Bayesian multivariate Beveridge–Nelson decomposition of I(1) and I(2) series with cointegration

Murasawa, Yasutomo

Konan University

5 February 2019

Bayesian multivariate Beveridge–Nelson decomposition of I(1) and I(2) series with cointegration*

Yasutomo Murasawa
Faculty of Economics, Konan University
8-9-1 Okamoto, Higashinada-ku, Kobe, Hyogo 658-8501, Japan
(yasutomo.murasawa@gmail.com)

First draft: December 2018
This version: February 5, 2019

Summary

The consumption Euler equation implies that the output growth rate and the real interest rate are of the same order of integration; thus if the real interest rate is I(1), then so is the output growth rate with possible cointegration, and log output is I(2). This paper extends the multivariate Beveridge–Nelson decomposition to such a case, and develops a Bayesian method to obtain error bands. The paper applies the method to US data to estimate the natural rates (or their permanent components) and gaps of output, inflation, interest, and unemployment jointly, and finds that allowing for cointegration gives much bigger estimates of all gaps.

Keywords Natural rate, Output gap, Trend–cycle decomposition, Trend inflation, Unit root, Vector error correction model (VECM)

JEL classification C11, C32, C82, E32

*This work was supported by JSPS KAKENHI Grant Number JP16K03605.

1 INTRODUCTION

Distinguishing between growth and cycles is fundamental in macroeconomics. One can define growth as the time-varying steady state, or the permanent component, and cycles as deviations from the steady state, or the transitory component. One may interpret the permanent and transitory components as the natural rate and gap, respectively, though some economists may disagree with such interpretation, in which case one can consider the permanent component of the natural rate.[1] If shocks affecting the two components differ, then policy prescriptions for promoting growth and stabilizing cycles differ. Thus it is useful to decompose economic fluctuations into the two components.

Among such decomposition methods, this paper focuses on the multivariate Beveridge–Nelson (B–N) decomposition, which decomposes a multivariate I(1) or CI(1,1) series into a random walk permanent component and an I(0) transitory component, assuming a linear time series model such as a VAR model or a vector error-correction model (VECM) for the differenced series. In practice, however, some series may be I(2), e.g., log output in some countries, in which case one must decompose I(1) and I(2) series jointly. Murasawa (2015) develops the multivariate B–N decomposition of I(1) and I(2) series.

As Murasawa (2015) shows, for non-US data, the B–N decomposition assuming I(1) log output often gives an unreasonable estimate of the output gap, perhaps because of possible structural breaks in the mean output growth rate; see Figure 1. Kamber, Morley, and Wong (2018, p. 563) explain,

> …if there is a large reduction in the long-run growth rate, a forecasting model that fails to account for it will keep anticipating faster growth

[1] Phelps (1995) defines the natural rate as the *current stationary rate* (p. 16) or the *equilibrium steady state path* (pp. 29–30). Woodford (2003, pp. 8–9) defines the natural rate as the *equilibrium rate under flexible prices*, which may not be in the steady state. Kiley (2013) compares alternative definitions empirically using a new Keynesian DSGE model for the US, and obtains similar estimates of the output gap. Laubach and Williams (2016) and Holston, Laubach, and Williams (2017) note that these long-run and short-run views are complementary. Del Negro, Giannone, Giannoni, and Tambalotti (2017) also distinguish the natural rate and its low-frequency component.

than actually occurs after the break, leading to a persistently negative estimate of the output gap based on the BN decomposition.[2]

Assuming I(2) log output and hence I(1) output growth rate, one introduces a stochastic trend in the output growth rate, which captures possible structural breaks in the mean growth rate automatically in real time without specifying break dates a priori. Thus the B–N decomposition assuming I(2) log output gives a more "reasonable" estimate of the output gap that fluctuates around 0; see Figure 1.

Figure 1

This paper extends Murasawa (2015) in two ways. First, we allow for cointegration in the multivariate B–N decomposition of I(1) and I(2) series. Recall that the consumption Euler equation in a simple macroeconomic model implies a dynamic IS equation such that for all t,

$$\mathrm{E}_t(\Delta y_{t+1}) = \frac{1}{\sigma}(r_t - \rho) \tag{1}$$

where y_t is log output, r_t is the real interest rate, ρ is the discount rate, and σ is the curvature of the utility of consumption; see Galí (2015, pp. 21–23). This equation implies that if $0 < \sigma < \infty$, then the output growth rate and the real interest rate are of the same order of integration; thus if the real interest rate is I(1), then so is the output growth rate with possible cointegration, and log output is I(2). This observation motivates our development of the multivariate B–N decomposition of I(1) and I(2) series with cointegration.

Second, we apply Bayesian analysis to obtain error bands for the components, building on recent developments in Bayesian analysis of a VECM. Since cointegrating

[2] In this quote, Kamber et al. (2018) seem to consider the *output growth rate gap*. If one fails to account for a large reduction in the true mean growth rate μ^*, then the output growth rate Δy_t tends to be below the assumed mean growth rate μ; i.e., the output growth rate gap $\Delta y_t - \mu$ indeed tends to be negative. If log output is I(1), so that the output growth rate is I(0), then with positive serial correlation, the future $\Delta y_t - \mu$ tends to be negative, implying that the current output gap is *positive*, as shown in Figure 1.

vectors require normalization, our parameter of interest is in fact the cointegrating space rather than cointegrating vectors. Strachan and Inder (2004) use a matrix angular central Gaussian (MACG) distribution proposed by Chikuse (1990) as a prior on the cointegrating space. Koop, León-González, and Strachan (2010) propose a collapsed Gibbs sampler for posterior simulation of such a model. Since one often has prior information on the steady state of a system, Villani (2009) specifies a prior on the steady state form of a VECM. Since some hyperparameters such as the tightness (shrinkage) hyperparameter on the VAR coefficients are difficult to choose, Giannone, Lenza, and Primiceri (2015) use a hierarchical prior. We utilize these ideas, and show how to simulate the joint posterior distribution of the components.

As an application, we simulate the joint posterior distribution of the natural rates (or their permanent components) and gaps of output, inflation, interest, and unemployment in the US during 1950Q1–2017Q4. To apply the Bayesian multivariate B–N decomposition of I(1) and I(2) series with cointegration, we assume a four-variate VAR model for the output growth rate, the CPI inflation rate, the short-term interest rate, and the unemployment rate, and estimate it in the VECM form. The Bayes factors give decisive evidences that the cointegrating rank is 2. The posterior medians of the gaps seem reasonable compared to previous works that focus on a particular natural rate or gap. The posterior probability of positive gap is useful when the sign of the gap is uncertain. The Phillips curve and Okun's law hold between the gaps, though we do not impose such relations. Comparisons of alternative model specifications show not only that assuming I(2) log output gives a more "reasonable" estimate of the output gap, but also that allowing for cointegration gives much bigger estimates of all gaps.

The paper proceeds as follows. Section 2 reviews the literature on the B–N decomposition. Section 3 derives the multivariate B–N decomposition of I(1) and I(2) series with cointegration. Section 4 specifies our model and prior, and explains our posterior simulation and model evaluation. Section 5 applies the method to US data. Section 6 discusses remaining issues. The Appendix gives the details of the derivation of our algorithm.

2 LITERATURE

Beveridge and Nelson (1981) give operational definitions of the permanent and transitory components, show that one can express any I(1) series as the sum of a random walk permanent component and an I(0) transitory component, and propose the B–N decomposition of a univariate I(1) series, assuming an ARIMA model.[3]

Multivariate extension of the B–N decomposition is straightforward. Evans (1989a, 1989b) and Evans and Reichlin (1994) apply the B–N decomposition to a multivariate series consisting of I(0) and I(1) series, assuming a VAR model for the stationarized series. Evans and Reichlin (1994) show that the transitory components are no smaller with the multivariate B–N decomposition than with the univariate one. This is because the transitory components are "forecastable movements" (Rotemberg and Woodford (1996)), and multivariate models forecast no worse than univariate models, using more information. King, Plosser, Stock, and Watson (1991) and Cochrane (1994) apply the B–N decomposition to a CI(1,1) series, assuming a VECM.

J. C. Morley (2002) gives a general framework for the B–N decomposition, using a state space representation of the assumed linear time series model. Garratt, Robertson, and Wright (2006) note that if the state vector is observable as in a VAR model or a VECM, then the transitory component is an explicit weighted sum of the observables given the model parameters; thus the multivariate B–N decomposition based on a VAR model or a VECM is transparent. They also note that the result of the multivariate B–N decomposition depends strongly on the assumed cointegrating rank.

[3] Beveridge and Nelson (1981) reverse the sign of the transitory component. Nelson (2008) explains why they had to do so.

The B–N decomposition also applies to an I(2) series. Newbold and Vougas (1996), Oh and Zivot (2006), and Oh, Zivot, and Creal (2008) extend the B–N decomposition to a univariate I(2) series. Murasawa (2015) extends the method to a multivariate series consisting of I(1) and I(2) series.

One can apply Bayesian analysis to obtain error bands for the components. This approach is useful especially when the state vector is observable as in a VAR model or a VECM, in which case the components are explicit functions of the model parameters and observables; thus the joint posterior distribution of the model parameters directly translates into that of the components. Murasawa (2014) uses a Bayesian VAR model to obtain error bands for the components.

Kiley (2013) uses a Bayesian DSGE model, but gives no error band for the components. Del Negro et al. (2017) use a Bayesian DSGE model and give error bands for the components. They also use a multivariate unobserved components (UC) model, where the permanent components have a factor structure and the transitory components follow a VAR model. Bayesian analysis of a UC model requires state smoothing, since the state vector is unobservable given the model parameters; thus it is less straightforward than that of a VAR model. J. Morley and Wong (2018) use a large Bayesian VAR model and give error bands for the components, but they do not take parameter uncertainty into account. There seems no previous work that uses a Bayesian VECM to obtain error bands for the components.[4]

[4] Cogley, Primiceri, and Sargent (2010) use a Bayesian time-varying parameter VAR model, which is a nonlinear time series model. They still apply the B–N decomposition and give error bands for the components, pretending at each date that the VAR coefficients no longer vary. They justify their approach as an approximation based on an "anticipated-utility" model. See also Cogley and Sargent (2002, 2005) and Cogley, Morozov, and Sargent (2005).

3 MODEL SPECIFICATION

3.1 VAR model

Let for $d = 1, 2$, $\{\boldsymbol{x}_{t,d}\}$ be an N_d-variate I(d) sequence. Let $N := N_1 + N_2$. Let for all t, $\boldsymbol{x}_t := (\boldsymbol{x}'_{t,1}, \boldsymbol{x}'_{t,2})'$, $\boldsymbol{y}_{t,1} := \boldsymbol{x}_{t,1}$, $\boldsymbol{y}_{t,2} := \Delta\boldsymbol{x}_{t,2}$, and $\boldsymbol{y}_t := (\boldsymbol{y}'_{t,1}, \boldsymbol{y}'_{t,2})'$, so that $\{\boldsymbol{y}_t\}$ is I(1). Assume also that $\{\boldsymbol{y}_t\}$ is CI(1,1) with cointegrating rank r. Let for $d = 1, 2$, $\boldsymbol{\mu}_d := \mathrm{E}(\Delta\boldsymbol{y}_{t,d})$. Let $\boldsymbol{\mu} := (\boldsymbol{\mu}'_1, \boldsymbol{\mu}'_2)'$. Let $\{\boldsymbol{y}^*_t\}$ be such that for all t,

$$\boldsymbol{y}_t = \boldsymbol{\alpha} + \boldsymbol{\mu}t + \boldsymbol{y}^*_t \tag{2}$$

Assume a VAR($p + 1$) model for $\{\boldsymbol{y}^*_t\}$ such that for all t,

$$\boldsymbol{\Pi}(\mathrm{L})\boldsymbol{y}^*_t = \boldsymbol{u}_t \tag{3}$$

$$\{\boldsymbol{u}_t\} \sim \mathrm{WN}(\boldsymbol{\Sigma}) \tag{4}$$

3.2 VECM representation

Write

$$\boldsymbol{\Pi}(\mathrm{L}) = \boldsymbol{\Pi}(1)\mathrm{L} + \boldsymbol{\Phi}(\mathrm{L})(1 - \mathrm{L})$$

where $\boldsymbol{\Phi}(\mathrm{L}) := (\boldsymbol{\Pi}(\mathrm{L}) - \boldsymbol{\Pi}(1)\mathrm{L})/(1 - \mathrm{L})$. Then we have a VECM of order p for $\{\Delta\boldsymbol{y}^*_t\}$ such that for all t,

$$\begin{aligned}\boldsymbol{\Phi}(\mathrm{L})\Delta\boldsymbol{y}^*_t &= -\boldsymbol{\Pi}(1)\boldsymbol{y}^*_{t-1} + \boldsymbol{u}_t \\ &= -\boldsymbol{\Lambda}\boldsymbol{\Gamma}'\boldsymbol{y}^*_{t-1} + \boldsymbol{u}_t \end{aligned} \tag{5}$$

where $\boldsymbol{\Lambda}, \boldsymbol{\Gamma} \in \mathbb{R}^{N \times r}$. Since $\{\boldsymbol{y}^*_t\}$ is CI(1,1), the roots of $\det(\boldsymbol{\Pi}(z)) = 0$ must lie on or outside the unit circle. This requirement gives an implicit restriction on the VECM parameters $(\boldsymbol{\Phi}(.), \boldsymbol{\Lambda}, \boldsymbol{\Gamma})$.

We can write for all t,

$$\begin{aligned}\boldsymbol{\Phi}(\mathrm{L})(\Delta \boldsymbol{y}_t - \boldsymbol{\mu}) &= -\boldsymbol{\Lambda}\boldsymbol{\Gamma}'[\boldsymbol{y}_{t-1} - \boldsymbol{\alpha} - \boldsymbol{\mu}(t-1)] + \boldsymbol{u}_t \\ &= -\boldsymbol{\Lambda}[\boldsymbol{\Gamma}'\boldsymbol{y}_{t-1} - \boldsymbol{\beta} - \boldsymbol{\delta}(t-1)] + \boldsymbol{u}_t \end{aligned} \tag{6}$$

where $\boldsymbol{\beta} := \boldsymbol{\Gamma}'\boldsymbol{\alpha}$ and $\boldsymbol{\delta} := \boldsymbol{\Gamma}'\boldsymbol{\mu}$. Though slightly different, the last expression is essentially a steady state VECM suggested by Villani (2009, p. 633). We have for all t,

$$\mathrm{E}(\Delta \boldsymbol{y}_t) = \boldsymbol{\mu} \tag{7}$$

$$\mathrm{E}(\boldsymbol{\Gamma}'\boldsymbol{y}_t) = \boldsymbol{\beta} + \boldsymbol{\delta}t \tag{8}$$

which help us to specify an informative prior on $(\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\delta})$. Thus a steady state VECM is useful for Bayesian analysis.

3.3 State space representation

Assume that $p \geq 1$. We have for all t,

$$\begin{aligned}\boldsymbol{\Gamma}'\boldsymbol{y}_t^* &= \boldsymbol{\Gamma}'(\boldsymbol{y}_{t-1}^* + \boldsymbol{\Phi}_1\Delta\boldsymbol{y}_{t-1}^* + \cdots + \boldsymbol{\Phi}_p\Delta\boldsymbol{y}_{t-p}^* - \boldsymbol{\Lambda}\boldsymbol{\Gamma}'\boldsymbol{y}_{t-1}^* + \boldsymbol{u}_t) \\ &= \boldsymbol{\Gamma}'\boldsymbol{\Phi}_1\Delta\boldsymbol{y}_{t-1}^* + \cdots + \boldsymbol{\Gamma}'\boldsymbol{\Phi}_p\Delta\boldsymbol{y}_{t-p}^* + (\boldsymbol{I}_r - \boldsymbol{\Gamma}'\boldsymbol{\Lambda})\boldsymbol{\Gamma}'\boldsymbol{y}_{t-1}^* + \boldsymbol{\Gamma}'\boldsymbol{u}_t\end{aligned}$$

or

$$\begin{aligned}\boldsymbol{\Gamma}'\boldsymbol{y}_t - \boldsymbol{\beta} - \boldsymbol{\delta}t &= \boldsymbol{\Gamma}'\boldsymbol{\Phi}_1(\Delta\boldsymbol{y}_{t-1} - \boldsymbol{\mu}) + \cdots + \boldsymbol{\Gamma}'\boldsymbol{\Phi}_p(\Delta\boldsymbol{y}_{t-p} - \boldsymbol{\mu}) \\ &\quad + (\boldsymbol{I}_r - \boldsymbol{\Gamma}'\boldsymbol{\Lambda})[\boldsymbol{\Gamma}'\boldsymbol{y}_{t-1} - \boldsymbol{\beta} - \boldsymbol{\delta}(t-1)] + \boldsymbol{\Gamma}'\boldsymbol{u}_t\end{aligned} \tag{9}$$

Let $\boldsymbol{s}_t$ be a state vector such that for all t,

$$\boldsymbol{s}_t := \begin{pmatrix} \Delta \boldsymbol{y}_t - \boldsymbol{\mu} \\ \vdots \\ \Delta \boldsymbol{y}_{t-p+1} - \boldsymbol{\mu} \\ \boldsymbol{\Gamma}' \boldsymbol{y}_t - \boldsymbol{\beta} - \boldsymbol{\delta} t \end{pmatrix}$$

which is I(0) and observable given the model parameters. A state space representation of the steady state VECM is for all t,

$$\boldsymbol{s}_t = \boldsymbol{A}\boldsymbol{s}_{t-1} + \boldsymbol{B}\boldsymbol{z}_t \tag{10}$$

$$\Delta \boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{C}\boldsymbol{s}_t \tag{11}$$

$$\{\boldsymbol{z}_t\} \sim \mathrm{WN}(\boldsymbol{I}_N) \tag{12}$$

where

$$\boldsymbol{A} := \begin{bmatrix} \boldsymbol{\Phi}_1 & \dots & \boldsymbol{\Phi}_p & -\boldsymbol{\Lambda} \\ & \boldsymbol{I}_{(p-1)N} & \mathbf{O}_{(p-1)N\times N} & \mathbf{O}_{(p-1)N\times r} \\ \boldsymbol{\Gamma}'\boldsymbol{\Phi}_1 & \dots & \boldsymbol{\Gamma}'\boldsymbol{\Phi}_p & \boldsymbol{I}_r - \boldsymbol{\Gamma}'\boldsymbol{\Lambda} \end{bmatrix}$$

$$\boldsymbol{B} := \begin{bmatrix} \boldsymbol{\Sigma}^{1/2} \\ \mathbf{O}_{(p-1)N\times N} \\ \boldsymbol{\Gamma}'\boldsymbol{\Sigma}^{1/2} \end{bmatrix}$$

$$\boldsymbol{C} := \begin{bmatrix} \boldsymbol{I}_N & \mathbf{O}_{N\times(p-1)N} & \mathbf{O}_{N\times r} \end{bmatrix}$$

Note that $\{\boldsymbol{s}_t\}$ is I(0) if and only if the eigenvalues of $\boldsymbol{A}$ lie inside the unit circle.

We have for all t, for $h \geq 0$,

$$\mathrm{E}_t(\Delta \boldsymbol{y}_{t+h}) = \boldsymbol{\mu} + \boldsymbol{C}\boldsymbol{A}^h \boldsymbol{s}_t$$

or

$$\mathrm{E}_t(\Delta \boldsymbol{x}_{t+h,1}) = \boldsymbol{\mu}_1 + \boldsymbol{C}_1 \boldsymbol{A}^h \boldsymbol{s}_t \tag{13}$$

$$\mathrm{E}_t\left(\Delta^2 \boldsymbol{x}_{t+h,2}\right) = \boldsymbol{\mu}_2 + \boldsymbol{C}_2 \boldsymbol{A}^h \boldsymbol{s}_t \tag{14}$$

where

$$\boldsymbol{C}_1 := \begin{bmatrix} \boldsymbol{I}_{N_1} & \mathbf{O}_{N_1 \times N_2} & \mathbf{O}_{N_1 \times (p-1)N} & \mathbf{O}_{N_1 \times r} \end{bmatrix}$$
$$\boldsymbol{C}_2 := \begin{bmatrix} \mathbf{O}_{N_2 \times N_1} & \boldsymbol{I}_{N_2} & \mathbf{O}_{N_2 \times (p-1)N} & \mathbf{O}_{N_2 \times r} \end{bmatrix}$$

3.4 Multivariate B–N decomposition

Introducing cointegration changes the state space model, but the formulae for the multivariate B–N decomposition of I(1) and I(2) series given by Murasawa (2015, Theorem 1) remain almost unchanged. Let $\boldsymbol{x}_t^*$ and $\boldsymbol{c}_t$ be the B–N permanent and transitory components in $\boldsymbol{x}_t$, respectively.

Theorem 1. *Suppose that the eigenvalues of $\boldsymbol{A}$ lie inside the unit circle. Then for all t,*

$$\boldsymbol{x}_{t,1}^* = \lim_{T\to\infty} (\mathrm{E}_t(\boldsymbol{x}_{t+T,1}) - T\boldsymbol{\mu}_1) \tag{15}$$

$$\boldsymbol{x}_{t,2}^* = \lim_{T\to\infty} \left\{ \mathrm{E}_t(\boldsymbol{x}_{t+T,2}) - T^2 \frac{\boldsymbol{\mu}_2}{2} - T\left[\frac{\boldsymbol{\mu}_2}{2} + \Delta \boldsymbol{x}_{t,2} + \boldsymbol{C}_2(\boldsymbol{I}_{pN+r} - \boldsymbol{A})^{-1}\boldsymbol{A}\boldsymbol{s}_t\right]\right\} \tag{16}$$

$$\boldsymbol{c}_{t,1} = -\boldsymbol{C}_1(\boldsymbol{I}_{pN+r} - \boldsymbol{A})^{-1}\boldsymbol{A}\boldsymbol{s}_t \tag{17}$$

$$\boldsymbol{c}_{t,2} = \boldsymbol{C}_2(\boldsymbol{I}_{pN+r} - \boldsymbol{A})^{-2}\boldsymbol{A}^2\boldsymbol{s}_t \tag{18}$$

Proof. See Murasawa (2015, pp. 158–159). □

Let

$$\boldsymbol{W} := \begin{bmatrix} -\boldsymbol{C}_1(\boldsymbol{I}_{pN+r} - \boldsymbol{A})^{-1}\boldsymbol{A} \\ \boldsymbol{C}_2(\boldsymbol{I}_{pN+r} - \boldsymbol{A})^{-2}\boldsymbol{A}^2 \end{bmatrix}$$

Then for all t,

$$\boldsymbol{c}_t = \boldsymbol{W}\boldsymbol{s}_t \tag{19}$$

where $\boldsymbol{W}$ depends only on the VECM coefficients and $\{\boldsymbol{s}_t\}$ is observable given the model parameters. This observation is useful for Bayesian analysis of $\{\boldsymbol{c}_t\}$.

4 BAYESIAN ANALYSIS

4.1 Conditional likelihood function

Assume Gaussian innovations for Bayesian analysis, and write the VECM as for all t,

$$\begin{aligned}\Delta\boldsymbol{y}_t - \boldsymbol{\mu} &= \boldsymbol{\Phi}_1(\Delta\boldsymbol{y}_{t-1} - \boldsymbol{\mu}) + \cdots + \boldsymbol{\Phi}_p(\Delta\boldsymbol{y}_{t-p} - \boldsymbol{\mu}) \\ &\quad - \boldsymbol{\Lambda}[\boldsymbol{\Gamma}'\boldsymbol{y}_{t-1} - \boldsymbol{\beta} - \boldsymbol{\Gamma}'\boldsymbol{\mu}(t-1)] + \boldsymbol{u}_t \qquad (20)\\ \{\boldsymbol{u}_t\} &\sim \mathrm{IN}_N\left(\boldsymbol{0}_N, \boldsymbol{P}^{-1}\right) \qquad (21)\end{aligned}$$

Let $\boldsymbol{\psi} := (\boldsymbol{\beta}', \boldsymbol{\mu}')'$, $\boldsymbol{\Phi} := [\boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_p]$, and $\boldsymbol{Y} := [\boldsymbol{y}_0, \dots, \boldsymbol{y}_T]$. By the prediction error decomposition, the joint pdf of $\boldsymbol{Y}$ is

$$\begin{aligned}p(\boldsymbol{Y}|\boldsymbol{\psi}, \boldsymbol{\Phi}, \boldsymbol{P}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}) &= p(\Delta\boldsymbol{y}_T, \dots, \Delta\boldsymbol{y}_{p+1}, \boldsymbol{s}_p|\boldsymbol{\psi}, \boldsymbol{\Phi}, \boldsymbol{P}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}) \\ &= \prod_{t=p+1}^{T} p(\Delta\boldsymbol{y}_t|\boldsymbol{s}_{t-1}, \boldsymbol{\psi}, \boldsymbol{\Phi}, \boldsymbol{P}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}) p(\boldsymbol{s}_p|\boldsymbol{\psi}, \boldsymbol{\Phi}, \boldsymbol{P}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}) \qquad (22)\end{aligned}$$

Our Bayesian analysis relies on $\prod_{t=p+1}^{T} p(\Delta\boldsymbol{y}_t|\boldsymbol{s}_{t-1}, \boldsymbol{\psi}, \boldsymbol{\Phi}, \boldsymbol{P}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma})$, the conditional likelihood function of $(\boldsymbol{\psi}, \boldsymbol{\Phi}, \boldsymbol{P}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma})$ given $\boldsymbol{s}_p$.

4.2 Identification

To specify a prior on the cointegrating space, we assume that $\boldsymbol{\Gamma}'\boldsymbol{\Gamma} = \boldsymbol{I}_r$. This restriction does not identify the sign of $\boldsymbol{\Gamma}$, however. If identification of $\boldsymbol{\Gamma}$ is necessary, then we can apply linear normalization. Write $\boldsymbol{\Gamma} := [\boldsymbol{\Gamma}_1', \boldsymbol{\Gamma}_2']'$, where $\boldsymbol{\Gamma}_1$ is $r \times r$ and $\boldsymbol{\Gamma}_2$ is $(N-r) \times r$. Let

$$\bar{\boldsymbol{\Lambda}} := \boldsymbol{\Lambda}\boldsymbol{\Gamma}_1$$
$$\bar{\boldsymbol{\Gamma}} := \boldsymbol{\Gamma}\boldsymbol{\Gamma}_1^{-1}$$

Then we can identify $\left(\bar{\boldsymbol{\Lambda}}, \bar{\boldsymbol{\Gamma}}\right)$. Let $\bar{\boldsymbol{\beta}} := \bar{\boldsymbol{\Gamma}}'\boldsymbol{\alpha} = \left(\boldsymbol{\Gamma}_1^{-1}\right)'\boldsymbol{\beta}$ and $\bar{\boldsymbol{\psi}} := \left(\bar{\boldsymbol{\beta}}', \boldsymbol{\mu}'\right)'$ correspondingly. Note that $\boldsymbol{\alpha}$ is not identifiable from the VECM.

4.3 Prior

4.3.1 Steady state parameters

Assume a normal prior on $(\boldsymbol{\alpha}, \boldsymbol{\mu})$ independent of $(\boldsymbol{\Phi}, \boldsymbol{P}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma})$ such that

$$\begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\mu} \end{pmatrix} \sim \mathrm{N}_{2N}\left(\begin{pmatrix} \boldsymbol{\alpha}_0 \\ \boldsymbol{\mu}_0 \end{pmatrix}, \begin{bmatrix} \boldsymbol{Q}_{0,\boldsymbol{\alpha}}^{-1} & \mathbf{O}_{N\times N} \\ \mathbf{O}_{N\times N} & \boldsymbol{Q}_{0,\boldsymbol{\mu}}^{-1} \end{bmatrix}\right) \tag{23}$$

Since $\boldsymbol{\beta} := \boldsymbol{\Gamma}'\boldsymbol{\alpha}$, this prior implies a prior on $\boldsymbol{\psi}$ such that

$$\boldsymbol{\psi}|\boldsymbol{\Gamma} \sim \mathrm{N}_{r+N}\left(\boldsymbol{\psi}_0, \boldsymbol{Q}_0^{-1}\right) \tag{24}$$

where

$$\boldsymbol{\psi}_0 := \begin{pmatrix} \boldsymbol{\Gamma}'\boldsymbol{\alpha}_0 \\ \boldsymbol{\mu}_0 \end{pmatrix}, \quad \boldsymbol{Q}_0 := \begin{bmatrix} \left(\boldsymbol{\Gamma}'\boldsymbol{Q}_{0,\boldsymbol{\alpha}}^{-1}\boldsymbol{\Gamma}\right)^{-1} & \mathbf{O}_{r\times r} \\ \mathbf{O}_{N\times N} & \boldsymbol{Q}_{0,\boldsymbol{\mu}} \end{bmatrix}$$

which depends on $\boldsymbol{\Gamma}$ in general.[5] The joint pdf of $\boldsymbol{\psi}$ is

$$p(\boldsymbol{\psi}) = (2\pi)^{-(r+N)/2} \det(\boldsymbol{Q}_0)^{1/2} \exp\left(-\frac{1}{2}(\boldsymbol{\psi} - \boldsymbol{\psi}_0)'\boldsymbol{Q}_0(\boldsymbol{\psi} - \boldsymbol{\psi}_0)\right) \tag{25}$$

[5] If $\boldsymbol{\alpha}_0 := \mathbf{0}_N$ and $\boldsymbol{Q}_{0,\boldsymbol{\alpha}} \propto \boldsymbol{I}_N$, then since $\boldsymbol{\Gamma}'\boldsymbol{\Gamma} = \boldsymbol{I}_r$, the prior on $\boldsymbol{\psi}$ becomes independent of $\boldsymbol{\Gamma}$.

4.3.2 VAR parameters

Let S be the set of $(\boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma})$ such that the eigenvalues of $\boldsymbol{A}$ lie inside the unit circle, so that $\{\boldsymbol{y}_t^*\}$ is CI(1,1). Assume a hierarchical normal–Wishart prior on $(\boldsymbol{\Phi}, \boldsymbol{P})$ independent of $\boldsymbol{\psi}$ but dependent on $(\boldsymbol{\Lambda}, \boldsymbol{\Gamma})$ such that

$$\boldsymbol{\Phi}|\boldsymbol{P}, \nu, \boldsymbol{\Lambda}, \boldsymbol{\Gamma} \sim \mathrm{N}_{N\times pN}\left(\boldsymbol{M}_0; \boldsymbol{P}^{-1}, (\nu\boldsymbol{D}_0)^{-1}\right)[(\boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}) \in S] \tag{26}$$

$$\boldsymbol{P} \sim \mathrm{W}_N\left(k_0; \boldsymbol{S}_0^{-1}\right) \tag{27}$$

$$\nu \sim \mathrm{Gam}\left(\frac{A_0}{2}, \frac{B_0}{2}\right) \tag{28}$$

where ν is a hyperparameter that controls the tightness of the prior on the VAR coefficients, which is often difficult to choose a priori, and we assume a gamma prior on ν. The joint pdf of $(\boldsymbol{\Phi}, \boldsymbol{P}, \nu)$ conditional on $(\boldsymbol{\Lambda}, \boldsymbol{\Gamma})$ is

$$p(\boldsymbol{\Phi}, \boldsymbol{P}, \nu|\boldsymbol{\Lambda}, \boldsymbol{\Gamma}) = p(\boldsymbol{\Phi}|\boldsymbol{P}, \nu, \boldsymbol{\Lambda}, \boldsymbol{\Gamma})p(\boldsymbol{P})p(\nu) \tag{29}$$

where

$$p(\boldsymbol{\Phi}|\boldsymbol{P}, \nu, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}) = \frac{\mathrm{etr}(-\boldsymbol{P}(\boldsymbol{\Phi} - \boldsymbol{M}_0)\nu\boldsymbol{D}_0(\boldsymbol{\Phi} - \boldsymbol{M}_0)'/2)}{(2\pi)^{pN^2/2}\det(\nu\boldsymbol{D}_0)^{-N/2}\det(\boldsymbol{P})^{-pN/2}}[(\boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}) \in S] \tag{30}$$

$$p(\boldsymbol{P}) = \frac{\det(\boldsymbol{P})^{(k_0-N-1)/2}/\,\mathrm{etr}((\boldsymbol{S}_0/2)\boldsymbol{P})}{\Gamma_N(k_0/2)\det(\boldsymbol{S}_0/2)^{-k_0/2}} \tag{31}$$

$$p(\nu) = \frac{\nu^{A_0/2-1}/\mathrm{e}^{(B_0/2)\nu}}{\Gamma(A_0/2)(B_0/2)^{-A_0/2}} \tag{32}$$

where $\Gamma_N(.)$ is the N-variate gamma function. See Gupta and Nagar (1999) on the pdfs of the matrix normal and Wishart distributions.

4.3.3 Cointegrating space

Let $V_r\left(\mathbb{R}^N\right)$ be the r-dimensional Steifel manifold in $\mathbb{R}^N$, i.e.,

$$V_r\left(\mathbb{R}^N\right) := \left\{\boldsymbol{H} \in \mathbb{R}^{N\times r} : \boldsymbol{H}'\boldsymbol{H} = \boldsymbol{I}_r\right\}$$

Its volume is

$$\mathrm{Vol}\left(V_r\left(\mathbb{R}^N\right)\right) = \frac{2^r \pi^{Nr/2}}{\Gamma_r(N/2)} \tag{33}$$

See Muirhead (1982, p. 70). Following Strachan and Inder (2004), we assume a prior not on the elements of $\boldsymbol{\Gamma}$ directly but on $V_r\left(\mathbb{R}^N\right)$. Let $\boldsymbol{H}_0 \in V_r\left(\mathbb{R}^N\right)$. Let $\boldsymbol{H}(.)$ be s.th. $\forall \tau \geq 0$,

$$\boldsymbol{H}(\tau) := \boldsymbol{H}_0 \boldsymbol{H}_0' + \tau \boldsymbol{H}_{0\perp} \boldsymbol{H}_{0\perp}'$$

so that $\boldsymbol{H}(0) = \boldsymbol{H}_0 \boldsymbol{H}_0'$ is of rank r and $\boldsymbol{H}(1) = \boldsymbol{I}_N$. Assume a prior on $(\boldsymbol{\Lambda}, \boldsymbol{\Gamma})$ conditionally independent of $(\boldsymbol{\psi}, \boldsymbol{P})$ given $\boldsymbol{\Phi}$ such that[6]

$$\boldsymbol{\Lambda}|\boldsymbol{\Gamma}, \boldsymbol{\Phi} \sim \mathrm{N}_{N\times r}\left(\boldsymbol{\Lambda}_0; \boldsymbol{G}_0^{-1}, (\boldsymbol{\Gamma}'\eta_0 \boldsymbol{H}(\tau_0)\boldsymbol{\Gamma})^{-1}\right)[(\boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}) \in S] \tag{34}$$

$$\boldsymbol{\Gamma}|\boldsymbol{\Phi} \sim \mathrm{MACG}_{N\times r}\left(\boldsymbol{H}(\tau_0)^{-1}\right) \tag{35}$$

The joint pdf of $(\boldsymbol{\Lambda}, \boldsymbol{\Gamma})$ conditional on $\boldsymbol{\Phi}$ is

$$p(\boldsymbol{\Lambda}, \boldsymbol{\Gamma}|\boldsymbol{\Phi}) = p(\boldsymbol{\Lambda}|\boldsymbol{\Gamma}, \boldsymbol{\Phi}) p(\boldsymbol{\Gamma}|\boldsymbol{\Phi}) \tag{36}$$

where

$$p(\boldsymbol{\Lambda}|\boldsymbol{\Gamma}, \boldsymbol{\Phi}) = \frac{\mathrm{etr}(-\boldsymbol{G}_0(\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)\boldsymbol{\Gamma}'\eta_0 \boldsymbol{H}(\tau_0)\boldsymbol{\Gamma}(\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)'/2)}{(2\pi)^{Nr/2} \det(\boldsymbol{\Gamma}'\eta_0 \boldsymbol{H}(\tau_0)\boldsymbol{\Gamma})^{-N/2} \det(\boldsymbol{G}_0)^{-r/2}}[(\boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}) \in S] \tag{37}$$

$$p(\boldsymbol{\Gamma}|\boldsymbol{\Phi}) = \frac{\det(\boldsymbol{\Gamma}'\boldsymbol{H}(\tau_0)\boldsymbol{\Gamma})^{-N/2} / \det(\boldsymbol{H}(\tau_0))^{-r/2}}{\mathrm{Vol}(V_r(\mathbb{R}^N))} \tag{38}$$

If $\tau_0 := 1$, then $p(\boldsymbol{\Gamma}) = 1/\mathrm{Vol}\left(V_r\left(\mathbb{R}^N\right)\right)$, i.e., the flat prior on $\boldsymbol{\Gamma}$. See Chikuse (1990) on the MACG distribution.

[6] Apart from the restriction that $(\boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}) \in S$, if $\tau_0 := 1$, then since $\boldsymbol{\Gamma}'\boldsymbol{\Gamma} = \boldsymbol{I}_r$, the prior on $\boldsymbol{\Lambda}$ becomes independent of $\boldsymbol{\Gamma}$.

The following transformation is useful for posterior simulation. Let

$$
\begin{aligned}
\boldsymbol{\Lambda}_* &:= (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)[(\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)'(\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)]^{-1/2} \\
\boldsymbol{\Gamma}_* &:= \boldsymbol{\Gamma}[(\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)'(\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)]^{1/2}
\end{aligned}
$$

Then $\boldsymbol{\Lambda}_*\boldsymbol{\Gamma}_*' = (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)\boldsymbol{\Gamma}'$ and $\boldsymbol{\Gamma}_*'\boldsymbol{\Gamma}_* = (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)'(\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)$. Following Koop et al. (2010, p.230), we have

$$
\begin{aligned}
p(\boldsymbol{\Lambda}, \boldsymbol{\Gamma}|\boldsymbol{\Phi}) &= p(\boldsymbol{\Lambda}_*, \boldsymbol{\Gamma}_*|\boldsymbol{\Phi}) \\
&= p(\boldsymbol{\Gamma}_*|\boldsymbol{\Lambda}_*, \boldsymbol{\Phi})p(\boldsymbol{\Lambda}_*|\boldsymbol{\Phi}) \qquad (39)
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{\Gamma}_*|\boldsymbol{\Lambda}_*, \boldsymbol{\Phi} &\sim \mathrm{N}_{N\times r}\left(\mathbf{O}_{N\times r}; (\eta_0\boldsymbol{H}(\tau_0))^{-1}, (\boldsymbol{\Lambda}_*'\boldsymbol{G}_0\boldsymbol{\Lambda}_*)^{-1}\right)[(\boldsymbol{\Phi}, \boldsymbol{\Lambda}_*, \boldsymbol{\Gamma}_*) \in S] \qquad (40) \\
\boldsymbol{\Lambda}_*|\boldsymbol{\Phi} &\sim \mathrm{MACG}_{N\times r}\left(\boldsymbol{G}_0^{-1}\right) \qquad (41)
\end{aligned}
$$

so that

$$
\begin{aligned}
p(\boldsymbol{\Gamma}_*|\boldsymbol{\Lambda}_*, \boldsymbol{\Phi}) &= \frac{\mathrm{etr}(-\eta_0\boldsymbol{H}(\tau_0)\boldsymbol{\Gamma}_*\boldsymbol{\Lambda}_*'\boldsymbol{G}_0\boldsymbol{\Lambda}_*\boldsymbol{\Gamma}_*/2)}{(2\pi)^{Nr/2}\det(\boldsymbol{\Lambda}_*'\boldsymbol{G}_0\boldsymbol{\Lambda}_*)^{-N/2}\det(\eta_0\boldsymbol{H}(\tau_0))^{-r/2}}[(\boldsymbol{\Phi}, \boldsymbol{\Lambda}_*, \boldsymbol{\Gamma}_*) \in S] \qquad (42) \\
p(\boldsymbol{\Lambda}_*|\boldsymbol{\Phi}) &= \frac{\det(\boldsymbol{\Lambda}_*'\boldsymbol{G}_0\boldsymbol{\Lambda}_*)^{-N/2}/\det(\boldsymbol{G}_0)^{-r/2}}{\mathrm{Vol}(V_r(\mathbb{R}^N))} \qquad (43)
\end{aligned}
$$

This is essentially because

$$
\begin{aligned}
p(\boldsymbol{\Lambda}|\boldsymbol{\Gamma}, \boldsymbol{\Phi})p(\boldsymbol{\Gamma}|\boldsymbol{\Phi}) &\propto \mathrm{etr}\left(-\frac{1}{2}\boldsymbol{G}_0(\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)\boldsymbol{\Gamma}'\eta_0\boldsymbol{H}(\tau_0)\boldsymbol{\Gamma}(\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)'\right)[(\boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}) \in S] \\
&= \mathrm{etr}\left(-\frac{1}{2}\boldsymbol{G}_0\boldsymbol{\Lambda}_*\boldsymbol{\Gamma}_*'\eta_0\boldsymbol{H}(\tau_0)\boldsymbol{\Gamma}_*\boldsymbol{\Lambda}_*'\right)[(\boldsymbol{\Phi}, \boldsymbol{\Lambda}_*, \boldsymbol{\Gamma}_*) \in S] \\
&= \mathrm{etr}\left(-\frac{1}{2}\eta_0\boldsymbol{H}(\tau_0)\boldsymbol{\Gamma}_*\boldsymbol{\Lambda}_*'\boldsymbol{G}_0\boldsymbol{\Lambda}_*\boldsymbol{\Gamma}_*'\right)[(\boldsymbol{\Phi}, \boldsymbol{\Lambda}_*, \boldsymbol{\Gamma}_*) \in S]
\end{aligned}
$$

4.4 Posterior simulation

We simulate $p(\boldsymbol{\psi}, \boldsymbol{\Phi}, \boldsymbol{P}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \nu|\boldsymbol{Y})$ by a Gibbs sampler consisting of five blocks:

1. Draw $\boldsymbol{\psi}$ from $p(\boldsymbol{\psi}|\boldsymbol{\Phi}, \boldsymbol{P}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \nu, \boldsymbol{Y}) = p(\boldsymbol{\psi}|\boldsymbol{\Phi}, \boldsymbol{P}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \boldsymbol{Y})$.

2. Draw $(\boldsymbol{\Phi}, \boldsymbol{P})$ from $p(\boldsymbol{\Phi}, \boldsymbol{P}|\boldsymbol{\psi}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \nu, \boldsymbol{Y})$.

3. Draw $\boldsymbol{\Lambda}$ from $p(\boldsymbol{\Lambda}|\boldsymbol{\psi}, \boldsymbol{\Phi}, \boldsymbol{P}, \boldsymbol{\Gamma}, \nu, \boldsymbol{Y}) = p(\boldsymbol{\Lambda}|\boldsymbol{\psi}, \boldsymbol{\Phi}, \boldsymbol{P}, \boldsymbol{\Gamma}, \boldsymbol{Y})$. Let $\boldsymbol{\Lambda}_* := (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)[(\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)'(\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)]^{-1/2}$.

4. Draw $\boldsymbol{\Gamma}_*$ from $p(\boldsymbol{\Gamma}_*|\boldsymbol{\psi}, \boldsymbol{\Phi}, \boldsymbol{P}, \boldsymbol{\Lambda}, \nu, \boldsymbol{Y}) = p(\boldsymbol{\Gamma}_*|\boldsymbol{\psi}, \boldsymbol{\Phi}, \boldsymbol{P}, \boldsymbol{\Lambda}, \boldsymbol{Y})$. Discard $\boldsymbol{\Lambda}$. Let $\boldsymbol{\Gamma} := \boldsymbol{\Gamma}_*(\boldsymbol{\Gamma}_*'\boldsymbol{\Gamma}_*)^{-1/2}$ and $\boldsymbol{\Lambda} := \boldsymbol{\Lambda}_0 + \boldsymbol{\Lambda}_*(\boldsymbol{\Gamma}_*'\boldsymbol{\Gamma}_*)^{1/2}$. Accept the draw if $(\boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}) \in S$; otherwise go back to step 2 and draw another $(\boldsymbol{\Phi}, \boldsymbol{P}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma})$.

5. Draw ν from $p(\nu|\boldsymbol{\psi}, \boldsymbol{\Phi}, \boldsymbol{P}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \boldsymbol{Y}) = p(\nu|\boldsymbol{\Phi}, \boldsymbol{P})$.

The first block builds on Villani (2009); the second block is standard; the third and fourth blocks come from the collapsed Gibbs sampler proposed by Koop et al. (2010); the fifth block is standard. See the Appendix for the details of each block.

4.5 Bayes factor

We use the Bayes factor for Bayesian model selection. When choosing between nested models with certain priors, the Savage–Dickey (S–D) density ratio gives the Bayes factor without estimating the marginal likelihoods; see Wagenmakers, Lodewyckx, Kuriyal, and Grasman (2010) for a tutorial on the S–D method.

We choose the cointegrating rank r. Consider comparing the following two models (hypotheses):

$$H_0 : \text{rk}(\boldsymbol{\Pi}) = 0 \quad \text{vs} \quad H_r : \text{rk}(\boldsymbol{\Pi}) = r \tag{44}$$

Koop, León-González, and Strachan (2008, pp. 451–452) note that the problem is the same as comparing the following two nested models:

$$H_0 : \boldsymbol{\Lambda} = \mathbf{O}_{N\times r} \quad \text{vs} \quad H_r : \boldsymbol{\Lambda} \neq \mathbf{O}_{N\times r} \tag{45}$$

Ignoring the constraint that $(\boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}) \in S$ for the moment and assuming that $\tau_0 := 1$, so that the priors on $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$ are independent,[7] the S–D density ratio for H_0 vs H_r is

$$B_{0,r} = \frac{p(\boldsymbol{\Lambda} = \mathbf{O}_{N \times r} | \boldsymbol{Y}; H_r)}{p(\boldsymbol{\Lambda} = \mathbf{O}_{N \times r} | H_r)} \tag{46}$$

The prior gives the denominator directly. For the numerator, we have

$$p(\boldsymbol{\Lambda} | \boldsymbol{Y}; H_r) = \mathrm{E}(p(\boldsymbol{\Lambda} | \boldsymbol{\psi}, \boldsymbol{\Phi}, \boldsymbol{P}, \boldsymbol{\Gamma}, \boldsymbol{Y}; H_r) | \boldsymbol{Y}; H_r) \tag{47}$$

Let $\{\boldsymbol{\psi}_l, \boldsymbol{\Phi}_l, \boldsymbol{P}_l, \boldsymbol{\Gamma}_l\}_{l=1}^{L}$ be posterior draws. Let

$$\hat{p}(\boldsymbol{\Lambda} = \mathbf{O}_{N \times r} | \boldsymbol{Y}; H_r) := \frac{1}{L} \sum_{l=1}^{L} p(\boldsymbol{\Lambda} = \mathbf{O}_{N \times r} | \boldsymbol{\psi}_l, \boldsymbol{\Phi}_l, \boldsymbol{P}_l, \boldsymbol{\Gamma}_l, \boldsymbol{Y}; H_r)$$

An estimator of the the S–D density ratio for H_0 vs H_r is

$$\hat{B}_{0,r} = \frac{\hat{p}(\boldsymbol{\Lambda} = \mathbf{O}_{N \times r} | \boldsymbol{Y}; H_r)}{p(\boldsymbol{\Lambda} = \mathbf{O}_{N \times r} | H_r)} \tag{48}$$

5 APPLICATION

5.1 Data

We consider joint estimation of the natural rates (or their permanent components) and gaps of the following four macroeconomic variables in the US:

Output Let Y_t be output. Assume that $\{\ln Y_t\}$ is I(2), so that $\{\Delta \ln Y_t\}$ is I(1).

Inflation rate Let P_t be the price level and $\pi_t := \ln(P_t/P_{t-1})$ be the inflation rate. Assume that $\{\pi_t\}$ is I(1).

Interest rate Let I_t be the 3-month nominal interest rate (annual rate in per cent), $i_t := \ln(1 + I_t/400)$, $r_t := i_t - \mathrm{E}_t(\pi_{t+1})$ be the ex ante real interest rate, and

[7] If the priors on $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$ are dependent, then we can use the generalized S–D density ratio proposed by Verdinelli and Wasserman (1995).

TABLE 1: Data

Variable	Description
Y_t	Real GDP (billions of chained 2009 dollars, SA, AR)
P_t	CPI for all urban consumers: all items (1982–84=100, SA)
i_t	3-month treasury bill: secondary market rate (%, AR)
L_t	Civilian labor force (thousands of persons, SA)
E_t	Civilian employment level (thousands of persons, SA)

Note: SA means 'seasonally-adjusted'; AR means 'annual rate'.

$\hat{r}_t := i_t - \pi_{t+1}$ be the ex post real interest rate. Assume that $\{r_t\}$ is I(1).[8]

Unemployment rate Let L_t be the labor force, E_t be employment, and $U_t := -\ln(E_t/L_t)$ be the unemployment rate. Assume that $\{U_t\}$ is I(1).

Table 1 describes the data, which are available from FRED (Federal Reserve Economic Data). When monthly series are available, i.e., except for real GDP, we take the 3-month arithmetic means of monthly series each quarter to obtain quarterly series, from which we construct the quarterly inflation, interest, and unemployment rates as defined above.

Let for all t,

$$\boldsymbol{x}_t := \begin{pmatrix} \pi_t \\ \hat{r}_t \\ U_t \\ \ln Y_t \end{pmatrix}, \quad \boldsymbol{y}_t := \begin{pmatrix} \pi_t \\ \hat{r}_t \\ U_t \\ \Delta \ln Y_t \end{pmatrix}$$

The sample period of $\{\boldsymbol{y}_t\}$ is 1948Q1–2017Q4 (280 observations).

5.2 Preliminary analyses

We perform some preliminary analyses to check our assumption that $\{\boldsymbol{y}_t\}$ is I(1).[9] Table 2 shows the results of the ADF and ADF-GLS tests for unit root, with or without constant and/or trend terms in the ADF regression. The results depend on

[8] We can estimate the interest rate gap even if we observe $\{\hat{r}_t\}$ instead of $\{r_t\}$. See Murasawa (2014, pp. 499–500).

[9] We use gretl 2018d for the preliminary analyses.

TABLE 2: Unit root tests

Variable	Const.	Trend	ADF			ADF-GLS	
			Lags	τ	p-value	Lags	τ
π_t	yes	yes	4	-3.71	.02	4	-3.38^{**}
$\hat{r}_t$	yes	yes	7	-3.74	.02	5	-2.57
U_t	yes	yes	13	-2.89	.17	12	-2.27
$\Delta \ln Y_t$	yes	yes	0	-11.49	.00	1	-8.42^{***}
π_t	yes	no	4	-3.68	.00	4	-2.09^{**}
$\hat{r}_t$	yes	no	7	-3.74	.00	5	-1.65^{*}
U_t	yes	no	13	-3.01	.03	12	-1.49
$\Delta \ln Y_t$	yes	no	0	-11.26	.00	1	-6.52^{***}
$\Delta\pi_t$	yes	no	3	-11.79	.00	0	-14.04^{***}
$\Delta\hat{r}_t$	yes	no	8	-8.70	.00	0	-21.49^{***}
ΔU_t	yes	no	12	-5.15	.00	0	-7.42^{***}
$\Delta^2 \ln Y_t$	yes	no	14	-7.55	.00	0	-24.30^{***}
$\Delta\pi_t$	no	no	3	-11.82	.00		
$\Delta\hat{r}_t$	no	no	8	-8.73	.00		
ΔU_t	no	no	12	-5.16	.00		
$\Delta^2 \ln Y_t$	no	no	14	-7.56	.00		

Note: For the ADF-GLS test, *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. For the number of lags included in the ADF regression, we use the default choice in gretl 2018d with maximum 15, where the lag order selection criteria are AIC for the ADF test, and a modified AIC using the Perron and Qu (2007) method for the ADF-GLS test. With no constant nor trend in the ADF regression, the ADF test is asymptotically point optimal; hence the ADF-GLS test is unnecessary.

the number of lags included in the ADF regression, since the ADF test suffers from size distortion with short lags and low power with long lags. The ADF-GLS test remedies the problem except when there is no constant nor trend term in the ADF regression, in which case the ADF test is asymptotically point optimal. The level .05 ADF-GLS test rejects $H_0 : \{y_{t,i}\} \sim \mathrm{I}(1)$ in favor of $H_1 : \{y_{t,i}\} \sim \mathrm{I}(0)$ for $\{\pi_t\}$ and $\{\Delta \ln Y_t\}$. Hence these unit root tests do not support our assumption that $\{\boldsymbol{y}_t\}$ is I(1).

Table 3 shows the results of the KPSS stationarity tests, with or without a trend term. The results depend on the lag truncation parameter for the Newey–West estimator of the long-run error variance. With a trend term, the level .05 KPSS test rejects $H_0 : \{y_{t,i}\} \sim \mathrm{I}(0)$ in favor of $H_1 : \{y_{t,i}\} \sim \mathrm{I}(1)$ except for $\{\Delta \ln Y_t\}$. With no

TABLE 3: KPSS stationarity tests

Variable	Trend	LM
π_t	yes	.53***
$\hat{r}_t$	yes	.33***
U_t	yes	.22***
$\Delta \ln Y_t$	yes	.03
π_t	no	.57**
$\hat{r}_t$	no	.34
U_t	no	.71**
$\Delta \ln Y_t$	no	.49**
$\Delta\pi_t$	no	.03
$\Delta\hat{r}_t$	no	.06
ΔU_t	no	.06
$\Delta^2 \ln Y_t$	no	.01

Note: ** and *** denote significance at the 5% and 1% levels, respectively. The lag truncation parameter for the Newey–West estimator of the long-run error variance is 5 (the default value for our sample length in gretl 2018d).

trend term, the test rejects $H_0 : \{y_{t,i}\} \sim \mathrm{I}(0)$ in favor of $H_1 : \{y_{t,i}\} \sim \mathrm{I}(1)$ except for $\{\hat{r}_t\}$. Though the results for $\{\hat{r}_t\}$ and $\{\Delta \ln Y_t\}$ are mixed, these stationarity tests support our assumption that $\{\boldsymbol{y}_t\}$ is I(1).

Overall, these unit root and stationarity tests confirm that $\{\Delta \boldsymbol{y}_t\}$ is I(0), but are inconclusive if each component of $\{\boldsymbol{y}_t\}$ is I(1). With no strong counter-evidence, we proceed with our prior belief that each component of $\{\boldsymbol{y}_t\}$ is I(1). See Murasawa (2014) for an analysis based on an alternative assumption.

We also perform some preliminary analyses of cointegration in $\{\boldsymbol{y}_t\}$. Table 4 shows the results of the Engle–Granger cointegration tests, with or without a trend term in the cointegrating regression. The results depend on the number of lags included in the ADF regression for the residual series, but overall, the tests fail to reject the null of no cointegration (the residual series is I(1)) against the alternative of cointegration (the residual series is I(0)).

Table 5 shows the results of Johansen's cointegration tests, with unrestricted constant and restricted trend terms in the VECM. The results depend on the number of lags included in the VECM, but overall, the tests suggest that the cointegrating

TABLE 4: Engle–Granger cointegration tests

Trend	τ	p-value
yes	−3.75	.22
no	−3.56	.17

Note: The number of lags included in the ADF regression is 4 (the default value for our sample length in gretl 2018d).

TABLE 5: Johansen's cointegration tests (unrestricted constant and restricted trend in the VECM)

Rank	Trace	p-value	λ-max	p-value
0	141.91	.00	83.45	.00
1	58.46	.00	40.08	.00
2	18.39	.33	12.50	.38
3	5.89	.48	5.89	.49

Note: The number of lags included in the VECM is 4 (the default value for our sample length in gretl 2018d).

rank is 2.

Since the results are mixed, we estimate models with different cointegrating ranks, and use the Bayes factor to choose the cointegrating rank. Table 6 shows summary statistics of $\{\boldsymbol{y}_t\}$ and $\{\Delta\boldsymbol{y}_t\}$ multiplied by 100. Note that $\{\pi_t\}$, $\{\hat{r}_t\}$, and $\{\Delta \ln Y_t\}$ are quarterly rates of change (not annualized).

5.3 Model specification

For our data, $N := 4$. To select p, we fit VAR models to $\{\boldsymbol{y}_t\}$ up to VAR(8), i.e., $p = 7$, and check model selection criteria. The common estimation period is

TABLE 6: Summary statistics

Variable	Min.	1st qu.	Median	Mean	3rd qu.	Max.
$100\pi_t$	−2.32	.39	.75	.85	1.13	3.98
$100\hat{r}_t$	−3.65	−.25	.23	.19	.57	2.69
$100U_t$	2.64	4.77	5.74	5.98	7.08	11.28
$100\Delta \ln Y_t$	−2.63	.32	.76	.78	1.28	3.91
$100\Delta\pi_t$	−3.85	−.26	.00	.00	.29	1.98
$100\Delta\hat{r}_t$	−1.95	−.27	.02	.00	.25	3.81
$100\Delta U_t$	−1.03	−.22	−.06	.00	.13	1.79
$100\Delta^2 \ln Y_t$	−2.84	−.71	−.02	.00	.00	4.81

TABLE 7: Lag order selection

Lag	Log-lik	LR	p-value	AIC	BIC	HQC
1	4329.73			−31.66	−31.34	−31.53
2	4676.38	693.31	.00	−34.09	−33.56*	−33.88
3	4705.68	58.61	.00	−34.19	−33.45	−33.89*
4	4723.70	36.02	.00	−34.20	−33.25	−33.82
5	4737.87	28.35	.03	−34.19	−33.02	−33.72
6	4757.53	39.31	.00	−34.22	−32.84	−33.66
7	4777.70	40.34	.00	−34.25*	−32.66	−33.61
8	4790.17	24.95	.07	−34.22	−32.42	−33.50

Note: For AIC, BIC, and HQC, * denotes the selected model. The LR test statistic for testing $H_0 : \{\boldsymbol{y}_t\} \sim \mathrm{VAR}(p-1)$ vs $H_1 : \{\boldsymbol{y}_t\} \sim \mathrm{VAR}(p)$ follows $\chi^2(16)$ under H_0.

1950Q1–2017Q4. Table 7 summarizes the results of lag order selection.[10] The level .05 LR test fails to reject $H_0 : \{\boldsymbol{y}_t\} \sim \mathrm{VAR}(7)$ against $H_1 : \{\boldsymbol{y}_t\} \sim \mathrm{VAR}(8)$ and AIC selects VAR(7), whereas BIC and HQC select much smaller models. Since a high-order VAR model covers low-order VAR models as special cases, to be conservative, we assume a VAR(8) model for $\{\boldsymbol{y}_t\}$, i.e., we choose $p = 7$, and impose a shrinkage prior on the VAR coefficients.

For the prior on $\boldsymbol{\alpha}$, we set $\boldsymbol{\alpha}_0 := \mathbf{0}_N$ and $\boldsymbol{Q}_{0,\boldsymbol{\alpha}} := \boldsymbol{I}_N$; hence the prior on $\boldsymbol{\beta}$ is $\mathrm{N}_r(\mathbf{0}_r, \boldsymbol{I}_r)$ independent of $\boldsymbol{\Gamma}$. For the prior on $\boldsymbol{\mu}$, we set $\boldsymbol{\mu}_0 := \hat{\boldsymbol{\mu}}$, where $\hat{\boldsymbol{\mu}}$ is the sample mean of $\{\Delta \ln \boldsymbol{y}_t\}$, and $\boldsymbol{Q}_{0,\boldsymbol{\mu}} := \boldsymbol{I}_N$.

Following Kadiyala and Karlsson (1997) and Bańbura, Giannone, and Reichlin (2010), we set

$$
\begin{aligned}
\boldsymbol{M}_0 &:= \mathbf{O}_{N \times pN} \\
\boldsymbol{D}_0 &:= \mathrm{diag}(1, \ldots, p)^2 \otimes \mathrm{diag}(s_1, \ldots, s_N)^2 \\
k_0 &:= N + 2 \\
\boldsymbol{S}_0 &:= (k_0 - N - 1)\, \mathrm{diag}(s_1, \ldots, s_N)^2
\end{aligned}
$$

[10] The standard lag order selection criteria are valid even when $\{\boldsymbol{y}_t\}$ is I(1); see Kilian and Lütkepohl (2017, pp. 99–100).

where for $i = 1, \dots, N$, s_i^2 is an estimate of $\mathrm{var}(u_{t,i})$ based on the univariate AR(p+1) model with constant and trend terms for $\{y_{t,i}\}$.

For the prior on $(\boldsymbol{\Lambda}, \boldsymbol{\Gamma})$, we set $\boldsymbol{\Lambda}_0 := \mathbf{O}_{N\times r}$, $\boldsymbol{G}_0 := \boldsymbol{I}_N$, $\eta_0 := .01$, and $\tau_0 := 1$. Since $\tau_0 := 1$, we have a flat prior on the cointegrating space, and the priors on $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$ are independent.

For the prior on ν, we set $A_0 := 1$ and $B_0 := 1$, i.e., $\nu \sim \chi^2(1)$; hence the tightness hyperparameter on the VAR coefficients tends to be small, implying potentially mild shrinkage toward $\boldsymbol{M}_0 := \mathbf{O}_{N\times pN}$.

Overall, our priors are weakly informative in the sense of Gelman et al. (2014, p. 55).

5.4 Bayesian computation

We run our Gibbs sampler on R 3.5.2 developed by R Core Team (2018). We use the ML estimate of $(\boldsymbol{\Phi}, \boldsymbol{P}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma})$ for their initial values.[11] With poor initial values, the restriction that the eigenvalues of $\boldsymbol{A}$ lie inside the unit circle does not hold, and the iteration cannot start. Hence the choice of initial values is important when one applies the B–N decomposition. Once the iteration starts, the restriction rarely binds for our sample.

To check convergence of the Markov chain generated by our Gibbs sampler to its stationary distribution, we perform convergence diagnoses discussed in Robert and Casella (2009, ch. 8) and available in the `coda` package for R. Given the diagnoses, we discard the initial 1,000 draws, and use the next 4,000 draws for the posterior inference.

To select the cointegrating rank r, we set $p := 7$, assume the above priors, and compute the S–D density ratios for $r = 1, 2, 3$, from which we compute the posterior probabilities of $r = 0, 1, 2, 3$, assuming equal prior probabilities. We find that the posterior probability of $r = 2$ is numerically 1, consistent with the results

[11] The `urca` package for R is useful for ML estimation of a VECM.

of Johansen's cointegration tests. Thus we set $r := 2$ in the following analysis.

5.5 Empirical results

Figure 2 plots the actual rates and our point estimates (posterior medians) of the natural rates (or their permanent components) of the four variables. For ease of comparison of the actual and natural rates, we omit error bands for the natural rates, which are identical to those for the gaps. Figure 3 plots our point estimates of the gaps and their 95% error bands.

Figure 2

Figure 3

Our estimate of the natural rate of inflation is smoother than typical univariate estimates of trend inflation in the CPI; e.g., Faust and Wright (2013, p. 22). It looks close to a recent estimate of trend inflation in the US CPI in J. Morley, Piger, and Rasche (2015, p. 894) based on a bivariate UC model for the inflation and unemployment rates, which assumes independent shocks to the trend and gap components and allows for structural breaks in the variances of these shocks. It is more volatile, however, than a recent estimate of trend inflation in the PCE price index by Chan, Clark, and Koop (2018, p. 21) that uses information in survey inflation expectations.

Our estimate of the natural rate of interest is more volatile than a recent estimate by Del Negro et al. (2017, p. 237) based on a VAR model with common trends, i.e., a multivariate UC model with independent shocks to the trend and gap components and a factor structure for the trend components, for short- and long-term interest rates, inflation and its survey expectations, and some other variables. Their estimate is smooth partly because they impose tight priors on the variances of the shocks to the trend components. Indeed, their estimate with the loosest possible prior is as volatile as ours; see Del Negro et al. (2017, p. 272). Interestingly, their estimate based on a DSGE model looks close to ours despite different definitions of the natural rate; see Del Negro et al. (2017, p. 237).[12] Estimates by Laubach and Williams (2016, p. 60), Holston et al. (2017, p. S61), and Lewis and Vazquez-Grande (2019) are close to trend output growth by construction, and quite different from estimates by Del Negro et al. (2017) and ours, especially before 1980.[13]

Our estimate of the natural rate of unemployment looks more volatile than a recent estimate in J. Morley et al. (2015, p. 898), obtained as a by-product of estimation of trend inflation. A possible reason for the difference is that they assume independent shocks to the trend and gap components, which may not hold in practice. If one allows for dependence between the shocks, then the two estimates may coincide, as J. C. Morley, Nelson, and Zivot (2003) show for the univariate trend–cycle decomposition. Our estimate of the unemployment rate gap looks close to the estimate in J. Morley et al. (2015, p. 901) in terms of the sign and magnitude, despite the difference in the volatility.

Our estimate of the output gap is at most about $\pm 5\%$ of the output level, and looks close to a recent estimate by J. Morley and Wong (2018, Fig. 2) based on a large Bayesian VAR(4) model with 23 variables. It also looks close to their estimate based on a Bayesian VAR(4) model with four variables using output growth, the unemployment rate, the CPI inflation rate, and the growth rate of industrial production, assuming that they are all I(0). Though output growth may or may not be I(0) in the US, it may be clearly I(1) in some countries or regions, in which case their method may give an unreasonable estimate of the output gap with a strong upward or downward trend. See Murasawa (2015) for such an example for the Japanese data.

[12] Del Negro et al. (2017) clearly distinguish the natural rate and its low-frequency component, estimating the former by a DSGE model and the latter by a VAR model.

[13] Holston et al. (2017, p. S63) writes,

> . . . we assume a one-for-one relationship between the trend growth rate of output and the natural rate of interest, which corresponds to assuming $\sigma = 1$ in Eq. (1).

where Eq. (1) in their paper is the consumption Euler equation in a steady state.

Figure 4 plots the posterior probability of positive gap for the four variables. This probability index is useful if the sign of the gap is of interest. Even if the 95% error band for the gap covers 0, the posterior probability of positive gap may be close to .025 or .975. Indeed, the probability index is often below .25 or above .75; hence we are often quite sure about the sign of the gap. Moreover, Figure 4 shows the relation between the gaps more clearly than Figure 3.

Figure 4

Figure 5 shows the relation between the gaps more directly. The left panels are the scatter plots of the posterior medians of the gaps in each quarter. The right panels are the posterior pdfs of the correlation coefficients between the gaps. We see that the Phillips curves and Okun's law hold between the gaps, though we do not impose such relations. Thus our estimates of the gaps seem mutually consistent from a macroeconomic point of view.

Figure 5

5.6 Comparison of alternative model specifications

Figure 6 compares point estimates of the gaps under three alternative assumptions, i.e., I(1) log output, I(2) log output with no cointegration, and I(2) log output with cointegration. For ease of comparison, we omit error bands here.

Figure 6

The result assuming I(1) log output is similar to that in Murasawa (2014), who uses a VAR model with no constant term for the differenced and centered series, sets $p := 12$, and chooses the tightness hyperparameter by the empirical Bayes method. In contrast to the result for the Japanese data in Murasawa (2015), for the US data, the multivariate B–N decomposition assuming I(1) log output gives "reasonable" estimates of the gaps, except that the output gap is persistently positive since 2010.

Assuming I(2) log output with no cointegration changes the estimate of the output gap, but hardly changes the estimates of other gaps, which is similar to the result for the Japanese data in Murasawa (2015). In particular, the output gap now keeps fluctuating around 0 since 2010, which seems more "reasonable".

Allowing for cointegration changes the estimates of all gaps, and we obtain much bigger gaps. The result makes sense because the B–N transitory components are "forecastable movements", and a VECM forecasts better than a VAR model, especially for our data. Moreover, these bigger gaps seem close to other recent estimates that focus on a particular gap, as already noted.

6 DISCUSSION

The consumption Euler equation implies that if the real interest rate is I(1), then so is the output growth rate with possible cointegration, and log output is I(2). We extend the multivariate B–N decomposition to such a case. To obtain error bands for the components, we apply Bayesian analysis. In particular, we assume hierarchical weakly informative priors, and develop a Gibbs sampler for posterior simulation. Application of the method to US data gives a reasonable joint estimate of the natural rates (or their permanent components) and gaps of output, inflation, interest, and unemployment.

The B–N decomposition assuming I(1) log output often gives an unreasonable estimate of the output gap, perhaps because of possible structural breaks in the mean output growth rate. Assuming I(2) log output, i.e., I(1) output growth rate, we introduce a stochastic trend in the output growth rate, which captures possible structural breaks in the mean growth rate automatically in real time without specifying break dates a priori, leading to a more reasonable estimate of the output gap that fluctuates around 0. Moreover, since the B–N transitory components are "forecastable movements" and a VECM forecasts no worse than a VAR model, allowing for cointegration gives larger estimates of all gaps.

Since a reduced-form VECM is the most basic forecasting model for cointegrated series, the multivariate B–N decomposition based on a VECM gives a benchmark joint estimate of the natural rates (or their permanent components) and gaps. One can compare this benchmark estimate with alternative estimates based on other forecasting models or DSGE models such as those in Del Negro et al. (2017). We conjecture that our method is useful especially for non-US data, where log output is often clearly I(2). Confirming this conjecture is an interesting and important issue for future work.

One can possibly refine our estimate in two ways. First, one can use a larger model with more variables, assuming a factor structure if necessary. Second, one can introduce Markov-switching, stochastic volatility, or more general time-varying parameters to our VECM. Since the B–N decomposition may not apply to nonlinear models, and nonlinear models may become unnecessary with more variables, the first direction seems more promising.

Lastly, to obtain a monthly instead of quarterly joint estimate of the natural rates and gaps of output and other variables, it seems straightforward to extend the B–N decomposition of mixed-frequency series proposed by Murasawa (2016) to I(1) and I(2) series with cointegration.

REFERENCES

Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, *25*, 71–92. doi: `10.1002/jae.1137`

Beveridge, S., & Nelson, C. R. (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. doi: 10.1016/0304-3932(81)90040-4

Chan, J. C. C., Clark, T. E., & Koop, G. (2018). A new model of inflation, trend inflation, and long-run inflation expectations. *Journal of Money, Credit and Banking*, *50*, 5–53. doi: 10.1111/jmcb.12452

Chikuse, Y. (1990). The matrix angular central Gaussian distribution. *Journal of Multivariate Analysis*, *33*, 265–274. doi: 10.1016/0047-259X(90)90050-R

Cochrane, J. H. (1994). Permanent and transitory components of GNP and stock prices. *Quarterly Journal of Economics*, *109*, 241–265. doi: 10.2307/2118434

Cogley, T., Morozov, S., & Sargent, T. J. (2005). Bayesian fan charts for U.K. inflation: Forecasting and sources of uncertainty in an evolving monetary system. *Journal of Economic Dynamics and Control*, *29*, 1893–1925. doi: 10.1016/j.jedc.2005.06.005

Cogley, T., Primiceri, G. E., & Sargent, T. J. (2010). Inflation-gap persistence in the US. *American Economic Journal: Macroeconomics*, *2*, 43–69. doi: 10.1257/mac.2.1.43

Cogley, T., & Sargent, T. J. (2002). Evolving post-World War II U.S. inflation dynamics. *NBER Macroeconomics Annual 2001*, *16*, 331–373. doi: 10.1086/654451

Cogley, T., & Sargent, T. J. (2005). Drifts and volatilities: Monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, *8*, 262–302. doi: 10.1016/j.red.2004.10.009

Del Negro, M., Giannone, D., Giannoni, M. P., & Tambalotti, A. (2017). Safety, liquidity, and the natural rate of interest. *Brookings Papers on Economic Activity*, *48*(1), 235–316.

Drèze, J. H., & Richard, J.-F. (1983). Bayesian analysis of simultaneous equation systems. In Z. Griliches & M. D. Intriligator (Eds.), *Handbook of econometrics* (Vol. 1, pp. 517–598). Elsevier. doi: 10.1016/s1573-4412(83)01013-2

Evans, G. W. (1989a). A measure of the U.S. output gap. *Economics Letters*, *29*, 285–289. doi: 10.1016/0165-1765(89)90202-4

Evans, G. W. (1989b). Output and unemployment dynamics in the United States: 1950–1985. *Journal of Applied Econometrics*, *4*, 213–237. doi: 10.1002/jae.3950040302

Evans, G. W., & Reichlin, L. (1994). Information, forecasts, and measurement of the business cycle. *Journal of Monetary Economics*, *33*, 233–254. doi: 10.1016/0304-3932(94)90002-7

Faust, J., & Wright, J. H. (2013). Forecasting inflation. In G. Elliott & A. Timmermann (Eds.), *Handbook of economic forecasting* (Vol. 2A, pp. 2–56). Elsevier. doi: 10.1016/b978-0-444-53683-9.00001-3

Galí, J. (2015). *Monetary policy, inflation, and the business cycle* (2nd ed.). Princeton University Press.

Garratt, A., Robertson, D., & Wright, S. (2006). Permanent vs transitory components and economic fundamentals. *Journal of Applied Econometrics*, *21*, 521–542. doi: 10.1002/jae.850

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). CRC Press.

Giannone, D., Lenza, M., & Primiceri, G. E. (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics*, *97*, 436–451. doi: 10.1162/rest_a_00483

Gupta, A. K., & Nagar, D. K. (1999). *Matrix variate distributions.* Taylor & Francis Ltd.

Holston, K., Laubach, T., & Williams, J. C. (2017). Measuring the natural rate of interest: International trends and determinants. *Journal of International Economics*, *108*, S59–S75. doi: 10.1016/j.jinteco.2017.01.004

Kadiyala, K. R., & Karlsson, S. (1997). Numerical methods for estimation and inference in Bayesian VAR-models. *Journal of Applied Econometrics*, *12*, 99–132. doi: 10.1002/(SICI)1099-1255(199703)12:2<99::AID-JAE429>3.0.CO;2-A

Kamber, G., Morley, J., & Wong, B. (2018). Intuitive and reliable estimates of the output gap from a Beveridge–Nelson filter. *Review of Economics and Statistics*, *100*, 550–566. doi: 10.1162/rest_a_00691

Kiley, M. T. (2013). Output gaps. *Journal of Macroeconomics*, *37*, 1–18. doi: 10.1016/j.jmacro.2013.04.002

Kilian, L., & Lütkepohl, H. (2017). *Structural vector autoregressive analysis.* Cambridge University Press.

King, R., Plosser, C., Stock, J., & Watson, M. (1991). Stochastic trends and economic fluctuations. *American Economic Review*, *81*, 819–840.

Koop, G. (2003). *Bayesian econometrics.* John Wiley & Sons.

Koop, G., León-González, R., & Strachan, R. W. (2008). Bayesian inference in a cointegrating panel data model. In S. Chib, W. Griffiths, G. Koop, & D. Terrell (Eds.), *Bayesian econometrics* (Vol. 23, pp. 433–469). Emerald. doi: 10.1016/s0731-9053(08)23013-6

Koop, G., León-González, R., & Strachan, R. W. (2010). Efficient posterior simulation for cointegrated models with priors on the cointegration space. *Econometric Reviews*, *29*, 224–242. doi: 10.1080/07474930903382208

Laubach, T., & Williams, J. C. (2016). Measuring the natural rate of interest redux. *Business Economics*, *51*, 57–67. doi: 10.1057/be.2016.23

Lewis, K. F., & Vazquez-Grande, F. (2019). Measuring the natural rate of interest: A note on transitory shocks. *Journal of Applied Econometrics*. (in press) doi: 10.1002/jae.2671

Morley, J., Piger, J., & Rasche, R. (2015). Inflation in the G7: Mind the gap(s)? *Macroeconomic Dynamics*, *19*, 883–912. doi: 10.1017/s1365100513000655

Morley, J., & Wong, B. (2018, September). *Estimating and accounting for the output gap with large Bayesian vector autoregressions* (Economics Working

Paper Series No. 2018-4). University of Sydney.

Morley, J. C. (2002). A state-space approach to calculating the Beveridge–Nelson decomposition. *Economics Letters*, *75*, 123–127. doi: 10.1016/s0165-1765(01)00581-x

Morley, J. C., Nelson, C. R., & Zivot, E. (2003). Why are the Beveridge–Nelson and unobserved-components decompositions of GDP so different? *Review of Economics and Statistics*, *85*, 235–243. doi: 10.1162/003465303765299765

Muirhead, R. J. (1982). *Aspects multivariate statistic theory.* John Wiley & Sons.

Murasawa, Y. (2014). Measuring the natural rates, gaps, and deviation cycles. *Empirical Economics*, *47*, 495–522. doi: 10.1007/s00181-013-0747-9

Murasawa, Y. (2015). The multivariate Beveridge–Nelson decomposition with I(1) and I(2) series. *Economics Letters*, *137*, 157–162. doi: 10.1016/j.econlet.2015.11.001

Murasawa, Y. (2016). The Beveridge–Nelson decomposition of mixed-frequency series. *Empirical Economics*, *51*, 1415–1441. doi: 10.1007/s00181-015-1061-5

Nelson, C. R. (2008). The Beveridge–Nelson decomposition in retrospect and prospect. *Journal of Econometrics*, *146*, 202–206. doi: 10.1016/j.jeconom.2008.08.008

Newbold, P., & Vougas, D. (1996). Beveridge–Nelson-type trends for I(2) and some seasonal models. *Journal of Time Series Analysis*, *17*, 151–169. doi: 10.1111/j.1467-9892.1996.tb00270.x

Oh, K. H., & Zivot, E. (2006, January). *The Clark model with correlated components.* (unpublished) doi: 10.2139/ssrn.877398

Oh, K. H., Zivot, E., & Creal, D. (2008). The relationship between the Beveridge–Nelson decomposition and other permanent–transitory decompositions that are popular in economics. *Journal of Econometrics*, *146*, 207–219. doi: 10.1016/j.jeconom.2008.08.021

Perron, P., & Qu, Z. (2007). A simple modification to improve the finite sample properties of Ng and Perron's unit root tests. *Economics Letters*. doi: 10.1016/j.econlet.2006.06.009

Phelps, E. S. (1995). The origins and further development of the natural rate of unemployment. In R. Cross (Ed.), *The natural rate of unemployment: Reflections on 25 years of the hypothesis* (pp. 15–31). Cambridge University Press.

R Core Team. (2018). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Robert, C. P., & Casella, G. (2009). *Introducing Monte Carlo methods with R.* Springer.

Rotemberg, J. J., & Woodford, M. (1996). Real-business-cycle models and the forecastable movements in output, hours, and consumption. *American Economic Review*, *86*, 71–89.

Strachan, R. W., & Inder, B. (2004). Bayesian analysis of the error correction model. *Journal of Econometrics*, *123*, 307–325. doi: 10.1016/j.jeconom.2003.12.004

Verdinelli, I., & Wasserman, L. (1995). Computing Bayes factors using a generalization of the Savage–Dickey density ratio. *Journal of the American Statistical Association*, *90*, 614–618. doi: 10.1080/01621459.1995.10476554

Villani, M. (2009). Steady-state priors for vector autoregressions. *Journal of Applied Econometrics*, *24*, 630–650. doi: 10.1002/jae.1065

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*, 158–189. doi: 10.1016/j.cogpsych.2009.12.001

Woodford, M. (2003). *Interest and prices.* Princeton University Press.

A APPENDIX: GIBBS SAMPLER

A.1 Useful lemmas

Our Gibbs sampler relies on the following two familiar results in Bayesian analysis of normal linear models, which we state as lemmas for ease of reference.

Lemma 1. *Suppose that*

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$$
$$\boldsymbol{u} \sim \mathrm{N}_n\left(\mathbf{0}_n, \boldsymbol{P}^{-1}\right)$$

and

$$\boldsymbol{\beta} \sim \mathrm{N}_k\left(\boldsymbol{\mu}_0, \boldsymbol{D}_0^{-1}\right)$$

Then

$$\boldsymbol{\beta}|\boldsymbol{P}, \boldsymbol{y}, \boldsymbol{X} \sim \mathrm{N}_k\left(\boldsymbol{\mu}_1, \boldsymbol{D}_1^{-1}\right)$$

where

$$\boldsymbol{D}_1 := \boldsymbol{X}'\boldsymbol{P}\boldsymbol{X} + \boldsymbol{D}_0$$
$$\boldsymbol{\mu}_1 := \boldsymbol{D}_1^{-1}(\boldsymbol{X}'\boldsymbol{P}\boldsymbol{X}\boldsymbol{b}_{\mathrm{GLS}} + \boldsymbol{D}_0\boldsymbol{\mu}_0)$$

with $\boldsymbol{b}_{\mathrm{GLS}} := (\boldsymbol{X}'\boldsymbol{P}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{P}\boldsymbol{y}$.

Proof. See Koop (2003, pp. 118–121). □

Lemma 2. *Suppose that*

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B}' + \boldsymbol{U}$$
$$\boldsymbol{U} \sim \mathrm{N}_{n\times m}\left(\mathbf{O}_{n\times m}; \boldsymbol{I}_n, \boldsymbol{P}^{-1}\right)$$

and

$$\begin{aligned}\boldsymbol{B}|\boldsymbol{P} &\sim \mathrm{N}_{m\times k}\left(\boldsymbol{M}_0; \boldsymbol{P}^{-1}, \boldsymbol{D}_0^{-1}\right)\\ \boldsymbol{P} &\sim \mathrm{W}_m\left(k_0; \boldsymbol{S}_0^{-1}\right)\end{aligned}$$

Then

$$\begin{aligned}\boldsymbol{B}|\boldsymbol{P},\boldsymbol{Y},\boldsymbol{X} &\sim \mathrm{N}_{m\times k}\left(\boldsymbol{M}_1; \boldsymbol{P}^{-1}, \boldsymbol{D}_1^{-1}\right)\\ \boldsymbol{P}|\boldsymbol{Y},\boldsymbol{X} &\sim \mathrm{W}_m\left(k_1; \boldsymbol{S}_1^{-1}\right)\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{D}_1 &:= \boldsymbol{X}'\boldsymbol{X} + \boldsymbol{D}_0\\ \boldsymbol{M}_1 &:= (\boldsymbol{B}_{\mathrm{OLS}}\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{M}_0\boldsymbol{D}_0)\boldsymbol{D}_1^{-1}\\ k_1 &:= n + k_0\\ \boldsymbol{S}_1 &:= (\boldsymbol{B}_{\mathrm{OLS}} - \boldsymbol{M}_0)\left[(\boldsymbol{X}'\boldsymbol{X})^{-1} + \boldsymbol{D}_0^{-1}\right]^{-1}(\boldsymbol{B}_{\mathrm{OLS}} - \boldsymbol{M}_0)' + \boldsymbol{S} + \boldsymbol{S}_0\end{aligned}$$

with $\boldsymbol{B}_{\mathrm{OLS}} := \boldsymbol{Y}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$ *and* $\boldsymbol{S} := (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}'_{\mathrm{OLS}})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}'_{\mathrm{OLS}})$.

Proof. See Drèze and Richard (1983, pp. 539–541). □

A.2 Steady state parameters

Write the VECM as for all t,

$$\boldsymbol{\Phi}(\mathrm{L})\Delta\boldsymbol{y}_t - \boldsymbol{\Phi}(1)\boldsymbol{\mu} = -\boldsymbol{\Lambda}[\boldsymbol{\Gamma}'\boldsymbol{y}_{t-1} - \boldsymbol{\beta} - \boldsymbol{\Gamma}'\boldsymbol{\mu}(t-1)] + \boldsymbol{u}_t$$

or

$$\boldsymbol{\Phi}(\mathrm{L})\Delta\boldsymbol{y}_t + \boldsymbol{\Lambda}\boldsymbol{\Gamma}'\boldsymbol{y}_{t-1} = \boldsymbol{\Lambda}\boldsymbol{\beta} + [\boldsymbol{\Phi}(1) + (t-1)\boldsymbol{\Lambda}\boldsymbol{\Gamma}']\boldsymbol{\mu} + \boldsymbol{u}_t \tag{49}$$

Let for all t,

$$\boldsymbol{w}_t := \boldsymbol{\Phi}(\mathrm{L})\Delta\boldsymbol{y}_t + \boldsymbol{\Lambda}\boldsymbol{\Gamma}'\boldsymbol{y}_{t-1}$$

$$\boldsymbol{Z}_t := \begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{\Phi}(1) + (t-1)\boldsymbol{\Lambda}\boldsymbol{\Gamma}' \end{bmatrix}$$

Then for all t,

$$\boldsymbol{w}_t = \boldsymbol{Z}_t\boldsymbol{\psi} + \boldsymbol{u}_t$$

Let

$$\boldsymbol{w} := \begin{pmatrix} \boldsymbol{w}_{p+1} \\ \vdots \\ \boldsymbol{w}_T \end{pmatrix}, \quad \boldsymbol{Z} := \begin{bmatrix} \boldsymbol{Z}_{p+1} \\ \vdots \\ \boldsymbol{Z}_T \end{bmatrix}, \quad \boldsymbol{u} := \begin{pmatrix} \boldsymbol{u}_{p+1} \\ \vdots \\ \boldsymbol{u}_T \end{pmatrix}$$

Then we have a normal linear model for $\boldsymbol{w}$ given $\boldsymbol{Z}$ such that

$$\boldsymbol{w} = \boldsymbol{Z}\boldsymbol{\psi} + \boldsymbol{u} \tag{50}$$

$$\boldsymbol{u} \sim \mathrm{N}_{N(T-p)}\left(\mathbf{0}_{N(T-p)}, (\boldsymbol{I}_{T-p} \otimes \boldsymbol{P})^{-1}\right) \tag{51}$$

Let $\boldsymbol{\psi}_{\mathrm{GLS}} := [\boldsymbol{Z}'(\boldsymbol{I}_{T-p} \otimes \boldsymbol{P})\boldsymbol{Z}]^{-1}\boldsymbol{Z}'(\boldsymbol{I}_{T-p} \otimes \boldsymbol{P})\boldsymbol{w}$.

Theorem 2.

$$\boldsymbol{\psi}|\boldsymbol{\Phi}, \boldsymbol{P}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \boldsymbol{w} \sim \mathrm{N}_{r+N}\left(\boldsymbol{\psi}_1, \boldsymbol{Q}_1^{-1}\right) \tag{52}$$

where

$$\boldsymbol{Q}_1 := \boldsymbol{Z}'(\boldsymbol{I}_{T-p} \otimes \boldsymbol{P})\boldsymbol{Z} + \boldsymbol{Q}_0$$

$$\boldsymbol{\psi}_1 := \boldsymbol{Q}_1^{-1}[\boldsymbol{Z}'(\boldsymbol{I}_{T-p} \otimes \boldsymbol{P})\boldsymbol{Z}\boldsymbol{\psi}_{\mathrm{GLS}} + \boldsymbol{Q}_0\boldsymbol{\psi}_0]$$

Proof. Apply Lemma 1. □

A.3 VAR parameters

Let for all t,

$$\boldsymbol{e}_t := \boldsymbol{\Gamma}'\boldsymbol{y}_t - \boldsymbol{\beta} - \boldsymbol{\delta}t, \quad \boldsymbol{s}_t^* := \begin{pmatrix} \Delta\boldsymbol{y}_t - \boldsymbol{\mu} \\ \vdots \\ \Delta\boldsymbol{y}_{t-p+1} - \boldsymbol{\mu} \end{pmatrix}$$

Write the VECM as for all t,

$$\Delta\boldsymbol{y}_t - \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{e}_{t-1} = \boldsymbol{\Phi}\boldsymbol{s}_{t-1}^* + \boldsymbol{u}_t \tag{53}$$

Let

$$\boldsymbol{Y}^* := \begin{bmatrix} (\Delta\boldsymbol{y}_{p+1} - \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{e}_p)' \\ \vdots \\ (\Delta\boldsymbol{y}_T - \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{e}_{T-1})' \end{bmatrix}, \quad \boldsymbol{X} := \begin{bmatrix} \boldsymbol{s^*}'_p \\ \vdots \\ \boldsymbol{s^*}'_{T-1} \end{bmatrix}, \quad \boldsymbol{U} := \begin{bmatrix} \boldsymbol{u}'_{p+1} \\ \vdots \\ \boldsymbol{u}'_T \end{bmatrix}$$

Then we have a normal linear model for $\boldsymbol{Y}^*$ given $\boldsymbol{X}$ such that

$$\boldsymbol{Y}^* = \boldsymbol{X}\boldsymbol{\Phi}' + \boldsymbol{U} \tag{54}$$

$$\boldsymbol{U} \sim \mathrm{N}_{(T-p)\times N}\left(\mathbf{O}_{(T-p)\times N}; \boldsymbol{I}_{T-p}, \boldsymbol{P}^{-1}\right) \tag{55}$$

Let $\boldsymbol{\Phi}_{\mathrm{OLS}} := \boldsymbol{Y}^{*\prime}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$ and $\boldsymbol{S} := (\boldsymbol{Y}^* - \boldsymbol{X}\boldsymbol{\Phi}'_{\mathrm{OLS}})'(\boldsymbol{Y}^* - \boldsymbol{X}\boldsymbol{\Phi}'_{\mathrm{OLS}})$.

Theorem 3.

$$\boldsymbol{\Phi}|\boldsymbol{P}, \boldsymbol{\psi}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \nu, \boldsymbol{Y}^*, \boldsymbol{X} \sim \mathrm{N}_{N\times pN}\left(\boldsymbol{M}_1; \boldsymbol{P}^{-1}, \boldsymbol{D}_1^{-1}\right)[(\boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}) \in S] \tag{56}$$

$$\boldsymbol{P}|\boldsymbol{\psi}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \nu, \boldsymbol{Y}^*, \boldsymbol{X} \sim \mathrm{W}_N\left(k_1; \boldsymbol{S}_1^{-1}\right) \tag{57}$$

where

$$
\begin{aligned}
\boldsymbol{D}_1 &:= \boldsymbol{X}'\boldsymbol{X} + \nu \boldsymbol{D}_0 \\
\boldsymbol{M}_1 &:= (\boldsymbol{\Phi}_{\mathrm{OLS}}\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{M}_0 \nu \boldsymbol{D}_0)\boldsymbol{D}_1^{-1} \\
k_1 &:= T - p + k_0 \\
\boldsymbol{S}_1 &:= (\boldsymbol{\Phi}_{\mathrm{OLS}} - \boldsymbol{M}_0)\left[(\boldsymbol{X}'\boldsymbol{X})^{-1} + (\nu \boldsymbol{D}_0)^{-1}\right]^{-1}(\boldsymbol{\Phi}_{\mathrm{OLS}} - \boldsymbol{M}_0)' + \boldsymbol{S} + \boldsymbol{S}_0
\end{aligned}
$$

Proof. Apply Lemma 2. □

A.4 Loading matrix

Write the VECM as for all t,

$$
\boldsymbol{\Phi}(\mathrm{L})(\Delta \boldsymbol{y}_t - \boldsymbol{\mu}) = -\boldsymbol{\Lambda} \boldsymbol{e}_{t-1} + \boldsymbol{u}_t \tag{58}
$$

Let

$$
\boldsymbol{W} := \begin{bmatrix} [\boldsymbol{\Phi}(\mathrm{L})(\Delta \boldsymbol{y}_{p+1} - \boldsymbol{\mu})]' \\ \vdots \\ [\boldsymbol{\Phi}(\mathrm{L})(\Delta \boldsymbol{y}_T - \boldsymbol{\mu})]' \end{bmatrix}, \quad \boldsymbol{E} := \begin{bmatrix} -\boldsymbol{e}_p' \\ \vdots \\ -\boldsymbol{e}_{T-1}' \end{bmatrix}
$$

Then we have a normal linear model for $\boldsymbol{W}$ given $\boldsymbol{E}$ such that

$$
\begin{aligned}
\boldsymbol{W} &= \boldsymbol{E}\boldsymbol{\Lambda}' + \boldsymbol{U} \\
\boldsymbol{U} &\sim \mathrm{N}_{(T-p)\times N}\left(\mathbf{O}_{(T-p)\times N}; \boldsymbol{I}_{T-p}, \boldsymbol{P}^{-1}\right)
\end{aligned}
$$

or

$$
\begin{aligned}
\boldsymbol{W}' &= \boldsymbol{\Lambda}\boldsymbol{E}' + \boldsymbol{U}' \\
\boldsymbol{U}' &\sim \mathrm{N}_{N\times(T-p)}\left(\mathbf{O}_{N\times(T-p)}; \boldsymbol{P}^{-1}, \boldsymbol{I}_{T-p}\right)
\end{aligned}
$$

Let $\boldsymbol{\lambda} := \text{vec}(\boldsymbol{\Lambda})$. Then we have a normal linear model for $\text{vec}(\boldsymbol{W}')$ such that

$$\text{vec}(\boldsymbol{W}') = (\boldsymbol{E} \otimes \boldsymbol{I}_N)\boldsymbol{\lambda} + \text{vec}(\boldsymbol{U}') \tag{59}$$

$$\text{vec}(\boldsymbol{U}') \sim \text{N}_{N(T-p)}\left(\boldsymbol{0}_{N(T-p)}, \boldsymbol{I}_{T-p} \otimes \boldsymbol{P}^{-1}\right) \tag{60}$$

Let $\boldsymbol{\lambda}_0 := \text{vec}(\boldsymbol{\Lambda}_0)$ and

$$\boldsymbol{U}_0 := \boldsymbol{\Gamma}'\eta_0\boldsymbol{H}(\tau_0)\boldsymbol{\Gamma} \otimes \boldsymbol{G}_0$$

so that

$$\boldsymbol{\lambda}|\boldsymbol{\Gamma}, \boldsymbol{\Phi} \sim \text{N}_{Nr}\left(\boldsymbol{\lambda}_0, \boldsymbol{U}_0^{-1}\right)[(\boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}) \in S] \tag{61}$$

Theorem 4.

$$\boldsymbol{\lambda}|\boldsymbol{\psi}, \boldsymbol{\Phi}, \boldsymbol{P}, \boldsymbol{\Gamma}, \boldsymbol{W}, \boldsymbol{E} \sim \text{N}_{Nr}\left(\boldsymbol{\lambda}_1, \boldsymbol{U}_1^{-1}\right)[(\boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}) \in S] \tag{62}$$

where

$$\boldsymbol{U}_1 = \boldsymbol{E}'\boldsymbol{E} \otimes \boldsymbol{P} + \boldsymbol{U}_0 \tag{63}$$

$$\boldsymbol{\lambda}_1 = \boldsymbol{U}_1^{-1}\,\text{vec}(\boldsymbol{P}\boldsymbol{W}'\boldsymbol{E} + \boldsymbol{G}_0\boldsymbol{\Lambda}_0\boldsymbol{\Gamma}'\eta_0\boldsymbol{H}(\tau_0)\boldsymbol{\Gamma}) \tag{64}$$

Proof. Let $\boldsymbol{\Lambda}_{\text{OLS}} := \boldsymbol{W}'\boldsymbol{E}(\boldsymbol{E}'\boldsymbol{E})^{-1}$ and $\boldsymbol{\lambda}_{\text{OLS}} := \text{vec}(\boldsymbol{\Lambda}_{\text{OLS}})$. By Lemma 1,

$$\begin{aligned}\boldsymbol{U}_1 &:= (\boldsymbol{E} \otimes \boldsymbol{I}_N)'(\boldsymbol{I}_{T-p} \otimes \boldsymbol{P})(\boldsymbol{E} \otimes \boldsymbol{I}_N) + \boldsymbol{U}_0 \\ &= \boldsymbol{E}'\boldsymbol{E} \otimes \boldsymbol{P} + \boldsymbol{U}_0 \\ \boldsymbol{\lambda}_1 &:= \boldsymbol{U}_1^{-1}[(\boldsymbol{E}'\boldsymbol{E} \otimes \boldsymbol{P})\boldsymbol{\lambda}_{\text{OLS}} + \boldsymbol{U}_0\boldsymbol{\lambda}_0]\end{aligned}$$

where

$$
\begin{aligned}
(\boldsymbol{E}'\boldsymbol{E} \otimes \boldsymbol{P})\boldsymbol{\lambda}_{\text{OLS}} &= (\boldsymbol{E}'\boldsymbol{E} \otimes \boldsymbol{P}) \operatorname{vec}\left(\boldsymbol{W}'\boldsymbol{E}(\boldsymbol{E}'\boldsymbol{E})^{-1}\right) \\
&= (\boldsymbol{E}'\boldsymbol{E} \otimes \boldsymbol{P}) \left[(\boldsymbol{E}'\boldsymbol{E})^{-1} \otimes \boldsymbol{W}'\right] \operatorname{vec}(\boldsymbol{E}) \\
&= (\boldsymbol{I}_r \otimes \boldsymbol{P}\boldsymbol{W}') \operatorname{vec}(\boldsymbol{E}) \\
&= \operatorname{vec}(\boldsymbol{P}\boldsymbol{W}'\boldsymbol{E}) \\
\boldsymbol{U}_0\boldsymbol{\lambda}_0 &= (\boldsymbol{\Gamma}'\eta_0\boldsymbol{H}(\tau_0)\boldsymbol{\Gamma} \otimes \boldsymbol{G}_0) \operatorname{vec}(\boldsymbol{\Lambda}_0) \\
&= \operatorname{vec}(\boldsymbol{G}_0\boldsymbol{\Lambda}_0\boldsymbol{\Gamma}'\eta_0\boldsymbol{H}(\tau_0)\boldsymbol{\Gamma})
\end{aligned}
$$

□

A.5 Cointegrating matrix

Write the VECM as for all t,

$$
\boldsymbol{\Phi}(\mathrm{L})(\Delta \boldsymbol{y}_t - \boldsymbol{\mu}) - \boldsymbol{\Lambda}\boldsymbol{\beta} = -\boldsymbol{\Lambda}_*\boldsymbol{\Gamma}_*'[\boldsymbol{y}_{t-1} - \boldsymbol{\mu}(t-1)] + \boldsymbol{u}_t \tag{65}
$$

Let

$$
\boldsymbol{W}_* := \begin{bmatrix} [\boldsymbol{\Phi}(\mathrm{L})(\Delta \boldsymbol{y}_{p+1} - \boldsymbol{\mu}) - \boldsymbol{\Lambda}\boldsymbol{\beta}]' \\ \vdots \\ [\boldsymbol{\Phi}(\mathrm{L})(\Delta \boldsymbol{y}_T - \boldsymbol{\mu}) - \boldsymbol{\Lambda}\boldsymbol{\beta}]' \end{bmatrix}, \quad \boldsymbol{Z}_* := \begin{bmatrix} -(\boldsymbol{y}_p - p\boldsymbol{\mu})' \\ \vdots \\ -[\boldsymbol{y}_{T-1} - (T-1)\boldsymbol{\mu}]' \end{bmatrix}
$$

Then we have a normal linear model for $\boldsymbol{W}_*$ given $\boldsymbol{Z}_*$ such that

$$
\begin{aligned}
\boldsymbol{W}_* &= \boldsymbol{Z}_*\boldsymbol{\Gamma}_*\boldsymbol{\Lambda}_*' + \boldsymbol{U} \\
\boldsymbol{U} &\sim \mathrm{N}_{(T-p)\times N}\left(\mathbf{O}_{(T-p)\times N}; \boldsymbol{I}_{T-p}, \boldsymbol{P}^{-1}\right)
\end{aligned}
$$

Let $\boldsymbol{\gamma}_* := \text{vec}(\boldsymbol{\Gamma}_*)$. Then we have a normal linear model for $\text{vec}(\boldsymbol{W}_*)$ such that

$$\text{vec}(\boldsymbol{W}_*) = (\boldsymbol{\Lambda}_* \otimes \boldsymbol{Z}_*)\boldsymbol{\gamma}_* + \text{vec}(\boldsymbol{U}) \tag{66}$$

$$\text{vec}(\boldsymbol{U}) \sim \text{N}_{N(T-p)}\left(\boldsymbol{0}_{N(T-p)}, \boldsymbol{P}^{-1} \otimes \boldsymbol{I}_{T-p}\right) \tag{67}$$

Let

$$\boldsymbol{V}_0 := \boldsymbol{\Lambda}_*' \boldsymbol{G}_0 \boldsymbol{\Lambda}_* \otimes \eta_0 \boldsymbol{H}(\tau_0)$$

so that

$$\boldsymbol{\gamma}_* | \boldsymbol{\Lambda}_*, \boldsymbol{\Phi} \sim \text{N}_{Nr}\left(\boldsymbol{0}_{Nr}, \boldsymbol{V}_0^{-1}\right) [(\boldsymbol{\Phi}, \boldsymbol{\Lambda}_*, \boldsymbol{\Gamma}_*) \in S] \tag{68}$$

Theorem 5.

$$\boldsymbol{\gamma}_* | \boldsymbol{\psi}, \boldsymbol{\Phi}, \boldsymbol{P}, \boldsymbol{\Lambda}_*, \boldsymbol{W}_*, \boldsymbol{Z}_* \sim \text{N}_{Nr}\left(\boldsymbol{\gamma}_1, \boldsymbol{V}_1^{-1}\right) [(\boldsymbol{\Phi}, \boldsymbol{\Lambda}_*, \boldsymbol{\Gamma}_*) \in S] \tag{69}$$

where

$$\boldsymbol{V}_1 = \boldsymbol{\Lambda}_*' \boldsymbol{P} \boldsymbol{\Lambda}_* \otimes \boldsymbol{Z}_*' \boldsymbol{Z}_* + \boldsymbol{V}_0 \tag{70}$$

$$\boldsymbol{\gamma}_1 = \boldsymbol{V}_1^{-1} \text{vec}(\boldsymbol{Z}_*' \boldsymbol{W}_* \boldsymbol{P} \boldsymbol{\Lambda}_*) \tag{71}$$

Proof. Let $\boldsymbol{\gamma}_{*,\text{GLS}}$ be the GLS estimator of $\boldsymbol{\gamma}_*$, i.e.,

$$\begin{aligned}\boldsymbol{\gamma}_{*,\text{GLS}} &= [(\boldsymbol{\Lambda}_* \otimes \boldsymbol{Z}_*)'(\boldsymbol{P} \otimes \boldsymbol{I}_{T-p})(\boldsymbol{\Lambda}_* \otimes \boldsymbol{Z}_*)]^{-1}(\boldsymbol{\Lambda}_* \otimes \boldsymbol{Z}_*)'(\boldsymbol{P} \otimes \boldsymbol{I}_{T-p}) \text{vec}(\boldsymbol{W}_*) \\ &= (\boldsymbol{\Lambda}_*' \boldsymbol{P}_* \boldsymbol{\Lambda}_* \otimes \boldsymbol{Z}_*' \boldsymbol{Z}_*)^{-1}(\boldsymbol{\Lambda}_*' \boldsymbol{P} \otimes \boldsymbol{Z}_*') \text{vec}(\boldsymbol{W}_*)\end{aligned}$$

By Lemma 1,

$$\begin{aligned}
\boldsymbol{V}_1 &:= (\boldsymbol{\Lambda}_* \otimes \boldsymbol{Z}_*)'(\boldsymbol{P} \otimes \boldsymbol{I}_{T-p})(\boldsymbol{\Lambda}_* \otimes \boldsymbol{Z}_*) + \boldsymbol{V}_0 \\
&= \boldsymbol{\Lambda}_*'\boldsymbol{P}\boldsymbol{\Lambda}_* \otimes \boldsymbol{Z}_*'\boldsymbol{Z}_* + \boldsymbol{V}_0 \\
\boldsymbol{\gamma}_1 &:= \boldsymbol{V}_1^{-1}(\boldsymbol{\Lambda}_*'\boldsymbol{P}\boldsymbol{\Lambda}_* \otimes \boldsymbol{Z}_*'\boldsymbol{Z}_*)\boldsymbol{\gamma}_{*,\mathrm{GLS}} \\
&= \boldsymbol{V}_1^{-1}(\boldsymbol{\Lambda}_*'\boldsymbol{P} \otimes \boldsymbol{Z}_*')\operatorname{vec}(\boldsymbol{W}_*) \\
&= \boldsymbol{V}_1^{-1}\operatorname{vec}(\boldsymbol{Z}_*'\boldsymbol{W}_*\boldsymbol{P}\boldsymbol{\Lambda}_*)
\end{aligned}$$

□

A.6 Tightness hyperparameter

We have

$$\begin{aligned}
&p(\nu|\boldsymbol{\psi},\boldsymbol{\Phi},\boldsymbol{P},\boldsymbol{\Lambda},\boldsymbol{\Gamma},\boldsymbol{Y}) \\
&\propto p(\boldsymbol{Y}|\boldsymbol{\psi},\boldsymbol{\Phi},\boldsymbol{P},\boldsymbol{\Lambda},\boldsymbol{\Gamma},\nu)p(\boldsymbol{\psi}|\boldsymbol{\Phi},\boldsymbol{P},\boldsymbol{\Lambda},\boldsymbol{\Gamma},\nu)p(\boldsymbol{\Lambda},\boldsymbol{\Gamma}|\boldsymbol{\Phi},\boldsymbol{P},\nu)p(\boldsymbol{\Phi},\boldsymbol{P}|\nu)p(\nu) \\
&= p(\boldsymbol{Y}|\boldsymbol{\psi},\boldsymbol{\Phi},\boldsymbol{P},\boldsymbol{\Lambda},\boldsymbol{\Gamma})p(\boldsymbol{\psi}|\boldsymbol{\Phi},\boldsymbol{P},\boldsymbol{\Lambda},\boldsymbol{\Gamma})p(\boldsymbol{\Lambda},\boldsymbol{\Gamma}|\boldsymbol{\Phi},\boldsymbol{P})p(\boldsymbol{\Phi},\boldsymbol{P}|\nu)p(\nu) \\
&\propto p(\boldsymbol{\Phi},\boldsymbol{P}|\nu)p(\nu)
\end{aligned}$$

Thus $p(\nu|\boldsymbol{\psi},\boldsymbol{\Phi},\boldsymbol{P},\boldsymbol{\Lambda},\boldsymbol{\Gamma},\boldsymbol{Y}) = p(\nu|\boldsymbol{\Phi},\boldsymbol{P})$.

Theorem 6.

$$\nu|\boldsymbol{\Phi},\boldsymbol{P} \sim \mathrm{Gam}\left(\frac{A_1}{2},\frac{B_1}{2}\right) \tag{72}$$

where

$$\begin{aligned}
A_1 &:= pN^2 + A_0 \\
B_1 &:= \operatorname{tr}(\boldsymbol{P}(\boldsymbol{\Phi}-\boldsymbol{M}_0)\boldsymbol{D}_0(\boldsymbol{\Phi}-\boldsymbol{M}_0)') + B_0
\end{aligned}$$

Proof. The result follows because

$$p(\boldsymbol{\Phi}, \boldsymbol{P}|\nu)p(\nu) \propto \nu^{pN^2/2} \operatorname{etr}\left(-\frac{\boldsymbol{P}(\boldsymbol{\Phi} - \boldsymbol{M}_0)\nu \boldsymbol{D}_0(\boldsymbol{\Phi} - \boldsymbol{M}_0)'}{2}\right) \frac{\nu^{A_0/2-1}}{\mathrm{e}^{(B_0/2)\nu}}$$

$$= \frac{\nu^{\left(pN^2+A_0\right)/2-1}}{\mathrm{e}^{\{[\operatorname{tr}(\boldsymbol{P}(\boldsymbol{\Phi}-\boldsymbol{M}_0)\boldsymbol{D}_0(\boldsymbol{\Phi}-\boldsymbol{M}_0)')+B_0]/2\}\nu}}$$

□

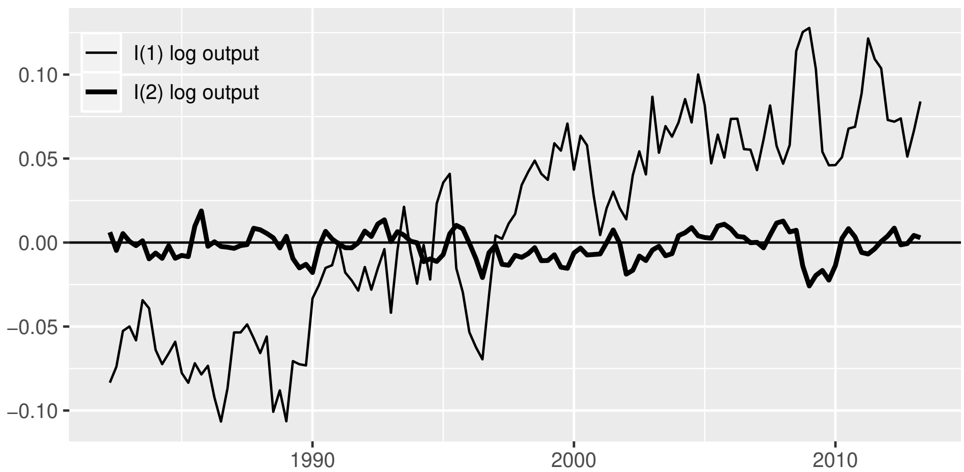

FIGURE 1: Output gap estimates in Japan given by the multivariate B–N decomposition assuming I(1) or I(2) log output. The plots replicate those in Murasawa (2015, Figures 2 and 3).

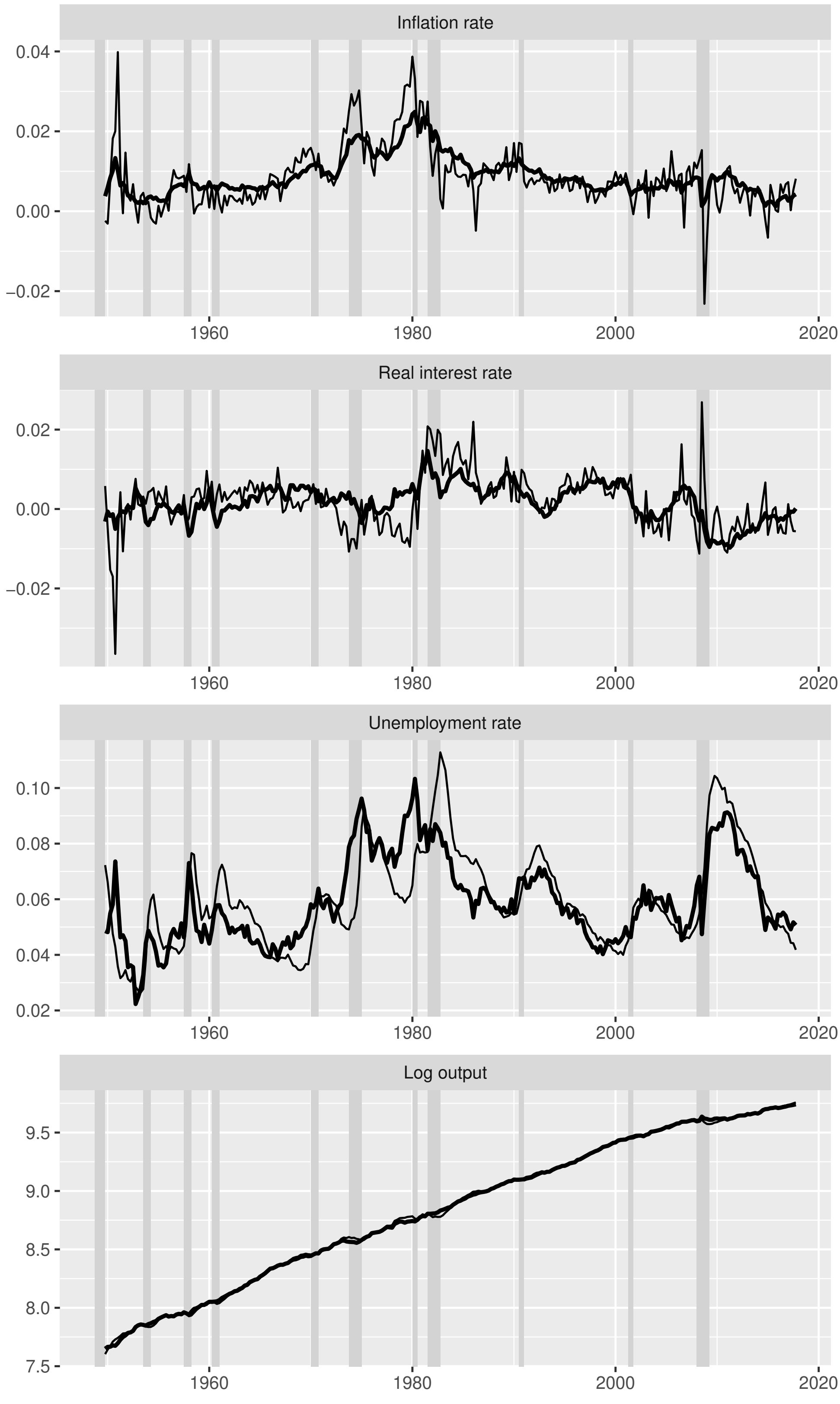

FIGURE 2: Actual rates (thin) and the posterior medians of the natural rates (thick). The shaded areas are the NBER recessions.

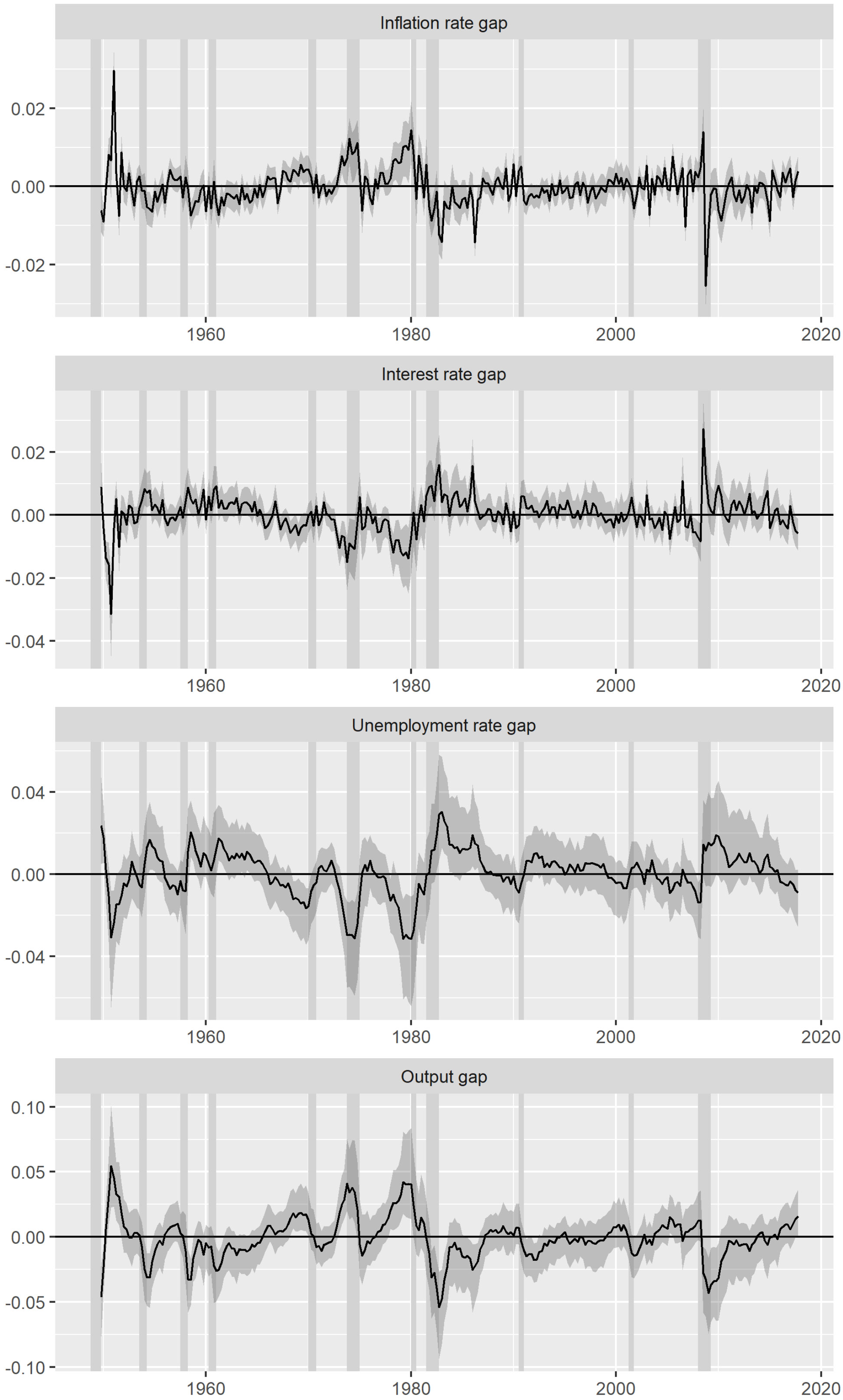

FIGURE 3: Posterior medians of the gaps and their 95% error bands (posterior .025- and .975-quantiles). The shaded areas are the NBER recessions.

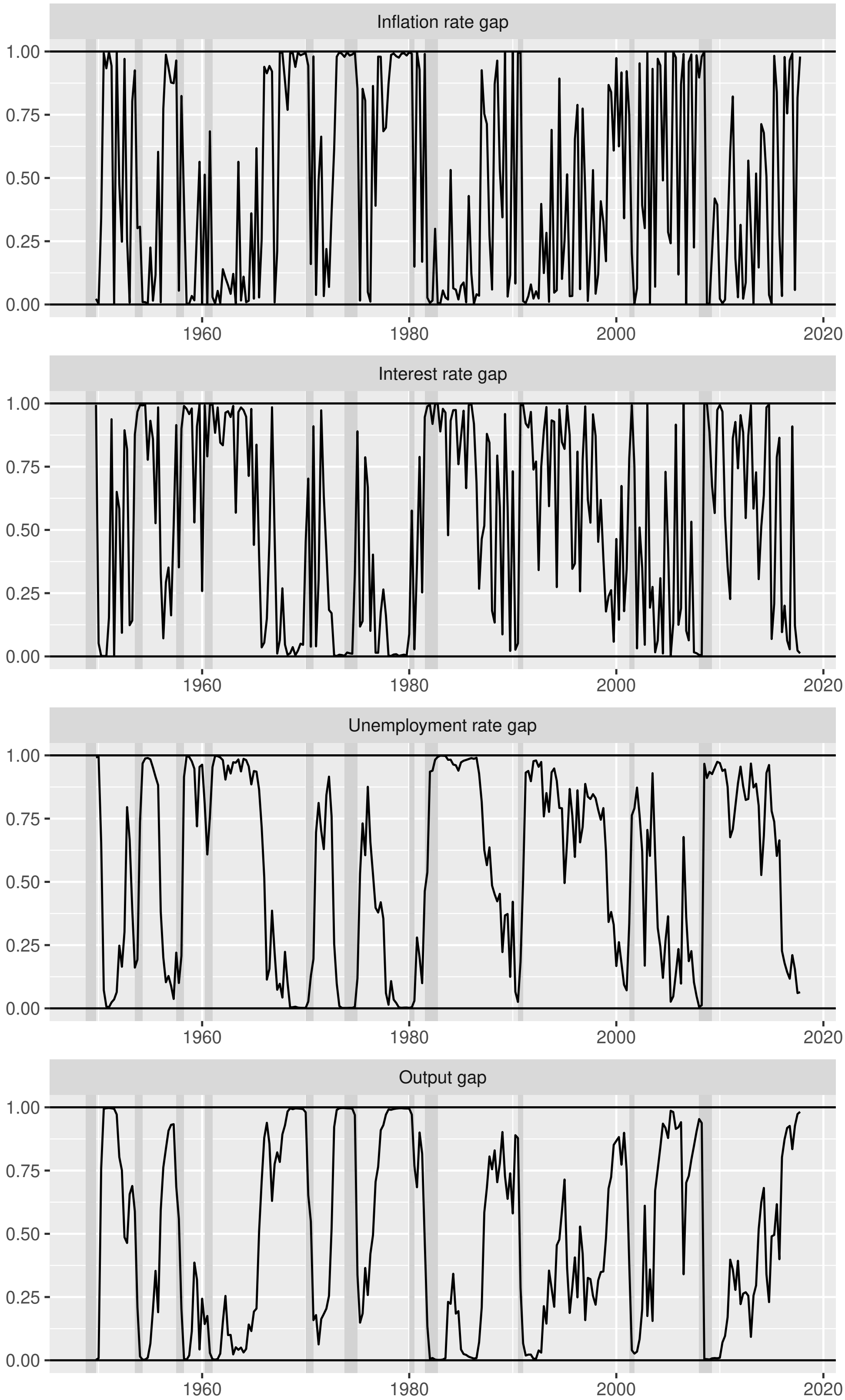

FIGURE 4: Posterior probability of positive gap. The shaded areas are the NBER recessions.

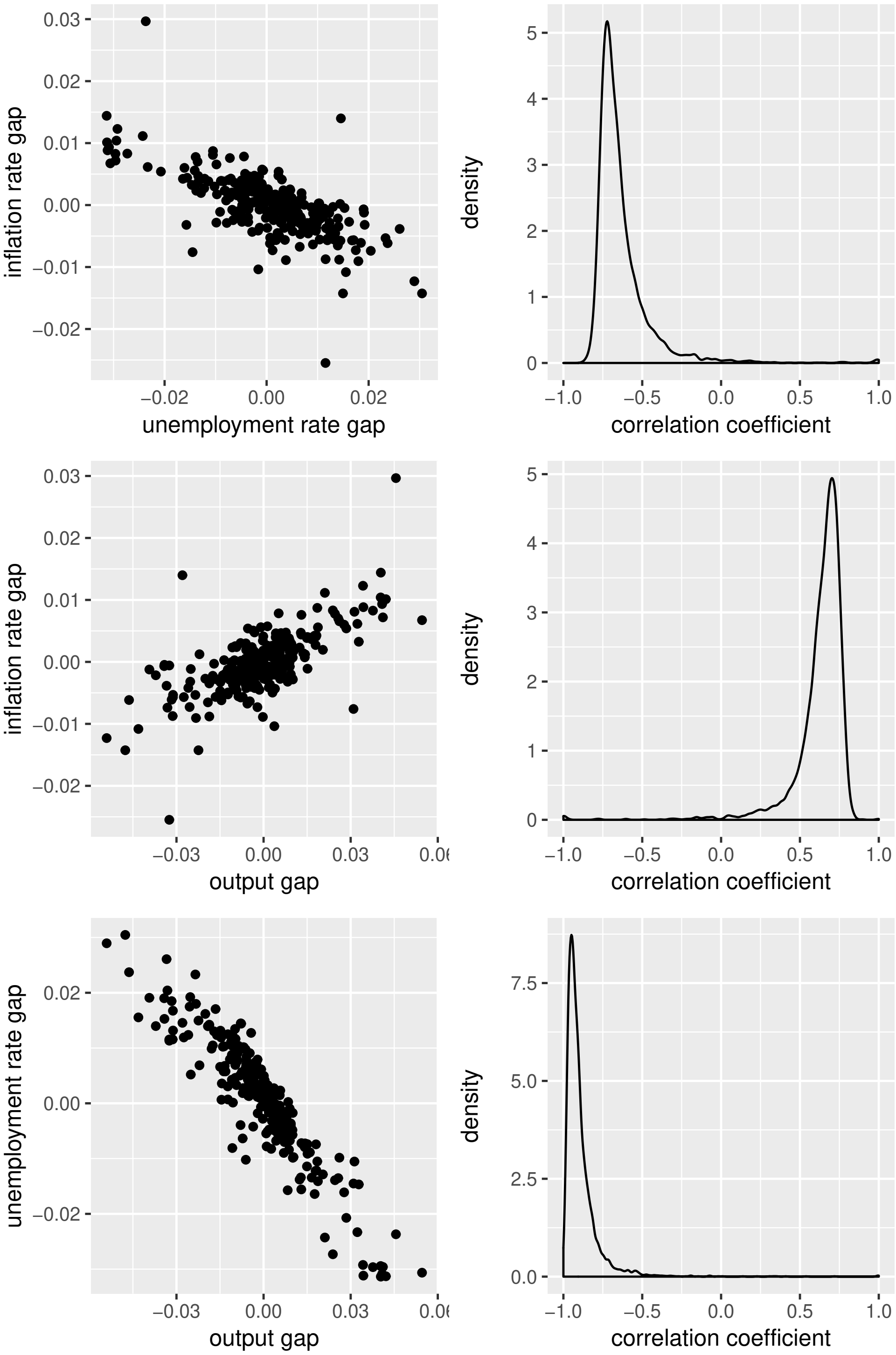

FIGURE 5: Phillips curves and Okun's law.

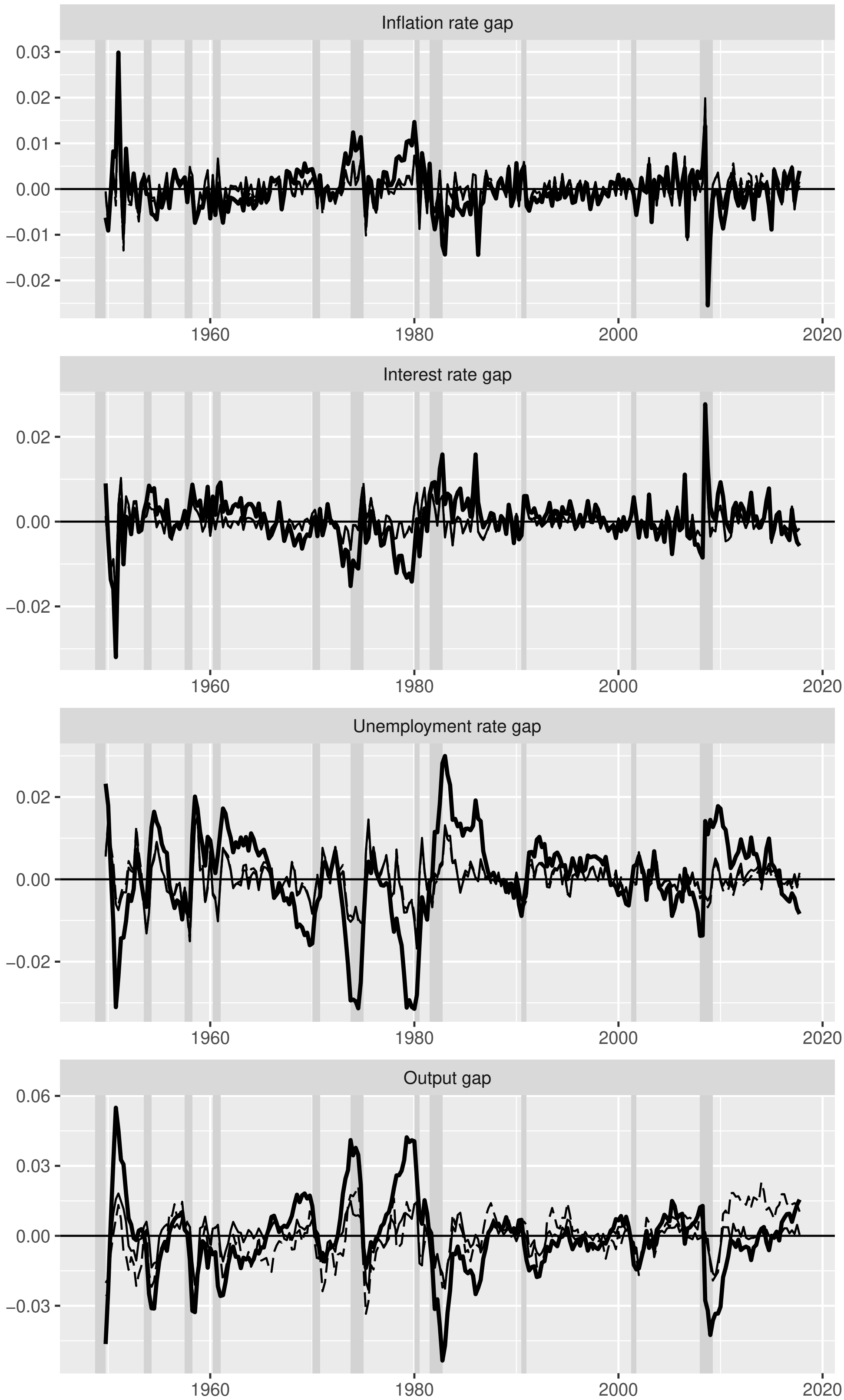

FIGURE 6: Posterior medians of the gaps assuming I(1) log output (dashed), I(2) log output with no cointegration (thin), and I(2) log output with cointegration (thick). The shaded areas are the NBER recessions.