Australian National University

Crawford School of Public Policy

CAMA

Centre for Applied Macroeconomic Analysis

A Bayesian Model Comparison for Trend-Cycle Decompositions of Output

CAMA Working Paper 31/2015
August 2015

Joshua C.C. Chan
Research School of Economics, ANU and
Centre for Applied Macroeconomic Analysis (CAMA), ANU

Angelia L. Grant
Research School of Economics, ANU

Abstract

We compare a number of widely used trend-cycle decompositions of output in a formal Bayesian model comparison exercise. This is motivated by the often markedly different results from these decompositions—different decompositions have broad implications for the relative importance of real versus nominal shocks in explaining variations in output. Using US quarterly real GDP, we find that the overall best model is an unobserved components model with two features: 1) a nonzero correlation between trend and cycle innovations; 2) a break in output growth in 2007. Under this specification, annualized trend output growth decreases from about 3.4% to 1.5% after the break. The results also indicate that real shocks are more important than nominal shocks.

Keywords

Bayesian model comparison, unobserved components, structural break, business cycle

JEL Classification

C11, C52, E32

Address for correspondence:

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

A Bayesian Model Comparison for Trend-Cycle Decompositions of Output

Joshua C.C. Chan*
Research School of Economics,
and Centre for Applied Macroeconomic Analysis,
Australian National University

Angelia L. Grant
Research School of Economics,
Australian National University

August 2015

Abstract

We compare a number of widely used trend-cycle decompositions of output in a formal Bayesian model comparison exercise. This is motivated by the often markedly different results from these decompositions—different decompositions have broad implications for the relative importance of real versus nominal shocks in explaining variations in output. Using US quarterly real GDP, we find that the overall best model is an unobserved components model with two features: 1) a nonzero correlation between trend and cycle innovations; 2) a break in output growth in 2007. Under this specification, annualized trend output growth decreases from about 3.4% to 1.5% after the break. The results also indicate that real shocks are more important than nominal shocks.

Keywords: Bayesian model comparison, unobserved components, structural break, business cycle

JEL classification: C11, C52, E32

*Joshua Chan would like to acknowledge financial support by the Australian Research Council via a Discovery Early Career Researcher Award (DE150100795).

1 Introduction

The decomposition of output into its trend and cyclical components is an important theoretical and empirical problem in the study of macroeconomic fluctuations, business cycles, and monetary and fiscal policy. Important early contributions to the literature include the Beveridge and Nelson (1981) decomposition using an unrestricted ARIMA model and the unobserved components models of Harvey (1985), Watson (1986) and Clark (1987). However, these two widely used trend-cycle decompositions yield markedly different results. The Beveridge-Nelson decomposition attributes most of the variance in output to the variation in trend—the cyclical component is small in amplitude and noisy. In contrast, the cyclical components from the unobserved components models are typically large, highly persistent and account for most of the variation in output.

These apparently conflicting results are reconciled in an important paper by Morley, Nelson, and Zivot (2003). They demonstrate that the difference is entirely due to one restriction imposed in the unobserved components model: the innovations to the trend and cycle are assumed to be uncorrelated. When this restriction is relaxed, they find that the two trend-cycle decompositions are identical (see also Morley, 2011). As a result, both estimation methods imply that real or permanent shocks are important and that cycles are small and noisy, and bear little resemblance to the business cycle chronology dated by the National Bureau of Economic Research (NBER). However, Perron and Wada (2009) argue that these features of the cycles are artifacts that arise from the neglect of a structural break in output growth. When a break in growth is allowed for, the cycle estimates are substantially more persistent and accord well with the NBER chronology. Importantly, their preferred model is one with a deterministic trend, which implies all the variation in output can be attributed to innovations to the cyclical component—i.e., real shocks are unimportant.

This brief overview of the literature underscores the sensitivity of cycle estimates to model choice, with differences in one or two key parameters giving starkly different trend-cycle decompositions. Other model specification choices—such as the dating of a break, which is fixed in 1973Q1 in Perron and Wada (2009)—can also be framed as a model selection problem. The model that is used has broad implications for the conclusions drawn about the relative importance of real versus nominal shocks in explaining variations in output. Hence, it is important to perform a model comparison exercise to select the best model (or average different model estimates across models), but it is seldom done in practice.

We take up this task and use a Bayesian model comparison framework to assess the adequacy of a variety of nested and nonnested models for decomposing US quarterly real GDP. In particular, we compare the unobserved components models of Clark (1987), Morley et al. (2003) and Perron and Wada (2009), and deterministic trend models (with or without a break). By treating each fixed break date as a separate model, we are able to date any change that might have occurred in output growth. A closely related paper is Morley and Piger (2012), who consider model comparison using information criteria and model averaging using an asymptotic approximation to the Bayes factors. In contrast, we

provide exact computations of the Bayes factors for either model comparison or model averaging. Consequently, there are notable differences between the results of the two approaches.

In order to compare the various trend-cycle decompositions, we develop new Bayesian estimation techniques using Markov chain Monte Carlo (MCMC) methods (see, e.g., Koop and Korobilis, 2010, for a general introduction of Bayesian methods for empirical macroeconomics). A key novel feature of our approach is that it builds upon the band and sparse matrix algorithms developed in Chan and Jeliazkov (2009), McCausland, Miller, and Pelletier (2011) and Chan (2013), which are shown to be more efficient than the conventional Kalman filter-based algorithms. In addition, due to the modular nature of MCMC algorithms, it is relatively straightforward to extend the estimation methods to regime-switching models or models with non-Gaussian innovations.

Our main results can be summarized as follows. First, allowing for a nonzero correlation between the permanent and transitory shocks substantially improves model fit. This is in line with the finding in Morley et al. (2003), who report an estimate of -0.9 for the correlation parameter. Second, the correlated unobserved components model of Morley et al. (2003) dominates any deterministic trend models with or without a break in growth. However, among correlated unobserved components models, a break in growth is likely to have occurred in 2007. In fact, the overall best model is one with a break in 2007Q1. Under this specification, annual trend output growth is estimated to have decreased from about 3.4% to 1.5% after the break. This model also indicates that real shocks are more important in explaining the variation in output relative to nominal shocks. Our model comparison results therefore complement Sinclair (2009), which confirms the robustness of the Morley et al. (2003) results to a bivariate model specification that includes GDP and unemployment.

The rest of this article is organized as follows. Section 2 first discusses a variety of trend-cycle decompositions of output that are widely used in the literature. It then outlines the Bayesian estimation methods used to fit these models. In Section 3 we give an overview of Bayesian model comparison using the marginal likelihood, as well as an importance sampling approach to compute this quantity. Then, in Section 4 we compare the performance of the various models in fitting US real GDP. Trend-cycle decompositions and other parameter estimates for selected models are also reported. Lastly, Section 5 concludes and briefly discusses some future research directions.

2 Trend-Cycle Decomposition using UC Models

In this section we discuss a variety of trend-cycle decompositions of output based on unobserved components models and the Bayesian estimation methods used to fit the models (see, e.g., Koop, Poirier, and Tobias, 2007, for a quick introduction of Bayesian computations in econometrics).

The estimation in the literature on trend-cycle decompositions using unobserved components models typically uses the maximum likelihood method. However, one issue with this approach is the so-called "pile-up" problem, whereby the maximum likelihood estimates take values at the boundary of the parameter space. This can occur, for example, if a variance parameter is estimated to be zero. The "pile-up" problem makes inference more difficult, as the usual asymptotic properties of the maximum likelihood estimator no longer hold.

Moreover, trend-cycle decompositions in the literature are typically obtained conditional on the maximum likelihood estimates. As such, there is no accounting for parameter uncertainty. However, given that trend-cycle decompositions can be sensitive to the values of a few key parameters, it is crucial to take parameter uncertainty into account. The models in Clark (1987) and Morley et al. (2003) highlight the importance of parameter uncertainty—the models differ in the value of only one parameter, but the trend-cycle decompositions from the two models are drastically different.

We adopt the Bayesian approach in which inference is based on the joint posterior distribution of the parameters. Using the posterior mean as the point estimate avoids the "pile up" problem—as long as a nondogmatic prior is used, by construction the posterior mean is away from the boundary of the parameter space. In addition, the trend-cycle decomposition is constructed by averaging the parameter values with respect to the joint posterior distribution of the parameters—hence, the decomposition does not depend on a particular set of parameter values. Furthermore, the Bayesian approach facilitates comparing nonnested models, which is discussed in more detail in Section 3.

2.1 Competing Models

The trend-cycle decomposition of aggregate output is motivated by the idea that it can be usefully viewed as the sum of two separate components: a nonstationary component that represents the long-term trend and a transitory deviation from the trend. More specifically, let y_t denote (100 times) the log of real GDP. Then y_t can be decomposed as

$$y_t = \tau_t + c_t, \tag{1}$$

where τ_t is the trend and c_t is the stationary, cyclical component. The nonstationary trend τ_t is modeled as a random walk with drift, whereas the cyclical component c_t is modeled as a zero mean stationary AR(p) process:

$$\tau_t = \mu_1 + \tau_{t-1} + u_t^{\tau}, \tag{2}$$
$$c_t = \phi_1 c_{t-1} + \cdots + \phi_p c_{t-p} + u_t^{y}, \tag{3}$$

where the initial τ_0 is treated as a parameter to be estimated and for simplicity we assume $c_{1-p} = \ldots = c_0 = 0$. Note that the drift μ_1 can be interpreted as the growth rate of trend GDP. Following Morley et al. (2003), we set $p = 2$ and assume innovations u_t^y and u_t^{τ} are

jointly normal

$$\begin{pmatrix} u_t^y \\ u_t^\tau \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \sigma_y^2 & \rho\sigma_y\sigma_\tau \\ \rho\sigma_y\sigma_\tau & \sigma_\tau^2 \end{pmatrix}\right).$$

We denote the model in (1)–(3) as UCUR. This model allows a nonzero correlation between the innovations u_t^y and u_t^τ. Hence, it includes the model of Clark (1987) as a special case with $\rho = 0$; this restricted model is denoted as UC0.

Perron and Wada (2009) point out that the trend-cycle decomposition might be sensitive to how the trend is modeled. In particular, they show that when a break is allowed for, the estimates of the cyclical component become larger in magnitude and are more persistent. Hence, we consider specifications with a break in the drift. Specifically, consider replacing (2) by a more general specification

$$\tau_t = \mu_1 1(t < t_0) + \mu_2 1(t \geqslant t_0) + \tau_{t-1} + u_t^\tau, \tag{4}$$

where $1(A)$ is the indicator function that takes the value 1 if the condition A is true and 0 otherwise, and t_0 is a known break point. In other words, the stochastic trend τ_t has a growth rate of μ_1 before the break t_0 and a growth rate of μ_2 after the break. In the model comparison exercise, we date the break point by comparing models with different t_0. We denote the UCUR model with a break at time t_0 as UCUR-t_0.

Lastly, we consider a set of models with deterministic trends. This is motivated by the findings in Perron and Wada (2009), where the preferred model is UC0 with a break point in 1973Q1. However, the variance of the innovation to the trend, σ_τ^2, is estimated to be zero, which is outside of the parameter space—the variance σ_τ^2 should be positive. To circumvent this difficulty, we consider instead the following deterministic trend

$$\tau_t = \mu_1 1(t < t_0) + \mu_2 1(t \geqslant t_0) + \tau_{t-1}. \tag{5}$$

The cyclical component c_t is modeled as in (3), with $u_t^y \sim \mathcal{N}(0, \sigma_y^2)$. This model is denoted as DT-t_0. We also consider a version without a break, which is denoted as DT.

2.2 Bayesian Estimation

In this section we outline the Bayesian estimation methods used to fit the UCUR model given in (1)–(3). More specifically, we develop a Markov sampler to obtain draws from the posterior distribution under the UCUR model. The other unobserved components models can be estimated similarly, and we leave the technical details to Appendix A. A key novel feature of our approach is that it builds upon the band and sparse matrix algorithms developed in Chan and Jeliazkov (2009) and Chan (2013). It is shown in McCausland et al. (2011) that this approach is more efficient compared to the conventional Kalman filter-based algorithms.

We assume proper but relatively noninformative priors for the model parameters $\boldsymbol{\phi} = (\phi_1, \phi_2)', \sigma_y^2, \sigma_\tau^2, \rho, \mu_1$ and τ_0. In particular, we consider a uniform prior on $(-1, 1)$ for ρ,

and identical uniform priors on $(0, 3)$ for σ_y^2 and σ_τ^2. The details of the priors are given in Appendix A. Since the marginal likelihood can be sensitive to prior specification, we use exactly the same priors for common parameters across models. For notational convenience, stack $\mathbf{y} = (y_1, \ldots, y_T)'$, and similarly define $\boldsymbol{\tau}$, $\mathbf{c}$, $\mathbf{u}^y$ and $\mathbf{u}^\tau$. Then, posterior draws can be obtained by sequentially sampling from the following densities: 1. $p(\boldsymbol{\tau} \,|\, \mathbf{y}, \boldsymbol{\phi}, \sigma_y^2, \sigma_\tau^2, \rho, \mu_1, \tau_0)$; 2. $p(\boldsymbol{\phi} \,|\, \mathbf{y}, \boldsymbol{\tau}, \sigma_y^2, \sigma_\tau^2, \rho, \mu_1, \tau_0)$; 3. $p(\sigma_y^2 \,|\, \mathbf{y}, \boldsymbol{\tau}, \boldsymbol{\phi}, \sigma_\tau^2, \rho, \mu_1, \tau_0)$; 4. $p(\sigma_\tau^2 \,|\, \mathbf{y}, \boldsymbol{\tau}, \boldsymbol{\phi}, \sigma_y^2, \rho, \mu_1, \tau_0)$; 5. $p(\rho \,|\, \mathbf{y}, \boldsymbol{\tau}, \boldsymbol{\phi}, \sigma_y^2, \sigma_\tau^2, \mu_1, \tau_0)$; and 6. $p(\tau_0, \mu_1 \,|\, \mathbf{y}, \boldsymbol{\tau}, \boldsymbol{\phi}, \sigma_y^2, \sigma_\tau^2, \rho)$. Here we discuss how Step 1 can be implemented; the details of other steps are given in Appendix A.

First we write the system (1)–(3) in the following matrix form:

$$\begin{aligned} \mathbf{y} &= \boldsymbol{\tau} + \mathbf{c}, \\ \mathbf{H}_{\boldsymbol{\phi}} \mathbf{c} &= \mathbf{u}^y, \\ \mathbf{H} \boldsymbol{\tau} &= \widetilde{\boldsymbol{\alpha}} + \mathbf{u}^\tau, \end{aligned}$$

where $\widetilde{\boldsymbol{\alpha}} = (\mu_1 + \tau_0, \mu_1, \ldots, \mu_1)'$ and

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

Note that both $\mathbf{H}$ and $\mathbf{H}_{\boldsymbol{\phi}}$ are band matrices with only a few nonzero elements arranged along the main diagonal. Further, since both are square matrices with unit determinant, they are invertible. Hence, given $\boldsymbol{\phi}$, σ_y^2, σ_τ^2, ρ and τ_0, we have

$$\begin{pmatrix} \mathbf{c} \\ \boldsymbol{\tau} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{0} \\ \boldsymbol{\alpha} \end{pmatrix}, \begin{pmatrix} \sigma_y^2 (\mathbf{H}_{\boldsymbol{\phi}}' \mathbf{H}_{\boldsymbol{\phi}})^{-1} & \rho \sigma_y \sigma_\tau (\mathbf{H}' \mathbf{H}_{\boldsymbol{\phi}})^{-1} \\ \rho \sigma_y \sigma_\tau (\mathbf{H}_{\boldsymbol{\phi}}' \mathbf{H})^{-1} & \sigma_\tau^2 (\mathbf{H}' \mathbf{H})^{-1} \end{pmatrix} \right),$$

where $\boldsymbol{\alpha} = \mathbf{H}^{-1} \widetilde{\boldsymbol{\alpha}}$. Using the properties of the Gaussian distributions (see, e.g., Kroese and Chan, 2014, Chapter 3.6), the marginal distribution of $\boldsymbol{\tau}$ (unconditional on $\mathbf{c}$) is

$$(\boldsymbol{\tau} \,|\, \sigma_\tau^2, \mu_1, \tau_0) \sim \mathcal{N}(\boldsymbol{\alpha}, \sigma_\tau^2 (\mathbf{H}' \mathbf{H})^{-1}),$$

and the conditional distribution of $\mathbf{y}$ given $\boldsymbol{\tau}$ and other parameters is given by

$$(\mathbf{y} \,|\, \boldsymbol{\tau}, \boldsymbol{\phi}, \sigma_y^2, \sigma_\tau^2, \rho, \mu_1, \tau_0) \sim \mathcal{N}\left(\mathbf{H}_{\boldsymbol{\phi}}^{-1} \mathbf{a} + \mathbf{H}_{\boldsymbol{\phi}}^{-1} \mathbf{B} \boldsymbol{\tau}, (1 - \rho^2) \sigma_y^2 (\mathbf{H}_{\boldsymbol{\phi}}' \mathbf{H}_{\boldsymbol{\phi}})^{-1} \right),$$

where

$$\mathbf{a} = -\frac{\rho \sigma_y}{\sigma_\tau} \mathbf{H} \boldsymbol{\alpha}, \quad \mathbf{B} = \mathbf{H}_{\boldsymbol{\phi}} + \frac{\rho \sigma_y}{\sigma_\tau} \mathbf{H}.$$

Therefore, the prior density of $\boldsymbol{\tau}$ and the *conditional* likelihood are given by

$$p(\boldsymbol{\tau} \,|\, \sigma_\tau^2, \mu_1, \tau_0) = (2\pi \sigma_\tau^2)^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2\sigma_\tau^2} (\boldsymbol{\tau} - \boldsymbol{\alpha})' \mathbf{H}' \mathbf{H} (\boldsymbol{\tau} - \boldsymbol{\alpha})} \tag{6}$$

$$p(\mathbf{y} \,|\, \boldsymbol{\tau}, \boldsymbol{\phi}, \sigma_y^2, \sigma_\tau^2, \rho, \mu_1, \tau_0) = (2\pi \sigma_y^2 (1 - \rho^2))^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2(1-\rho^2)\sigma_y^2} (\mathbf{H}_{\boldsymbol{\phi}} \mathbf{y} - \mathbf{a} - \mathbf{B} \boldsymbol{\tau})' (\mathbf{H}_{\boldsymbol{\phi}} \mathbf{y} - \mathbf{a} - \mathbf{B} \boldsymbol{\tau})}. \tag{7}$$

Then, by standard linear regression results (see, e.g., Kroese and Chan, 2014, pp. 237-240), we have

$$(\boldsymbol{\tau} \mid \mathbf{y}, \boldsymbol{\phi}, \sigma_y^2, \sigma_\tau^2, \mu_1, \rho, \tau_0) \sim \mathcal{N}(\widehat{\boldsymbol{\tau}}, \mathbf{K}_{\boldsymbol{\tau}}^{-1}),$$

where

$$\mathbf{K}_{\boldsymbol{\tau}} = \frac{1}{\sigma_\tau^2}\mathbf{H}'\mathbf{H} + \frac{1}{(1-\rho^2)\sigma_y^2}\mathbf{B}'\mathbf{B}, \quad \widehat{\boldsymbol{\tau}} = \mathbf{K}_{\boldsymbol{\tau}}^{-1}\left(\frac{1}{\sigma_\tau^2}\mathbf{H}'\mathbf{H}\boldsymbol{\alpha} + \frac{1}{(1-\rho^2)\sigma_y^2}\mathbf{B}'(\mathbf{H}_{\boldsymbol{\phi}}\mathbf{y} - \mathbf{a})\right).$$

Since $\mathbf{H}$, $\mathbf{H}_{\boldsymbol{\phi}}$ and $\mathbf{B}$ are all band matrices, so is the precision matrix $\mathbf{K}_{\boldsymbol{\tau}}$. As such, the precision sampler of Chan and Jeliazkov (2009) can be used to sample $\boldsymbol{\tau}$ efficiently. We leave the details of Steps 2-6 to Appendix A.

In addition, this approach allows us to derive an analytical expression of the *integrated* or *observed-data* likelihood $p(\mathbf{y} \mid \boldsymbol{\phi}, \sigma_y^2, \sigma_\tau^2, \rho, \mu_1, \tau_0)$, which is a crucial quantity for model comparison. We refer the readers to Appendix B for the exact expression and the derivations. Using this analytical expression, the integrated likelihood can then be evaluated quickly using band matrix routines, which is more efficient than using the Kalman filter.

3 Model Comparison via the Marginal Likelihood

In this section, we give an overview of Bayesian model comparison using the marginal likelihood. Then, we outline an importance sampling approach based on the improved cross-entropy method to compute the marginal likelihood.

Suppose we wish to compare a set of possibly nonnested models $\{M_1, \ldots, M_K\}$. Each model M_k is formally defined by two components: a likelihood function $p(\mathbf{y} \mid \boldsymbol{\theta}_k, M_k)$ that depends on the model-specific parameter vector $\boldsymbol{\theta}_k$ and a prior density $p(\boldsymbol{\theta}_k \mid M_k)$. The marginal likelihood under model M_k is defined as

$$p(\mathbf{y} \mid M_k) = \int p(\mathbf{y} \mid \boldsymbol{\theta}_k, M_k) p(\boldsymbol{\theta}_k \mid M_k) \mathrm{d}\boldsymbol{\theta}_k. \tag{8}$$

This marginal likelihood can be interpreted as a density forecast of the data under model M_k evaluated at the actual observed data $\mathbf{y}$. Hence, if the observed data are likely under the model, the associated marginal likelihood would be "large". Since the marginal likelihood is essentially a density forecast evaluation, it has a built-in penalty for model complexity.

Given two models M_i and M_j, if the marginal likelihood under model M_i is larger than that under M_j—i.e., the observed data are more likely under model M_i compared to model M_j—then it is viewed as evidence in favor of model M_i. The weight of evidence can be gauged by the *posterior odds ratio* between the two models, which can be written as follows:

$$\frac{\mathbb{P}(M_i \mid \mathbf{y})}{\mathbb{P}(M_j \mid \mathbf{y})} = \frac{\mathbb{P}(M_i)}{\mathbb{P}(M_j)} \times \frac{p(\mathbf{y} \mid M_i)}{p(\mathbf{y} \mid M_j)},$$

where $\mathbb{P}(M_i)/\mathbb{P}(M_j)$ is the prior odds ratio and the ratio of the marginal likelihoods $p(\mathbf{y} \mid M_i)/p(\mathbf{y} \mid M_j)$ is called the *Bayes factor* in favor of model M_i against M_j. If both models are equally probable *a priori*, i.e., the prior odds ratio is one, the posterior odds ratio between the two models is then equal to the Bayes factor. Then, if, for example, $\text{BF}_{ij} = 50$, it implies model M_i is 50 times more likely than model M_j given the data. For a more detailed discussion of the Bayes factor, we refer the readers to Koop (2003). Next, we outline an importance sampling method for calculating the marginal likelihoods under the various unobserved components models discussed in the previous section.

The computation of the marginal likelihood is in general nontrivial—the integral in (8) is often high-dimensional and cannot be obtained analytically. Here we adopt an improved version of the classic cross-entropy method to estimate the marginal likelihood. The classic cross-entropy method was originally developed for rare-event simulation by Rubinstein (1997, 1999) using a multi-level procedure to construct the optimal importance sampling density (see also Rubinstein and Kroese, 2004, for a book-length treatment). Chan and Kroese (2012) later show that the optimal importance sampling density can be obtained more accurately in one step using MCMC methods. This new variant is applied in Chan and Eisenstat (2015) for marginal likelihood estimation, which is outlined as follows.

Suppose we wish to estimate $p(\mathbf{y} \mid M_k)$, the marginal likelihood under model M_k. For notational convenience we drop the model index M_k, and write the marginal likelihood, likelihood and prior as $p(\mathbf{y})$, $p(\mathbf{y} \mid \boldsymbol{\theta})$ and $p(\boldsymbol{\theta})$, respectively. The ideal zero-variance importance sampling density for this estimation problem is the posterior density $p(\boldsymbol{\theta} \mid \mathbf{y})$. Unfortunately, this density is only known up to a constant and therefore cannot be used directly in practice. Nevertheless, it provides a good benchmark to obtain a suitable importance sampling density.

The idea is to locate a density that is "close" to the ideal importance sampling density. Operationally, we find the density within a convenient family of distributions such that its Kullback-Leibler divergence—or the cross-entropy distance—to the ideal density is minimized. Once the optimal density, say, $g(\cdot)$, is obtained, it is used to construct the importance sampling estimator:

$$\widehat{p(\mathbf{y})} = \frac{1}{R} \sum_{r=1}^{R} \frac{p(\mathbf{y} \mid \boldsymbol{\theta}^{(r)}) p(\boldsymbol{\theta}^{(r)})}{g(\boldsymbol{\theta}^{(r)})}, \tag{9}$$

where $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(R)}$ are draws from the importance sampling density $g(\boldsymbol{\theta})$. The main advantage of this importance sampling approach is that it is easy to implement and the numerical standard error of the estimator is readily available. We refer the readers to Chan and Eisenstat (2015) for technical details. For the unobserved components models discussed in Section 2.1, the likelihood—or more accurately, the *integrated* likelihood or *observed-data* likelihood (the density of the data marginal of the latent states)—can in principle be evaluated using the Kalman filter. Here we adopt a more efficient approach, which is substantially faster than the Kalman filter. Specifically, we first derive analytical expressions for the integrated likelihoods under the various unobserved components models. These expressions are then evaluated using band and sparse matrix routines.

The technical details are given in Appendix B.

4 Empirical Results

In this section we compare the performance of the various unobserved components models discussed in Section 2 in fitting US real GDP. The main goal of this exercise is to establish the types of model features that are useful in trend-cycle decompositions. For example, does allowing for correlation between permanent and transitory shocks substantially improve model fit? Or is it more important to allow for a break in GDP growth? If yes, when is the break date?

We use US quarterly real GDP for our analysis, which is sourced from the Federal Reserve Bank of St. Louis economic database. The data are then transformed by taking the logs and multiplying by 100. We first report the model comparison results in Section 4.1. Trend-cycle decompositions and other parameter estimates for selected models are reported in Section 4.2.

4.1 Model Comparison Results

All the models are estimated using the new sampling approach based on band matrix routines discussed in Section 2.2 and Appendix A. The marginal likelihoods are computed using the improved cross-entropy method of Chan and Eisenstat (2015), which is outlined in Section 3. Each set of results is based on 100000 posterior draws after a burn-in period of 10000. For computing each marginal likelihood value, we use 50000 importance sampling draws.

We first address the timing of a break in GDP growth. Motivated by the results in Perron and Wada (2009)—in which the preferred specification is DT-73, i.e., a deterministic trend model with a break in GDP growth in 1973Q1—we consider two classes of models with a break, namely, UCUR-t_0 and DT-t_0. For each class of models, we consider 10 possible break dates: every first quarter from 1971Q1 to 1975Q1, and from 2005Q1 to 2009Q1. The latter dates are chosen so that we can determine whether trend GDP growth has changed following the Global Financial Crisis. We only consider a break in the first quarter of each year, given that the difference between models with breaks in consecutive quarters is expected to be small. Hence, it might be more useful to think of the break occurring in that year than in that particular quarter.

The model comparison results are reported in Table 1. By comparing UCUR-t_0 and DT-t_0 for each break date, we conclude that UCUR-t_0 uniformly outperforms DT-t_0 in fitting the US data. For example, the Bayes factor in favor of UCUR-07 against DT-07 is about 28—if we assume both models are equally likely *a priori*, the former becomes 28 times more likely given the data—indicating strong evidence in favor of UCUR-07.

In addition, for both classes of models, a break date in 2007Q1 is most favored by the data among the 10 possible break dates. This indicates that there seems to be a structural break in trend GDP growth in the early stages of the Global Financial Crisis.

Table 1: Log marginal likelihoods of the UCUR-t_0 and DT-t_0 models with various break dates. Numerical standard errors are in parenthesis.

DT-71	DT-72	DT-73	DT-74	DT-75
-368.96	-368.52	-367.95	-367.50	-368.20
(0.008)	(0.010)	(0.005)	(0.009)	(0.019)
DT-05	DT-06	DT-07	DT-08	DT-09
-367.43	-367.55	**-367.37**	-367.62	-369.86
(0.006)	(0.006)	(0.007)	(0.006)	(0.006)
UCUR-71	UCUR-72	UCUR-73	UCUR-74	UCUR-75
-366.25	-366.03	-365.69	-365.39	-365.99
(0.046)	(0.077)	(0.046)	(0.090)	(0.049)
UCUR-05	UCUR-06	UCUR-07	UCUR-08	UCUR-09
-364.45	-364.29	**-364.04**	-364.29	-365.60
(0.045)	(0.065)	(0.039)	(0.036)	(0.050)

The above results show that *if* a break is assumed, the break is most likely to have occurred in 2007. Next, we address the question of whether a break is indeed needed by comparing models with and without a break. In addition, by directly comparing UC0 and UCUR, we investigate whether or not allowing for a nonzero correlation between the permanent and transitory shocks substantially improves model fit. The model comparison results are reported in Table 2.

Table 2: Log marginal likelihoods of competing models with and without a break. Numerical standard errors are in parenthesis.

DT	UC0	UCUR	DT-07	UCUR-07
-370.63	-370.54	-365.02	-367.37	**-364.04**
(0.004)	(0.030)	(0.026)	(0.007)	(0.039)

Among the three models without a break, the UCUR model receives the most support from the data, followed by UC0 and the deterministic trend model DT. For instance, the Bayes factor in favor of UCUR against UC0 is about 250, indicating overwhelming support for the former. This is consistent with the findings in Morley et al. (2003)—the maximum likelihood estimate of the correlation ρ is reported to be about -0.9 with a relatively small standard error. In contrast, using information criteria to compare models, Morley and Piger (2012) find mixed results: Akaike information criterion weakly prefers UCUR but Bayesian information criterion slightly favors UC0.

It is also interesting to note that the best model among UC0, DT and DT-73 is DT-73,

which is in line with the results in Perron and Wada (2009), who prefer a deterministic trend model with a break in 1973Q1. However, among all the models we consider, the overall best model is UCUR-07. For example, compared to DT-73, the Bayes factor in favor of UCUR-07 is about 50, showing strong support for the latter model. Moreover, this also indicates that a break in trend GDP growth is likely to have occurred in the early stages of the Global Financial Crisis.

4.2 Trend-Cycle Estimates and Variance Decomposition

In this section we report the trend and cycle estimates under three models: DT-73, the deterministic trend model with a break in growth in 1973Q1; UCUR, the correlated unobserved components model of Morley et al. (2003); and UCUR-07, an extension of UCUR with a break in growth in 2007Q1.

Figure 1 plots the trend and cycle estimates under DT-73.[1] It can be seen in the left panel that there is a kink in the trend in 1973Q1, reflecting a slower estimated trend growth rate after the break date. Under this deterministic trend model, all variation in output is attributed to the cyclical component. Consequently, the cycles are large and persistent. For example, output started to outpace its trend from the early 1990s, and the cyclical component reached a peak in the new millennium. Output was substantially above trend until the Global Financial Crisis—since then it has dropped below trend. Interestingly, the output gap seems to have widened since 2010, reflecting the slower growth in GDP compared to the historical trend.

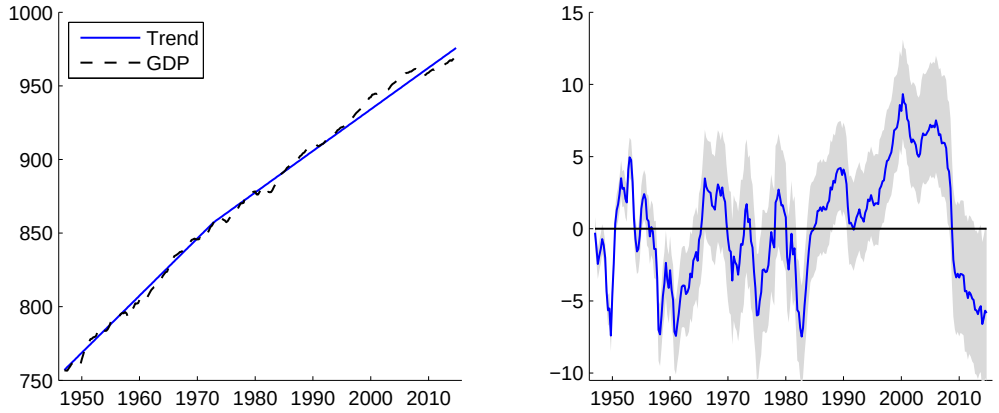

Figure 1: Estimates of trend (left panel) and cycle (right panel) under DT-73. The shaded region represents the 10% and 90% quantiles.

We report the trend and cycle estimates of UCUR and UCUR-07 in Figures 2 and 3.

[1]The trend estimates reported in this section are *smoothed* values obtained conditional on the whole sample. More specifically, the point estimates are the posterior means $\mathbb{E}(\boldsymbol{\tau} \mid \mathbf{y})$. The cycle estimates are then given by $\mathbf{y} - \mathbb{E}(\boldsymbol{\tau} \mid \mathbf{y})$.

Compared to those of DT-73, the cycle estimates are much smaller in magnitude and less persistent. This highlights the sensitivity of trend and cycle estimates to model specification, and hence the importance of model comparison. The cycle estimates also suggest that output is above trend in the run-up to the Global Financial Crisis, even though not to the same magnitude as in DT-73.

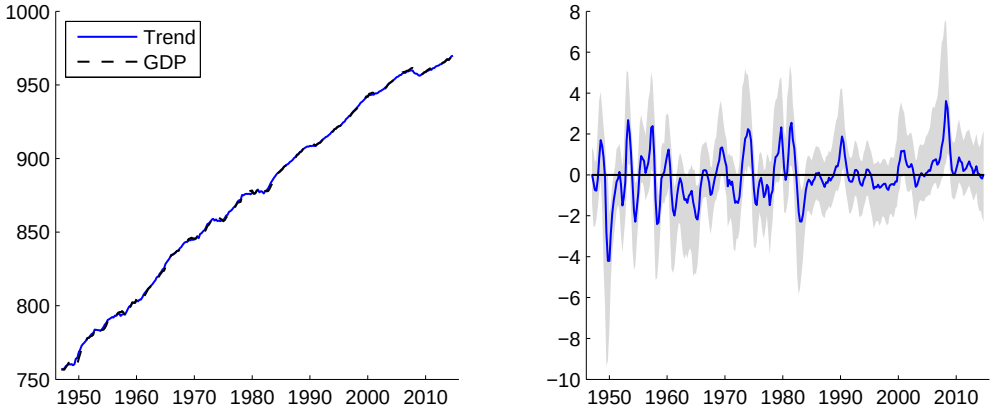

Figure 2: Estimates of trend (left panel) and cycle (right panel) under UCUR. The shaded region represents the 10% and 90% quantiles.

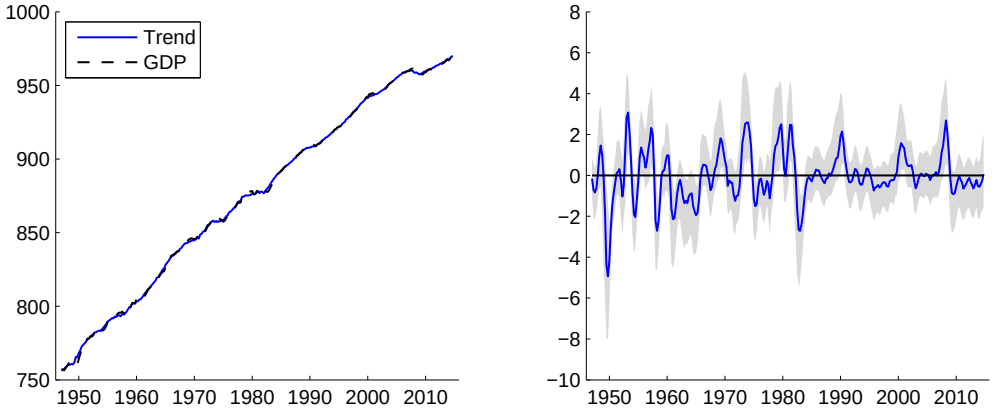

Figure 3: Estimates of trend (left panel) and cycle (right panel) under UCUR-07. The shaded region represents the 10% and 90% quantiles.

Not surprisingly, the cycle estimates of UCUR and UCUR-07 are fairly similar. The only noticeable difference occurs after the Global Financial Crisis: the former model suggests that output has been at trend since 2009 and remains there, whereas the latter model indicates that output has been slightly below trend since 2009. Figure 4 plots the trend estimates of the three models from 2005Q1 to 2014Q4.

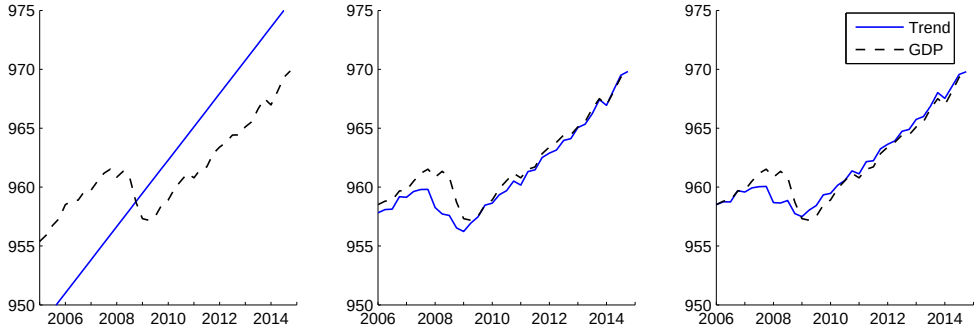

Figure 4: Estimates of trend from 2005Q1 to 2014Q4 under the DT-73 model (left panel), UCUR model (middle panel) and UCUR-07 model (right panel).

In addition to the different trend-cycle decompositions, these models also differ in their conclusions regarding the relative importance of permanent and transitory shocks. As mentioned above, DT-73 has a deterministic trend and therefore all variation of output is due to the transitory shocks. In contrast, both UCUR and UCUR-07 allow us to decompose the variance of output into the portions contributed by the permanent and transitory shocks. Table 3 reports the parameter estimates of the three models.

Table 3: Estimated posterior means under DT-73, UCUR and UCUR-07. Numerical standard errors are in parenthesis.

	DT-73	UCUR	UCUR-07
μ_1	0.97	0.78	0.84
	(0.039)	(0.082)	(0.077)
μ_2	0.70	–	0.37
	(0.039)		(0.199)
ϕ_1	1.34	0.95	1.10
	(0.057)	(0.343)	(0.361)
ϕ_2	-0.37	-0.36	-0.44
	(0.057)	(0.184)	(0.180)
σ_y^2	0.79	1.12	0.90
	(0.069)	(0.553)	(0.486)
σ_τ^2	–	1.85	1.42
		(0.494)	(0.593)
ρ	–	-0.87	-0.76
		(0.071)	(0.246)
$\mathbb{P}(\sigma_\tau^2 > \sigma_y^2 \mid \mathbf{y})$	–	0.92	0.83
$\mathbb{P}(\mu_1 > \mu_2 \mid \mathbf{y})$	1.00	–	0.98

Under UCUR-07 the estimates of σ_τ^2 and σ_y^2 are 1.42 and 0.9 respectively, giving a ratio of about 1.5. This indicates that permanent shocks are relatively more important compared

to transitory shocks, which is in line with the conclusion in Morley et al. (2003). However, since the variance parameters are not precisely estimated, we need to take account of parameter uncertainty.

In Figure 5 we plot the posterior densities of the variance ratio σ_τ^2/σ_y^2 under the two models.[2] Most of the mass for both densities is in regions that are larger than unity. In fact, the posterior probabilities $\mathbb{P}(\sigma_\tau^2 > \sigma_y^2 \,|\, \mathbf{y})$ are 0.92 and 0.83 for UCUR and UCUR-07, respectively. These results show that despite the high parameter uncertainty, both models conclude that real shocks are relatively much more important in explaining the variation in output.

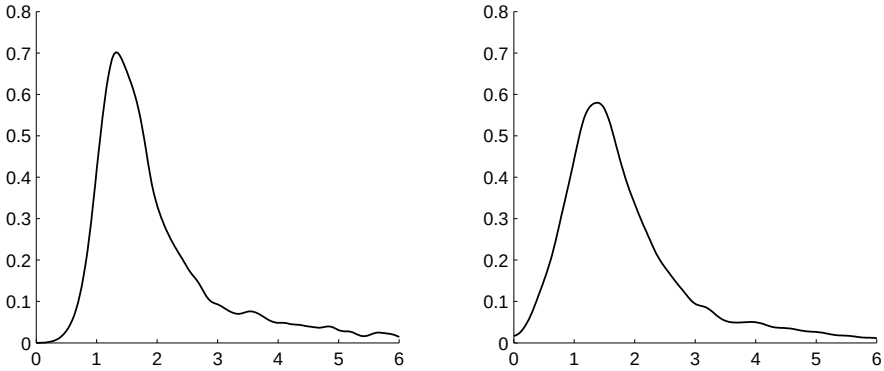

Figure 5: Posterior densities of the variance ratio σ_τ^2/σ_y^2 under UCUR (left panel) and UCUR-07 (right panel).

Next, we consider the difference in growth rates for models that allow for a break. Recall the model comparison results in Section 4.1 that show between the two deterministic trend models DT-73 and DT, the data favor the former. The parameter estimates of μ_1 and μ_2 under DT-73 support this conclusion. In particular, the annualized trend growth rate dropped from 3.88% before 1973 to 2.80% after 1973. The left panel of Figure 6 plots the posterior density of $\mu_1 - \mu_2$, which has virtually no mass below zero.

However, recall also that the overall best model is UCUR-07, which allows for a break in 2007. Under UCUR-07, the annualized trend growth rate more than halves before and after 2007—dropping from 3.36% to 1.48%. The low estimate of μ_2 is partly due to the influence of the Global Financial Crisis—we only have eight years of data after the break. Even so, our results support the view that growth has slowed after the financial crisis.

The posterior density of $\mu_1 - \mu_2$ in Figure 6 shows that there is more parameter uncertainty under UCUR-07 compared to DT-73, reflecting the difficulty in estimating the growth of a stochastic trend as opposed to a deterministic one. Nevertheless, the density has little

[2]For each model, posterior draws of the ratio σ_τ^2/σ_y^2 are first obtained using the MCMC sampler described in Appendix A. These draws are then used to compute the density using the kernel density estimator of Botev, Grotowski, and Kroese (2010).

mass below zero, which is consistent with the model comparison results that show the data prefer UCUR-07 compared to UCUR without a break.

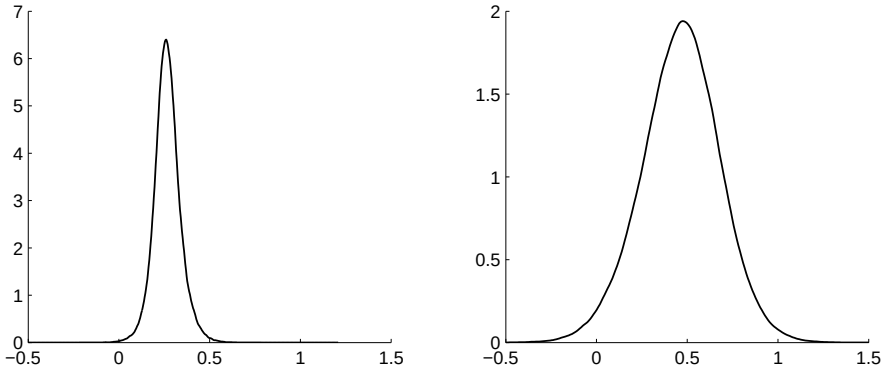

Figure 6: Posterior densities of the difference $\mu_1 - \mu_2$ under DT-73 (left panel) and UCUR-07 (right panel).

For both UCUR and UCUR-07, the estimates of ρ are negative and large in magnitude. In particular, ρ is estimated to be -0.87 and -0.76 for UCUR and UCUR-07, respectively. The posterior densities of ρ are plotted in Figure 7. The posterior modes of both densities are near -0.9, which is in line with the results in Morley et al. (2003). In addition, both densities have little mass near 0, showing the empirical relevance of allowing for nonzero correlation between the permanent and transitory shocks.

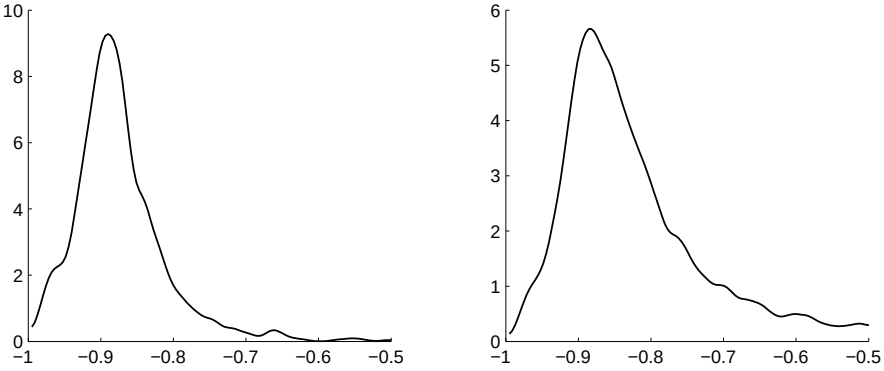

Figure 7: Posterior densities of ρ under UCUR (left panel) and UCUR-07 (right panel).

5 Concluding Remarks and Future Research

We have undertaken a formal Bayesian model comparison exercise to assess a number of models for decomposing US output into its trend and cyclical components. We find that

it is empirically important to allow for correlation between permanent and transitory shocks. The correlated unobserved components model dominates any deterministic trend models with or without a break in growth. The overall best model is the correlated unobserved components model with a break in 2007. Under this specification, annualized trend output growth decreases from about 3.4% to 1.5% after the break. This model also indicates that permanent shocks are relatively more important in explaining the variation in output compared to transitory shocks. It would be interesting to see if these conclusions remain true if a broader set of nonlinear and asymmetric models are included, such as those in Morley and Piger (2008), Sinclair (2010) and Morley and Piger (2012).

Many recent papers, including Stock and Watson (2007), Chan (2013) and Clark and Doh (2014), have demonstrated the importance of allowing for stochastic volatility in modeling inflation using unobserved components models. For future research, it would also be worthwhile to investigate if stochastic volatility is useful for decomposing output. In particular, the stochastic volatility framework allows us to assess if the contributions of permanent and transitory shocks have changed over time in explaining the variation in output.

Appendix A: Estimation Details

This appendix discusses the priors and provides the estimation details of the unobserved components models discussed in Section 2.

Estimation of the UCUR Model

For the UCUR model in (1)–(3), the model parameters are $\boldsymbol{\phi} = (\phi_1, \phi_2)', \sigma_y^2, \sigma_\tau^2, \rho, \mu_1$ and τ_0. We assume standard independent priors for $\boldsymbol{\phi}$, μ_1 and τ_0:

$$\boldsymbol{\phi} \sim \mathcal{N}(\boldsymbol{\phi}_0, \mathbf{V}_{\boldsymbol{\phi}}) 1(\boldsymbol{\phi} \in \mathbf{R}), \quad \mu_1 \sim \mathcal{N}(\mu_0, V_\mu), \quad \tau_0 \sim \mathcal{N}(\tau_{00}, V_\tau),$$

where $\mathbf{R}$ is the stationarity region. We assume relatively large prior variances with $\mathbf{V}_{\boldsymbol{\phi}} = \mathbf{I}_2$, $V_\mu = 1$ and $V_\tau = 100$. For the prior means, we set $\boldsymbol{\phi}_0 = (1.3, -0.7)'$, $\tau_{00} = 750$ and $\mu_0 = 0.75$. In particular, these values imply that the prior mean of the annualized growth rate is 3% and the AR(2) process of the transitory component has two complex roots. Next, σ_y^2, σ_τ^2 and ρ have uniform priors:

$$\sigma_y^2 \sim \mathcal{U}(0, b_y), \quad \sigma_\tau^2 \sim \mathcal{U}(0, b_\tau), \quad \rho \sim \mathcal{U}(-1, 1),$$

where we set the upper bounds as $b_y = b_\tau = 3$.

Posterior draws are obtained by sequentially sampling from: 1. $p(\boldsymbol{\tau} \,|\, \mathbf{y}, \boldsymbol{\phi}, \sigma_y^2, \sigma_\tau^2, \rho, \mu_1, \tau_0)$; 2. $p(\boldsymbol{\phi} \,|\, \mathbf{y}, \boldsymbol{\tau}, \sigma_y^2, \sigma_\tau^2, \rho, \mu_1, \tau_0)$; 3. $p(\sigma_y^2 \,|\, \mathbf{y}, \boldsymbol{\tau}, \boldsymbol{\phi}, \sigma_\tau^2, \rho, \mu_1, \tau_0)$; 4. $p(\sigma_\tau^2 \,|\, \mathbf{y}, \boldsymbol{\tau}, \boldsymbol{\phi}, \sigma_y^2, \rho, \mu_1, \tau_0)$; 5. $p(\rho \,|\, \mathbf{y}, \boldsymbol{\tau}, \boldsymbol{\phi}, \sigma_y^2, \sigma_\tau^2, \mu_1, \tau_0)$; and 6. $p(\tau_0, \mu_1 \,|\, \mathbf{y}, \boldsymbol{\tau}, \boldsymbol{\phi}, \sigma_y^2, \sigma_\tau^2, \rho)$. The implementation of Step 1 is discussed in Section 2.2. Here we provide the details of the other steps.

To sample $\boldsymbol{\phi}$ in Step 2, recall that $\mathbf{u}^y$ and $\boldsymbol{\tau}$ are jointly normal:

$$\begin{pmatrix} \mathbf{u}^y \\ \boldsymbol{\tau} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{0} \\ \boldsymbol{\alpha} \end{pmatrix}, \begin{pmatrix} \sigma_y^2 \mathbf{I}_T & \rho\sigma_y\sigma_\tau(\mathbf{H}')^{-1} \\ \rho\sigma_y\sigma_\tau\mathbf{H}^{-1} & \sigma_\tau^2(\mathbf{H}'\mathbf{H})^{-1} \end{pmatrix} \right), \tag{10}$$

where $\boldsymbol{\alpha} = \mathbf{H}^{-1}\widetilde{\boldsymbol{\alpha}}$ with $\widetilde{\boldsymbol{\alpha}} = (\mu_1 + \tau_0, \mu_1, \ldots, \mu_1)'$. Hence, the conditional distribution of $\mathbf{u}^y$ given $\boldsymbol{\tau}$ and other parameters is

$$(\mathbf{u}^y \,|\, \boldsymbol{\tau}, \sigma_y^2, \sigma_\tau^2, \rho, \mu_1, \tau_0) \sim \mathcal{N}\left(\frac{\rho\sigma_y}{\sigma_\tau}\mathbf{H}(\boldsymbol{\tau} - \boldsymbol{\alpha}), (1-\rho^2)\sigma_y^2\mathbf{I}_T \right).$$

Next, we write (3) as

$$\mathbf{c} = \mathbf{X}_{\boldsymbol{\phi}}\boldsymbol{\phi} + \mathbf{u}^y,$$

where $\mathbf{X}_{\boldsymbol{\phi}}$ is a $T \times 2$ matrix consisting of lagged values of c_t. Then, by standard regression results, we have

$$(\boldsymbol{\phi} \,|\, \mathbf{y}, \boldsymbol{\tau}, \sigma_y^2, \sigma_\tau^2, \rho, \mu_1, \tau_0) \sim \mathcal{N}(\widehat{\boldsymbol{\phi}}, \mathbf{K}_{\boldsymbol{\phi}}^{-1}) 1(\boldsymbol{\phi} \in \mathbf{R}),$$

where

$$\mathbf{K}_{\boldsymbol{\phi}} = \mathbf{V}_{\boldsymbol{\phi}}^{-1} + \frac{1}{(1-\rho^2)\sigma_y^2}\mathbf{X}_{\boldsymbol{\phi}}'\mathbf{X}_{\boldsymbol{\phi}},$$
$$\widehat{\boldsymbol{\phi}} = \mathbf{K}_{\boldsymbol{\phi}}^{-1}\left(\mathbf{V}_{\boldsymbol{\phi}}^{-1}\boldsymbol{\phi}_0 + \frac{1}{(1-\rho^2)\sigma_y^2}\mathbf{X}_{\boldsymbol{\phi}}'\left(\mathbf{c} - \frac{\rho\sigma_y}{\sigma_\tau}\mathbf{H}(\boldsymbol{\tau}-\boldsymbol{\alpha})\right)\right).$$

A draw from this truncated normal distribution can be obtained by the acceptance-rejection method, i.e., keep sampling from $\mathcal{N}(\widehat{\boldsymbol{\phi}}, \mathbf{K}_{\boldsymbol{\phi}}^{-1})$ until $\boldsymbol{\phi} \in \mathbf{R}$.

To implement Steps 3 to 5, we first derive the joint density of $\mathbf{u}^y$ and $\mathbf{u}^\tau$. To that end, note that given σ_y^2, σ_τ^2 and ρ, we can factorize (u_t^y, u_t^τ) as:

$$u_t^\tau \sim \mathcal{N}(0, \sigma_\tau^2), \quad (u_t^y \,|\, u_t^\tau) \sim \mathcal{N}\left(\frac{\rho\sigma_y}{\sigma_\tau}u_t^\tau, (1-\rho^2)\sigma_y^2\right).$$

Hence, the joint density of $\mathbf{u}^y$ and $\mathbf{u}^\tau$ is given by

$$\begin{aligned} p(\mathbf{u}^y, \mathbf{u}^\tau \,|\, \sigma_y^2, \sigma_\tau^2, \rho) &\propto (\sigma_\tau^2)^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2\sigma_\tau^2}\sum_{t=1}^T (u_t^\tau)^2} ((1-\rho^2)\sigma_y^2)^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2(1-\rho^2)\sigma_y^2}\sum_{t=1}^T \left(u_t^y - \frac{\rho\sigma_y}{\sigma_\tau}u_t^\tau\right)^2}, \\ &= ((1-\rho^2)\sigma_y^2\sigma_\tau^2)^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2\sigma_\tau^2}k_3 - \frac{1}{2(1-\rho^2)\sigma_y^2}\left(k_1 - \frac{2\rho\sigma_y}{\sigma_\tau}k_2 + \frac{\rho^2\sigma_y^2}{\sigma_\tau^2}k_3\right)}, \end{aligned} \tag{11}$$

where $k_1 = \sum_{t=1}^T (u_t^y)^2$, $k_2 = \sum_{t=1}^T u_t^y u_t^\tau$ and $k_3 = \sum_{t=1}^T (u_t^\tau)^2$. It follows from (11) that

$$p(\sigma_y^2 \,|\, \mathbf{y}, \boldsymbol{\tau}, \boldsymbol{\phi}, \sigma_\tau^2, \rho, \mu_1, \tau_0) \propto p(\sigma_y^2) \times (\sigma_y^2)^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2(1-\rho^2)\sigma_y^2}\left(k_1 - \frac{2\rho\sigma_y}{\sigma_\tau}k_2 + \frac{\rho^2\sigma_y^2}{\sigma_\tau^2}k_3\right)},$$

where $p(\sigma_y^2)$ is the truncated normal prior specified above. This full conditional density of σ_y^2 is not a standard density and we sample from it using a Griddy-Gibbs step. That is, we evaluate the full conditional density on a fine grid, and obtain a draw from the density using the inverse-transform method (see, e.g., Kroese, Taimre, and Botev, 2011, pp. 45–47). Steps 4 and 5 can be similarly implemented by noting that

$$p(\sigma_\tau^2 \,|\, \mathbf{y}, \boldsymbol{\tau}, \boldsymbol{\phi}, \sigma_y^2, \rho, \mu_1, \tau_0) \propto p(\sigma_\tau^2) \times (\sigma_\tau^2)^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2\sigma_\tau^2}k_3 - \frac{1}{2(1-\rho^2)\sigma_y^2}\left(k_1 - \frac{2\rho\sigma_y}{\sigma_\tau}k_2 + \frac{\rho^2\sigma_y^2}{\sigma_\tau^2}k_3\right)}$$
$$p(\rho \,|\, \mathbf{y}, \boldsymbol{\tau}, \boldsymbol{\phi}, \sigma_y^2, \sigma_\tau^2, \mu_1, \tau_0) \propto p(\rho) \times (1-\rho^2)^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2(1-\rho^2)\sigma_y^2}\left(k_1 - \frac{2\rho\sigma_y}{\sigma_\tau}k_2 + \frac{\rho^2\sigma_y^2}{\sigma_\tau^2}k_3\right)},$$

where $p(\sigma_\tau^2)$ and $p(\rho)$ are the priors for σ_τ^2 and ρ respectively.

Lastly, to jointly sample τ_0 and μ_1, note that we can write $\boldsymbol{\alpha} = \tau_0\mathbf{1}_T + \mu_1\mathbf{H}^{-1}\mathbf{1}_T = \mathbf{X}_{\boldsymbol{\delta}}\boldsymbol{\delta}$, where $\mathbf{1}_T$ is a $T \times 1$ column of ones, $\mathbf{X}_{\boldsymbol{\delta}} = (\mathbf{1}_T, \mathbf{H}^{-1}\mathbf{1}_T)$ and $\boldsymbol{\delta} = (\tau_0, \mu_1)'$. It follows from (10) that the conditional distribution of $\boldsymbol{\tau}$ given $\mathbf{u}^y$ and other parameters is

$$(\boldsymbol{\tau} \,|\, \mathbf{u}^y, \sigma_y^2, \sigma_\tau^2, \rho, \mu_1, \tau_0) \sim \mathcal{N}\left(\mathbf{X}_{\boldsymbol{\delta}}\boldsymbol{\delta} + \frac{\rho\sigma_\tau}{\sigma_y}\mathbf{H}^{-1}\mathbf{u}^y, (1-\rho^2)\sigma_\tau^2(\mathbf{H}'\mathbf{H})^{-1}\right).$$

Then, by standard regression results, we have

$$(\tau_0, \mu_1 \,|\, \mathbf{y}, \boldsymbol{\tau}, \sigma_y^2, \sigma_\tau^2, \rho, \boldsymbol{\phi}, \tau_0) \sim \mathcal{N}(\widehat{\boldsymbol{\delta}}, \mathbf{K}_{\boldsymbol{\delta}}^{-1}),$$

where

$$\mathbf{K}_{\boldsymbol{\delta}} = \mathbf{V}_{\boldsymbol{\delta}}^{-1} + \frac{1}{(1-\rho^2)\sigma_\tau^2}\mathbf{X}_{\boldsymbol{\delta}}'\mathbf{H}'\mathbf{H}\mathbf{X}_{\boldsymbol{\delta}},$$
$$\widehat{\boldsymbol{\delta}} = \mathbf{K}_{\boldsymbol{\delta}}^{-1}\left(\mathbf{V}_{\boldsymbol{\delta}}^{-1}\boldsymbol{\delta}_0 + \frac{1}{(1-\rho^2)\sigma_\tau^2}\mathbf{X}_{\boldsymbol{\delta}}'\mathbf{H}'\mathbf{H}\left(\boldsymbol{\tau} - \frac{\rho\sigma_\tau}{\sigma_y}\mathbf{H}^{-1}\mathbf{u}^y\right)\right),$$

where $\mathbf{V}_{\boldsymbol{\delta}} = \text{diag}(V_\tau, V_\mu)$ and $\boldsymbol{\delta}_0 = (\tau_{00}, \mu_0)'$.

Estimation of the UCUR-t_0 Model

We now consider an extension of the UCUR model that allows for a break in the growth rate of the trend at time t_0. Specifically, we replace (4) with (5), which is reproduced below:

$$\tau_t = \mu_1 1(t < t_0) + \mu_2 1(t \geqslant t_0) + \tau_{t-1} + u_t^\tau,$$

where $1(A)$ is the indicator function that takes the value 1 if the condition A is true and 0 otherwise. Compared to the UCUR model, the only additional parameter is μ_2. Its prior is assumed to be the same as that of μ_1, i.e., $\mu_2 \sim \mathcal{N}(\mu_0, V_\mu)$ with $\mu_0 = 0.75$ and $V_\mu = 1$. For the common parameters, we assume exactly the same priors as in the UCUR model.

Only minor modifications of the sampler for the UCUR model are needed to fit this extension. For example, if we redefine

$$\widetilde{\boldsymbol{\alpha}} = (\underbrace{\mu_1 + \tau_0, \mu_1, \ldots, \mu_1}_{t_0-1}, \underbrace{\mu_2, \ldots, \mu_2}_{T-t_0+1})'$$

and $\boldsymbol{\alpha} = \mathbf{H}^{-1}\widetilde{\boldsymbol{\alpha}}$, then Steps 1 and 2 can be implemented exactly as before. Similarly, if we compute $\mathbf{u}^\tau$ using $\mathbf{u}^\tau = \mathbf{H}\boldsymbol{\tau} - \widetilde{\boldsymbol{\alpha}}$, then Steps 3–5 remain the same as before. Lastly, to sample τ_0, μ_1 and μ_2 jointly, write $\boldsymbol{\alpha} = \tau_0\mathbf{1}_T + \mu_1\mathbf{H}^{-1}\mathbf{d}_1 + \mu_2\mathbf{H}^{-1}\mathbf{d}_2 = \mathbf{X}_{\boldsymbol{\delta}}\boldsymbol{\delta}$, where $\mathbf{d}_1$ is a $T \times 1$ vector of dummy variables where the first $t_0 - 1$ elements are 1 and the rest are 0, and $\mathbf{d}_2$ is defined so that $\mathbf{d}_1 + \mathbf{d}_2 = \mathbf{1}_T$. Note that $\mathbf{X}_{\boldsymbol{\delta}}$ and $\boldsymbol{\delta}$ are redefined as $\mathbf{X}_{\boldsymbol{\delta}} = (\mathbf{1}_T, \mathbf{H}^{-1}\mathbf{d}_1, \mathbf{H}^{-1}\mathbf{d}_2)$ and $\boldsymbol{\delta} = (\tau_0, \mu_1, \mu_2)'$. Then, the last step is implemented as before.

Estimation of the DT-t_0 Model

For the deterministic trend model with the trend

$$\tau_t = \mu_1 1(t < t_0) + \mu_2 1(t \geqslant t_0) + \tau_{t-1},$$

the model parameters are $\boldsymbol{\phi}, \sigma_y^2, \mu_1, \mu_2$ and τ_0. We adopt the same priors as in the UCUR-t_0 model. Note that we can write this model as

$$\begin{aligned} \mathbf{y} &= \mathbf{X}_{\boldsymbol{\delta}}\boldsymbol{\delta} + \mathbf{c}, \\ \mathbf{c} &= \mathbf{H}_{\boldsymbol{\phi}}^{-1}\mathbf{u}^y, \end{aligned} \tag{12}$$

where $\mathbf{X}_{\boldsymbol{\delta}} = (\mathbf{1}_T, \mathbf{H}^{-1}\mathbf{d}_1, \mathbf{H}^{-1}\mathbf{d}_2)$ and $\boldsymbol{\delta} = (\tau_0, \mu_1, \mu_2)'$.

Posterior draws can be obtained by sequentially sampling from: 1. $p(\boldsymbol{\phi} \,|\, \mathbf{y}, \sigma_y^2, \mu_1, \mu_2, \tau_0)$; 2. $p(\sigma_y^2 \,|\, \mathbf{y}, \boldsymbol{\phi}, \mu_1, \mu_2, \tau_0)$; and 3. $p(\tau_0, \mu_1, \mu_2 \,|\, \mathbf{y}, \boldsymbol{\phi}, \sigma_y^2)$. To implement Step 1, note that $(\mathbf{u}^y \,|\, \sigma_y^2) \sim \mathcal{N}(\mathbf{0}, \sigma_y^2 \mathbf{I}_T)$. Hence, we have

$$(\boldsymbol{\phi} \,|\, \mathbf{y}, \sigma_y^2, \mu_1, \mu_2, \tau_0) \sim \mathcal{N}(\widehat{\boldsymbol{\phi}}, \mathbf{K}_{\boldsymbol{\phi}}^{-1}) 1(\boldsymbol{\phi} \in \mathbf{R}),$$

where

$$\mathbf{K}_{\boldsymbol{\phi}} = \mathbf{V}_{\boldsymbol{\phi}}^{-1} + \frac{1}{\sigma_y^2}\mathbf{X}_{\boldsymbol{\phi}}'\mathbf{X}_{\boldsymbol{\phi}}, \quad \widehat{\boldsymbol{\phi}} = \mathbf{K}_{\boldsymbol{\phi}}^{-1}\left(\mathbf{V}_{\boldsymbol{\phi}}^{-1}\boldsymbol{\phi}_0 + \frac{1}{\sigma_y^2}\mathbf{X}_{\boldsymbol{\phi}}'\mathbf{c}\right).$$

A draw from this truncated normal distribution can be obtained by using the acceptance-rejection method.

Next, the full conditional density of σ_y^2 is given by

$$p(\sigma_y^2 \,|\, \mathbf{y}, \boldsymbol{\phi}, \mu_1, \mu_2, \tau_0) \propto p(\sigma_y^2) \times (\sigma_y^2)^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2\sigma_y^2}\sum_{t=1}^T (u_t^y)^2},$$

where $\mathbf{u}^y$ can be computed by $\mathbf{u}^y = \mathbf{H}_{\boldsymbol{\phi}}(\mathbf{y} - \mathbf{X}_{\boldsymbol{\delta}}\boldsymbol{\delta})$. As before, a draw of σ_y^2 can be obtained using the Griddy-Gibbs step. Lastly, it follows from (12) that

$$(\tau_0, \mu_1, \mu_2 \,|\, \mathbf{y}, \sigma_y^2, \boldsymbol{\phi}) \sim \mathcal{N}(\widehat{\boldsymbol{\delta}}, \mathbf{K}_{\boldsymbol{\delta}}^{-1}),$$

where

$$\mathbf{K}_{\boldsymbol{\delta}} = \mathbf{V}_{\boldsymbol{\delta}}^{-1} + \frac{1}{\sigma_y^2}\mathbf{X}_{\boldsymbol{\delta}}'\mathbf{H}_{\boldsymbol{\phi}}'\mathbf{H}_{\boldsymbol{\phi}}\mathbf{X}_{\boldsymbol{\delta}}, \quad \widehat{\boldsymbol{\delta}} = \mathbf{K}_{\boldsymbol{\delta}}^{-1}\left(\mathbf{V}_{\boldsymbol{\delta}}^{-1}\boldsymbol{\delta}_0 + \frac{1}{\sigma_y^2}\mathbf{X}_{\boldsymbol{\delta}}'\mathbf{H}_{\boldsymbol{\phi}}'\mathbf{H}_{\boldsymbol{\phi}}\mathbf{y}\right).$$

Appendix B: Integrated Likelihood Evaluation

This appendix derives analytical expressions for the integrated likelihoods under the unobserved components models discussed in Section 2. These integrated likelihoods can then be evaluated using band matrix routines, which are more efficient than using the conventional Kalman filter.

The UCUR Model

Recall that the prior density of $\boldsymbol{\tau}$ and the conditional likelihood under the UCUR model are given by

$$p(\boldsymbol{\tau} \,|\, \sigma_\tau^2, \mu_1, \tau_0) = (2\pi\sigma_\tau^2)^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2\sigma_\tau^2}(\boldsymbol{\tau}-\boldsymbol{\alpha})'\mathbf{H}'\mathbf{H}(\boldsymbol{\tau}-\boldsymbol{\alpha})}$$
$$p(\mathbf{y} \,|\, \boldsymbol{\tau}, \boldsymbol{\phi}, \sigma_y^2, \sigma_\tau^2, \rho, \mu_1, \tau_0) = (2\pi\sigma_y^2(1-\rho^2))^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2(1-\rho^2)\sigma_y^2}(\mathbf{H}_{\boldsymbol{\phi}}\mathbf{y}-\mathbf{a}-\mathbf{B}\boldsymbol{\tau})'(\mathbf{H}_{\boldsymbol{\phi}}\mathbf{y}-\mathbf{a}-\mathbf{B}\boldsymbol{\tau})},$$

where

$$\mathbf{a} = -\frac{\rho\sigma_y}{\sigma_\tau}\mathbf{H}\boldsymbol{\alpha}, \quad \mathbf{B} = \mathbf{H}_{\boldsymbol{\phi}} + \frac{\rho\sigma_y}{\sigma_\tau}\mathbf{H}.$$

Let $k_4 = (2\pi)^{-T}((1-\rho^2)\sigma_y^2\sigma_\tau^2)^{-\frac{T}{2}}$. Then, the integrated likelihood can be derived as follows:

$$\begin{aligned}
&p(\mathbf{y} \,|\, \boldsymbol{\phi}, \sigma_y^2, \sigma_\tau^2, \rho, \mu_1, \tau_0) = \int p(\mathbf{y} \,|\, \boldsymbol{\tau}, \boldsymbol{\phi}, \sigma_y^2, \sigma_\tau^2, \rho, \mu_1, \tau_0) p(\boldsymbol{\tau} \,|\, \sigma_\tau^2, \mu_1, \tau_0) \mathrm{d}\boldsymbol{\tau} \\
&= k_4 \int \mathrm{e}^{-\frac{1}{2(1-\rho^2)\sigma_y^2}(\mathbf{H}_{\boldsymbol{\phi}}\mathbf{y}-\mathbf{a}-\mathbf{B}\boldsymbol{\tau})'(\mathbf{H}_{\boldsymbol{\phi}}\mathbf{y}-\mathbf{a}-\mathbf{B}\boldsymbol{\tau})} \mathrm{e}^{-\frac{1}{2\sigma_\tau^2}(\boldsymbol{\tau}-\boldsymbol{\alpha})'\mathbf{H}'\mathbf{H}(\boldsymbol{\tau}-\boldsymbol{\alpha})} \mathrm{d}\boldsymbol{\tau} \\
&= k_4 \int \mathrm{e}^{-\frac{1}{2}\left(\frac{1}{(1-\rho^2)\sigma_y^2}\left((\mathbf{H}_{\boldsymbol{\phi}}\mathbf{y}-\mathbf{a})'(\mathbf{H}_{\boldsymbol{\phi}}\mathbf{y}-\mathbf{a}) - 2\boldsymbol{\tau}'\mathbf{B}'(\mathbf{H}_{\boldsymbol{\phi}}\mathbf{y}-\mathbf{a}) + \boldsymbol{\tau}'\mathbf{B}'\mathbf{B}\boldsymbol{\tau}\right) + \frac{1}{\sigma_\tau^2}(\boldsymbol{\tau}'\mathbf{H}'\mathbf{H}\boldsymbol{\tau} - 2\boldsymbol{\tau}'\mathbf{H}'\mathbf{H}\boldsymbol{\alpha} + \boldsymbol{\alpha}'\mathbf{H}'\mathbf{H}\boldsymbol{\alpha})\right)} \mathrm{d}\boldsymbol{\tau} \\
&= k_4 \mathrm{e}^{-\frac{1}{2}\left(\frac{1}{(1-\rho^2)\sigma_y^2}(\mathbf{H}_{\boldsymbol{\phi}}\mathbf{y}-\mathbf{a})'(\mathbf{H}_{\boldsymbol{\phi}}\mathbf{y}-\mathbf{a}) + \frac{1}{\sigma_\tau^2}\boldsymbol{\alpha}'\mathbf{H}'\mathbf{H}\boldsymbol{\alpha}\right)} \int \mathrm{e}^{-\frac{1}{2}(\boldsymbol{\tau}'\mathbf{K}_{\boldsymbol{\tau}}\boldsymbol{\tau} - 2\boldsymbol{\tau}'\mathbf{d}_{\boldsymbol{\tau}})} \mathrm{d}\boldsymbol{\tau} \\
&= k_4 \mathrm{e}^{-\frac{1}{2}\left(\frac{1}{(1-\rho^2)\sigma_y^2}(\mathbf{H}_{\boldsymbol{\phi}}\mathbf{y}-\mathbf{a})'(\mathbf{H}_{\boldsymbol{\phi}}\mathbf{y}-\mathbf{a}) + \frac{1}{\sigma_\tau^2}\boldsymbol{\alpha}'\mathbf{H}'\mathbf{H}\boldsymbol{\alpha} - \mathbf{d}_{\boldsymbol{\tau}}'\mathbf{K}_{\boldsymbol{\tau}}^{-1}\mathbf{d}_{\boldsymbol{\tau}}\right)} \int \mathrm{e}^{-\frac{1}{2}\left((\boldsymbol{\tau}-\mathbf{K}_{\boldsymbol{\tau}}^{-1}\mathbf{d}_{\boldsymbol{\tau}})'\mathbf{K}_{\boldsymbol{\tau}}(\boldsymbol{\tau}-\mathbf{K}_{\boldsymbol{\tau}}^{-1}\mathbf{d}_{\boldsymbol{\tau}})\right)} \mathrm{d}\boldsymbol{\tau} \\
&= k_4 \mathrm{e}^{-\frac{1}{2}\left(\frac{1}{(1-\rho^2)\sigma_y^2}(\mathbf{H}_{\boldsymbol{\phi}}\mathbf{y}-\mathbf{a})'(\mathbf{H}_{\boldsymbol{\phi}}\mathbf{y}-\mathbf{a}) + \frac{1}{\sigma_\tau^2}\boldsymbol{\alpha}'\mathbf{H}'\mathbf{H}\boldsymbol{\alpha} - \mathbf{d}_{\boldsymbol{\tau}}'\mathbf{K}_{\boldsymbol{\tau}}^{-1}\mathbf{d}_{\boldsymbol{\tau}}\right)} (2\pi)^{\frac{T}{2}} |\mathbf{K}_{\boldsymbol{\tau}}|^{-\frac{1}{2}} \\
&= (2\pi(1-\rho^2)\sigma_y^2\sigma_\tau^2)^{-\frac{T}{2}} |\mathbf{K}_{\boldsymbol{\tau}}|^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}\left(\frac{1}{(1-\rho^2)\sigma_y^2}(\mathbf{H}_{\boldsymbol{\phi}}\mathbf{y}-\mathbf{a})'(\mathbf{H}_{\boldsymbol{\phi}}\mathbf{y}-\mathbf{a}) + \frac{1}{\sigma_\tau^2}\boldsymbol{\alpha}'\mathbf{H}'\mathbf{H}\boldsymbol{\alpha} - \mathbf{d}_{\boldsymbol{\tau}}'\mathbf{K}_{\boldsymbol{\tau}}^{-1}\mathbf{d}_{\boldsymbol{\tau}}\right)},
\end{aligned}$$

where

$$\mathbf{K}_{\boldsymbol{\tau}} = \frac{1}{\sigma_\tau^2}\mathbf{H}'\mathbf{H} + \frac{1}{(1-\rho^2)\sigma_y^2}\mathbf{B}'\mathbf{B}, \quad \mathbf{d}_{\boldsymbol{\tau}} = \frac{1}{\sigma_\tau^2}\mathbf{H}'\mathbf{H}\boldsymbol{\alpha} + \frac{1}{(1-\rho^2)\sigma_y^2}\mathbf{B}'(\mathbf{H}_{\boldsymbol{\phi}}\mathbf{y} - \mathbf{a}).$$

Since $\mathbf{H}, \mathbf{H}_{\boldsymbol{\phi}}$ and $\mathbf{K}_{\boldsymbol{\tau}}$ are band matrices, this integrated likelihood can be evaluated quickly using the band matrix algorithms discussed Chan and Grant (2014).

The UCUR-t_0 Model

For the extension of the UCUR model that allows for a break in the growth rate of the trend at time t_0, only minor modifications are needed. In particular, if we redefine

$$\widetilde{\boldsymbol{\alpha}} = (\underbrace{\mu_1 + \tau_0, \mu_1, \ldots, \mu_1}_{t_0-1}, \underbrace{\mu_2, \ldots, \mu_2}_{T-t_0+1})'$$

and $\boldsymbol{\alpha} = \mathbf{H}^{-1}\widetilde{\boldsymbol{\alpha}}$, then the integrated likelihood of this generalization is exactly the same as that of the UCUR model.

The DT-t_0 Model

For the deterministic trend model, recall that from (12) we have

$$\mathbf{y} = \mathbf{X}_{\boldsymbol{\delta}}\boldsymbol{\delta} + \mathbf{H}_{\boldsymbol{\phi}}^{-1}\mathbf{u}^y,$$

where $\mathbf{u}^y \sim \mathcal{N}(\mathbf{0}, \sigma_y^2\mathbf{I}_T)$, $\mathbf{X}_{\boldsymbol{\delta}} = (\mathbf{1}_T, \mathbf{H}^{-1}\mathbf{d}_1, \mathbf{H}^{-1}\mathbf{d}_2)$ and $\boldsymbol{\delta} = (\tau_0, \mu_1, \mu_2)'$. Hence, the likelihood is given by

$$p(\mathbf{y} \mid \boldsymbol{\phi}, \sigma_y^2, \mu_1, \mu_2, \tau_0) = (2\pi\sigma_y^2)^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2\sigma_y^2}(\mathbf{y}-\mathbf{X}_{\boldsymbol{\delta}}\boldsymbol{\delta})'\mathbf{H}_{\boldsymbol{\phi}}'\mathbf{H}_{\boldsymbol{\phi}}(\mathbf{y}-\mathbf{X}_{\boldsymbol{\delta}}\boldsymbol{\delta})}.$$

References

S. Beveridge and C. R. Nelson. A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. *Journal of Monetary Economics*, 7(2):151 – 174, 1981.

Z. I. Botev, J. F. Grotowski, and D. P. Kroese. Kernel density estimation via diffusion. *The Annals of Statistics*, 38(5):2916–2957, 10 2010.

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 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 and D. P. Kroese. Improved cross-entropy method for estimation. *Statistics and Computing*, 22(5):1031–1040, 2012.

P. K. Clark. The cyclical component of US economic activity. *The Quarterly Journal of Economics*, 102(4):797–814, 1987.

T. E. Clark and T. Doh. A Bayesian evaluation of alternative models of trend inflation. *International Journal of Forecasting*, 30(3):426–448, 2014.

A. C. Harvey. Trends and cycles in macroeconomic time series. *Journal of Business and Economic Statistics*, 3(3):216–227, 1985.

G. Koop. *Bayesian Econometrics*. Wiley & Sons, New York, 2003.

G. Koop and D. Korobilis. Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics*, 3(4):267–358, 2010.

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

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

D. P. Kroese, T. Taimre, and Z. I. Botev. *Handbook of Monte Carlo Methods*. John Wiley and Sons, New York, 2011.

W. J. McCausland, S. Miller, and D. Pelletier. Simulation smoothing for state-space models: A computational efficiency analysis. *Computational Statistics and Data Analysis*, 55(1):199–212, 2011.

J. C. Morley. The two interpretations of the Beveridge–Nelson decomposition. *Macroeconomic Dynamics*, 15(03):419–439, 2011.

J. C. Morley and J. Piger. Trend/cycle decomposition of regime-switching processes. *Journal of Econometrics*, 146(2):220 – 226, 2008.

J. C. Morley and J. Piger. The asymmetric business cycle. *Review of Economics and Statistics*, 94(1):208–221, 2012.

J. C. Morley, C. R. Nelson, and E. Zivot. Why are the Beveridge-Nelson and unobserved-components decompositions of GDP so different? *Review of Economics and Statistics*, 85(2):235–243, 2003.

P. Perron and T. Wada. Let's take a break: Trends and cycles in US real GDP. *Journal of Monetary Economics*, 56(6):749–765, 2009.

R .Y. Rubinstein. Optimization of computer simulation models with rare events. *European Journal of Operational Research*, 99:89–112, 1997.

R. Y. Rubinstein. The cross-entropy method for combinatorial and continuous optimization. *Methodology and Computing in Applied Probability*, 2:127–190, 1999.

R. Y. Rubinstein and D. P. Kroese. *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization Monte-Carlo Simulation, and Machine Learning*. Springer-Verlag, New York, 2004.

T. M. Sinclair. The relationships between permanent and transitory movements in US output and the unemployment rate. *Journal of Money, Credit and Banking*, 41(2-3): 529–542, 2009.

T. M. Sinclair. Asymmetry in the business cycle: Friedman's plucking model with correlated innovations. *Studies in Nonlinear Dynamics and Econometrics*, 14(1):Article 3, 2010.

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.

M. W. Watson. Univariate detrending methods with stochastic trends. *Journal of Monetary Economics*, 18(1):49–75, 1986.