	HOUSTS	ln, Δ	
Housing Starts in West Census Region	HOUSTW	ln, Δ	

Personal Consumption Expenditures: Chain-type Price Index	PCECTPI	ln, Δ^2	x
Personal Consumption Expenditures Excluding Food and Energy (Chain-Type Price Index)	PCEPILFE	ln, Δ^2	
Gross Domestic Product: Chain-type Price Index	GDPCTPI	ln, Δ^2	x
Gross Private Domestic Investment: Chain-type Price Index	GPDICTPI	ln, Δ	
Business Sector: Implicit Price Deflator	IPDBS	ln, Δ^2	
Personal consumption expenditures: Goods (chain-type price index)	DGDSRG3Q086SBEA	ln, Δ	
Personal consumption expenditures: Services (chain-type price index)	DSERRG3Q086SBEA	ln, Δ^2	
Consumer Price Index for All Urban Consumers: All Items	CPIAUCSL	ln, Δ	x
Producer Price Index for All Commodities	PPIACO	ln, Δ	x
Producer Price Index by Commodity Industrial Commodities	PPIIDC	ln, Δ	
Producer Price Index by Commodity for Fuels and Related Products and Power: Crude Petroleum (Domestic Production)	WPU0561	ln, Δ	
Average Hourly Earnings of Production and Nonsupervisory Employees: Construction	CES2000000008x	ln, Δ^2	
Average Hourly Earnings of Production and Nonsupervisory Employees: Manufacturing	CES3000000008x	ln, Δ^2	x
Nonfarm Business Sector: Real Compensation Per Hour	COMPRNFB	ln, Δ	
Business Sector: Real Compensation Per Hour	RCPHBS	ln, Δ	
Nonfarm Business Sector: Real Output Per Hour of All Persons	OPHNFB	ln, Δ	x
Business Sector: Real Output Per Hour of All Persons	OPHPBS	ln, Δ	
Business Sector: Unit Labor Cost	ULCBS	ln, Δ	
Nonfarm Business Sector: Unit Labor Cost	ULCNFB	ln, Δ	
Nonfarm Business Sector: Unit Nonlabor Payments	UNLPNBS	ln, Δ	
Producer Price Index by Commodity Metals and metal products: Primary nonferrous metals	PPICMM	ln, Δ	
Consumer Price Index for All Urban Consumers: Apparel	CPIAPPSL	ln, Δ	
Consumer Price Index for All Urban Consumers: Transportation	CPITRNSL	ln, Δ	
Consumer Price Index for All Urban Consumers: Medical Care	CPIMEDSL	ln, Δ	
Consumer Price Index for All Urban Consumers: Commodities	CUSR0000SAC	ln, Δ	
Consumer Price Index for All Urban Consumers: Durables	CUUR0000SAD	ln, Δ^2	
Consumer Price Index for All Urban Consumers: Services	CUSR0000SAS	ln, Δ^2	
Consumer Price Index for All Urban Consumers: All Items Less Food	CPIULFSL	ln, Δ	
Consumer Price Index for All Urban Consumers: All items less shelter	CUUR0000SA0L2	ln, Δ	
Consumer Price Index for All Urban Consumers: All items less medical care	CUSR0000SA0L5	ln, Δ	
Average Hourly Earnings of Production and Nonsupervisory Employees: Goods-Producing	CES0600000008	ln, Δ^2	
Consumer Motor Vehicle Loans Owned by Finance Companies, Outstanding	DTCOLNVHFNM	ln, Δ	

Effective Federal Funds Rate	FEDFUNDS	Δ	**x**
3-Month Treasury Bill: Secondary Market Rate	TB3MS	Δ	
6-Month Treasury Bill: Secondary Market Rate	TB6MS	Δ	
1-Year Treasury Constant Maturity Rate	GS1	Δ	
10-Year Treasury Constant Maturity Rate	GS10	Δ	
Moody's Seasoned Aaa Corporate Bond Yield	AAA	Δ	
Moody's Seasoned Baa Corporate Bond Yield	BAA	Δ	
Moody's Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity	BAA10YM	Δ	
6-Month Treasury Bill Minus Federal Funds Rate	TB6SMFFM	Δ	
10-Year Treasury Constant Maturity Minus Federal Funds Rate	T10YFFM	Δ	**x**
Real St. Louis Adjusted Monetary Base	AMBSLREALx	ln, Δ	
Real M1 Money Stock	M1REALx	ln, Δ	**x**
Real M2 Money Stock	M2REALx	ln, Δ	**x**
Real MZM Money Stock	MZMREALx	ln, Δ	
Commercial and Industrial Loans, All Commercial Banks	BUSLOANSx	ln, Δ	
Consumer Loans at All Commercial Banks	CONSUMERx	ln, Δ	
Total Nonrevolving Credit Owned and Securitized, Outstanding	NONREVSLx	ln, Δ	
Real Estate Loans, All Commercial Banks	REALLNx	ln, Δ	
Total Consumer Credit Owned and Securitized, Outstanding	TOTALSLx	ln, Δ	
Households and Nonprofit Organizations; Total Assets, Level	TABSHNOx	ln, Δ	
Households and Nonprofit Organizations; Total Liabilities, Level	TLBSHNOx	ln, Δ	
Households and Nonprofit Organizations; Credit Market Instruments; Liability, Level	CMDEBT	ln, Δ^2	
Households and Nonprofit Organizations; Net Worth, Level	TNWBSHNOx	ln, Δ	
Households and Nonprofit Organizations; Total Financial Assets, Level	TFAABSHNO	ln, Δ	
Households and nonprofit organizations; real estate at market value, Level	HNOREMQ027Sx	ln, Δ	
Households and Nonprofit Organizations; Total Financial Assets, Level	TFAABSHNOx	ln, Δ	
Shares of gross domestic product: Exports of goods and services	B020RE1Q156NBEA	Δ	
Shares of gross domestic product: Imports of goods and services	B021RE1Q156NBEA	Δ	
Industrial Production: Manufacturing (SIC)	IPMANSICS	ln, Δ	
Industrial Production: Residential utilities	IPB51222S	ln, Δ	
Industrial Production: Fuels	IPFUELS	ln, Δ	
Average (Mean) Duration of Unemployment	UEMPMEAN	ln, Δ	

Average Weekly Hours of Production and Nonsupervisory Employees: Goods-Producing	CES0600000007	ln, Δ	
Total Reserves of Depository Institutions	TOTRESNS	ln, Δ	x
Reserves of Depository Institutions, Nonborrowed	NONBORRES	ln, Δ	x
5-Year Treasury Constant Maturity Rate	GS5	Δ	
3-Month Treasury Bill Minus Federal Funds Rate	TB3SMFFM	Δ	
5-Year Treasury Constant Maturity Minus Federal Funds Rate	T5YFFM	Δ	
Moody's Seasoned Aaa Corporate Bond Minus Federal Funds Rate	AAAFFM	Δ	
Total Consumer Loans and Leases Owned and Securitized by Finance Companies, Outstanding	DTCTHFNM	ln, Δ	
Securities in Bank Credit at All Commercial Banks	INVEST	ln, Δ	
Nikkei Stock Average, Nikkei 225	NIKKEI225	ln, Δ	
Nonfinancial Corporate Business; Total Liabilities, Level	TLBSNNCBx	ln, Δ	
Nonfinancial Corporate Business; Nonfinancial Assets, Level	TTAABSNNCBx	ln, Δ	
Nonfinancial Corporate Business; Net Worth, Level	TNWMVBSNNCBx	ln, Δ	
Nonfinancial noncorporate business; total liabilities, Level	NNBTILQ027Sx	ln, Δ	
Nonfinancial noncorporate business; total assets, Level	NNBTASQ027Sx	ln, Δ	
Nonfinancial Noncorporate Business; Proprietors' Equity in Noncorporate Business (Net Worth), Level	TNWBSNNBx	ln, Δ	
Corporate Net Cash Flow with IVA	CNCFx	ln, Δ	
U.S.: Industrial Share Prices (2010=100)	IFS	ln, Δ	x

A2 Bayesian Estimation via Dummy Observations

To conduct Bayesian estimation of the model, we cast the VAR in equation (8) of the manuscript into a system of multivariate regressions:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \tag{A1}$$

where $\mathbf{Y} = [\mathbf{Y_1}, \ldots, \mathbf{Y_T}]'$, $\mathbf{X} = [\mathbf{X_1}, \ldots, \mathbf{X_T}]'$ with $\mathbf{X_t} = [\mathbf{Y'_{t-1}}, \ldots, \mathbf{Y'_{t-p}}]'$, and $\mathbf{u} = [\mathbf{u_1}, \ldots, \mathbf{u_T}]'$. Our prior structure is a Normal-Inverse Wishart prior, which has the form

$$vec(\boldsymbol{\beta})|\Sigma \sim N(vec(\boldsymbol{\beta}_0), \Sigma \otimes \Omega_0) \quad \text{and} \quad \Sigma \sim IW(S_0, \alpha_0), \tag{A2}$$

where the prior parameters $\boldsymbol{\beta}_0$, Ω_0, S_0, and α_0 are set to be consistent with equations (9) and (10) in the manuscript and the expectation of Σ being $diag(\sigma_1^2, \ldots \sigma_n^2)$. The prior from (A2) can then be implemented by choosing the following dummy observations in order to match the moments of the prior (see, e.g., Del Negro and Schorfheide, 2011; Woźniak, 2016):

$$Y_d = \begin{pmatrix} 0_{np,n} \\ diag(\sigma_1 \ldots \sigma_n) \end{pmatrix}, X_d = \begin{pmatrix} J_p \otimes diag(\sigma_1 \ldots \sigma_n)/\lambda \\ 0_{n,np} \end{pmatrix}, \tag{A3}$$

where Y_d and X_d are dummy observations, $J_p = diag(1, \ldots p)$, $S_0 = (Y_d - X_d B_0)'(Y_d - X_d B_0)$, $B_0 = (X_d' X_d)^{-1} X_d' Y_d$, $\Omega_0 = (X_d' X_d)^{-1}$, and $\alpha_0 = T_d - np$, where T_d is the number of rows for both Y_d and X_d.[1] The first block of dummy observations places the prior on all of the individual VAR slope coefficients and the second block imposes the priors on the covariance matrix.

Augmenting the regression in equation (A1) with the dummy observations gives the following:

$$\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\beta} + \mathbf{u}^*, \tag{A4}$$

where $\mathbf{Y}^* = [\mathbf{Y}', \mathbf{Y_d}']'$, $\mathbf{X}^* = [\mathbf{X}', \mathbf{X_d}']'$ and $\mathbf{u}^* = [\mathbf{u}', \mathbf{u_d}']'$. Estimating the BVAR then simply amounts to conducting least squares regression of Y^* on X^*. Therefore, the posterior distribution has the form

$$vec(\boldsymbol{\beta})|\Sigma, \mathbf{Y} \sim N(\mathbf{vec}(\tilde{\boldsymbol{\beta}}, \boldsymbol{\Sigma} \otimes (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-\mathbf{1}}) \tag{A5}$$

$$\Sigma|\mathbf{Y} \sim IW(\tilde{\Sigma}, T_d + T - np + 2), \tag{A6}$$

where $\tilde{\boldsymbol{\beta}} = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-\mathbf{1}}\mathbf{X}^{*\prime}\mathbf{Y}^{*\prime}$ and $\tilde{\Sigma} = (\mathbf{Y}^* - \mathbf{X}^*\tilde{\boldsymbol{\beta}})'(\mathbf{Y}^* - \mathbf{X}^*\tilde{\boldsymbol{\beta}})$.

[1] Note that because we demean all the variables prior to estimation, we do not include a constant in our BVAR. Thus the number of parameters in each equation is np, not $n \times (p+1)$.

A3 Causal Determinants of the U.S. Output Gap and Trend Growth

The empirical application within the manuscript largely abstracts from causal analysis and focuses on associating movements in the estimated output gap with different sources of information. Here, we outline and conduct a straightforward extension of the main empirical analysis to demonstrate how to conduct structural analysis by decomposing the estimated trend output and output gap into identified structural shocks.

We use standard SVAR analysis for two widely-considered structural shocks: a monetary policy shock and an oil price shock. The monetary policy shock is identified by ordering the federal funds rate after 'slow moving' variables, but before 'fast moving' ones in a Cholesky decomposition. This identification strategy is similar in spirit to work by, *inter alia*, Christiano et al. (1999) and Bernanke et al. (2005), where the idea is that financial market variables are in the fast moving block because they can respond contemporaneously to monetary policy shocks, while slow moving variables take at least a quarter to respond. The fast moving variables in our benchmark 23 variable specification are real M1 and M2, stock prices, non-borrowed reserves, total reserves, and the slope of the yield curve. The oil price shock is identified by drawing from Kilian and Vega (2011), who show that oil prices do not appear responsive to macroeconomic news and thus can be taken to be pre-determined. This in essence orders the oil price first in a Cholesky decomposition and also has precedence in the wider SVAR literature studying oil price shocks (e.g., see Edelstein and Kilian, 2009; Wong, 2015). Our system is partially identified in the sense that we only identify two out of 23 potential structural shocks in our benchmark system and we do not attempt to disentangle any of the remaining 21 unidentified shocks. However, assumed orthogonality of structural shocks makes this partial identification possible.

We consider how much a given structural shock has driven the historical BN trend and cycle by performing a variance decomposition. To set up a variance decomposition for the BN cycle, we first note that $\mathbb{E}\boldsymbol{e_t} = \mathbf{0}$. Working off equation (7) in the manuscript, it can be verified that the difference between the actual h-step-ahead BN cycle and the conditional expectation of the BN cycle at time $t-1$ is

$$\begin{aligned}\mathbf{c_{t+h}} - \mathbb{E}_{\mathbf{t-1}}\mathbf{c_{t+h}} &= \sum_{i=0}^{h} \boldsymbol{\Gamma}_{\boldsymbol{i+1}}\mathbf{H}\boldsymbol{e}_{\boldsymbol{t+h-i}} &\text{(A7)}\\ &= \sum_{i=0}^{h} \boldsymbol{\Gamma}_{\boldsymbol{i+1}}\mathbf{H}\mathbf{A}\boldsymbol{\varepsilon}_{\boldsymbol{t+h-i}}, &\text{(A8)}\end{aligned}$$

where the second equality follows from the identification associated with the structural shocks from the SVAR. Because $\mathbb{E}(\boldsymbol{e_t'}\boldsymbol{e_{t-i}}) = 0, i > 0$, the total variance can therefore be written as

$$Var(\mathbf{c_{t+h}} - \mathbb{E}_{\mathbf{t-1}}\mathbf{c_{t+h}}) = \sum_{\mathbf{i=0}}^{\mathbf{h}} \boldsymbol{\Gamma}_{\boldsymbol{i+1}}\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}'\boldsymbol{\Gamma}_{\boldsymbol{i+1}}'. \tag{A9}$$

It follows, then, that a variance decomposition of the h-step-ahead variation in the BN cycle of the l^{th}-ordered target variable can be calculated using equations (A8) and (A9):

$$FEVD_{k,h}^{c} = \frac{\left[\sum_{i=0}^{h} \mathbf{s_{np,l}}\boldsymbol{\Gamma}_{\boldsymbol{i+1}}\mathbf{s_{n,k}}\mathbf{s_{n,k}}'\mathbf{H}\mathbf{a_k}\right]^2}{\mathbf{s_{np,l}}\left[\sum_{\mathbf{i=0}}^{\mathbf{h}} \boldsymbol{\Gamma}_{\boldsymbol{i+1}}\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}'\boldsymbol{\Gamma}_{\boldsymbol{i+1}}'\right]\mathbf{s}_{\mathbf{np,l}}'}, \tag{A10}$$

where $FEVD_{k,h}^{c}$ is the h-step-ahead share of the variance of the BN cycle of the target variable due to the k^{th} structural shock that is identified using the k^{th} column, $\mathbf{a_k}$, of $\mathbf{A}$. Similarly, to perform a variance decomposition of trend growth for a target variable, it is straightforward to verify from equation (9) in the manuscript that the variance of the change in trend can be written as

$$Var(\boldsymbol{\Delta\tau_t} - \mathbb{E}_{\boldsymbol{t-1}}\boldsymbol{\Delta\tau_t}) = \boldsymbol{\Gamma_0}\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}'\boldsymbol{\Gamma_0'} \tag{A11}$$

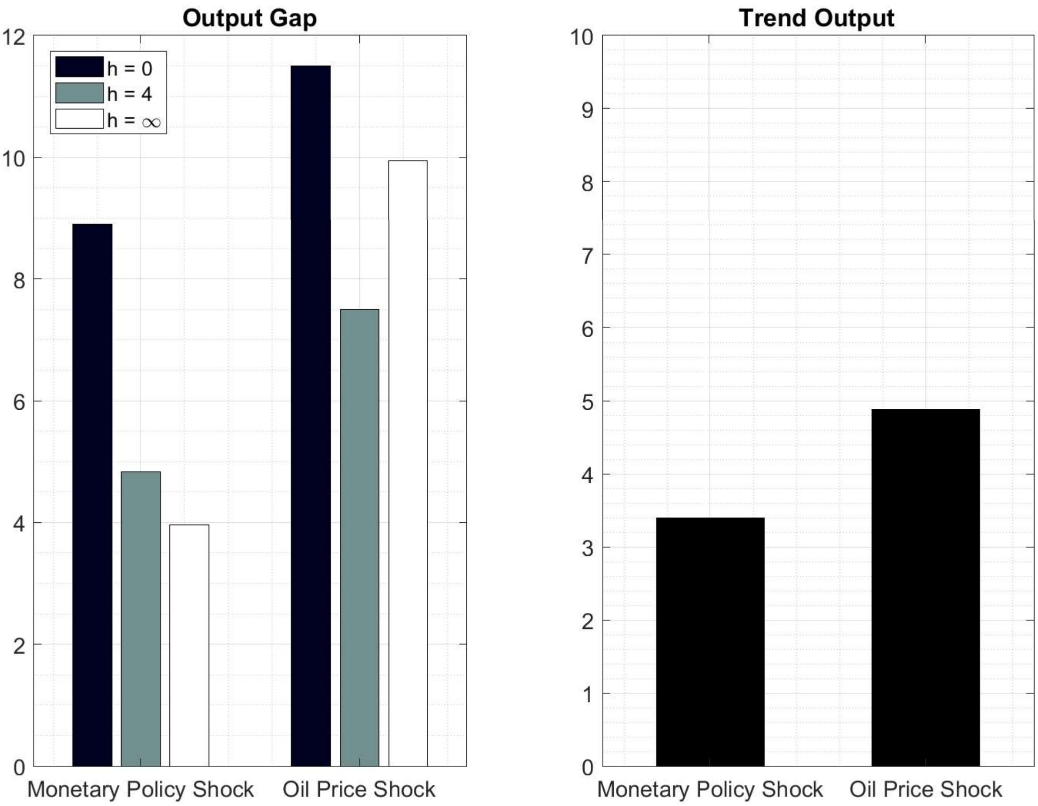
and the share of the variance can be similarly decomposed as

$$FEVD_{j}^{\tau} = \frac{\left[\sum_{i=0}^{h} \mathbf{s_{np,l}}\boldsymbol{\Gamma_0}\mathbf{s_{n,k}}\mathbf{s_{n,k}}'\mathbf{H}\mathbf{a_k}\right]^2}{\mathbf{s_{np,l}}\left[\boldsymbol{\Gamma_0}\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}'\boldsymbol{\Gamma_0'}\right]\mathbf{s}_{\mathbf{np,l}}'}. \tag{A12}$$

Note that due to the random walk trend, the variance of trend is unbounded as the time horizon goes to infinity. Consequently, a decomposition of the contemporaneous variance of the change in the trend is sufficient to provide insight into how much of the variation of trend growth is due to the various identified structural shocks.

Figure A1 presents a variance decomposition of the output gap and output trend growth. For the output gap, we present the share of monetary policy shocks and oil price shocks at horizons $h = 0$, $h = 4$, and $h = \infty$. Neither the monetary policy shock nor the oil price shock explain more than 10% of the variance of the output gap at any horizon. While the monetary policy shock explains about 7% of the variance of the output gap contemporaneously, its share quickly dissipates and it only explains about 4% of the unconditional variance. Therefore, it appears that the role of the monetary policy shock in driving the output gap is limited and relatively short lived. This finding is consistent with the wider SVAR literature, which often reports that monetary policy shocks explain only a small part of real economic activity. The oil price shock explains a somewhat larger share at about 10% of the variance of the output gap over all horizons. Meanwhile, consistent with traditional theories of growth that assume technology shocks are the main determinant of the long-run level of output, neither of these shocks explains much of output trend growth, with shares of about 5% for the oil price shock and less than 4% for the monetary policy shock. Notably, the latter result is reflective of the idea of long-run money neutrality, which suggests monetary policy should not have any permanent effects on the level of output.

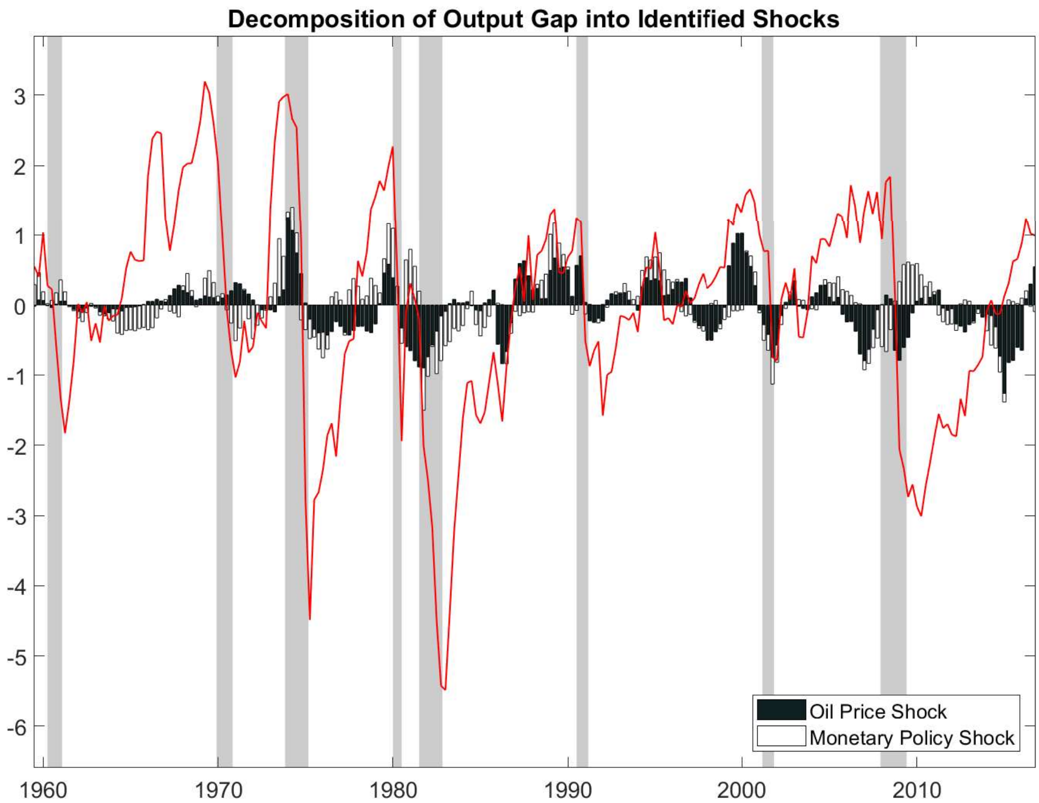
Figure A1: Variance Decompositions of Estimated U.S. Output Gap and Output Trend Growth



Notes: Results are for 23-variable benchmark BVAR. Units are percentage of total variation

Although variance decompositions are useful to gain an overall perspective on the relative importance of shocks, we can also calculate a historical contribution of shocks to the output gap to better understand specific historical episodes. This analysis is displayed in Figure A2. For realized monetary policy shocks, we can observe that they explain a large share of the positive output gap before the 1980 recession, consistent with anecdotal evidence that the Fed may have been overstimulating the economy in the 1970s. Although we can see that monetary policy shocks contributed to some of the negative output gap in the early 1980s, consistent with the Volcker disinflation, the overall output gap in the early 1980s was estimated to be large and negative, with monetary policy shocks only contributing to part of the negative gap rather than being the dominant cause. Meanwhile, a recent interpretation of the events leading to the Great Recession argues that the Fed was perhaps running the economy too hot before 2008 (e.g., see Taylor, 2012). Our historical decomposition does not support this story. We find that, while monetary policy shocks did contribute modestly to a rising positive output gap in the early 2000s, this contribution largely turned negative by 2005, while the estimated output gap continued to increase up until the advent of the Great Recession. Meanwhile, we find that realized oil price shocks tend to contribute positively to the output gap when oil prices are low and contribute negatively when oil prices are high. This can be seen from the negative contribution of oil price shocks throughout the 2000s and the positive contribution in the late 1990s. We also observe a positive contribution turning negative around 1990, consistent with the timing when the First Gulf War caused oil prices to rise from a low starting level. Furthermore, oil price shocks contributed negatively to the output gap around 1979 and 1980, consistent with the timing of the Iranian hostage crisis and the start of the Iraq-Iran War. Overall, we find that the contributions of realized monetary policy and oil price shocks line up with many well-understood historical events.

Figure A2: Historical Decomposition of the Estimated U.S. Output Gap



Notes: Units are 100 times natural log deviation from trend. Shaded bars correspond to NBER recession dates.

A4 Prior on the Signal-to-Noise Ratio

For completeness, we discuss how to implement a prior on the implied signal-to-noise ratio in terms of the variance of trend shocks relative to the variance of forecast errors. Typical BVAR methods would shrink a variable like log real GDP towards to a random walk process with an implied signal-to-noise ratio of 1. The underlying idea is that because a random walk provides a competitive forecast for many macroeconomic variables, shrinking towards a random walk balances overfitting, which worsens the forecasting performance of the model, with a more parsimonious and accurate forecasting model. However, a slight concern is that with larger models requiring more shrinkage, as shown by Banbura et al. (2010) and our baseline empirical analysis has shown, there is a possibility that as the number of time series relative to time series observations gets large, the model shrinks too much towards a random walk, creating an possible upward bias in the implied signal-to-noise ratio.

If one were concerned about such a possibility, it is possible to consider shrinking the target variable not towards a random walk, but towards a pre-specified signal-to-noise ratio δ, building on work by Kamber et al. (2018). To interpret this signal-to-noise ratio, $\delta = 0.01x$ implies $x\%$ of the variance of a forecast error for Δy_t is due to permanent shocks to y_t. Kamber et al. (2018) demonstrate how to perform a univariate BN decomposition with a pre-specified δ because there is a direct mapping from δ to the AR coefficients in an AR(p) model. In particular, letting ρ be the sum of AR coefficients in an AR(p) regression of output growth, the mapping between the two is $\rho = 1 - 1/\sqrt{\delta}$. In Kamber et al. (2018), the estimation of the output gap from a univariate AR(p) model of output growth treats ρ as being fixed and so can be viewed as a dogmatic prior on the signal-to-noise ratio. Here, in the multivariate environment, we place a prior on δ, but we do not make it dogmatic to allow the multivariate information to move the posterior away from the prior depending on how well the multivariate information helps to forecast Δy_t. A prior on δ amounts to placing a prior on the sum of the autoregressive coefficients in the target variable equation, which we label $\rho(\delta)$.

Recalling that Δy_t is the l^{th} variable in our BVAR and letting $\rho(\bar{\delta})$ be the sum of the autoregressive coefficients in the target variable equation consistent with a pre-specified $\bar{\delta}$. Implementing the prior on the signal-to-noise ratio implies:

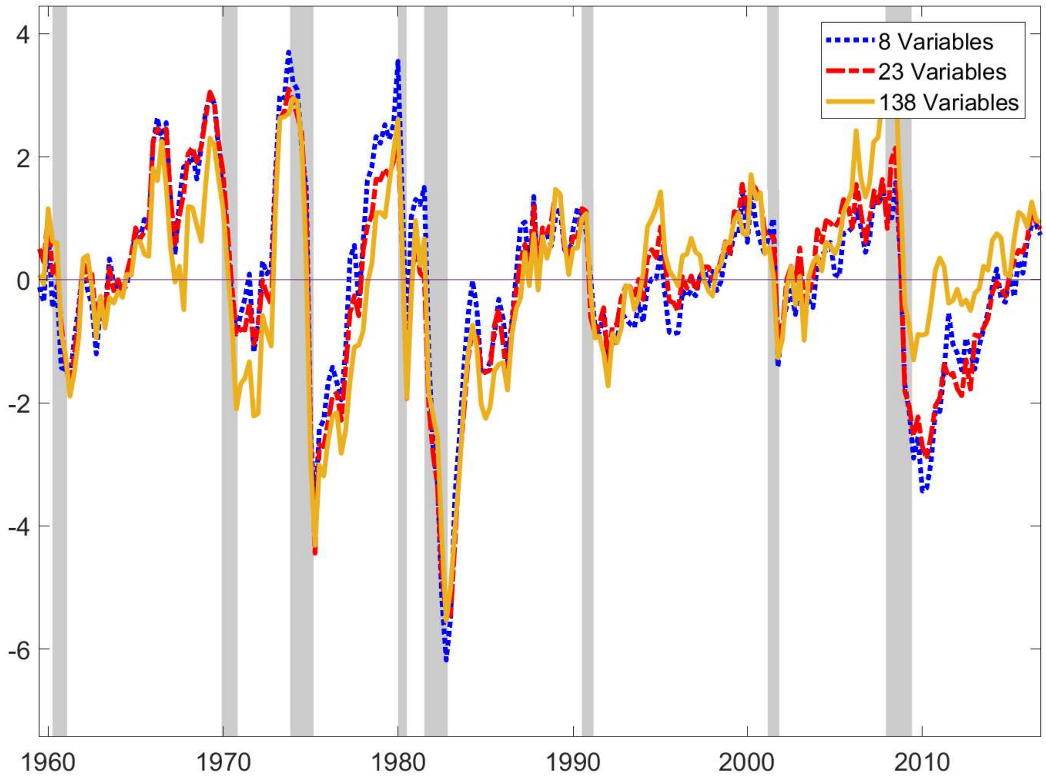
$$\mathbb{E}[\sum_{i=1}^{p} \beta_i^{ll}] = \rho(\bar{\delta}) \tag{A13}$$

$$Var[\sum_{i=1}^{p} \beta_i^{ll}] = \chi^2, \tag{A14}$$

where we set $\bar{\delta} = 0.25$ based on Kamber et al. (2018) and $\chi = \lambda/10$ to make the prior relatively informative compared to the usual Minnesota prior. The prior on the signal-to-noise ratio can be readily implemented using dummy observations. In particular, this will append the rows $\begin{bmatrix} 0_{1,l-1} & \rho/\chi & 0_{1,n-l} \end{bmatrix}$ and $\begin{bmatrix} 1_{1,n} \otimes \begin{pmatrix} 0_{1,l-1} & 1/\chi & 0_{1,n-l} \end{pmatrix} \end{bmatrix}$ to the Y_d and X_d matrices, respectively.

Figure A3 plots the estimated output gap for the eight-, 23- and 138-variable systems using the prior on the signal-to-noise ratio, with $\delta = 0.25$, and once again choosing the shrinkage parameter λ by optimizing on the pseudo-out-of-sample forecast performance. The results are similar to those in Figure 4 in the main text based on a Minnesota prior, suggesting that the likelihood dominates the prior on δ, at least for our empirical application. Thus, imposing such a prior would only matter for smaller samples when wanting to avoid an upward bias in the implied signal-to-noise ratio.

Figure A3: Estimated U.S. Output Gap for Various-Sized BVARs with Prior on Signal-to-Noise Ratio



Notes: Units are 100 times natural log deviation from trend. Shaded bars correspond to NBER recession dates.

References

Banbura M, Giannone D, Reichlin L. 2010. Large Bayesian Vector Auto Regressions. *Journal of Applied Econometrics* **25**: 71–92.

Bernanke BS, Boivin J, Eliasz P. 2005. Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach. *The Quarterly Journal of Economics* **120**: 387–422.

Christiano LJ, Eichenbaum M, Evans CL. 1999. Monetary Policy Shocks: What Have We Learned and to What End? In Taylor JB, Woodford M (eds.) *Handbook of Macroeconomics*, volume 1 of *Handbook of Macroeconomics*, chapter 2. Elsevier, 65–148.

Del Negro M, Schorfheide F. 2011. Bayesian Macroeconometrics. In Geweke J, Koop G, Van Dijk H (eds.) *The Oxford Handbook of Bayesian Econometrics*. Oxford University Press, 293–389.

Edelstein P, Kilian L. 2009. How Sensitive Are Consumer Expenditures to Retail Energy Prices? *Journal of Monetary Economics* **56**: 766–779.

Kamber G, Morley J, Wong B. 2018. Intuitive and Reliable Estimates of the Output Gap from a Beveridge-Nelson Filter. *The Review of Economics and Statistics* **100**: 550–566.

Kilian L, Vega C. 2011. Do Energy Prices Respond to U.S. Macroeconomic News? A Test of the Hypothesis of Predetermined Energy Prices. *The Review of Economics and Statistics* **93**: 660–671.

Taylor JB. 2012. Monetary Policy Rules Work and Discretion Doesn't: A Tale of Two Eras. *Journal of Money, Credit and Banking* **44**: 1017–1032, doi: j.1538-4616.2012.00521.x.

Wong B. 2015. Do Inflation Expectations Propagate the Inflationary Impact of Real Oil Price Shocks?: Evidence from the Michigan Survey. *Journal of Money, Credit and Banking* **47**: 1673–1689.

Woźniak T. 2016. Bayesian Vector Autoregressions. *Australian Economic Review* **49**: 365–380.