Mapping the Danish Business Cycle

Advanced Empirical Macroeconomic Analysis

Martin A. Kildemark*

December 2024

*University of Copenhagen

Contents

1 Introduction

Knowledge of past business cycles, and how they evolve, is valuable for policy makers when assessing stabilization policies and the general evolvement of economic growth and volatility. When conducting stabilization policies, knowledge of the economic fluctuations and trends is crucial, but these can not be observed directly in economic data. As such, policies are based on best estimates of the state of the economy, and methods that can provide these estimates are needed.

In this paper I estimate the Danish business cycles and trend growth from 1991 to 2024. I do this by estimating structural time series models using Bayesian methods. First I consider a linear model with a 2^{nd} order cycle formulation as introduced in Harvey and Trimbur (2003). Second I consider a non-linear Markov switching model as in Kim and Nelson (1999). I consider two versions of this model. One with a cycle assumed to follow an AR(2) process and one with a cycle assumed to be a 2^{nd} order cycle. I compare the estimated trends and cycles to those obtained from standard filters: the Hodrick-Prescott filter and the Butterworth filter.

I find that the filters and structural models generally result in similar trends and cycles. The period 1991-2008 is a calm period with small fluctuations in the cycle and steady trend growth, albeit 2001-2003 saw a period of stagnation in GDP. This stagnation is caused by a negative output gab in the HP-filter and linear structural model, but by slowing trend growth in the Butterworth filter and Switching models. 2009 and 2019-20 are periods with negative cycles, but in most models these cycles are short. In 2009 the output gab is closed be a stagnation in the trend, and in 2021 the output gab turns positive due to large increases in the cycles. In the switching model with an AR(2)-cycle the trend does not stagnate in 2009, and the economy remains below trend output until 2021 when the output gab is finally and quite suddenly closed and becomes positive.

2 Related Literature

To be done

3 State Space Models and Bayesian Methods

3.1 State Space Models and the Kalman Filter

In this section I provide a brief walk through of Gaussian linear state space models as presented in Harvey (1990). State Space Models (SSM) are models where an $N \times 1$ outcome vector today $\boldsymbol{y}_t$ is given by a set of underlying states $\boldsymbol{\alpha}_t$ through a *measurement equation* given by

$$\boldsymbol{y}_t = \boldsymbol{Z}\boldsymbol{\alpha}_t + \boldsymbol{d} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \boldsymbol{H}) \tag{1}$$

where $\boldsymbol{Z}$, $\boldsymbol{d}$ and $\boldsymbol{H}$ are non-stochastic. The state-vector $\boldsymbol{\alpha}_t$ are generally not observable, but follow a first-order Markov process resulting in a *transition equation*

$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} + \boldsymbol{c} + \boldsymbol{R}\boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathcal{N}(0, \boldsymbol{Q}) \tag{2}$$

such that the expected states tomorrow are independent of all other states than the current. $\boldsymbol{c}$, $\boldsymbol{R}$ and $\boldsymbol{Q}$ are non-stochastic and are together with $\boldsymbol{Z}$, $\boldsymbol{d}$ and $\boldsymbol{H}$ referred to as the *system matrices*. These matrices are given by the structure and parameters $\boldsymbol{\theta}$ of the underlying model. The State Space model is closed with two final assumptions:

- There is an initial state vector with the distribution $\boldsymbol{\alpha}_0 \sim \mathcal{N}(\boldsymbol{a}_0, \boldsymbol{P}_0)$.
- Disturbances $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$ are uncorrelated with each other and the initial state $\boldsymbol{\alpha}_0$ across all periods.

Once a model is formulated in this form it can be estimated using the *Kalman filter*.

The Kalman filter is a recursive algorithm which provides the mean square error minimizing estimates of the state vector at time t given the available information at time t. This information consists of the observed outcomes $\boldsymbol{y}_0, ..., \boldsymbol{y}_t$, together with the system matrices and the distribution of the initial states, $\boldsymbol{\alpha}_0 \sim \mathcal{N}(\boldsymbol{a}_0, \boldsymbol{P}_0)$, which are assumed to be known. The estimate of $\boldsymbol{\alpha}_t$ is denoted as $\boldsymbol{a}_t$, and $\boldsymbol{P}_t$ denotes the covariance matrix of the estimation error:

$$\boldsymbol{P}_{t-1} = E[(\boldsymbol{\alpha}_{t-1} - \boldsymbol{a}_{t-1})(\boldsymbol{\alpha}_{t-1} - \boldsymbol{a}_{t-1})'] \tag{3}$$

Given $\boldsymbol{a}_{t-1}$ and $\boldsymbol{P}_{t-1}$, the estimate of $\boldsymbol{\alpha}_t$ given available information at $t-1$ follows from the state space form at gives the *prediction equations*:

$$\boldsymbol{a}_{t|t-1} = \boldsymbol{T}\boldsymbol{a}_{t-1} + \boldsymbol{c} \tag{4}$$

$$\boldsymbol{P}_{t|t-1} = \boldsymbol{T}\boldsymbol{P}_{t-1}\boldsymbol{T}' + \boldsymbol{R}\boldsymbol{Q}\boldsymbol{R}' \tag{5}$$

Then at time t the observation of $\boldsymbol{y}_t$ becomes available and the state estimate is updated through the *updating equations*:

$$\mathbf{y}_{t|t-1} = \boldsymbol{Z}\boldsymbol{a}_{t|t-1} + \boldsymbol{d} \tag{6}$$

$$\boldsymbol{a}_t = \boldsymbol{a}_{t|t-1} + \boldsymbol{P}_{t|t-1}\boldsymbol{Z}'\boldsymbol{F}_t^{-1}(\boldsymbol{y}_t - \mathbf{y}_{t|t-1}) \tag{7}$$

$$\boldsymbol{P}_t = \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{Z}'\boldsymbol{F}_t^{-1}\boldsymbol{Z}\boldsymbol{P}_{t|t-1} \tag{8}$$

where

$$\boldsymbol{F}_t = \boldsymbol{Z}\boldsymbol{P}_{t|t-1}\boldsymbol{Z}' + \boldsymbol{H}$$

The updating equation intuitively works as follows: if $\boldsymbol{H}$ is large, observations are noisy, and the new observations $\boldsymbol{y}_t$ are weighted less when updating the state estimate, the the estimate relies more on the model structure. If $\boldsymbol{P}_{t|t-1}$ is large their is more uncertainty in the prediction $\boldsymbol{a}_{t|t-1}$ and the updated estimate relies more on the new observations.

Conditional on information available at time $t-1$ the estimate for $\boldsymbol{y}_t$ is given by $\mathbf{y}_{t|t-1}$ and the conditional variance is given by $\boldsymbol{F}_t$. As disturbances and the initial state vector are normal and independently distributed $\boldsymbol{y}_t$ is normal and independently distributed when conditioned on $\boldsymbol{y}_{t-1}, ...\boldsymbol{y}_1$. The likelihood function is then given by

$$L(\boldsymbol{\theta}|\boldsymbol{y}) = \prod_{t=1}^{T} p(\boldsymbol{y_t}|\boldsymbol{y}_{t-1}, ...\boldsymbol{y}_1) \tag{9}$$

$$\Leftrightarrow logL = -\frac{NT}{2}log2\pi - \frac{1}{2}\sum_{t=1}^{T} log|\boldsymbol{F}_t| - \frac{1}{2}\sum_{t=1}^{T} \boldsymbol{v}_t'\boldsymbol{F}_t^{-1}\boldsymbol{v}_t \tag{10}$$

where $\boldsymbol{v}$ are prediction errors $\boldsymbol{v}_t = \boldsymbol{y}_t - \mathbf{y}_{t|t-1}$. As such estimation of the model parameters $\boldsymbol{\theta}$ is possible by maximizing the likelihood function which consists of inputs obtained from running the Kalman filter.

When running the filter, the optimal estimate of the state vector at every point given information available at that point is obtained. To obtain more efficient estimates of the unobserved states at time t, the estimates can be conditioned on the entire information set and not just the information set available at time t. This is done with a *Kalman smoother*. There are multiple algorithms to obtain the smoothed estimates. I apply the *fixed-interval smoothing* algorithm, which is run after the Kalman filter, and iterates backwards through stored estimates of $\boldsymbol{a}_t$ and $\boldsymbol{P}_t$ as follows

$$\begin{aligned}
\boldsymbol{P}_t^* &= \boldsymbol{P}_t \boldsymbol{T}' \boldsymbol{P}_{t+1|t}^{-1} \\
\boldsymbol{a}_{t|T} &= \boldsymbol{a}_t + \boldsymbol{P}_t^* (\boldsymbol{a}_{t+1|T} - \boldsymbol{T} \boldsymbol{a}_t) \\
\boldsymbol{P}_{t|T} &= \boldsymbol{P}_t + \boldsymbol{P}_t^* (\boldsymbol{P}_{t+1|T} - \boldsymbol{P}_{t+1|t}) \boldsymbol{P}_t^{*\prime}
\end{aligned}$$

3.2 Bayesian Inference

Estimation and inference of unknown model parameters $\boldsymbol{\theta}$ in a Gaussian state space model is possible using maximum likelihood and the Kalman filter as seen above. It is also possible to obtain the optimal estimate of the unobserved states for every period through the Kalman smoother. The Kalman Smoother does not say anything about the underlying distribution of the states other than the mean, and inference on the obtained state estimates is therefore not possible. This challenge can be dealt with by applying Bayesian methods, which estimate the entire distribution of the unobserved states and parameters. Bayesian methods also make it possible to include prior knowledge about parameters in the estimation, which can be useful in complex models, where maximizing the likelihood function may be challenging, or leads to extreme estimates as discussed in Harvey et al. (2007).

When applying a Bayesian approach distributions are derived from Bayes' theorem:

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{y})} = \frac{L(\boldsymbol{\theta}|\boldsymbol{y})p(\boldsymbol{\theta})}{p(\boldsymbol{y})}$$

where $p(\boldsymbol{\theta}|\boldsymbol{y})$ is the *posterior* distribution and $p(\boldsymbol{\theta})$ is the *prior* distribution, given by a belief about the distribution before estimating the model. The prior can also be chosen to be uninformative, such that prior beliefs don't affect the estimated posterior. $L(\boldsymbol{\theta}|\boldsymbol{y})$ is the likelihood and can be obtained from the Kalman filter for Gaussian models. $p(\boldsymbol{y})$ is known as the *marginal likelihood* and given by $p(\boldsymbol{y}) = \int_{\boldsymbol{\Theta}} p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$. The marginal likelihood is not available analytically, so the posterior distribution is unknown. To estimate the posterior distribution *Markov Chain Monte Carlo (MCMC)* methods can be used.

3.2.1 Markov Chain Monte Carlo

Markov Chain Monte Carlo is a broader class of algorithms that create a Markov chain of draws of $\boldsymbol{\theta}$. That is, the distribution of the current draw is given only be the previous draw. After a sufficient amount of draws, the distribution of the draws will match the true posterior probability distribution, which can be used for probabilistic inference. The coefficient estimates are then simply given by the mean of the draws and the 95% credible interval is between the 2.5th and 97.5th percentile of the draws. To obtain the Markov chain of draws the algorithms Metropolis-Hastings sampling and Gibbs sampling are typically used.

3.2.2 Metropolis-Hastings Sampling

In the Metropolis-Hastings algorithm a Markov chain is constructed by by drawing a proposed sample $\boldsymbol{\theta}_i^p$ based only on the previous sample $\boldsymbol{\theta}_{i-1}$ through a Markov transition density function $q(\boldsymbol{\theta}_i^p|\boldsymbol{\theta}_{i-1}) = q(\boldsymbol{\theta}_i^p|\boldsymbol{\theta}_{i-1}, ..., \boldsymbol{\theta}_0)$. The proposed draw is then accepted setting $\boldsymbol{\theta}_i = \boldsymbol{\theta}_i^p$ with probability γ, and rejected setting $\boldsymbol{\theta}_i = \boldsymbol{\theta}_{i-1}$ with probability $1-\gamma$ where:

$$\gamma = \min\left(1, \frac{p(\boldsymbol{\theta}_i^p|\boldsymbol{y})q(\boldsymbol{\theta}_{i-1}|\boldsymbol{\theta}_i^p)}{p(\boldsymbol{\theta}_{i-1}|\boldsymbol{y})q(\boldsymbol{\theta}_i^p|\boldsymbol{\theta}_{i-1})}\right) \tag{11}$$

A better draw is therefore always accepted and a worse draw is accepted with some probability given by the relative fit of the draws. The marginal likelihood $p(\boldsymbol{y})$ cancels out of the fraction, and as such the probability of keeping a proposed draw is given by the likelihoods $L(\boldsymbol{\theta}|\boldsymbol{y})$, the priors $p(\boldsymbol{\theta})$ and the transition densities q which are all available. A common transition density is a random walk such that $\boldsymbol{\theta}_i^p = \boldsymbol{\theta}_{i-1} + \epsilon_i$. If ϵ is normally distributed, then $q(\boldsymbol{\theta}_i^p|\boldsymbol{\theta}_{i-1}) = q(\boldsymbol{\theta}_{i-1}|\boldsymbol{\theta}_i^p)$ and the transition densities cancel out of the fraction as well.

3.2.3 Gibbs sampling

Another method to obtain a Markov chain of draws is by Gibbs sampling which can be used if the *conditional distributions* of the elements of $\boldsymbol{\theta}$ are known. $\boldsymbol{\theta}$ consists of k unknowns, such that $\boldsymbol{\theta} = \{\theta^{(1)}, ..., \theta^{(k)}\}$. The conditional distributions are then

$$\begin{gathered} p(\theta^{(1)}|\boldsymbol{y}, \theta^{(2)}, \dots, \theta^{(k)}) \\ \vdots \\ p(\theta^{(k)}|\boldsymbol{y}, \theta^{(1)}, \dots, \theta^{(k-1)}) \end{gathered}$$

A Markov chain of draws is then achieved by sampling one θ-element at a time from the conditional distribution conditioned on all other newest draws:

$$\begin{aligned} \theta_i^{(1)} &\sim p(\theta_i^{(1)}|\boldsymbol{y}, \theta_{i-1}^{(2)}, \theta_{i-1}^{(3)}, \dots, \theta_{i-1}^{(k)}) \\ \theta_i^{(2)} &\sim p(\theta_i^{(2)}|\boldsymbol{y}, \theta_i^{(1)}, \theta_{i-1}^{(3)}, \dots, \theta_{i-1}^{(k)}) \\ &\vdots \\ \theta_i^{(k)} &\sim p(\theta_i^{(k)}|\boldsymbol{y}, \theta_i^{(1)}, \theta_i^{(2)}, \dots, \theta_i^{(k-1)}) \end{aligned}$$

Metropolis-Hastings and Gibbs sampling can also be combined if the conditional distributions are only available for some subset of $\boldsymbol{\theta}$. Some subset of $\boldsymbol{\theta}_i$ is drawn using Metropolis-Hastings, and the rest are then drawn from their conditional distributions conditional on the newest draw of each θ.

3.2.4 Sampling Unknown Parameters and Unobserved States

In state space models the unknown model parameters $\boldsymbol{\theta}$ can be sampled using Metropolis-Hastings. For each draw of $\boldsymbol{\theta}$, the unobserved state vector $\boldsymbol{\alpha}$ can be sampled using Gibbs sampling from the conditional distribution $p(\boldsymbol{\alpha}|\boldsymbol{\theta}_i, \boldsymbol{y})$. This is done using algorithms called *simulation smoothers*. The commonly used algorithm is presented in Durbin and Koopman (2002) and is given by:

1. Draw $\boldsymbol{\alpha}^+$ and $\boldsymbol{y}^+$ as:

(a) Initialize $\boldsymbol{\alpha}_1^+ \sim N(\mathbf{0}, \boldsymbol{P}_1)$.

(b) Iterate forwards through 4 and 6 to generate $\boldsymbol{\alpha}^+$ and $\boldsymbol{y}^+$.

2. Construct an artificial series $\boldsymbol{y}^* = \boldsymbol{y} - \boldsymbol{y}^+$.

3. Compute $\hat{\boldsymbol{\alpha}}^* = E(\boldsymbol{\alpha}|\boldsymbol{y}^*)$ by applying the Kalman smoother to $\boldsymbol{y}^*$.

4. Compute $\tilde{\boldsymbol{\alpha}} = \hat{\boldsymbol{\alpha}}^* + \boldsymbol{\alpha}^+$.

5. Then $\tilde{\boldsymbol{\alpha}}$ is a draw from $p(\boldsymbol{\alpha}|\boldsymbol{\theta}, \boldsymbol{y})$.

As the draws of $\boldsymbol{\theta}$ match the true distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$, $\boldsymbol{\theta}$ is integrated out after enough draws: $p(\boldsymbol{\alpha}|\boldsymbol{y}) = \int_\Theta p(\boldsymbol{\alpha}|\boldsymbol{\theta}, \boldsymbol{y})p(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}$. When starting the Markov chain an initial $\boldsymbol{\theta}_0$ is chosen arbitrarily, and as such it can take some time for the chain to converge towards the true distribution. Therefore initial draws of θ are discarded. This is called the burn-in period. After the burn-in period the chain of draws should have converged and samples of $\boldsymbol{\theta}$ reflect the true distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$ and draws of $\boldsymbol{\alpha}$ reflect $p(\boldsymbol{\alpha}|\boldsymbol{y})$. As such the entire distribution of the state vector for every period t is obtained when sampling the states, whereas only the mean of the distribution was obtained when directly applying the Kalman smoother.

4 A Gaussian Structural Time Series Model

In this section I estimate a univariat structural time series model as in Harvey et al. (2007) on Danish GDP data, with cycles and trend being unobserved states. I compare the estimated cycles and trend with estimates obtained from non-structural filters: the Hodrick-Prescott filter and the Butterworth filter.

4.1 The Model

Obervations y_t are given by:

$$y_t = u_t + \psi_{n,t} + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon)$$

The stochastic trend is an integrated random walk:

$$\begin{aligned} u_t &= u_{t-1} + \beta_{t-1} \\ \beta_t &= \beta_{t-1} + \xi_t, \quad \xi_t \sim \mathcal{N}(0, \sigma_\xi) \end{aligned}$$

The n^{th} order cycle is given by:

$$\begin{pmatrix} \psi_{n,t} \\ \psi^*_{n,t} \end{pmatrix} = \rho \begin{pmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{pmatrix} \begin{pmatrix} \psi_{n,t-1} \\ \psi^*_{n,t-1} \end{pmatrix} + \begin{pmatrix} \psi_{n-1,t-1} \\ \psi^*_{n-1,t-1} \end{pmatrix}$$

$$\begin{pmatrix} \psi_{1,t} \\ \psi^*_{1,t} \end{pmatrix} = \rho \begin{pmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{pmatrix} \begin{pmatrix} \psi_{1,t-1} \\ \psi^*_{1,t-1} \end{pmatrix} + \begin{pmatrix} \kappa_t \\ \kappa^*_t \end{pmatrix}$$

with $\kappa_t \sim \mathcal{N}(0, \sigma_\kappa)$ and $\kappa^*_t \sim \mathcal{N}(0, \sigma_\kappa)$. I always apply the 2nd order cycle $n = 2$.

In this model ϵ are noise shocks with no persistence, κ are transitory shocks affecting the cycle, and ξ are permanent shocks affecting the trend. The model can be formulated in the state space form, and system matrices are then obtained. This is done in appendix A.1 for a 2nd order cycle.

4.2 Estimating the Model

To estimate the model I follow the approach used in Hasenzagl et al. (2022). This involves an initialization phase and a recursion phase. In the initialization phase unknown parameters are drawn using the Metropolis-Hastings algorithm. I construct a chain of 40.000 draws, and discard the first 20.000. I then compute the variance-covariance matrix of the accepted draws, Σ. In the recursion phase parameter proposals are then based on Σ to achieve more efficient draws. Again I take 40.000 draws, and after a burn-in period of 20.000 I use the simulation smoother from Durbin and Koopman (2002) to obtain samples of the states. The parameters are bounded in their support, which creates bounds on which draws can be made. To handle this I transform the parameters to have unbounded support, and then draw from the unbounded parameters.

4.2.1 Initialization

For s = 1,..,40,000 draws:

Draw unbounded candidate vector Γ^* from multivariate normal distribution with mean Γ_{s-1} and variance-covariance ωI where ω is a scaling constant to get acceptance rate around 25-35 pct.

New draw is accepted with probability γ

$$\Gamma_s = \begin{cases} \Gamma_* & \text{with probability } \gamma \\ \Gamma_{s-1} & \text{with probability } 1-\gamma \end{cases}$$

where

$$\eta = min\left(1, \frac{p(y|\theta^*)p(\theta^*)J(\Gamma^*)}{p(y|\theta_{s-1})p(\theta_{s-1})J(\Gamma_{s-1})}\right)$$

Where $J(\Gamma)$ are the Jacobians of the transformed variables: $\frac{d\theta}{d\Gamma}$.

First half of the initialization draws are discarded

4.2.2 Recursion

for q = 1,..., 40,000 draws:

Set Σ to sample covariance of the chain Γ_s from the initialization step. A candidate vector for the unbounded paramters Γ^* is drawn form a multivariate distribution with mean Γ_{q-1} and variance $\omega\Sigma$ with ω being a scaling constant to get 25-35 acceptance rate. New draw is accepted as in the initialization. Discard draws from burn in period.

For $q > 20,000$: After burn-in period sample unobserved states α using simulation smoother from Durbin and Koopman (2002)

4.2.3 Priors and Jacobians

Draws of θ are drawn from the prior distribution specified by the researcher. I use uninformative priors w.r.t the support of the underlying parameter. A variance must be positive, ρ must be lower than 1 and λ must be lower than π for the cycle to be staionary.

Name	Support	Density	Parameter 1	Parameter 2
σ^2	$(0, \infty)$	Inverse-Gamma	1e-6	1e-6
ρ	$[0.001, 0.970]$	Uniform	0.001	0.970
λ	$[0.001, \pi]$	Uniform	0.001	π

Table 1: Prior Distributions

To get unbounded support the following transformations are used:

$$\begin{aligned} \Gamma^{U,\beta} &= \ln\left(\frac{\theta^{U,\beta} - a}{b - \theta^{U,\beta}}\right) \\ \Gamma^{IG} &= \ln(\theta^{IG} - a) \end{aligned}$$

where a is the lower bound and b is the upper bound of their support. The inverse of the transformations are then

$$\begin{aligned} \theta^{U,\beta} &= \frac{a + bexp(\Gamma^{U,\beta})}{1 + \exp(\Gamma^{U,\beta})} \\ \theta^{IG} &= \exp(\Gamma^{IG}) + a \end{aligned}$$

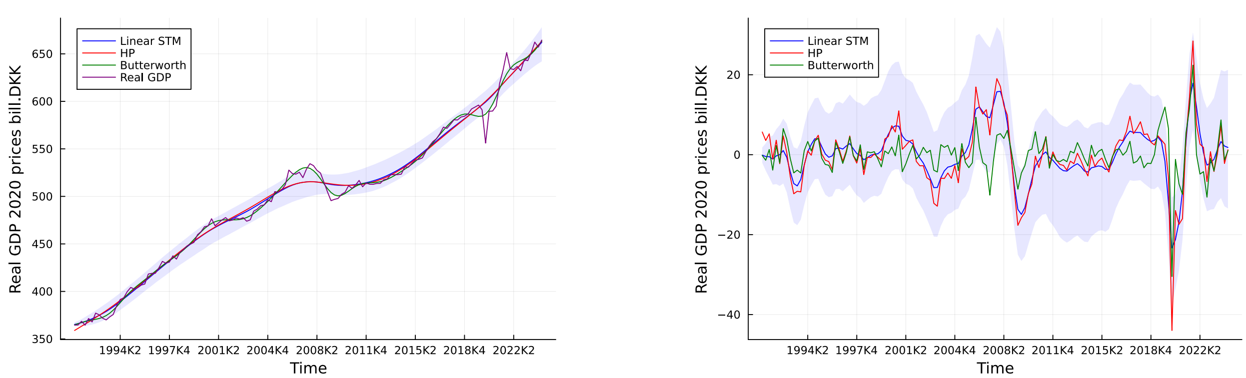
and the log of the derivatives are

$$\begin{aligned} \ln\left(\frac{d\theta^{U,\beta}}{d\Gamma^{U,\beta}}\right) &= \ln(b - a) + \Gamma^{U,\beta} - 2\ln(1 + \exp(\Gamma^{U,\beta})) \\ \ln\left(\frac{d\theta^{IG}}{d\Gamma^{IG}}\right) &= \Gamma^{IG} \end{aligned}$$

4.3 Results

I estimate the model on real quarterly Danish GDP data covering the period 1991Q1-2024Q2, and thus a total of 134 observations. The data is obtained from Danmarks Statistik table NKN1. The results are seen in figure 1, where the estimated trend and cycle is plotted together with trends and cycles from the non-structural Hodrick-Prescott filter and the Butterworth filter (described in Appendix B). The estimated parameter values and distributions are seen in Appendix C along with trace plots that show the draws of the Markov chain. The chain is stationary and not autocorrelated, which indicates succesful convergence to the true posterior distribution. The results show that the estimated states from the structural model follow the trend and cycles obtained from the HP-filter closely, while the Butterworth filter deviates from the two others especially around 2007. This is slightly surprising as the higher order cycle model converges to a Butterworth filter when the order of the cycle increases, and is as such a special case of the Butterworth filter as mentioned in Harvey and Trimbur (2003). Grant and Chan (2017) show that the HP-filter is a special case of a structural time series model as above, but with the restriction that the cyclical components are uncorrelated. This restriction implies fixing $\rho = 0$ in the model above. Grant and Chan (2017) use this as an argument against using the HP-filter as cyclical components are found to be correlated in data. This is also the case in my estimation, where the distribution of ρ is centered nicely around 0.7, with very little mass below 0.5 as seen in Appendix C.

When looking at the credible interval of the trend and cycle estimates, the cycle is only significantly different from 0 at the 95% confidence level in 2008-2009 and 2019-2020. In 2008 before the financial crisis the cycle is significantly above 0, and in 2009 as the financial crisis hits, the cycle is significantly below zero. In 2019 the cycle is signficantly below zero as the country enters lockdown during Covid-19, and significantly above zero in 2021 when the country and economy reopens. The cycle is generally negative in the early 2000s, but the slowdown could also be due to a slight decrease in trend growth. It is also striking how the financial crisis creates a period of multiple year with no to slightly decreasing trend growth until 2012 where trend growth resumes.



(a) Estimated Trend

(b) Estimated Cycle

Figure 1: Unobserved States in Model and Standard Filters

5 A Non-linear Markov-switching Model

Now I consider a Markov Switching State Space Model where the economy can switch between structural compositions. This could be times of high volatility or recession when $S_t = 1$ and normal times $S_t = 0$ as in Kim and Nelson (1999). Observations y_t are given by:

$$y_t = u_t + \psi_{n,t}$$

The stochastic trend is an integrated random walk:

$$\begin{aligned} u_t &= u_{t-1} + \beta_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma_{\epsilon,S_t}) \\ \sigma_{\epsilon,S_t} &= \sigma_{\epsilon,0}(1 - S_t) + \sigma_{\epsilon,1} S_t \\ \beta_t &= \beta_{t-1} + \xi_t, \quad \xi_t \sim \mathcal{N}(0, \sigma_\xi) \end{aligned}$$

The cycle is either given by an 2^{nd} order cycle as in Harvey et al. (2007) or by an AR(2) process as in Kim and Nelson (1999).

The n^{th} order cycle is given by:

$$\begin{pmatrix} \psi_{n,t} \\ \psi^*_{n,t} \end{pmatrix} = \rho \begin{pmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{pmatrix} \begin{pmatrix} \psi_{n,t-1} \\ \psi^*_{n,t-1} \end{pmatrix} + \begin{pmatrix} \psi_{n-1,t-1} \\ \psi^*_{n-1,t-1} \end{pmatrix}$$

$$\begin{pmatrix} \psi_{1,t} \\ \psi^*_{1,t} \end{pmatrix} = \rho \begin{pmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{pmatrix} \begin{pmatrix} \psi_{1,t-1} \\ \psi^*_{1,t-1} \end{pmatrix} + \begin{pmatrix} \omega S_t \\ \omega S_t \end{pmatrix} + \begin{pmatrix} \kappa_t \\ \kappa^*_t \end{pmatrix}$$

with $\kappa_t \sim \mathcal{N}(0, \sigma_{\kappa,S_t})$ and $\kappa^*_t \sim \mathcal{N}(0, \sigma_{\kappa,S_t})$ where $\sigma_{\kappa,S_t} = \sigma_{\kappa,0}(1 - S_t) + \sigma_{\kappa,1} S_t$.

The AR(2) process is given by

$$\psi_t = \phi_1 \psi_{t-1} + \phi_2 \psi_{t-2} + \omega S_t + \kappa_t$$

S_t evolves according to a first order Markov process:

$$p(S_t = 1|S_{t-1} = 1) = p$$
$$p(S_t = 0|S_{t-1} = 0) = q$$

In this model there are no longer disturbances in the observation, but rather in the trend. ϵ_t affects u_t, and not y_t directly as in the previous model. As such ϵ_t becomes a permanent shock, where it was just noise in the observations in the previous model, with no persistency.

5.1 Estimating the Model

I run the same initialization and recursion algorithm used in Hasenzagl et al. (2022) to estimate the model, with the priors given in 2. I now use an informative prior on σ^2_{ϵ,S_t}. I do this because the chain otherwise included some extreme draws of the parameter, and all variation in GDP was matched by variation in the trend, eliminating all cycles.

Name	**Support**	**Density**	**Parameter 1**	**Parameter 2**
σ^2_ϵ	$(0, \infty)$	Inverse-Gamma	3	1
σ^2	$(0, \infty)$	Inverse-Gamma	1e-6	1e-6
ρ	$[0.001, 0.970]$	Uniform	0.001	0.970
λ	$[0.001, \pi]$	Uniform	0.001	π
ψ	$[0.001, 0.99]$	Uniform	0.001	0.99
ω	$(-\inf, \inf)$	Uniform	-100	0

Table 2: Prior Distributions

Estimating the model with the MCMC algorithm is straight forward when the Kalman filter and smoother can be used. The switching structure of the model prevents this. As such I follow Kim and Nelson (1999), and implement an approximate Kalman filter.

5.1.1 The Filter

Conditional on $S_{t-1} = i$ and $S_t = j$ the Kalman filter is

$$\begin{aligned}
a_{t|t-1}^{i,j} &= Ta_{t-1}^{i} + c_t^{j} \\
P_{t|t-1}^{i,j} &= TP_{t-1}^{i}T' + RQ^{j}R' \\
v_t^{i,j} &= y_t - Za_{t|t-1}^{i,j} \\
F_t^{i,j} &= ZP_{t|t-1}^{i,j}Z' + H \\
a_t^{i,j} &= a_{t|t-1}^{i} + P_{t|t-1}^{i,j}Z'(F_t^{i,j})^{-1}v_t^{i} \\
P_t^{i,j} &= P_{t|t-1}^{i,j} - P_{t|t-1}^{i,j}Z'(F_t^{i,j})^{-1}ZP_{t|t-1}^{i,j}
\end{aligned}$$

If I iterate through the filter, I have to consider the entire history of S. As S can take two values: 0 and 1, I would have to consider 2^T cases, which is not feasible. As such after every iteration I reduce the dimension of $a_t^{i,j}$ and $P_t^{i,j}$ at each iteration from 2×2 matrices to 2×1 vectors thus obtaining a_t^j and P_t^j as done in Kim and Nelson (1999). This is done using the following approximations:

$$a_t^j = \frac{\sum_{i=0}^{1} p(S_{t-1} = i, S_t = j|\Psi_t)a_t^{i,j}}{p(S_t = j|\Psi_t)}$$

$$P_t^j = \frac{\sum_{i=0}^{1} p(S_{t-1} = i, S_t = j|\Psi_t)\left(P_t^{i,j} + (a_t^j - a_t^{i,j})(a_t^j - a_t^{i,j})'\right)}{p(S_t = j|\Psi_t)}$$

where Ψ_t represent information available at time t. The probabilities are given by:

$$p(S_t = j|\Psi_t) = \sum_{i=0}^{1} p(S_{t-1} = i, S_t = j|\Psi_t)$$

and

$$p(S_{t-1} = i, S_t = j|\Psi_t) = \frac{p(y_t|S_{t-1} = i, S_t = j, \Psi_{t-1}) \cdot p(S_{t-1} = i, S_t = j|\Psi_{t-1})}{p(y_t|\Psi_{t-1})}$$

where

$$
\begin{aligned}
p(y_t|S_{t-1}=i,S_t=j,\Psi_{t-1}) &= \frac{1}{\sqrt{2\pi F_t^{i,j}}} exp\left(-\frac{(v_t^i)^2}{2F_t^{i,j}}\right) \\
p(y_t|\Psi_{t-1}) &= \sum_{i=0}^{1}\sum_{j=0}^{1} p(y_t, S_{t-1}=i, S_t=j|\Psi_{t-1}) \\
&= \sum_{i=0}^{1}\sum_{j=0}^{1} p(y_t|S_{t-1}=i,S_t=j,\Psi_{t-1})\cdot p(S_{t-1}=i,S_t=j|\Psi_{t-1})
\end{aligned}
$$

and

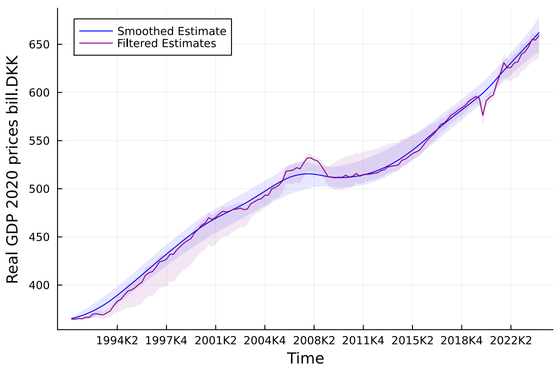
$$
p(S_{t-1}=i,S_t=j|\Psi_{t-1}) = p(S_t=j|S_{t-1}=i)\cdot p(S_{t-1}=i|\Psi_{t-1})
$$

Where $p(S_{t-1}=i|\Psi_{t-1})$ is obtained in the previous iteration. In the initial iteration set $p(S_0=0|\Psi_0)=\frac{1-p}{2-p-q}$ and $p(S_0=1|\Psi_0)=\frac{1-q}{2-p-q}$. The log-likelihood is then obtained through the iterations and given by

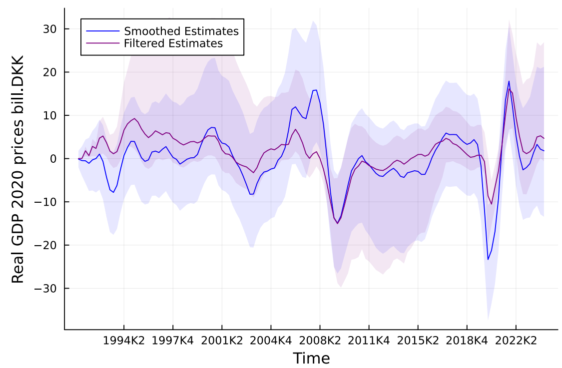
$$
logL = \sum_{t=1}^{T} log\left(p(y_t|\Psi_{t-1})\right)
$$

Smoothed estimates can no longer be obtained with the Kalman smoother due to the non-linearity caused by the switching structure. As such the simulation smoother from Durbin and Koopman (2002) can not be used to sample the unobserved states. Instead I simply use the filtered state estimates from every draw of $\boldsymbol{\theta}$. To obtain the estimated states at time t for every draw I compute $a_t=\sum_{j=0}^{1}\sum_{i=0}^{1} p(S_{t-1}=i,S_t=j|\Psi_t)a_t^{i,j}$. As discussed, the Kalman filter computes the optimal estimate of the states at time t given information available at time t, which results in a significant precession loss compared to the smoothed estimates. When sampling the estimated states from the filter at every iteration I obtain $p(\boldsymbol{a}_{t|t}|\boldsymbol{y})$ instead of $p(\boldsymbol{\alpha}_t|\boldsymbol{y})$ As I obtain estimates of the mean of the state, and not draws from the distribution of the state, I am unable to do inference on the unobserved states, but can only say something about the certainty of the estimate of the mean of the states. Using the filtered state estimates conditioned on the smaller information set has non-negligible effects, which is seen in figure 6, where the state estimates from the linear model obtained with the simulation smoother are compared to estimated obtained with the Kalman filter.

The trend obtained from the smoothed estimates is a lot smother, while the trend from the filtered estimates fluctuates mores, and 'swallows' small variations in GDP.



(a) Estimated Trend



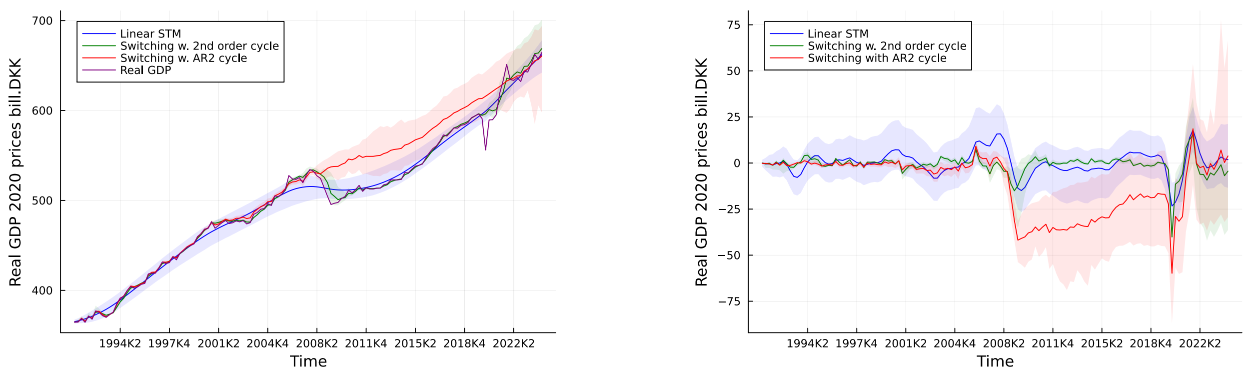
(b) Estimated Cycle

Figure 2: Smoothed VS Filtered State Estimates in Linear Model

5.2 Results

The estimated parameter values and distributions are seen in Appendix C along with trace plots that show the draws of the Markov chain. The chains are stationary, and not to auto-correlated, but more than the case in the linear model. The posteriors of Σ_ϵ^2 follow the prior distribution closely indicating that these parameters are not very well identified in the data. The parameter estimates show that when σ_ξ^2 are similar across both cycle formulations and also similar to the estimate from the linear model. When $S_t = 0$, σ_κ is similar to or lower than that of the linear model, but when $S_t = 1$ σ_κ is a lot higher. ω is a parameter to allow for deeper recessions than booms, but it is not significantly different from 0. As such $S_t = 0$ can be seen as normal times, and $S_t = 1$ can be seen as time with large fluctuations caused by large disturbances to the cycle, which are both positive and negative.

The estimated trend, cycles, and probability of being in the volatile state $S_t = 1$ are shown in figures 3 and 4 below. The relative deviation between observations and the estimated trends are shown in 7 in Appendix D. In these figures credible interval bands are plotted for all models, but for the switching models, they don't resemble the distribution of the state, but rather the distribution of the estimated means of the state. As such the error bands for the linear model shows the range where the model believes the true state at that time is distributed, whereas the error bands for the switching models show where the model believes the mean of the true state distribution to be. The most noticeable result is that the switching model with the AR2 process results in a very different trend, although the uncertainty is large. The trend in this model does not experience the long period of stagnation during in 2009-2012, but continues growing, resulting in a large negative output gap which persists until 2021. The switching model with the 2nd order cycle follows the linear model more closely, but with a non-smooth trend, resulting especially in smaller booms in the cycle compared to the linear model, while busts in the cycles are comparable. Figure 4 shows that the switching models assign high probabilities of being in the volatile state in 2009, 2019 and 2021, but also in 2006, and to some degree in 2001. In 2006 and 2021 the higher volatility in the cycle is caused by large positive transitory shocks, and in the other cases by large negative transitory shocks.



(a) Estimated Trend

(b) Estimated Cycle

Figure 3: Unobserved States in Linear and Switching Models

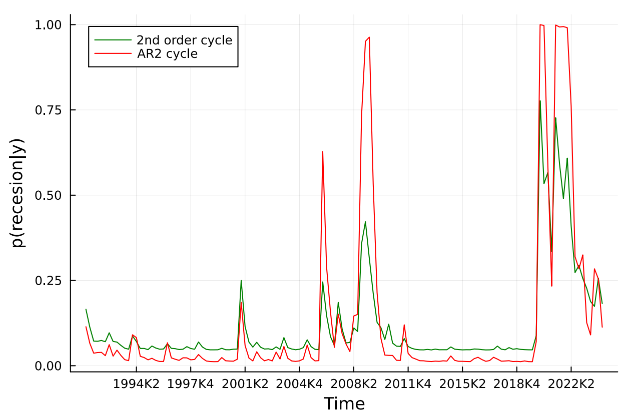


Figure 4: Model Implied Probability of Being in Volatile State

6 Discussion

6.1 Implication of Results

Generally the models and filters show an increasing trend, but with a long period of stagnation from 2009-2012, whereas the switching AR(2)-model shows a stably increasing trend but a prolonged period with a negative output gap from 2009 to 2021. The policy implications are very different in these two cases. In the first case policy makers should focus on the cause of the stagnation in the trend, and conduct policies that encourage innovation and growth. In the second case policy makers should focus on stimulating the economy (as done in 2020-2022) and return to the trend level quickly, as time spent beneath the trend output is a large and seemingly unnecessary welfare cost. As such the second case appeals to economists and policy makers looking back at the 2010s who call the tight fiscal policy of the period a big mistake, whereas the first case of a slowing trend appeals to economists leaning towards secular stagnation.

6.2 Advantages of Structural Cycle Estimation

My results show that simple filters such as the Hp and Butterworth filters generate trends and cycles not far from those generated from the more demanding linear structural model. Structural models do however allow for much more flexibility. I propose an univariat model to map the business cycle, but a multivariate much larger model can be implemented in the exact same way which is done in Hasenzagl et al. (2022). Then fluctuations in GDP can be linked to fluctuations in other variables such as inflation and employment, and these links can be studied. The structural model also allows for regime-switching as done in this paper. Implementing these switching models and estimating them reliably turned out to be difficult.

6.3 Estimation Challenges in the Non-linear Case

When allowing for regime switching the model is no longer Gaussian, but conditional on the current and latest regime it is Gaussian, which was utilized to be able to estimate the model. I validated this approach on simulated data where parameters and states where estimated reliably. In the real data some parameters where difficult to estimate, and the states seem highly affected by not being estimated using the entire information set available.

How to estimate switching models has become a large area of research, with one popular new approach being the use of particle filters in a MCMC algorithm as proposed in Andrieu et al. (2010). Such algorithms called PMCMC can be used to estimate switching models as in this paper as discussed in Whiteley et al. (2010), allowing for efficient estimation of parameters and sampling of the unobserved states. The routine is quite complicated and I have not been able to do it yet.

7 Conclusion

In this paper I have applied filters and structural models to map the Danish business cycle from 1991 to 2024. The filters and models generally show that the period 1991-2008 is a calm period with small fluctuations in the cycle and steady trend growth, albeit 2001-2003 saw a period of stagnation in GDP. This stagnation is caused by a negative output gab in the HP-filter and linear structural model, but by slowing trend growth in the Butterworth filter and Switching models. 2009 and 2019-20 are periods with negative cycles, but in most models these cycles are short. In 2009 the output gab is closed be a stagnation in the trend, and in 2021 the output gab turns positive due to large increases in the cycles. In the switching model with an AR(2)-cycle the trend does not stagnate in 2009, and the economy remains below trend output until 2021 when the outputgab is finally and quite suddenly closed and becomes positive.

A State Space Form

A.1 The Gaussian Structural Time Series Model

The model can be written in the state space form with the state vector α (following Harvey 1989):

$$y_t = Z\alpha_t + \epsilon_t, \quad Var(\epsilon_t) = H$$
$$\alpha_t = T\alpha_{t-1} + R\nu_t, \quad Var(\nu_t) = Q$$

The state vector α_t is:

$$\alpha_t = \begin{pmatrix} \mu_t \\ \beta_t \\ \psi_{1,t} \\ \psi^*_{1,t} \\ \psi_{2,t} \\ \psi^*_{2,t} \end{pmatrix}$$

The disturbances are

$$\nu_t = \begin{pmatrix} \xi_t \\ \kappa_t \\ \kappa^*_t \end{pmatrix}$$

:

Matrices/vectors/numbers Z, H, T, R and Q are referred to as system matrices. They are assumed non-stochastic. Putting the model in state space form with a second order cycle yields the following system matrices:

$$Z = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

The transition matrix T is:

$$T = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \rho\cos\lambda_c & \rho\sin\lambda_c & 0 & 0 \\ 0 & 0 & -\rho\sin\lambda_c & \rho\cos\lambda_c & 0 & 0 \\ 0 & 0 & 1 & 0 & \rho\cos\lambda_c & \rho\sin\lambda_c \\ 0 & 0 & 0 & 1 & -\rho\sin\lambda_c & \rho\cos\lambda_c \end{pmatrix}$$

The matrix R that links the disturbances to the states is:

$$R = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

The disturbance variance-covarinace:

$$Q = \begin{pmatrix} \sigma_\xi^2 & 0 & 0 \\ 0 & \sigma_\kappa^2 & 0 \\ 0 & 0 & \sigma_\kappa^2 \end{pmatrix}$$

$$H = \sigma_\epsilon^2$$

B Non-model based Filters

to be done

B.1 The Hodrick-Prescott Filter

B.2 The Butterworth Filter

C Estimated Parameters in Structural Models

C.1 Parameter Estimates

Table 3: Parameter Estimates

Parameter	**1**	**2**	**3**
ρ	0.69	0.21	
λ_c	0.25	0.33	
σ^2_ξ	0.37	0.68	0.04
σ^2_κ	8.65		
$\sigma^2_{\kappa,0}$		9.37	14.60
$\sigma^2_{\kappa,1}$		384.79	310.56
σ^2_ϵ	14.84		
$\sigma^2_{\epsilon,0}$		0.50	0.47
$\sigma^2_{\epsilon,1}$		0.61	0.46
ω		-7.80	-4.64
p		0.78	0.75
q		0.96	0.96
ψ_1			0.76
ψ_2			0.22

C.2 Posteriors Distributions and Trace Plots

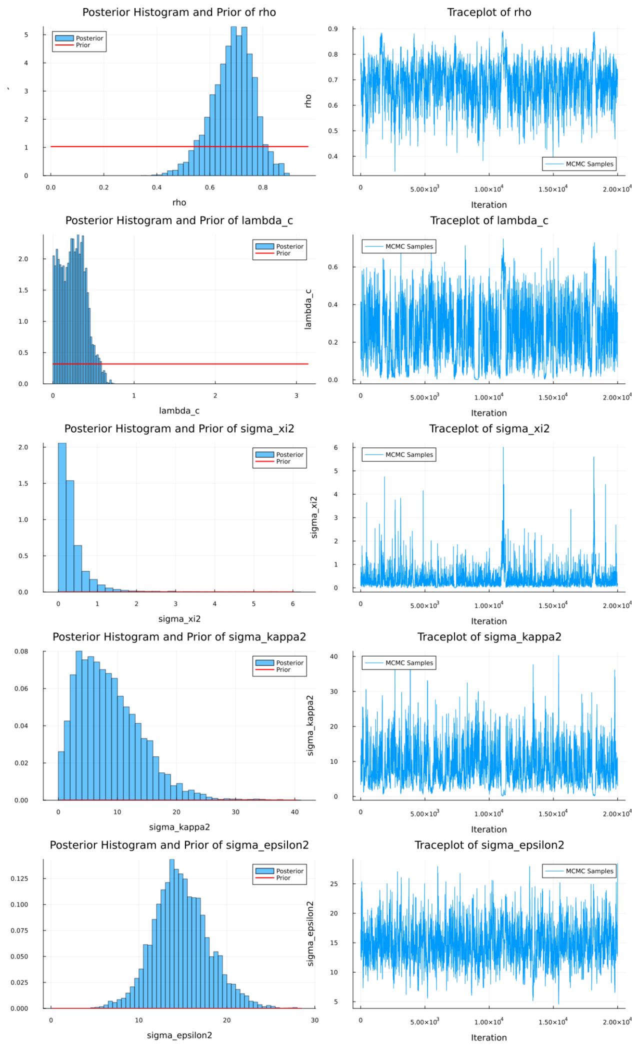
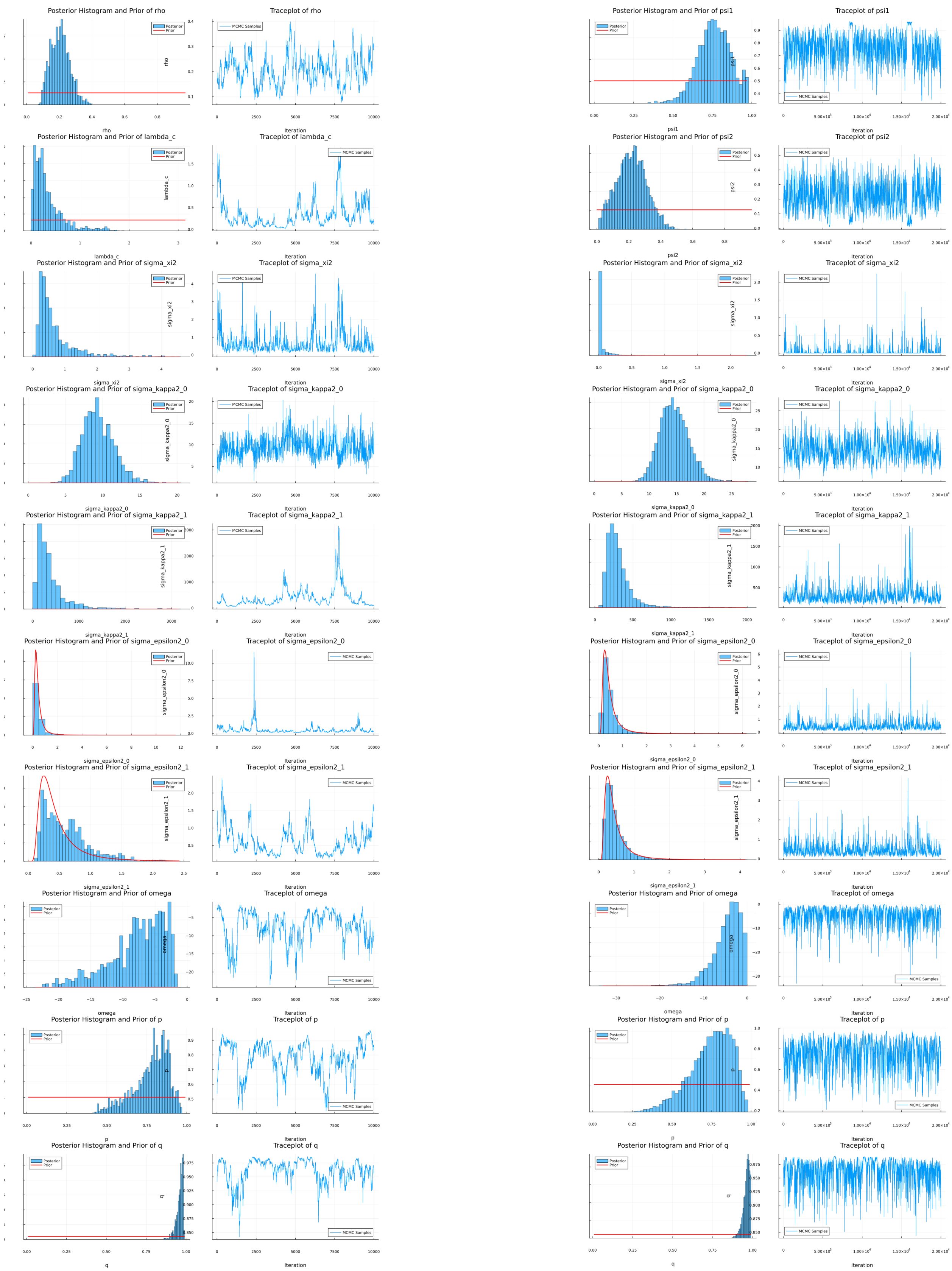


Figure 5: Posteriors from Linear Model



(a) 2nd Order Cycle

(b) AR2 Cycle

Figure 6: Posteriors from Switching Models

D Relative Cycles

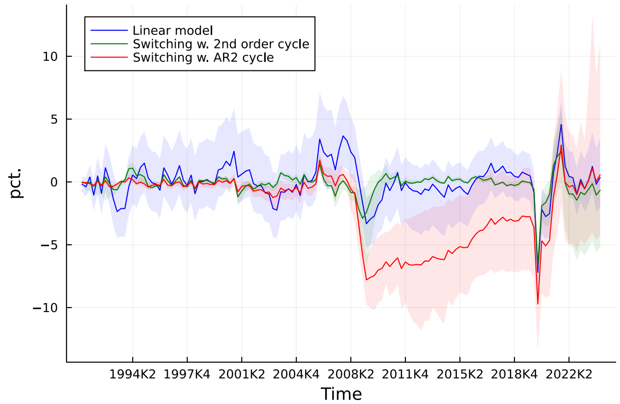


Figure 7: Deviations from Trend

References

Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 72(3):269–342.

Durbin, J. and Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3):603–616.

Grant, A. L. and Chan, J. C. (2017). Reconciling output gaps: Unobserved components model and hodrick–prescott filter. *Journal of Economic Dynamics and Control*, 75:114–121.

Hamilton, J. D. (2018). Why you should never use the hodrick-prescott filter. *Review of Economics and Statistics*, 100(5):831–843.

Harvey, A. C. (1990). Forecasting, structural time series models and the kalman filter, chapter 3.

Harvey, A. C. and Trimbur, T. M. (2003). General model-based filters for extracting cycles and trends in economic time series. *Review of Economics and Statistics*, 85(2):244–255.

Harvey, A. C., Trimbur, T. M., and Van Dijk, H. K. (2007). Trends and cycles in economic time series: A bayesian approach. *Journal of Econometrics*, 140(2):618–649.

Hasenzagl, T., Pellegrino, F., Reichlin, L., and Ricco, G. (2022). A model of the fed's view on inflation. *Review of Economics and Statistics*, 104(4):686–704.

Jarociński, M. (2015). A note on implementing the durbin and koopman simulation smoother. *Computational Statistics & Data Analysis*, 91:1–3.

Kim, C.-J. and Nelson, C. R. (1999). Friedman's plucking model of business fluctuations: tests and estimates of permanent and transitory components. *Journal of Money, Credit and Banking*, pages 317–334.

Phillips, P. C. and Shi, Z. (2021). Boosting: Why you can use the hp filter. *International Economic Review*, 62(2):521–570.

Whiteley, N., Andrieu, C., and Doucet, A. (2010). Efficient bayesian inference for switching state-space models using discrete particle markov chain monte carlo methods. *arXiv preprint arXiv:1011.2437*.