Specification Tests for Time-Varying Parameter Models with Stochastic Volatility

Joshua C.C. Chan[*]
Research School of Economics,
Australian National University

February 2016

Abstract

We propose an easy technique to test for time-variation in coefficients and volatilities. Specifically, by using a noncentered parameterization for state space models, we develop a method to directly calculate the relevant Bayes factor using the Savage-Dickey density ratio—thus avoiding the computation of the marginal likelihood altogether. The proposed methodology is illustrated via two empirical applications. In the first application we test for time-variation in the volatility of inflation in the G7 countries. The second application investigates if there is substantial time-variation in the NAIRU in the US.

Keywords: Bayesian model comparison, state space, inflation uncertainty, NAIRU

JEL classification: C11; C32; E31; E52

1   Introduction

A voluminous literature has demonstrated the importance of allowing for time-varying parameters and volatilities in modeling macroeconomic data. In fact, following the seminal work of Cogley and Sargent (2005) and Primiceri (2005), the time-varying parameter vector autoregression (TVP-VAR) model with stochastic volatility has become a benchmark in the literature (recent papers include Benati, 2008; Koop, Leon-Gonzalez, and Strachan, 2009; Koop and Korobilis, 2013; Liu and Morley, 2014, among many others). In addition, models with time-varying parameters and stochastic volatility are often found to forecast better than their constant-coefficient counterparts; see, e.g., Clark (2011), D'Agostino, Gambetti, and Giannone (2013) and Clark and Ravazzolo (2015).

Despite the empirical success of these flexible time-varying models, an emerging literature has highlighted concerns about their potential over-parameterization. In particular, a variety of time-varying shrinkage methods—such as those in Chan, Koop, Leon-Gonzalez, and Strachan (2012), Nakajima and West (2013), Belmonte, Koop, and Korobilis (2014), Eisenstat, Chan, and Strachan (2015) and Kalli and Griffin (2014)—have been developed to address this over-parameterization concern. One key feature of these methods is that they *a priori* shrink each time-varying parameter to a constant—unless the data prefer time-variation in that coefficient. In a formal model comparison exercise, Chan and Grant (2014) find evidence that data favor a VAR where some coefficients are constant while others are time-varying, relative to both a full-fledged TVP-VAR and a constant-coefficient VAR.

These recent developments highlight the need for specification tests for time-varying parameter models with stochastic volatility. In principle, this can be done by simply computing the marginal likelihoods under a variety of models and selecting the model with the largest marginal likelihood value. In practice, however, it is rarely done in empirical macroeconomics because computing the marginal likelihoods for multivariate models with stochastic volatility is typically difficult.[1]

The main contribution of this paper is to develop an easy method to compare time-varying parameter models with stochastic volatility. Specifically, by using a noncentered parameterization for state space models proposed in Frühwirth-Schnatter and Wagner (2010), we directly compute the Bayes factor—the ratio of the marginal likelihoods of two competing models—using the Savage-Dickey density ratio, thus avoiding the estimation of the marginal likelihood altogether.

The proposed methodology is illustrated via two applications. In the first application

---

[1]Another popular model selection criterion is the deviance information criterion (DIC) introduced by Spiegelhalter, Best, Carlin, and van der Linde (2002). It has been used to compare a wide variety of stochastic volatility models (Berg, Meyer, and Yu, 2004; Abanto-Valle, Bandyopadhyay, Lachos, and Enriquez, 2010; Vo, 2011; Brooks and Prokopczuk, 2013; Wang, Choy, and Chan, 2013). All of these papers use a version of the DIC based on the conditional likelihood—the density of the data given the parameters and the states—that can be readily evaluated. Recently, however, Chan and Grant (2016) provide Monte Carlo evidence that shows this conditional DIC tends to pick overfitted models.

we test if inflation in the G7 countries exhibits time-varying volatility. We find that for all countries there is overwhelming support for stochastic volatility in either the trend inflation or the transitory component. The second application investigates if there is time-variation in the non-accelerating inflation rate of unemployment, or NAIRU, using US inflation and unemployment data. We find strong evidence that the NAIRU is time-varying.

The rest of the paper is organized as follows. Section 2 discusses the noncentered parameterization of Frühwirth-Schnatter and Wagner (2010) and how it can be used to compute the relevant Savage-Dickey density ratio. Sections 3 and 4 illustrate the methodology through two applications: the first tests for time-variation in inflation volatility in the G7 countries; the second investigates time-variation in the NAIRU in the US.

2 Testing for Time-Variation

2.1 Bayes Factor and Savage-Dickey Density Ratio

We illustrate the proposed method using a simple unobserved components model with stochastic volatility. It is straightforward to extend the approach to multivariate settings, as shown in the second empirical application.

Consider the following model:

$$y_t = \tau_t + \mathrm{e}^{\frac{1}{2}h_t}\varepsilon_t^y, \tag{1}$$

$$\tau_t = \tau_{t-1} + \omega_\tau \varepsilon_t^\tau, \tag{2}$$

$$h_t = h_{t-1} + \omega_h \varepsilon_t^h, \tag{3}$$

where ε_t^y, ε_t^τ and ε_t^h are independent $\mathcal{N}(0,1)$. The state equations (2) and (3) are initialized with $\tau_1 \sim \mathcal{N}(\tau_0, \omega_\tau^2 V_\tau)$ and $h_1 \sim \mathcal{N}(h_0, \omega_h^2 V_h)$ respectively. We treat τ_0 and h_0 as parameters to be estimated whereas the variances V_τ and V_h are known constants.[2]

It is often of interest to formally test if the time-variation in the intercept or the volatility is needed. For concreteness, suppose we wish to test if there is indeed time-variation in the variance. It amounts to comparing the model in (1)–(3) to a restricted version where the variance is constant, i.e., $h_0 = h_1 = \cdots = h_T$. Denote the former model as M_1 and the latter as M_2. To formally compare these two models, one popular model comparison

---

[2]We assume the log volatility h_t follows a random walk process in (3), which is a popular specification in empirical macroeconomics (see, e.g., Cogley and Sargent, 2005; Primiceri, 2005; Koop et al., 2009). An alternative specification is a stationary autoregressive model that is commonly used in finance. Using US inflation, Eisenstat and Strachan (2015) compare the two specifications and conclude that the estimates differ very little across the two volatility specifications. Since both our applications involve macroeconomic series (inflation, GDP growth and unemployment), we expect both specifications would give very similar results.

criterion is the *Bayes factor* in favor of M_1 against M_2, defined as

$$\mathrm{BF}_{12} = \frac{p(\mathbf{y} \mid M_1)}{p(\mathbf{y} \mid M_2)},$$

where $p(\mathbf{y} \mid M_i)$ is the *marginal likelihood* for M_i, $i = 1, 2$, which is simply the marginal data density under model M_i evaluated at the observed data $\mathbf{y}$. It is related to the posterior odds ratio as follows:

$$\frac{\mathbb{P}(M_1 \mid \mathbf{y})}{\mathbb{P}(M_2 \mid \mathbf{y})} = \frac{\mathbb{P}(M_1)}{\mathbb{P}(M_2)} \times \mathrm{BF}_{12},$$

where $\mathbb{P}(M_1)/\mathbb{P}(M_2)$ is the prior odds ratio. Thus, under equal prior model probabilities, i.e., $\mathbb{P}(M_1) = \mathbb{P}(M_2)$, the posterior odds ratio in favor of M_1 reduces to the Bayes factor BF_{12}. In that case, if, for example, $\mathrm{BF}_{12} = 10$, then model M_1 is 10 times more likely than model M_2 given the data.

The Bayes factor has a natural interpretation and is commonly used to compare models. However, in our settings where we wish to test for time-variation, one key difficulty is the computation of the marginal likelihood of the model with time-varying parameters. Marginal likelihood estimation has generated a large literature; see, e.g., Friel and Wyse (2012) and Ardia, Baştürk, Hoogerheide, and van Dijk (2012) for a recent review. There are several papers dealing specifically with marginal likelihood estimation for Gaussian and non-Gaussian state space models using importance sampling (Frühwirth-Schnatter, 1995; Chan and Eisenstat, 2015) or auxiliary mixture sampling (Frühwirth-Schnatter and Wagner, 2008). In particular, Chan and Grant (2016) and Kastner (2015) discuss marginal likelihood computation for various univariate stochastic volatility models.

Despite recent advances, marginal likelihood computation for nonlinear state space models with multiple states remains a nontrivial task. The main challenge is to evaluate the integrated likelihood that involves integrating out all the states. For example, to compute the integrated likelihood for the model in (1)–(3), one needs to integrate out both $\boldsymbol{\tau}$ and $\mathbf{h}$. For models in the applications with three or four types of states, the computation of the integrated likelihood is even more difficult.[3] That perhaps explains why despite the popularity of state space models, there are few empirical studies that explicitly test for time-variation.

Fortunately, one much simpler approach is available when one wants to compute the Bayes factor for nested models. Specifically, for nested models, the Bayes factor can be calculated using the Savage-Dickey density ratio (Verdinelli and Wasserman, 1995),

---

[3] One could in principle compute the marginal likelihood using the conditional likelihood instead of the integrated likelihood. For instance, one could estimate the marginal likelihood using the Chib's method (Chib, 1995) based on the conditional likelihood. However, Frühwirth-Schnatter and Wagner (2008) find that "this estimator is extremely inaccurate" and "an upward bias seems to be present". Chan and Grant (2015) reach a similar conclusion when the marginal likelihood is estimated using a modified harmonic mean (Gelfand and Dey, 1994) of the conditional likelihood—they find that this approach does not work well in practice, as the resulting estimates have substantial bias and tend to select the wrong model.

which requires only the estimation of the unrestricted model. More importantly, no explicit computation of the marginal likelihood is needed. This approach has been used to compute the Bayes factor in many empirical applications, including Koop and Potter (1999), Deborah and Strachan (2009), Koop, Leon-Gonzalez, and Strachan (2010) and Chan (2013).

For concreteness, suppose we wish to compare the unobserved components model with stochastic volatility in (1)–(3) (M_1) against a version where the variance is constant over time (M_2). Formally, the latter is a restricted version of the former model with $\omega_h^2 = 0$, then the variance becomes e^{h_0}. Ignoring some technical difficulties (which are discussed below), the Bayes factor in favor of the unrestricted model M_1 can be obtained using the Savage-Dickey density ratio as

$$\mathrm{BF}_{12} = \frac{p(\omega_h^2 = 0)}{p(\omega_h^2 = 0 \,|\, \mathbf{y})},$$

where the numerator is the marginal prior density of ω_h^2 evaluated at 0 and the denominator is the marginal posterior evaluated at 0.[4] Intuitively, if ω_h^2 is more unlikely to be zero under the posterior density relative to the prior density, then it is viewed as evidence in favor of the time-varying model. Hence, to compute the relevant Bayes factor, one only needs to evaluate two univariate densities at a point, which is often easy to do.

Nevertheless, this easier approach cannot be directly applied in our setting due to two related issues. First, the value 0 is at the boundary of parameter space of ω_h^2—therefore the Savage-Dickey density ratio approach is not applicable. Second, one typically assumes an inverse-gamma prior for ω_h^2, which has zero density at zero. Our proposed approach overcomes these two difficulties by using the noncentered parameterization of Frühwirth-Schnatter and Wagner (2010)—instead of the error variance ω_h^2, we work with the standard derivation ω_h that is *defined* to have support on the whole real line. With these modifications, one can use the Savage-Dickey density ratio to compute the relevant Bayes factor. We discuss the details in the next section.

2.2   Noncentered Parameterization

Following Frühwirth-Schnatter and Wagner (2010), we first define $\widetilde{\tau}_t = (\tau_t - \tau_0)/\omega_\tau$ and $\widetilde{h}_t = (h_t - h_0)/\omega_h$. Then the state space model in (1)–(3) can be rewritten as follows:

$$y_t = \tau_0 + \omega_\tau \widetilde{\tau}_t + \mathrm{e}^{\frac{1}{2}(h_0 + \omega_h \widetilde{h}_t)} \varepsilon_t^y, \tag{4}$$

$$\widetilde{\tau}_t = \widetilde{\tau}_{t-1} + \varepsilon_t^\tau, \tag{5}$$

$$\widetilde{h}_t = \widetilde{h}_{t-1} + \varepsilon_t^h, \tag{6}$$

---

[4]For the Savage-Dickey density ratio identity to hold, the priors under the restricted and unrestricted models need to satisfy a certain condition. Throughout we assume the sufficient condition that the restricted and unrestricted parameters are independent *a priori*. In our example it implies $p(\tau_0, h_0, \omega_\tau^2, \omega_h^2) = p(\omega_h^2)p(\tau_0, h_0, \omega_\tau^2)$.

where ε_t^y, ε_t^τ and ε_t^h are independent $\mathcal{N}(0,1)$. The state equations (5) and (6) are initialized with $\widetilde{\tau}_1 \sim \mathcal{N}(0, V_\tau)$ and $\widetilde{h}_1 \sim \mathcal{N}(0, V_h)$ respectively. Frühwirth-Schnatter and Wagner (2010) consider a normal prior centered at 0 for ω_τ, say, $\omega_\tau \sim \mathcal{N}(0, V_{\omega_\tau})$. This normal prior on the standard deviation ω_τ has two main advantages over the usual inverse-gamma prior on the variance ω_τ^2.

First, by a change of variable (see, e.g., Kroese and Chan, 2014, Section 3.5), it can be shown that the implied prior for ω_τ^2 is $\mathcal{G}(\frac{1}{2}, \frac{1}{2V_{\omega_\tau}})$, where $\mathcal{G}(\alpha, \beta)$ denotes the Gamma distribution with mean α/β. Compared to the conventional inverse-gamma prior, this gamma prior has more mass concentrated around small values of ω_τ^2. Hence, this prior provides shrinkage—*a priori* it favors the more parsimonious constant-coefficient model.

The second advantage is that it is a conjugate prior for ω_τ under the noncentered parameterization—it therefore facilitates computation. We note that the sign of ω_τ is not identified—the prior of ω_τ is symmetric around 0 and changing both the signs of ω_τ and $\widetilde{\tau}_t$ does not alter the likelihood value.

Frühwirth-Schnatter and Wagner (2010) only consider a noncentered parameterization for the unobserved components with constant variance and develop a Gibbs sampler to estimate the model. Kastner and Frühwirth-Schnatter (2014) later extend this approach to the case where the stochastic volatility is also in the noncentered parameterization. In addition, they provide an efficient algorithm for drawing the log volatilities and ω_h by adopting the auxiliary mixture sampler in Kim, Shepherd, and Chib (1998). Specifically, letting $y_t^* = \log[(y_t - \tau_0 - \omega_\tau \widetilde{\tau}_t)^2]$, then (4) becomes

$$y_t^* = h_0 + \omega_h \widetilde{h}_t + \varepsilon_t^{y*}, \tag{7}$$

where $\varepsilon_t^{y*} = \log[(\varepsilon_t^y)^2]$. Using a mixture of Gaussian distributions to approximate the distribution of ε_t^{y*}, (6) and (7) define a conditionally linear Gaussian state space model. As such, standard algorithms can be used to sample $\widetilde{\mathbf{h}} = (\widetilde{h}_1, \ldots, \widetilde{h}_T)'$. In addition, given $\widetilde{\mathbf{h}}$ and other model parameters, the full conditional distribution of ω_h is Gaussian. Note that similar to ω_τ, the sign of ω_h is not identified and its density can be bimodal. To fully explore this possibly bimodal density, Frühwirth-Schnatter and Wagner (2010) suggest performing a random sign switch for ω_h and $\widetilde{\mathbf{h}}$: with probability 0.5, the draws of ω_h and $\widetilde{\mathbf{h}}$ remain unchanged; they are substituted by $-\omega_h$ and $-\widetilde{\mathbf{h}}$ with the same probability.

Since the conditional distribution of ω_h is Gaussian, its density $p(\omega_h \,|\, \mathbf{y})$ can be easily estimated using Monte Carlo methods; see Appendix A for details. Hence, we can estimate the Bayes factor in favor of the stochastic volatility model against the constant-variance model using the Savage-Dickey density ratio $p(\omega_h = 0)/p(\omega_h = 0 \,|\, \mathbf{y})$.[5]

---

[5]Another approach to estimate posterior model probabilities without explicit computation of the marginal likelihoods is the stochastic model specification search approach proposed in Frühwirth-Schnatter and Wagner (2010). The key step of this approach is to sample the model indicators. Frühwirth-Schnatter and Wagner (2010) recommend sampling the indicators jointly, and those parameters for which the variable selection is carried out should be integrated out to improve efficiency. This can be done in linear Gaussian state space models, but it is more difficult for nonlinear models, such as the one in (4)–(6).

2.3 Computational Issues

Next, we discuss some computational issues that arise in estimating the Bayes factor using the Savage-Dickey density ratio $p(\omega_\tau = 0)/p(\omega_\tau = 0 \,|\, \mathbf{y})$.[6] If one assumes the Gaussian prior $\omega_\tau \sim \mathcal{N}(0, V_{\omega_\tau})$ as suggested, then $p(\omega_\tau = 0)$ can be evaluated analytically, and it suffices to estimate only $p(\omega_\tau = 0 \,|\, \mathbf{y})$. In Appendix A we consider a Monte Carlo estimator for approximating the latter quantity. Even though that estimator is simulation-consistent, its numerical accuracy with finite simulation size is not guaranteed. In fact, in the context of marginal likelihood estimation, some simulation-consistent estimators are known to perform poorly in practice (see the discussion in footnote 3). Hence, it is crucial to understand under what conditions the proposed method is likely to yield unreliable results.

Since the Monte Carlo estimator depends on the full conditional density of ω_τ evaluated at 0, the estimator is likely to be unstable if the density has little mass at 0—when the restricted model is unlikely compared to the unrestricted model. In those cases, the Savage-Dickey density ratio would tend to be large, but the exact value is unlikely to be accurately estimated. Reporting numerical standard errors would help gauge the accuracy. In addition, we can also corroborate the conclusion by visually inspecting the prior and posterior densities of ω_τ. For example, if the posterior density has little mass around 0 relative to the prior density, this can be viewed as evidence against the restricted model.

In the next two sections we illustrate the proposed methodology through two empirical applications. The first involves testing for time-variation in inflation volatility in the G7 countries. The second investigates if there is substantial time-variation in the NAIRU in the US.

3 Application 1: Time-Varying Volatility in Inflation

In the first application we revisit the unobserved components model considered in Stock and Watson (2007) for modeling inflation. Specifically, they decompose the inflation series into a trend and transitory component, where each component follows an independent stochastic volatility process. A large and growing literature has shown that allowing for stochastic volatility substantially improves forecasts (see, e.g., Clark, 2011; Koop and Korobilis, 2012; Chan, 2013, among many others). Here we evaluate the in-sample fit and formally test if both stochastic volatility processes are needed for trend inflation in the G7 countries.

To that end, we first rewrite the unobserved components model in the noncentered pa-

---

[6]We thank an anonymous reviewer for raising some of the issues discussed in this section.

rameterization as follows:

$$\begin{aligned}
\pi_t &= \tau_t + \mathrm{e}^{\frac{1}{2}(h_0 + \omega_h \widetilde{h}_t)} \varepsilon_t^{\pi}, \\
\tau_t &= \tau_{t-1} + \mathrm{e}^{\frac{1}{2}(g_0 + \omega_g \widetilde{g}_t)} \varepsilon_t^{\tau}, \\
\widetilde{h}_t &= \widetilde{h}_{t-1} + \varepsilon_t^{h}, \\
\widetilde{g}_t &= \widetilde{g}_{t-1} + \varepsilon_t^{g},
\end{aligned} \tag{8}$$

where π_t is the inflation rate, ε_t^{π}, ε_t^{τ}, ε_t^{h} and ε_t^{g} are independent $\mathcal{N}(0,1)$, and the state equations are initialized with $\tau_1 \sim \mathcal{N}(\tau_0, V_\tau \exp(g_0 + \omega_g \widetilde{g}_1))$, $\widetilde{h}_1 \sim \mathcal{N}(0, V_h)$ and $\widetilde{g}_1 \sim \mathcal{N}(0, V_g)$. We set $V_\tau = V_h = V_g = 10$ and $\tau_0 = 0$. The model parameters are h_0, g_0, ω_g and ω_h.

Hence, using this setup, we can turn off the stochastic volatility in the trend and the transitory component by assuming that $\omega_g = 0$ and $\omega_h = 0$ respectively. In their analysis, Stock and Watson (2007) fix the parameters $\omega_h^2 = \omega_g^2 = 0.2$. In a recent study, Moura and Turatti (2014) estimate the two parameters while maintaining the assumption that $\omega_h^2 = \omega_g^2 = \omega^2$. Using inflation in the G7 countries, they find that the estimates of ω^2 are statistically different from the calibrated value of 0.2 for a few countries. Here we estimate both ω_h^2 and ω_g^2 and allow them to be different.

The data are the quarterly consumer price indices (CPI) for the G7 countries from 1955Q1 to 2013Q4 obtained from the OECD statistics database. We transform the indices to annualized growth rates. Due to space constraint, in Figure 1 we only plot the CPI inflation for Germany and the US. It is evident from the figure that compared to the US, inflation in Germany is relatively low and more stable.
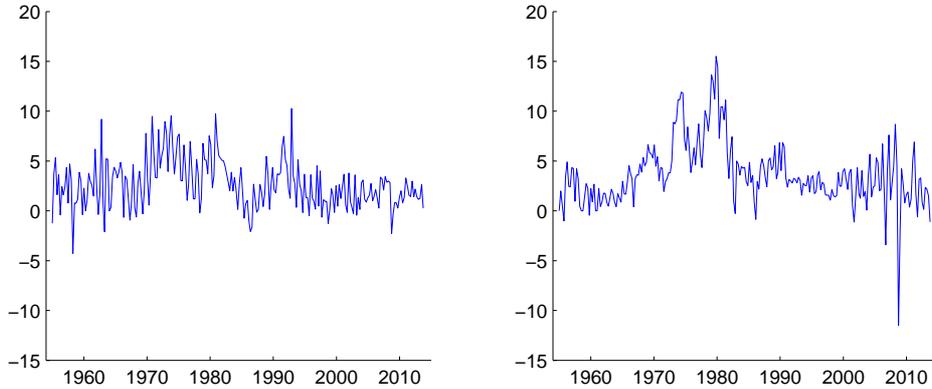
Figure 1: CPI inflation for Germany (left) and the US (right) from 1955Q1 to 2013Q4.

As discussed above, we assume normal priors for ω_g and ω_h: $\omega_g \sim \mathcal{N}(0, V_{\omega_g})$ and $\omega_h \sim \mathcal{N}(0, V_{\omega_h})$. We set $V_{\omega_h} = V_{\omega_g} = 0.2$ so that the implied prior means are $\mathbb{E}\omega_h^2 = \mathbb{E}\omega_g^2 = 0.2$, which are the same as the calibrated value in Stock and Watson (2007). We further assume that $h_0 \sim \mathcal{N}(b_0, V_{h_0})$ and $g_0 \sim \mathcal{N}(c_0, V_{g_0})$ with $b_0 = c_0 = 0$, and $V_{h_0} = V_{g_0} = 10$. Posterior results are based on 100000 draws after a burn-in period of 10000.

8

We report in Figure 2 the estimated time-varying standard deviations of the transitory and trend components for Germany inflation. As the figure shows, the variance of the transitory component has been decreasing since the 1960s, which is only interrupted briefly following the unification of West and East Germany in 1990. On the other hand, the variance of the trend component exhibits much less movement, and has been low and stable throughout the sample period.
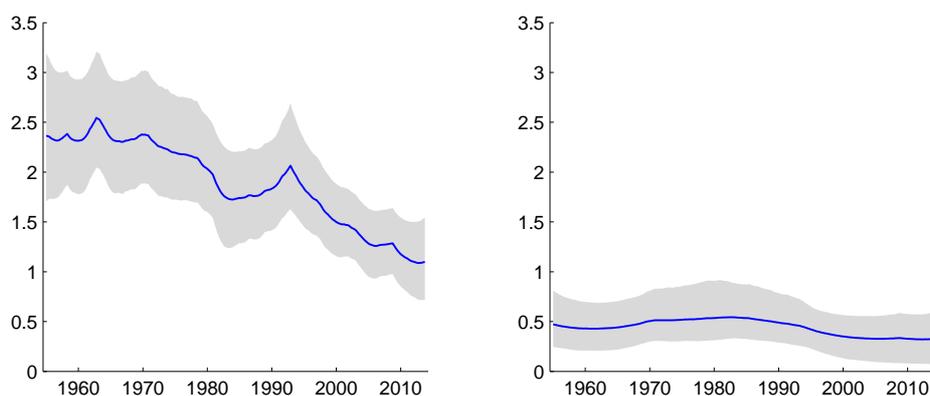
Figure 2: The estimates of $\exp(h_t/2)$ (left panel) and $\exp(g_t/2)$ (right panel) for Germany inflation. The shaded areas represent the 90% credible intervals.

In contrast, the corresponding estimates for US inflation, reported in Figure 3, depict a very different inflation experience. In particular, there is more time-variation in the standard deviations of both the transitory and trend components. For instance, the standard deviation of the trend component doubles in late 1970s compared to a decade earlier. It is also interesting to note that the negative spike in US inflation in late 2008 is accommodated by a sharp rise in the variance of the transitory component, whereas the variance of the trend component remains stable.
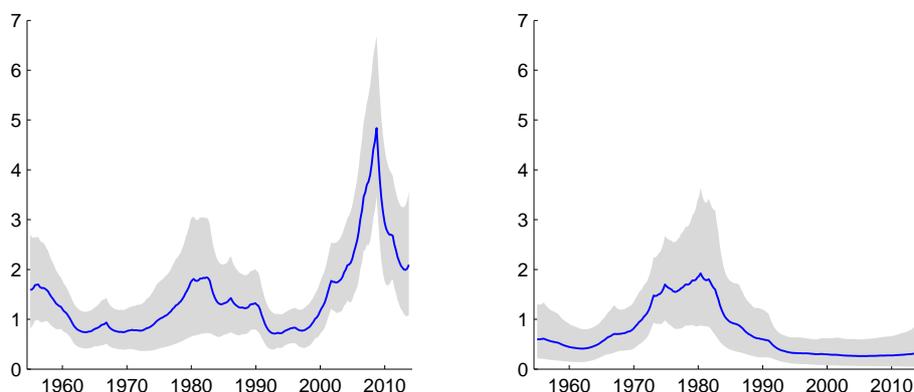
Figure 3: The estimates of $\exp(h_t/2)$ (left panel) and $\exp(g_t/2)$ (right panel) for US inflation. The shaded areas represent the 90% credible intervals.

Posterior means of ω_h^2 and ω_g^2, as well as the associated 90% credible intervals are reported in Table 1. The estimated posterior means of ω_h^2 are all much smaller than 0.2, the calibrated value used in Stock and Watson (2007). In fact, most of the 90% credible intervals exclude 0.2. On the other hand, the estimates of ω_g^2 are larger, and with the exception of Germany, all the 90% credible intervals include 0.2. Moreover, it is more difficult to accurately estimate ω_g^2 compared to ω_h^2. This is not surprising as ω_h^2 controls the stochastic volatility of the inflation π_t, whereas ω_g^2 controls the stochastic volatility of the underlying trend τ_t, which is not observed.

Table 1: Posterior means and 90% credible intervals of ω_h^2 and ω_g^2.

	ω_h^2	ω_g^2
Canada	0.07	0.15
	(0.017, 0.170)	(0.018, 0.405)
France	0.03	0.15
	(0.004, 0.086)	(0.037, 0.362)
Germany	0.02	0.03
	(0.002, 0.046)	(0.001, 0.101)
Italy	0.13	0.15
	(0.020, 0.315)	(0.014, 0.384)
Japan	0.04	0.18
	(0.004, 0.121)	(0.009, 0.515)
UK	0.04	0.13
	(0.007, 0.091)	(0.011, 0.362)
US	0.11	0.11
	(0.033, 0.229)	(0.029, 0.266)

The estimates of ω_h^2 and ω_g^2 are always positive and provide little information on the posterior mass near zero. In fact, some results even look puzzling. For example, the 90% credible intervals of ω_h^2 and ω_g^2 for Germany are respectively $(0.002, 0.046)$ and $(0.001, 0.101)$, even though the estimates of $\exp(h_t/2)$ and $\exp(g_t/2)$ in Figure 2 suggest that there is more time-variation in h_t relative to g_t. To shed some light on this issue, we plot the posterior densities of ω_h and ω_g in Figures 4 and 5.[7]

Since the signs of both ω_h and ω_g are not identified, the posterior densities are symmetric with respect to the origin; see also the discussion in Frühwirth-Schnatter and Wagner (2010). For Germany, the posterior density of ω_h is clearly bimodal. In addition, it has little mass around 0 compared to the prior density. Since the Bayes factor in favor of the unrestricted model against the restricted version where $\omega_h = 0$ is simply the ratio $p(\omega_h = 0)/p(\omega_h = 0 \,|\, \mathbf{y})$, this suggests that the relevant Bayes factor is very large. In contrast, the posterior density of ω_g looks almost unimodal and has more mass around 0 relative to the prior density, corroborating the relatively flat estimates of $\exp(g_t/2)$ reported in Figure 3. Given these results, the reason why the 90% credible interval of ω_g^2

---

[7]We thank an anonymous reviewer for this suggestion.

is wider than that of ω_h^2 seems to be mainly due to the higher posterior uncertainty for ω_g. For US data, the posterior densities of ω_h and ω_g are clearly bimodal and have little mass near 0, suggesting that both stochastic volatility processes are important.
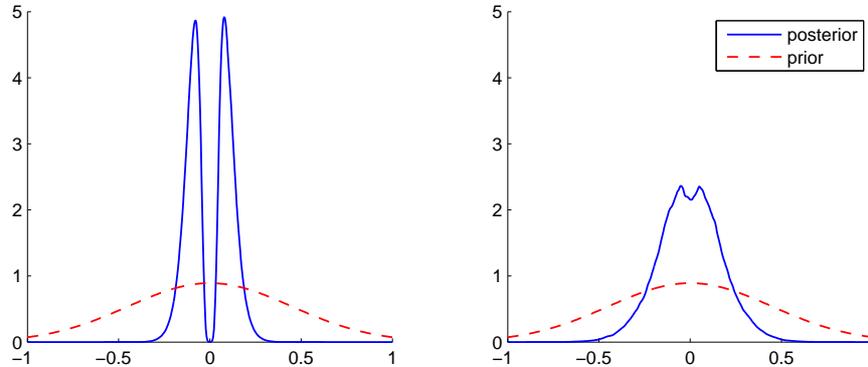
Figure 4: The prior and posterior densities of ω_h (left panel) and ω_g (right panel) for Germany inflation.
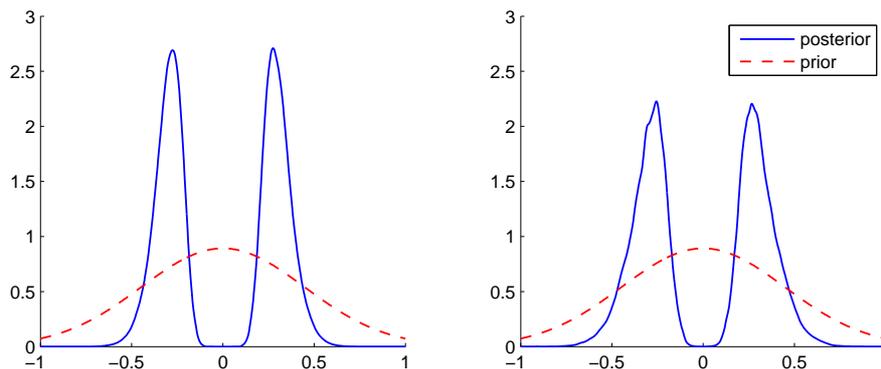
Figure 5: The prior and posterior densities of ω_h (left panel) and ω_g (right panel) for US inflation.

Next, we formally test if both stochastic volatility processes are necessary. To that end, let M_u denote the unrestricted model in (8), and let M_h and M_g represent respectively the restricted versions where $\omega_h = 0$ and $\omega_g = 0$. Moreover, we also consider the restricted model with only homoscedastic innovations, i.e., both $\omega_h = \omega_g = 0$, and we call this version M_{hg}. Note that the Bayes factor in favor of M_u against M_{hg} can be obtained using the Savage-Dickey density ratio $p(\omega_h = \omega_g = 0)/p(\omega_h = \omega_g = 0 \,|\, \mathbf{y})$. Its computation is discussed in Appendix A. The estimated log Bayes factors are reported in Table 2; the numerical standard errors are computed using 10 parallel chains, each of which is of length 100000 with a burn-in of 10000.

11

Table 2: The estimated log Bayes factors and the numerical standard errors.

	Canada	France	Germany	Italy	Japan	UK	US
$\log \text{BF}_{uh}$	5.96	2.90	6.21	2.88	7.89	6.15	67.09
	(0.33)	(0.06)	(0.14)	(0.18)	(0.32)	(0.26)	(10.36)
$\log \text{BF}_{ug}$	3.13	29.47	-0.82	2.60	1.81	2.73	11.35
	(0.04)	(4.45)	(0.02)	(0.10)	(0.10)	(0.13)	(2.28)
$\log \text{BF}_{u,hg}$	20.36	71.16	8.00	234.94	75.69	76.69	195.34
	(2.45)	(7.71)	(0.50)	(7.32)	(3.57)	(2.26)	(4.53)

Notes: BF_{uh} and BF_{ug} are the Bayes factors in favor of having the stochastic volatility process in the transitory and trend components, respectively. $\text{BF}_{u,hg}$ is the Bayes factor in favor of having both stochastic volatility processes against the restricted version without any stochastic volatility.

Recall that a positive log Bayes factor represents evidence in support of the unrestricted model against the restricted version—i.e., the model with some form of stochastic volatility is preferred. Table 2 shows that all the estimated log Bayes factors $\log \text{BF}_{u,hg}$ are large and positive, indicating that the data generally prefer having at least one stochastic volatility component. Also note that when the log Bayes factor estimate is large, the associated numerical standard error also tends to be large—since in those cases it is more difficult to accurately estimate the relevant posterior density at 0 as discussed in Section 2.3.

In general, there is stronger evidence in favor of having stochastic volatility in the transitory component than in the trend component (except for France). For example, for Canada the Bayes factors in favor of the stochastic volatility in the transitory and trend components are respectively $e^{5.96} \approx 388$ and $e^{3.13} \approx 23$. In addition, Germany presents the only example where the stochastic volatility in the trend component is not favored by the data, though the evidence is weak.[8] These results are in line with the conclusions drawn from the densities of ω_h and ω_g given in Figures 4 and 5.

4   Application 2: Time Variation in the NAIRU

Since the seminal work of Milton Friedman, there has been considerable interest in the estimation of the NAIRU, or non-accelerating inflation rate of unemployment. The NAIRU is typically treated as a constant parameter, but a few studies have found evidence that it might be time-varying; see, e.g., Staiger, Stock, and Watson (1997) and Morley, Piger, and Rasche (2013). Here we formally assess if there is substantial time-variation in the NAIRU. Following Chan, Koop, and Potter (2015), consider the following bivariate model

---

[8]It is worth noting that $\log \text{BF}_{u,hg}$ is not simply the sum of $\log \text{BF}_{uh}$ and $\log \text{BF}_{ug}$. For example, for Italy both $\log \text{BF}_{uh}$ and $\log \text{BF}_{ug}$ are less than 3, but $\log \text{BF}_{u,hg}$ is about 235. This can happen if the marginal densities of ω_h and ω_g have nonnegligible mass around 0, but the joint density of ω_h and ω_g has little mass around (0,0)—it is highly unlikely that both ω_h and ω_g take on small values jointly.

for inflation, π_t, and unemployment, u_t:

$$(\pi_t - \tau_t) = \lambda(u_t - \nu_t) + \varepsilon_t^\pi, \tag{9}$$

$$(u_t - \nu_t) = \varepsilon_t^u, \tag{10}$$

where τ_t and ν_t are respectively the underlying inflation trend and the NAIRU. We assume $\varepsilon_t^\pi \sim \mathcal{N}(0, e^{h_t})$ and the underlying inflation trend τ_t follows a random walk process with stochastic volatility described below. The innovation ε_t^u follows an AR(2) process:

$$\varepsilon_t^u = \phi_1 \varepsilon_{t-1}^u + \phi_2 \varepsilon_{t-2}^u + \zeta_t, \tag{11}$$

where $\zeta_t \sim \mathcal{N}(0, \sigma_u^2)$ and $\varepsilon_0^u = \varepsilon_{-1}^u = 0$.

The inflation equation (9) states that $\pi_t - \tau_t$, the deviation of the inflation from its underlying trend, depends on the "gap" between the unemployment rate and the NAIRU—when the unemployment rate is at the NAIRU, the inflation equals its underlying trend plus a transitory shock. The unemployment equation (10) models the deviation of the unemployment rate from the NAIRU as an AR(2) process. In addition, the NAIRU ν_t follows the following random walk process:

$$\nu_t = \nu_{t-1} + \omega_\nu \varepsilon_t^\nu,$$

where $\varepsilon_t^\nu \sim \mathcal{N}(0, 1)$ and the process is initialized with $\nu_1 \sim \mathcal{N}(\nu_0, \omega_\nu^2 V_\nu)$.

Our focus is to assess whether the NAIRU is time-varying. To that end, we rewrite the system (9)–(10) in the noncentered parameterization with $\nu_t = \nu_0 + \omega_\nu \widetilde{\nu}_t$ as follows:

$$(\pi_t - \tau_t) = \lambda(u_t - \nu_0 - \omega_\nu \widetilde{\nu}_t) + \varepsilon_t^\pi, \tag{12}$$

$$(u_t - \nu_0 - \omega_\nu \widetilde{\nu}_t) = \varepsilon_t^u. \tag{13}$$

There are altogether four state equations, which are:

$$\widetilde{\nu}_t = \widetilde{\nu}_{t-1} + \varepsilon_t^\nu \tag{14}$$

$$\tau_t = \tau_{t-1} + e^{\frac{1}{2}(g_0 + \omega_g \widetilde{g}_t)} \varepsilon_t^\tau,$$

$$\widetilde{h}_t = \widetilde{h}_{t-1} + \varepsilon_t^h,$$

$$\widetilde{g}_t = \widetilde{g}_{t-1} + \varepsilon_t^g,$$

where ε_t^ν, ε_t^τ, ε_t^h and ε_t^g are independent $\mathcal{N}(0, 1)$ random variables. The states are initialized by $\widetilde{\nu}_1 \sim \mathcal{N}(0, V_\nu)$, $\tau_1 \sim \mathcal{N}(\tau_0, V_\tau \exp(g_0 + \omega_g \widetilde{g}_1))$, $\widetilde{h}_1 \sim \mathcal{N}(0, V_h)$ and $\widetilde{g}_1 \sim \mathcal{N}(0, V_g)$ with $V_\nu = V_\tau = V_h = V_g = 10$ and $\tau_0 = 0$. Then, the Bayes factor in favor of a time-varying NAIRU against its constant counterpart can be evaluated using the Savage-Dickey density ratio $\text{BF}_{u\nu} = p(\omega_\nu = 0)/p(\omega_\nu = 0 \mid \boldsymbol{\pi}, \mathbf{u})$.

The data consist of the CPI inflation rate and the civilian unemployment rate from 1948Q1 to 2013Q4 obtained from the Federal Reserve Bank of St. Louis economic database. The priors and the estimation details are given in Appendix B. Posterior

analysis is based on 10 parallel chains each of which is of length 100000 after a burn-in period of 10000. The posterior means and the associated 90% credible intervals of the main parameters are reported in Table 3.

Table 3: Posterior means and 90% credible intervals of selected parameters.

	posterior mean	90% credible interval
λ	-0.57	(-0.855, -0.304)
ϕ_1	1.64	(1.557, 1.715)
ϕ_2	-0.70	(-0.779, -0.615)
σ_u^2	0.07	(0.056, 0.089)
ω_h^2	0.27	(0.101, 0.506)
ω_g^2	0.12	(0.022, 0.312)
ω_ν^2	0.01	(0.005, 0.025)

The estimate of λ is -0.57 with a 90% credible interval that excludes 0, indicating a downward sloping Philips curve. In particular, if the unemployment rate is 1% below the NAIRU, the inflation rate would be on average 0.57% above its underlying trend. The deviation of the unemployment rate from the NAIRU seems to be highly persistent—the AR coefficients ϕ_1 and ϕ_2 are estimated to be 1.64 and -0.70, respectively.
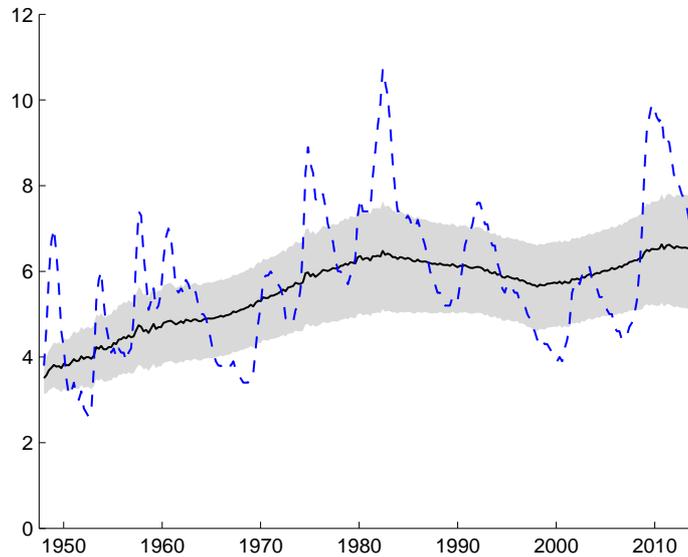
Figure 6: The estimated time-varying NAIRU (solid line) and the unemployment rate (dashed line). The shaded area represents the 90% credible intervals of the NAIRU.

Figure 6 plots the estimated time-varying NAIRU and the associated 90% credible intervals. The figure shows a gradual increase in the NAIRU from about 4% in 1950 to about

6.5% in the early 1980s. Then, it drops steadily throughout the 1980s and 1990s, until it picks up again since the 2000s. Seemingly, the NAIRU has experienced considerable variation in the past six decades.

Next, we report in Figure 7 the estimate of the marginal posterior density $p(\omega_\nu \,|\, \boldsymbol{\pi}, \mathbf{u})$. It is evident from the figure that this posterior density is bimodal and has little mass around 0 compared to the prior density, which can be viewed as evidence in support of a time-varying NAIRU.
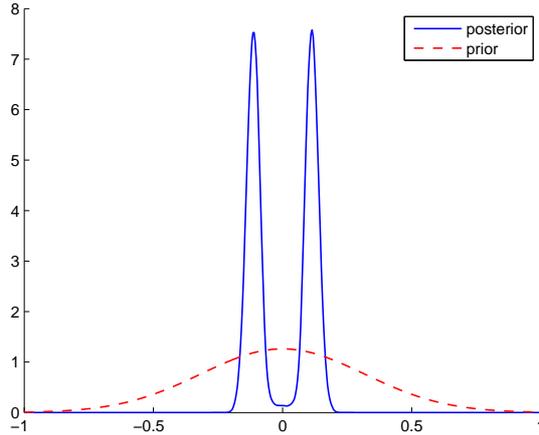
Figure 7: The prior and posterior densities of ω_ν.

To formally test if there is substantial time-variation in the NAIRU, we compute the Bayes factor in favor of the unrestricted model against the restricted version where the NAIRU is a constant ($\omega_\nu = 0$). The log Bayes factor is estimated to be 2.7 with a numerical standard error of 0.18. That is, the Bayes factor in favor of the time-varying NAIRU model is about 15. In other words, if we assume *a priori* that it is equally likely that the NAIRU is time-varying and time-invariant, given the data the former possibility is 15 times more likely than the latter.

5   Concluding Remarks and Future Research

We have developed an easy technique to test for time-variation in coefficients and volatilities. The proposed method was used to address two questions: does inflation display time-varying volatility? Is the NAIRU time-varying or is it constant? This technique is especially useful in the model building stage, where the modeler is faced with many different modeling choices—various specification tests can be done by simply estimating one unrestricted model.

To test for time-variation in multiple coefficients, the proposed approach involves estimat-

ing the quantity of the form $p(\boldsymbol{\omega} = \mathbf{0} \,|\, \mathbf{y})$. In Application 1 we showed an example where $\boldsymbol{\omega} = (\omega_h, \omega_g)'$ is two-dimensional. In cases where the dimension of $\boldsymbol{\omega}$ is high, estimating the quantity $p(\boldsymbol{\omega} = \mathbf{0} \,|\, \mathbf{y})$ accurately presents a challenge. In those cases, one might need to consider alternative methods to compute the relevant Bayes factor.

Appendix A: Evaluating Marginal and Joint Densities

In this appendix we first outline the auxiliary mixture sampler to jointly sample the transformed log volatilities $\widetilde{\mathbf{h}}$. We next describe how one can evaluate the densities $p(\omega_h \,|\, \mathbf{y})$ and $p(\omega_h, \omega_g \,|\, \mathbf{y})$ at zero.

First, it follows from $\varepsilon_t^y \sim \mathcal{N}(0, 1)$ that $\varepsilon_t^{y*} = \log[(\varepsilon_t^y)^2]$ has a log χ_1^2 distribution. Kim et al. (1998) show that the log χ_1^2 distribution can be well approximated using a seven-component Gaussian mixture density with fixed parameters. By introducing the mixture component indicators $s_t \in \{1, \ldots, 7\}$ for $t = 1, \ldots, T$, (7) can be approximated by a conditionally linear Gaussian state space model:

$$\mathbf{y}^* = h_0 \mathbf{1} + \omega_h \widetilde{\mathbf{h}} + \boldsymbol{\varepsilon}^{y*}, \tag{15}$$

where $\mathbf{y}^* = (y_1^*, \ldots, y_T^*)'$, $\boldsymbol{\varepsilon}^{y*} = (\varepsilon_1^{y*}, \ldots, \varepsilon_T^{y*})'$, $\mathbf{1}$ is a $T \times 1$ column of ones, $\boldsymbol{\varepsilon}^{y*} \sim \mathcal{N}(\mathbf{d_s}, \boldsymbol{\Omega_s})$ with $\mathbf{d_s}$ and $\boldsymbol{\Omega_s}$ being constant matrices—depending on $\mathbf{s} = (s_1, \ldots, s_T)'$—obtained from the Gaussian mixture approximation of the log χ_1^2 distribution. We then use the precision sampler of Chan and Jeliazkov (2009) to sample $\widetilde{\mathbf{h}}$.

Next, we outline how one can evaluate the marginal density $p(\omega_h \,|\, \mathbf{y})$ at 0. To that end, rewrite (15) as

$$\mathbf{y}^* = \mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}^{y*},$$

where $\boldsymbol{\gamma} = (h_0, \omega_h)'$ and

$$\mathbf{X} = \begin{pmatrix} 1 & \widetilde{h}_1 \\ \vdots & \vdots \\ 1 & \widetilde{h}_T \end{pmatrix}.$$

Recall that the prior for ω_h is $\omega_h \sim \mathcal{N}(0, V_{\omega_h})$ and we further assume that $h_0 \sim \mathcal{N}(b_0, V_{h_0})$. Let $\boldsymbol{\gamma}_0 = (b_0, 0)'$ and $\mathbf{V}_{\boldsymbol{\gamma}} = \mathrm{diag}(V_h, V_{\omega_h})$. Then by standard linear regression results (see, e.g., Koop, Poirier, and Tobias, 2007, pp. 191-193), we have

$$(\boldsymbol{\gamma} \,|\, \mathbf{y}, \mathbf{s}, \widetilde{\mathbf{h}}) \sim \mathcal{N}(\widehat{\boldsymbol{\gamma}}, \mathbf{D}_{\boldsymbol{\gamma}}),$$

where

$$\mathbf{D}_{\boldsymbol{\gamma}} = \left(\mathbf{V}_{\boldsymbol{\gamma}}^{-1} + \mathbf{X}' \boldsymbol{\Omega_s}^{-1} \mathbf{X} \right)^{-1}, \qquad \widehat{\boldsymbol{\gamma}} = \mathbf{D}_{\boldsymbol{\gamma}} (\mathbf{V}_{\boldsymbol{\gamma}}^{-1} \boldsymbol{\gamma}_0 + \mathbf{X}' \boldsymbol{\Omega_s}^{-1} (\mathbf{y}^* - \mathbf{d_s})).$$

It follows that the marginal distribution of ω_h given $\mathbf{y}, \mathbf{s}$ and $\widetilde{\mathbf{h}}$ is $\mathcal{N}(\widehat{\boldsymbol{\gamma}}_2, \mathbf{D}_{\boldsymbol{\gamma},22})$, where $\widehat{\boldsymbol{\gamma}}_2$ is the second element of $\widehat{\boldsymbol{\gamma}}$ and $\mathbf{D}_{\boldsymbol{\gamma},22}$ is the (2,2) element of $\mathbf{D}_{\boldsymbol{\gamma}}$. Hence, the density $p(\omega_h \,|\, \mathbf{y}, \boldsymbol{\tau}, \mathbf{s}, \widetilde{\mathbf{h}})$ can be evaluated exactly. Finally, $p(\omega_h = 0 \,|\, \mathbf{y})$ can be estimated using the Monte Carlo average:

$$\widehat{p}(\omega_h = 0 \,|\, \mathbf{y}) = \frac{1}{R} \sum_{i=1}^{R} p(\omega_h = 0 \,|\, \mathbf{y}, \boldsymbol{\tau}^{(i)}, \mathbf{s}^{(i)}, \widetilde{\mathbf{h}}^{(i)}),$$

where $(\boldsymbol{\tau}^{(1)}, \mathbf{s}^{(1)}, \widetilde{\mathbf{h}}^{(1)}), \ldots, (\boldsymbol{\tau}^{(R)}, \mathbf{s}^{(R)}, \widetilde{\mathbf{h}}^{(R)})$ are posterior draws.

Next, we describe how one can evaluate $p(\omega_h, \omega_g \,|\, \mathbf{y})$ at $(0,0)$. Let $\mathbf{s}$ denote the vector of label indicators for $\widetilde{\mathbf{h}}$ as before. Similarly, let $\mathbf{r}$ denote the label indicators associated with $\widetilde{\mathbf{g}}$. Note that ω_h and ω_g are conditional independent given the parameters and states, i.e.,

$$p(\omega_h, \omega_g \,|\, \mathbf{y}, \boldsymbol{\tau}, \mathbf{s}, \widetilde{\mathbf{h}}, \mathbf{r}, \widetilde{\mathbf{g}}) = p(\omega_h \,|\, \mathbf{y}, \boldsymbol{\tau}, \mathbf{s}, \widetilde{\mathbf{h}}) p(\omega_g \,|\, \mathbf{y}, \boldsymbol{\tau}, \mathbf{r}, \widetilde{\mathbf{g}}).$$

In addition, both densities on the right-hand side are univariate Gaussian, and they can be evaluated at 0 easily. Then, we can estimate $p(\omega_h = 0, \omega_g = 0 \,|\, \mathbf{y})$ using the Monte Carlo average:

$$\widehat{p}(\omega_h = 0, \omega_g = 0 \,|\, \mathbf{y}) = \frac{1}{R} \sum_{i=1}^{R} p(\omega_h = 0 \,|\, \mathbf{y}, \boldsymbol{\tau}^{(i)}, \mathbf{s}^{(i)}, \widetilde{\mathbf{h}}^{(i)}) p(\omega_g = 0 \,|\, \mathbf{y}, \boldsymbol{\tau}^{(i)}, \mathbf{r}^{(i)}, \widetilde{\mathbf{g}}^{(i)}),$$

where $(\boldsymbol{\tau}^{(1)}, \mathbf{s}^{(1)}, \widetilde{\mathbf{h}}^{(1)}, \mathbf{r}^{(1)}, \widetilde{\mathbf{g}}^{(1)}), \dots, (\boldsymbol{\tau}^{(R)}, \mathbf{s}^{(R)}, \widetilde{\mathbf{h}}^{(R)}, \mathbf{r}^{(R)}, \widetilde{\mathbf{g}}^{(R)})$ are posterior draws.

Appendix B: Estimation Details

In this appendix we provide the details of the priors and estimation for the bivariate unobserved components model with a time-varying NAIRU in Section 4. For notational convenience, stack $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_T)'$ and $\mathbf{u} = (u_1, \ldots, u_T)'$, and similarly define $\boldsymbol{\tau}$, $\widetilde{\mathbf{h}}$, $\widetilde{\mathbf{g}}$, $\boldsymbol{\phi}$, $\widetilde{\boldsymbol{\nu}}$, $\boldsymbol{\varepsilon}^\pi$, $\boldsymbol{\varepsilon}^u$, $\boldsymbol{\varepsilon}^\nu$ and $\boldsymbol{\zeta}$.

Priors

We assume the following independent priors:

$$\nu_0 \sim \mathcal{N}(a_0, V_{\nu_0}), \qquad h_0 \sim \mathcal{N}(b_0, V_{h_0}), \qquad g_0 \sim \mathcal{N}(c_0, V_{g_0}),$$
$$\omega_\nu \sim \mathcal{N}(0, V_{\omega_\nu}), \qquad \omega_h \sim \mathcal{N}(0, V_{\omega_h}), \qquad \omega_g \sim \mathcal{N}(0, V_{\omega_g}),$$
$$\boldsymbol{\phi} \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_\phi), \qquad \lambda \sim \mathcal{N}(0, V_\lambda), \qquad \sigma_u^2 \sim \mathcal{IG}(\nu_u, S_u).$$

We set $a_0 = b_0 = c_0 = 0$ and $V_{\nu_0} = V_{h_0} = V_{g_0} = 10$ so that the initial states are all centered at 0 with relatively large variances. Moreover, we assume $V_{\omega_\nu} = 0.1$ and $V_{\omega_h} = V_{\omega_g} = 0.2$, so that the implied prior means are $\mathbb{E}\omega_\nu^2 = 0.1$ and $\mathbb{E}\omega_h^2 = \mathbb{E}\omega_g^2 = 0.2$. We assume relatively large variances for $\boldsymbol{\phi}$ and λ: $\mathbf{V}_\phi = \mathbf{I}_2$, $V_\lambda = 10$. Finally, we set $\nu_u = 5$ and $S_u = 0.1$, which imply $\mathbb{E}\sigma_u^2 = 0.025$.

Likelihood

To derive the (conditional) likelihood, we first rewrite (12) and (13) as

$$\boldsymbol{\pi} = \boldsymbol{\tau} + \lambda(\mathbf{u} - \nu_0 \mathbf{1}) - \lambda \omega_\nu \widetilde{\boldsymbol{\nu}} + \boldsymbol{\varepsilon}^\pi \tag{16}$$
$$\mathbf{u} = \nu_0 \mathbf{1} + \omega_\nu \widetilde{\boldsymbol{\nu}} + \boldsymbol{\varepsilon}^u, \tag{17}$$

where $\boldsymbol{\varepsilon}^\pi \sim \mathcal{N}(\mathbf{0}, \mathbf{S}_\pi)$ and $\mathbf{S}_\pi = \text{diag}(\mathrm{e}^{h_0 + \omega_h \widetilde{h}_1}, \ldots, \mathrm{e}^{h_0 + \omega_h \widetilde{h}_T})$.

The first measurement equation (16) implies that the log density of $\boldsymbol{\pi}$ given $\mathbf{u}$, the states and other parameters is given by

$$\log p(\boldsymbol{\pi} \mid \mathbf{u}, \boldsymbol{\tau}, \widetilde{\boldsymbol{\nu}}, \widetilde{\mathbf{h}}, \lambda, \nu_0, \omega_\nu) \propto -\frac{1}{2} \mathbf{1}'(h_0 \mathbf{1} + \omega_h \widetilde{\mathbf{h}})$$
$$-\frac{1}{2}(\boldsymbol{\pi} - \boldsymbol{\tau} - \lambda(\mathbf{u} - \nu_0 \mathbf{1}) + \lambda \omega_\nu \widetilde{\boldsymbol{\nu}})' \mathbf{S}_\pi^{-1}(\boldsymbol{\pi} - \boldsymbol{\tau} - \lambda(\mathbf{u} - \nu_0 \mathbf{1}) + \lambda \omega_\nu \widetilde{\boldsymbol{\nu}}). \tag{18}$$

Next, we derive the joint density of $\mathbf{u}$. To that end, let

$$\mathbf{H}_\phi = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ -\phi_1 & 1 & 0 & 0 & \cdots & 0 \\ -\phi_2 & -\phi_1 & 1 & 0 & \cdots & 0 \\ 0 & -\phi_2 & -\phi_1 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -\phi_2 & -\phi_1 & 1 \end{pmatrix}.$$

Note that $\mathbf{H}_\phi$ is a $T \times T$ band matrix with unit determinant, and therefore it is invertible. Then, we rewrite the AR(2) process in (11) as

$$\mathbf{H}_\phi \boldsymbol{\varepsilon}^u = \boldsymbol{\zeta},$$

where $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \sigma_u^2 \mathbf{I}_T)$. Hence, we have $\boldsymbol{\varepsilon}^u \sim \mathcal{N}(\mathbf{0}, \sigma_u^2 (\mathbf{H}_\phi' \mathbf{H}_\phi)^{-1})$. It follows that the second measurement equation (17) implies

$$\log p(\mathbf{u} \,|\, \widetilde{\boldsymbol{\nu}}, \sigma_u^2, \nu_0, \omega_\nu, \boldsymbol{\phi}, \sigma_u^2) \propto -\frac{T}{2} \log \sigma_u^2 - \frac{1}{2\sigma_u^2} (\mathbf{u} - \nu_0 \mathbf{1} - \omega_\nu \widetilde{\boldsymbol{\nu}})' \mathbf{H}_\phi' \mathbf{H}_\phi (\mathbf{u} - \nu_0 \mathbf{1} - \omega_\nu \widetilde{\boldsymbol{\nu}}). \quad (19)$$

Posterior Sampler

Given the priors and the (conditional) likelihood derived in the previous section, posterior draws can be obtained by sequentially sampling from:

1. $p(\widetilde{\boldsymbol{\nu}} \,|\, \boldsymbol{\pi}, \mathbf{u}, \boldsymbol{\tau}, \widetilde{\mathbf{h}}, \lambda, \nu_0, \omega_\nu, \boldsymbol{\phi}, \sigma_u^2)$;

2. $p(\boldsymbol{\tau} \,|\, \boldsymbol{\pi}, \mathbf{u}, \widetilde{\boldsymbol{\nu}}, \widetilde{\mathbf{h}}, \widetilde{\mathbf{g}}, \lambda, \nu_0, \omega_\nu, \omega_g, g_0, \boldsymbol{\phi})$;

3. $p(\widetilde{\mathbf{h}}, \widetilde{\mathbf{g}} \,|\, \boldsymbol{\pi}, \mathbf{u}, \boldsymbol{\tau}, \widetilde{\boldsymbol{\nu}}, \lambda, g_0, h_0, \nu_0, \omega_g, \omega_h, \omega_\nu) = p(\widetilde{\mathbf{h}} \,|\, \boldsymbol{\pi}, \mathbf{u}, \boldsymbol{\tau}, \widetilde{\boldsymbol{\nu}}, \lambda, h_0, \nu_0, \omega_h, \omega_\nu) p(\widetilde{\mathbf{g}} \,|\, \boldsymbol{\tau}, g_0, \omega_g)$;

4. $p(h_0, g_0, \nu_0, \omega_g, \omega_h, \omega_\nu \,|\, \boldsymbol{\pi}, \mathbf{u}, \boldsymbol{\tau}, \widetilde{\boldsymbol{\nu}}, \widetilde{\mathbf{h}}, \widetilde{\mathbf{g}}, \lambda, \boldsymbol{\phi}, \sigma_u^2) = p(\nu_0, \omega_\nu \,|\, \boldsymbol{\pi}, \mathbf{u}, \boldsymbol{\tau}, \widetilde{\boldsymbol{\nu}}, \widetilde{\mathbf{h}}, \lambda, \boldsymbol{\phi}, \sigma_u^2)$
   $\times\, p(h_0, \omega_h \,|\, \nu_0, \omega_\nu, \boldsymbol{\pi}, \mathbf{u}, \boldsymbol{\tau}, \widetilde{\boldsymbol{\nu}}, \widetilde{\mathbf{h}}, \lambda) p(g_0, \omega_g \,|\, \boldsymbol{\tau}, \widetilde{\mathbf{g}})$;

5. $p(\boldsymbol{\phi} \,|\, \mathbf{u}, \widetilde{\boldsymbol{\nu}}, \nu_0, \omega_\nu, \sigma_u^2)$;

6. $p(\lambda \,|\, \boldsymbol{\pi}, \mathbf{u}, \boldsymbol{\tau}, \widetilde{\boldsymbol{\nu}}, \widetilde{\mathbf{h}}, \nu_0, \omega_\nu)$;

7. $p(\sigma_u^2 \,|\, \mathbf{u}, \widetilde{\boldsymbol{\nu}}, \nu_0, \omega_\nu, \boldsymbol{\phi})$.

To implement Step 1, note that

$$p(\widetilde{\boldsymbol{\nu}} \,|\, \boldsymbol{\pi}, \mathbf{u}, \boldsymbol{\tau}, \widetilde{\mathbf{h}}, \lambda, \nu_0, \omega_\nu, \boldsymbol{\phi}, \sigma_u^2) \propto p(\boldsymbol{\pi} \,|\, \mathbf{u}, \boldsymbol{\tau}, \widetilde{\boldsymbol{\nu}}, \widetilde{\mathbf{h}}, \lambda, \nu_0, \omega_\nu) p(\mathbf{u} \,|\, \widetilde{\boldsymbol{\nu}}, \sigma_u^2, \nu_0, \omega_\nu, \boldsymbol{\phi}, \sigma_u^2) p(\widetilde{\boldsymbol{\nu}}).$$

The first two densities on the right-hand side are given in (18) and (19), respectively. Next, we derive the prior density $p(\widetilde{\boldsymbol{\nu}})$. To that end, we rewrite (14) in matrix form:

$$\mathbf{H} \widetilde{\boldsymbol{\nu}} = \boldsymbol{\varepsilon}^\nu,$$

where $\boldsymbol{\varepsilon}^\nu \sim \mathcal{N}(\mathbf{0}, \mathbf{S}_\nu)$, $\mathbf{S}_\nu = \mathrm{diag}(V_\nu, 1, \ldots, 1)$ and

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}$$

Hence, the prior density of $\widetilde{\boldsymbol{\nu}}$ is given by

$$\log p(\widetilde{\boldsymbol{\nu}}) \propto -\frac{1}{2}\widetilde{\boldsymbol{\nu}}'\mathbf{H}'\mathbf{S}_\nu^{-1}\mathbf{H}\widetilde{\boldsymbol{\nu}}. \tag{20}$$

Hence, it follows from (18), (19) and (20) that

$$
\begin{aligned}
\log p(\widetilde{\boldsymbol{\nu}} \,|\, \boldsymbol{\pi}, \mathbf{u}, \boldsymbol{\tau}, \widetilde{\mathbf{h}}, \lambda, \nu_0, \omega_\nu, \boldsymbol{\phi}, \sigma_u^2) \\
\propto{}& \log p(\boldsymbol{\pi} \,|\, \mathbf{u}, \boldsymbol{\tau}, \widetilde{\boldsymbol{\nu}}, \widetilde{\mathbf{h}}, \lambda, \nu_0, \omega_\nu) + \log p(\mathbf{u} \,|\, \widetilde{\boldsymbol{\nu}}, \sigma_u^2, \nu_0, \omega_\nu, \boldsymbol{\phi}, \sigma_u^2) + \log p(\widetilde{\boldsymbol{\nu}}) \\
\propto{}& -\frac{1}{2}\left(\lambda^2\omega_\nu^2\widetilde{\boldsymbol{\nu}}'\mathbf{S}_\pi^{-1}\widetilde{\boldsymbol{\nu}} + 2\lambda\omega_\nu\widetilde{\boldsymbol{\nu}}'\mathbf{S}_\pi^{-1}(\boldsymbol{\pi} - \boldsymbol{\tau} - \lambda(\mathbf{u} - \nu_0\mathbf{1}))\right) \\
&-\frac{1}{2\sigma_u^2}(\omega_\nu^2\widetilde{\boldsymbol{\nu}}'\mathbf{H}_\phi'\mathbf{H}_\phi\widetilde{\boldsymbol{\nu}} - 2\omega_\nu\widetilde{\boldsymbol{\nu}}'\mathbf{H}_\phi'\mathbf{H}_\phi(\mathbf{u} - \nu_0\mathbf{1})) - \frac{1}{2}\widetilde{\boldsymbol{\nu}}'\mathbf{H}'\mathbf{S}_\nu^{-1}\mathbf{H}\widetilde{\boldsymbol{\nu}} \\
\propto{}& -\frac{1}{2}(\widetilde{\boldsymbol{\nu}} - \widehat{\boldsymbol{\nu}})'\mathbf{K}_{\widetilde{\nu}}(\widetilde{\boldsymbol{\nu}} - \widehat{\boldsymbol{\nu}}),
\end{aligned}
$$

where

$$\mathbf{K}_{\widetilde{\nu}} = \lambda^2\omega_\nu^2\mathbf{S}_\pi^{-1} + \frac{\omega_\nu^2}{\sigma_u^2}\mathbf{H}_\phi'\mathbf{H}_\phi + \mathbf{H}'\mathbf{S}_\nu^{-1}\mathbf{H},$$

$$\widehat{\boldsymbol{\nu}} = \mathbf{K}_{\widetilde{\nu}}^{-1}\left(\frac{\omega_\nu}{\sigma_u^2}\mathbf{H}_\phi'\mathbf{H}_\phi(\mathbf{u} - \nu_0\mathbf{1}) - \lambda\omega_\nu\mathbf{S}_\pi^{-1}(\boldsymbol{\pi} - \boldsymbol{\tau} - \lambda(\mathbf{u} - \nu_0\mathbf{1}))\right).$$

In other words, $(\widetilde{\boldsymbol{\nu}} \,|\, \boldsymbol{\pi}, \mathbf{u}, \boldsymbol{\tau}, \widetilde{\mathbf{h}}, \lambda, \nu_0, \omega_\nu, \boldsymbol{\phi}, \sigma_u^2) \sim \mathcal{N}(\widehat{\boldsymbol{\nu}}, \mathbf{K}_{\widetilde{\nu}}^{-1})$. Since $\mathbf{S}_\pi$, $\mathbf{S}_\nu$, $\mathbf{H}_\phi$ and $\mathbf{H}$ are all band matrices, so is the precision matrix $\mathbf{K}_{\widetilde{\nu}}$. Hence, the precision sampler of Chan and Jeliazkov (2009) can be used to sample $\widetilde{\boldsymbol{\nu}}$ efficiently.

To implement Step 2, note that

$$p(\boldsymbol{\tau} \,|\, \boldsymbol{\pi}, \mathbf{u}, \widetilde{\boldsymbol{\nu}}, \widetilde{\mathbf{h}}, \widetilde{\mathbf{g}}, \lambda, \nu_0, \omega_\nu, \omega_g, g_0, \boldsymbol{\phi}) \propto p(\boldsymbol{\pi} \,|\, \mathbf{u}, \boldsymbol{\tau}, \widetilde{\boldsymbol{\nu}}, \widetilde{\mathbf{h}}, \lambda, \nu_0, \omega_\nu)p(\boldsymbol{\tau} \,|\, \widetilde{\mathbf{g}}),$$

where the first density on the right-hand side is given in (18). The state equation for $\boldsymbol{\tau}$ can be written as

$$\mathbf{H}\boldsymbol{\tau} = \widetilde{\boldsymbol{\alpha}} + \boldsymbol{\varepsilon}^\tau,$$

where $\widetilde{\boldsymbol{\alpha}} = (\tau_0, 0, \ldots, 0)'$, $\boldsymbol{\varepsilon}^\tau \sim \mathcal{N}(\mathbf{0}, \mathbf{S}_\tau)$ with $\mathbf{S}_\tau = \operatorname{diag}(V_\tau \mathrm{e}^{g_0 + \omega_g\widetilde{g}_1}, \mathrm{e}^{g_0 + \omega_g\widetilde{g}_2}, \ldots, \mathrm{e}^{g_0 + \omega_g\widetilde{g}_T})$. Using a similar argument as before, it can be shown that the conditional distribution of $\boldsymbol{\tau}$ is $\mathcal{N}(\widehat{\boldsymbol{\tau}}, \mathbf{K}_\tau^{-1})$, where

$$\mathbf{K}_\tau = \mathbf{S}_\pi^{-1} + \mathbf{H}'\mathbf{S}_\tau^{-1}\mathbf{H}, \quad \widehat{\boldsymbol{\tau}} = \mathbf{K}_\tau^{-1}\left(\mathbf{H}'\mathbf{S}_\tau^{-1}\widetilde{\boldsymbol{\alpha}} + \mathbf{S}_\pi^{-1}(\boldsymbol{\pi} - \lambda(\mathbf{u} - \nu_0\mathbf{1} - \omega_\nu\widetilde{\boldsymbol{\nu}}))\right).$$

Once again, since the precision matrix $\mathbf{K}_\tau$ is a band matrix, the precision sampler of Chan and Jeliazkov (2009) can be applied to sample $\boldsymbol{\tau}$ efficiently.

Steps 3 and 4 can be carried out as described in Appendix A. The remaining steps are standard and involve drawing from standard distributions.

References

C. A. Abanto-Valle, D. Bandyopadhyay, V. H. Lachos, and I. Enriquez. Robust Bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions. *Computational Statistics and Data Analysis*, 54(12):2883–2898, 2010.

D. Ardia, N. Baştürk, L. Hoogerheide, and H. K. van Dijk. A comparative study of Monte Carlo methods for efficient evaluation of marginal likelihood. *Computational Statistics and Data Analysis*, 56(11):3398–3414, 2012.

M. Belmonte, G. Koop, and D. Korobilis. Hierarchical shrinkage in time-varying parameter models. *Journal of Forecasting*, 33(1):80–94, 2014.

L. Benati. The great moderation in the United Kingdom. *Journal of Money, Credit and Banking*, 40(1):121–147, 2008.

A. Berg, R. Meyer, and J. Yu. Deviance information criterion for comparing stochastic volatility models. *Journal of Business and Economic Statistics*, 22(1):107–120, 2004.

C. Brooks and M. Prokopczuk. The dynamics of commodity prices. *Quantitative Finance*, 13(4):527–542, 2013.

J. C. C. Chan. Moving average stochastic volatility models with application to inflation forecast. *Journal of Econometrics*, 176(2):162–172, 2013.

J. C. C. Chan and E. Eisenstat. Marginal likelihood estimation with the Cross-Entropy method. *Econometric Reviews*, 34(3):256–285, 2015.

J. C. C. Chan and A. L. Grant. Fast computation of the deviance information criterion for latent variable models. *Computational Statistics and Data Analysis*, 2014. Forthcoming.

J. C. C. Chan and A. L. Grant. Pitfalls of estimating the marginal likelihood using the modified harmonic mean. *Economics Letters*, 131:29–33, 2015.

J. C. C. Chan and A. L. Grant. On the observed-data deviance information criterion for volatility modeling. *Working Paper*, 2016.

J. C. C. Chan and I. Jeliazkov. Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1:101–120, 2009.

J. C. C. Chan, G. Koop, R. Leon-Gonzalez, and R. W. Strachan. Time varying dimension models. *Journal of Business and Economic Statistics*, 30(3):358–367, 2012.

J. C. C. Chan, G. Koop, and S. M. Potter. A bounded model of time variation in trend inflation, NAIRU and the Phillips curve. *Journal of Applied Econometrics*, 2015. Forthcoming.

S. Chib. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 90:1313–1321, 1995.

T. E. Clark. Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business and Economic Statistics*, 29(3):327–341, 2011.

T. E. Clark and F. Ravazzolo. Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, 30(4):551–575, 2015.

T. Cogley and T. J. Sargent. Drifts and volatilities: Monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, 8(2):262–302, 2005.

A. D'Agostino, L. Gambetti, and D. Giannone. Macroeconomic forecasting and structural change. *Journal of Applied Econometrics*, 28(1):82–101, 2013.

G. Deborah and R. W. Strachan. Nonlinear impacts of international business cycles on the U.K.—A Bayesian smooth transition VAR approach. *Studies in Nonlinear Dynamics and Econometrics*, 14(1):1–33, 2009.

E. Eisenstat and R. W. Strachan. Modelling inflation volatility. *Journal of Applied Econometrics*, 2015. Forthcoming.

E. Eisenstat, J. C. C. Chan, and R. W. Strachan. Stochastic model specification search for time-varying parameter VARs. *Econometric Reviews*, 2015. Forthcoming.

N. Friel and J. Wyse. Estimating the evidence—a review. *Statistica Neerlandica*, 66(3):288–308, 2012.

S. Frühwirth-Schnatter. Bayesian model discrimination and Bayes factors for linear Gaussian state space models. *Journal of the Royal Statistical Society Series B*, 57(1):237–246, 1995.

S. Frühwirth-Schnatter and H. Wagner. Marginal likelihoods for non-Gaussian models using auxiliary mixture sampling. *Computational Statistics and Data Analysis*, 52(10):4608–4624, 2008.

S. Frühwirth-Schnatter and H. Wagner. Stochastic model specification search for Gaussian and partial non-Gaussian state space models. *Journal of Econometrics*, 154(1):85–100, 2010.

A. E. Gelfand and D. K. Dey. Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society Series B*, 56(3):501–514, 1994.

M. Kalli and J. E. Griffin. Time-varying sparsity in dynamic regression models. *Journal of Econometrics*, 178(2):779–793, 2014.

G. Kastner. Dealing with stochastic volatility in time series using the R package stochvol. *Journal of Statistical Software*, 2015. Forthcoming.

G. Kastner and S. Frühwirth-Schnatter. Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics and Data Analysis*, 76:408–423, 2014.

S. Kim, N. Shepherd, and S. Chib. Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies*, 65(3):361–393, 1998.

G. Koop and D. Korobilis. Forecasting inflation using dynamic model averaging. *International Economic Review*, 53(3):867–886, 2012.

G. Koop and D. Korobilis. Large time-varying parameter VARs. *Journal of Econometrics*, 177(2):185–198, 2013.

G. Koop and S. M. Potter. Bayes factors and nonlinearity: Evidence from economic time series. *Journal of Econometrics*, 88(2):251–281, 1999.

G. Koop, D. J. Poirier, and J. L. Tobias. *Bayesian Econometric Methods*. Cambridge University Press, 2007.

G. Koop, R. Leon-Gonzalez, and R. W. Strachan. On the evolution of the monetary policy transmission mechanism. *Journal of Economic Dynamics and Control*, 33(4): 997–1017, 2009.

G. Koop, R. Leon-Gonzalez, and R. W. Strachan. Dynamic probabilities of restrictions in state space models: an application to the Phillips curve. *Journal of Business and Economic Statistics*, 28(3):370–379, 2010.

D. P. Kroese and J. C. C. Chan. *Statistical Modeling and Computation*. Springer, New York, 2014.

Y. Liu and J. Morley. Structural evolution of the U.S. economy. *Journal of Economic Dynamics and Control*, 42:50–68, 2014.

J. Morley, J. Piger, and R. Rasche. Inflation in the G7: Mind the gap(s). *Macroeconomic Dynamics*, 2013. Forthcoming.

G. V. Moura and D. E. Turatti. Efficient estimation of conditionally linear and Gaussian state space models. *Economics Letters*, 124(3):494–499, 2014.

J. Nakajima and M. West. Bayesian analysis of latent threshold dynamic models. *Journal of Business and Economic Statistics*, 31(2):151–164, 2013.

G. E. Primiceri. Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, 72(3):821–852, 2005.

D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B*, 64(4): 583–639, 2002.

D. Staiger, J. H. Stock, and M. W. Watson. The NAIRU, unemployment and monetary policy. *The Journal of Economic Perspectives*, 11(1):33–49, 1997.

J. H. Stock and M. W. Watson. Why has U.S. inflation become harder to forecast? *Journal of Money Credit and Banking*, 39(s1):3–33, 2007.

I. Verdinelli and L. Wasserman. Computing Bayes factors using a generalization of the Savage-Dickey density ratio. *Journal of the American Statistical Association*, 90(430): 614–618, 1995.

M. Vo. Oil and stock market volatility: A multivariate stochastic volatility perspective. *Energy Economics*, 33(5):956–965, 2011.

J. J. J. Wang, S. T. B. Choy, and J. S. K. Chan. Modelling stochastic volatility using generalized t distribution. *Journal of Statistical Computation and Simulation*, 83(2): 340–354, 2013.