State Space Time Series Analysis: Various applications

A. Kahali (2542764)
L. Scholte (2500114)

Vrije Universiteit Amsterdam (FEWEB)

Time Series Econometrics - MSc Econometrics and Operations Research
Supervisor: Prof. Dr. S.J. Koopman & M. Li
April 1, 2016

Contents

1 The Kalman filter for the Local Level model **3**
1.1 Filtering . . . 3
1.2 Sate smoothing . . . 3
1.3 Disturbance smoothing . . . 4
1.4 Missing observations . . . 5
1.5 Forecasting . . . 5
1.6 Diagnostic checking . . . 7
1.7 Simulation smoothing . . . 7
1.8 Filtering by Importance Sampling . . . 9

2 Formulation of stochastic cycle components **13**
2.1 Deterministic and stochastic cycles . . . 13
2.2 State space form . . . 14
2.3 Illustration . . . 15

3 Application: Cambridge boat race **17**
3.1 Mode estimation . . . 17
3.2 Importance sampling . . . 18
3.3 Results . . . 18

4 Final Assignment: Stochastic Volatility and NAIS **22**
4.1 Part (a) . . . 22
4.2 Part (b) . . . 22
4.3 Part (c) . . . 24
4.4 Part (d) . . . 25
4.5 Part (e) . . . 26
4.6 Part (f) . . . 26
4.7 Part (g) . . . 28
4.8 Part (h) . . . 29
4.9 Part (i) . . . 30
4.10 Part (j) . . . 31
4.11 Part (k) . . . 32
4.12 Part (l) . . . 32

References **35**

1 The Kalman filter for the Local Level model

The Local Level model is defined as

$$\begin{aligned} y_t &= \alpha_t + \epsilon_t, & \epsilon_t &\sim \mathcal{N}(0, \sigma_\epsilon^2), \\ \alpha_{t+1} &= \alpha_t + \eta_t, & \eta_t &\sim \mathcal{N}(0, \sigma_\eta^2), \\ \alpha_1 &\sim \mathcal{N}(a_1, P_1), \end{aligned}$$

for $t = 1, ..., n$ where the ϵ_t's and the η_t's are mutually independent and are independent of α_1. We use the Nile data set of Durbin and Koopman (2012), which consists of a series of readings of the annual flow volume at Aswan from 1871 to 1970.

1.1 Filtering

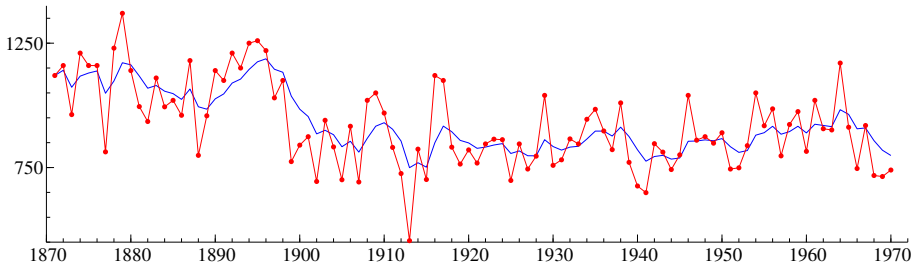
Let Y_{t_1} be the vector of observations $(y_1, \ldots, y_{t-1})'$ for $t = 2, 3, \ldots$. We use that the conditional distribution of $\alpha_t | Y_{t-1} \sim \mathcal{N}(a_t, P_t)$ is normal with mean $\mathbb{E}(\alpha_t | Y_{t-1}) = a_t$ and variance $\mathbb{V}\text{ar}(\alpha_t | Y_{t-1}) = P_t$. Further we have $\alpha_t | Y_t \sim \mathcal{N}(a_{t|t}, P_{t|t})$ and let $v_t = y_t - a_t$ be the prediction error with associated variance F_t. Then the Kalman filter recursion equations are given by

$$\begin{aligned} v_t &= y_t - a_t, & F_t &= P_t + \sigma_\epsilon^2, \\ a_{t|t} &= a_t + K_t v_t, & P_{t|t} &= P_t(1 - K_t), \\ a_{t+1} &= a_t + K_t v_t, & P_{t+1} &= P_t(1 - K_t) + \sigma_\eta^2, \\ K_t &= P_t / F_t, \end{aligned}$$

for $t = 1, ..., n$, where K_t is referred to as the Kalman gain.

For implementation of the Kalman filter, we proceed with the maximum likelihood estimates $\sigma_\epsilon^2 = 15099$, $\sigma_\eta^2 = 1469.1$ and use diffuse initialisation of the Kalman filter. That is, we set $a_1 = 0$ and let $P_1 \to \infty$ which we implement by setting $P_1 = 10^7$. In Figure 1 we present the filtered state $a_{t|t}$.

Figure 1: Nile data (dotted line) and filtered state a_t (solid line).

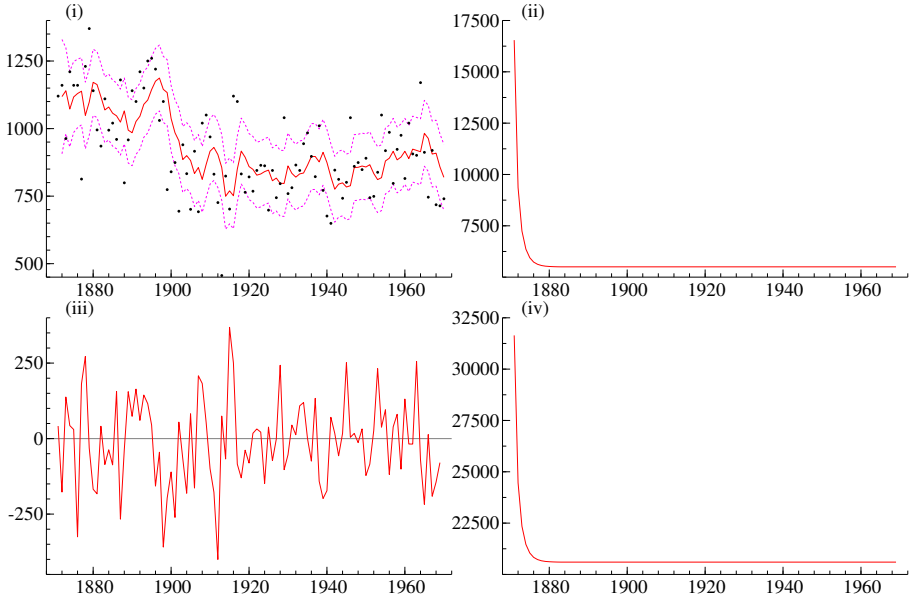


In Figure 2 we have plotted the the filtered state a_t together with its 90% confidence intervals, the filtered variance P_t, the prediction error v_t, the prediction error variance F_t. The figure illustrates the fact that P_t and F_t converge rapidly to constant values, confirming that the local level model has a steady state solution.

1.2 Sate smoothing

Since all distributions are normal, the conditional density of α_t given Y_n is $\mathcal{N}(\hat{\alpha}_t)$. Here the smoothed state is defined as $\hat{\alpha}_t = \mathbb{E}(\alpha_t | Y_n)$ and the smoothed state variance as $V_t = \mathbb{V}\text{ar}(\alpha_t | Y_n)$. According to

Figure 2: Nile data and output of Kalman filter: (i) data (dots), filtered state a_t (solid red line) and its 90% confidence intervals (green and blue lines); (ii) filtered state variance P_t; (iii) prediction errors v_t; (iv) prediction variance F_t.



the regression lemma for the bivariate normal distribution, we have that

$$\hat{\alpha}_t = \mathbb{E}(\alpha_t|v_t, ..., v_n, Y_{t-1}) = a_t + \sum_{j=t}^{n} \mathbb{C}\text{ov}(\alpha_t, v_j)F_j^{-1}v_j,$$

$$V_t = \mathbb{V}\text{ar}(\alpha_t|v_t, ..., v_n, Y_{t-1}) = P_t - \sum_{j=t}^{n} [\mathbb{C}\text{ov}(\alpha_t, v_j)]^2 F_j^{-1}.$$

We use these functions to obtain the state smoothing recursion

$$r_{t-1} = F_t^{-1}v_t + L_t r_t, \qquad \hat{\alpha}_t = a_t + P_t r_{t-1},$$
$$N_{t-1} = F_t^{-1} + L_t^2 N_t, \qquad V_t = P_t - P_t^2 N_{t-1},$$

for $t = n, ..., 1$, where $L_t = 1 - K_t$, r_t is the weighted usm of innovations after t and N_t is a weighted sum of the inverse variances of innovations after time t. Since there are no observations after time n, we have that $r_n = N_n = 0$.

From the standard error $\sqrt{V_t}$ of $\hat{\alpha}_t$ we can construct confidence intervals for the smoothed state α_t. In Figure 3 we have displayed the smoothed state α_t, the smoothed state variance V_t, the smoothing cumulant r_t and the smoothing variance cumulant N_t.

1.3 Disturbance smoothing

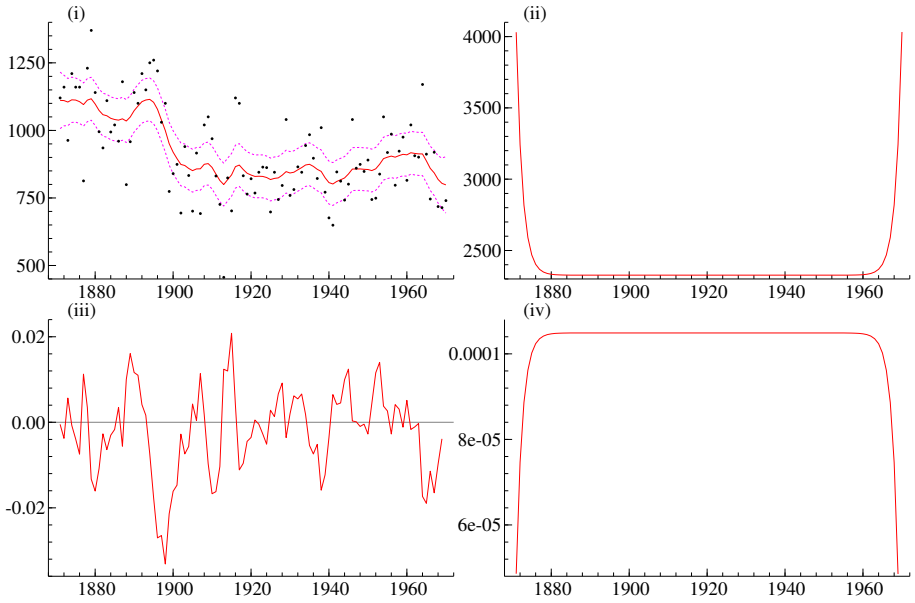
We consider the calculation of the smoothed observation disturbance $\hat{\epsilon}_t = E(\epsilon_t|Y_n) = y_t - \hat{\alpha}_t$ and the smoothed state disturbances $\hat{\eta}_t = E(\eta_t|Y_n) = \hat{\alpha}_{t+1} - \hat{\alpha}_t$ together with their error variances. As Durbin and Koopman (2012) argue, these can be calculated directly from r_t and N_t without first calculating the smoothed states $\hat{\alpha}_t$, saving computational time. Namely, we have the following disturbance smoothing

recursion

$$\hat{\epsilon}_t = \sigma_\epsilon^2(F_t^{-1}v_t - K_t r_t), \qquad \mathbb{V}\text{ar}(\epsilon|Y_n) = \sigma_\epsilon^2 - \sigma_\epsilon^4 D_t,$$
$$\hat{\eta}_t = \sigma_\eta^2 r_t, \qquad \mathbb{V}\text{ar}(\eta_t|Y_n) = \sigma_\eta^2 - \sigma_\eta^4 N_t,$$
$$D_t = F_t^{-1} + K_t^2 N_t,$$

for $t = n, ..., 1$. In Figure 4 we display the output of the disturbance smoothing recursions, i.e. the observation error $\hat{\epsilon}_t$ and the state error $\hat{\eta}_t$ along with their variances.

Figure 3: Nile data and output of state smoothing recursion: (i) data (dots), smoothed state $\hat{\alpha}_t$ and its 90% confidence intervals; (ii) smoothed state variance V_t; (iii) smoothing cumulant r_t, (iv) smoothing variance cumulant N_t.



1.4 Missing observations

It is easy to correct for missing values in the Kalman filter. We consider the case where the observations for the Nile data set at time points $21, ..., 40$ and $61, ..., 80$ are missing. We use the following general approach for missing values. If observations y_t for $t = \tau, ..., \tau^* - 1$ are missing, we put $v_t = 0$ and $F_t \to \infty$ such that $K_t = 0$. The implication for the Local Level model is that $a_{t|t} = a_{t+1} = a_t$, $P_{t|t} = P_t$ and $P_{t+1} = P_t + \sigma_\eta^2$ for $t = \tau, ..., \tau^* - 1$. For the state smoothing, we obtain the following recursion

$$r_{t-1} = r_t, \qquad \hat{\alpha}_t = a_t + P_t r_{t-1},$$

for $t = \tau, ..., \tau^* - 1$. Hence, we can use the original filter and smoother for all t by taking $K_t = 0$ at the missing time points. In Figure 5 we show the filtering and smoothing output when observations are missing.

1.5 Forecasting

The Kalman filter can be used for forecasting by treating the observations at times $n+1, \ldots, n+J$ as missing. Consequently, we apply the Kalman filter with $K_t = 0$ for $t = n+1, ..., n+30$. It follows that

Figure 4: Output of disturbance smoothing recursion: (i) observation error $\hat{\epsilon}_t$; (ii) observation error variance $\mathbb{V}\text{ar}(\epsilon_t|Y_n)$; (iii) state error $\hat{\eta}_t$; (iv) state error variance $\mathbb{V}\text{ar}(\eta_t|Y_n)$.

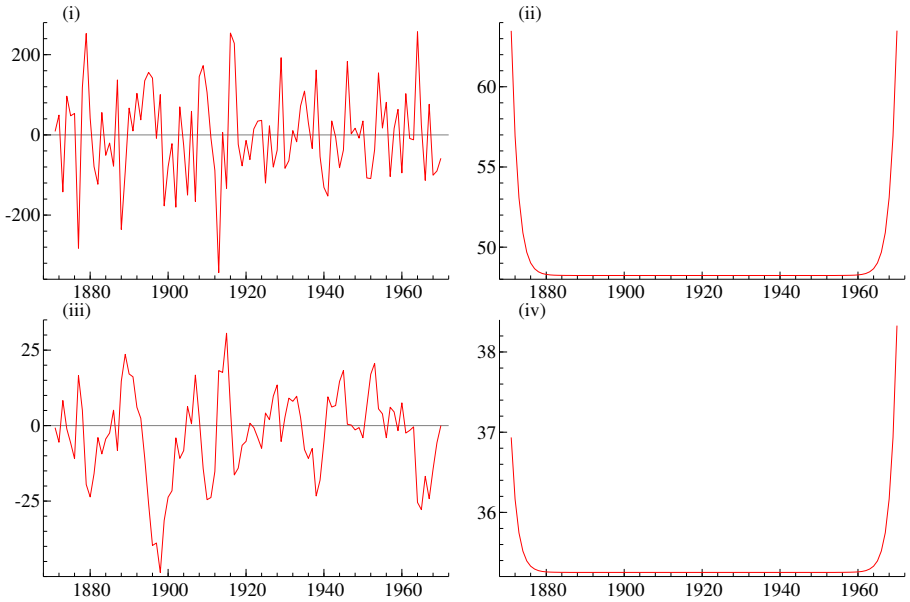
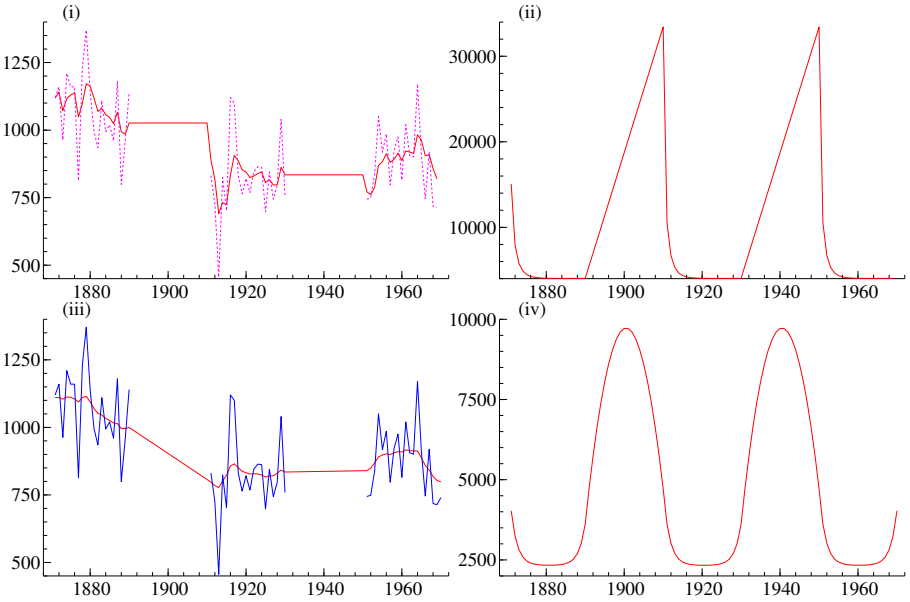


Figure 5: Filtering and smoothing output when observations are missing: (i) data and filtered state a_t (extrapolation); (ii) filtered state variance P_t; (iii) data and smoothed state $\hat{\alpha}_t$ (interpolation); (iv) smoothed state variance V_t.



for $j = 1, ..., J$

$$\mathbb{E}(y_{n+j}|Y_n) = \mathbb{E}(\alpha_{n+j}|Y_n), \qquad \mathbb{V}\text{ar}(y_{n+j}|Y_n) = \mathbb{V}\text{ar}(\alpha_{n+j}|Y_n) + \sigma_\epsilon^2,$$
$$\mathbb{E}(\alpha_{n+j+1}|Y_n) = \mathbb{E}(\alpha_{n+j}|Y_n), \qquad \mathbb{V}\text{ar}(\alpha_{n+j+1}|Y_n) = \mathbb{V}\text{ar}(\alpha_{n+j}|Y_n) + \sigma_\eta^2.$$

In Figure 6 we have computed 30-step-ahead forecasts for the Nile data set by creating 30 missing observations which allow the computation of forecasts for the observations $t = n + 1, ..., n + J$ with $J = 30$.

1.6 Diagnostic checking

For the local level model we assume that ϵ_t and η_t are normally distributed and serially independent with constant variances. With these assumptions the one-step ahead forecast errors

$$e_t = v_t F_t^{-\frac{1}{2}},$$

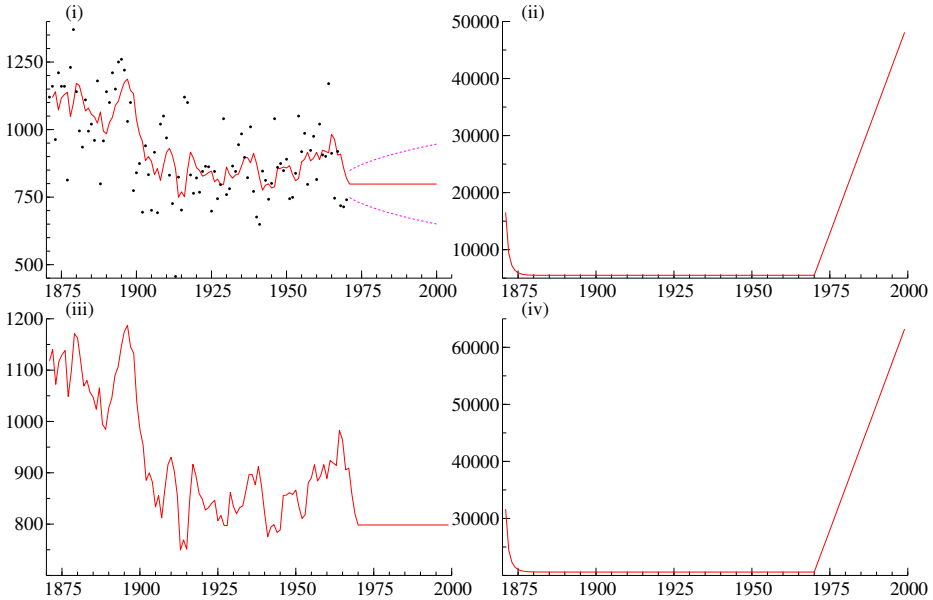
for $t = 1, ..., n$ are also normally distributed and serially independent with unit variance. To check if these properties hold we can construct the histogram, QQ plot and correlogram of e_t, see Figure 7. The QQ plot and histogram indicate that the standardized one-step ahead forecasts errors are normally distributed. The ACF plot confirms the serial independence.

We can also obtain the standardized smoothed residuals. These are given by

$$u_t^* = \hat{\epsilon}_t / \sqrt{\mathbb{V}\text{ar}(\hat{\epsilon}_t)} = D_t^{-\frac{1}{2}} u_t,$$
$$r_t^* = \hat{\eta}_t / \sqrt{\mathbb{V}\text{ar}(\hat{\eta}_t)} = N_t^{-\frac{1}{2}} r_t,$$

In Figure 8 we show the plots of u_t^* and r_t^* together with their histograms. It shows indicate outliers in 1913 and 1918 and a level break in 1899. The plot of the Nile data confirms these findings.

Figure 6: Nile data and output of forecasting: (i) data (dots), state forecast a_t and 50% confidence intervals; (ii) state variance P_t; (iii) observation forecast $\mathbb{E}(y_t|Y_{t-1})$; (iv) observation forecast variance F_t



1.7 Simulation smoothing

Simulation smoothing requires us to draw samples generated by the Local Level model (unconditional simulation) and samples generated by the model conditional on the observed time series $y_1, ..., y_n$

Figure 7: Diagnostic plots for standardised prediction errors: (i) standardised residual; (ii) histogram plus estimated density; (iii) ordered residuals; (iv) correlogram.

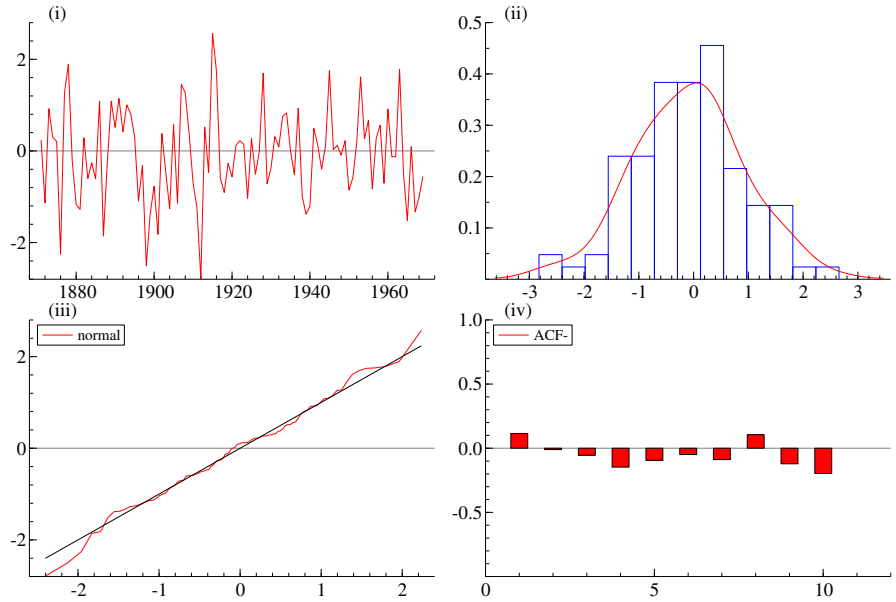
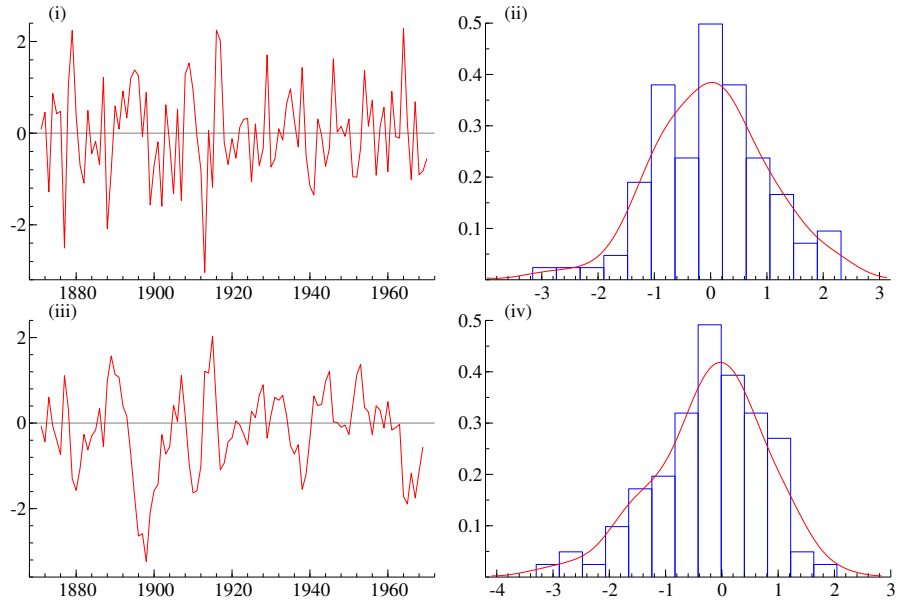


Figure 8: Diagnostic plots for auxiliary residuals: (i) observation residual u_t^*; (ii) histogram and estimated density for u_t^*; (iii) state residual r_t^*; (iv) histogram and estimated density for r_t^*.



(conditional simulation). Since for the linear Gaussian state space model we know that $p(\alpha|Y_n)$ is $\mathcal{N}(\mathbb{E}(\alpha|Y_n), \mathbb{V}\text{ar}(\alpha|Y_n))$, the simulation smoothing gives a computational advantage for large samples. That is, with the simulation smoother, you do not need to invert the huge conditional variance $\mathbb{V}\text{ar}(\alpha|Y_n)$.

The unconditional simulation is performed by drawing normally distributed random variables

$$\epsilon_t^+ \sim N(0, \sigma_\epsilon^2), \qquad \eta_t^+ \sim N(0, \sigma_\eta^2),$$

for $t = 1, \ldots, n$. Next, by drawing α_1^+ from $\mathcal{N}(a_1, P_1)$, we use the Local Level recursion equations to obtain for $t = 1, \ldots, n$

$$y_t^+ = \alpha_t^+ + \epsilon_t^+, \qquad \alpha_{t+1}^+ = \alpha_t^+ + \eta_t^+.$$

In the conditional simulation we draw samples of state or disturbance vectors conditional on the observations $y_1, ..., y_n$. This can be achieved by the method of mean corrections. Define $\tilde{\alpha}$ as a vector drawn from the density $p(\alpha|Y_n)$. It follows that $\tilde{\alpha} = \alpha^+ - \hat{\alpha}^+ + \hat{\alpha}$, where $\hat{\alpha}^+ = \mathbb{E}(\alpha|y^+)$ and $\hat{\alpha} = \mathbb{E}(\alpha|Y_n)$ are obtained by the Kalman filter and the smoothing recursion (KFS). Hence $\tilde{\alpha}$ is a draw from density $p(\alpha|Y_n)$. The above holds since $\tilde{\alpha} \sim p(\alpha|Y_n) \equiv \mathcal{N}(\hat{\alpha}, V)$ as Y_n and α are normally distributed and since for the multivariate normal distribution the conditional variance is independent of the y we have that

$$\begin{aligned}
\mathbb{E}(\tilde{\alpha}|Y_n) &= \mathbb{E}(\alpha^+ - \hat{\alpha}^+ + \hat{\alpha}|Y_n) \\
&= \hat{\alpha} + \mathbb{E}(\alpha^+ - \hat{\alpha}^+|Y_n) \\
&= \hat{\alpha} + \mathbb{E}(\alpha^+|Y_n) - \hat{\alpha}^+ \\
&= \hat{\alpha}
\end{aligned} \qquad
\begin{aligned}
\mathbb{V}\mathrm{ar}(\tilde{\alpha}|Y_n) &= \mathbb{E}((\tilde{\alpha} - \mathbb{E}(\tilde{\alpha}|Y_n))(\tilde{\alpha} - \mathbb{E}(\tilde{\alpha}|Y_n))'|Y_n) \\
&= \mathbb{E}((\tilde{\alpha} - \hat{\alpha})(\tilde{\alpha} - \hat{\alpha})'|Y_n) \\
&= \mathbb{E}((\alpha^+ - \hat{\alpha}^+)(\alpha^+ - \hat{\alpha}^+)'|Y_n) \\
&= \mathbb{V}\mathrm{ar}(\alpha^+|Y_n) = V.
\end{aligned}$$

The algorithm for the unconditional simulation can be summarized as follows:

1. Compute the smoothed estimates $\hat{\alpha}_1, \ldots, \hat{\alpha}_n$ using observations $y_1, \ldots, y_n$ by performing KFS.
2. Apply the unconditional simulation to obtain y_t^+ and α_t^+ for $t = 1, \ldots, n$.
3. Use KFS to obtain the smoothed levels $\hat{\alpha}_1^+, \ldots, \hat{\alpha}_n^+$ using the simulated observations $y_1^+, \ldots, y_n^+$.
4. Calculate $\tilde{\alpha} = \alpha^+ - \hat{\alpha}^+ + \hat{\alpha}$. According to the method of mean correction, this corresponds to a draw from $p(\alpha|Y_n)$.

Alternatively, one could choose to draw from $p(\epsilon|Y_n)$ to obtain $\tilde{\alpha}$. This requires the drawing of samples ϵ_t^+ and η_t^+ and using them to generate y_t. It follows that a conditional draw for ϵ_t given Y_n is given by the equation for $\tilde{\epsilon}_t$, where

$$\begin{aligned}
y_t^+ &= \alpha_t^+ + \epsilon_t^+, \\
\tilde{\epsilon}_t &= \epsilon_t^+ - \hat{\epsilon}_t^+ + \hat{\epsilon}_t, \\
\hat{\epsilon}_t &= \mathbb{E}(\epsilon_t|Y_n),
\end{aligned} \qquad
\begin{aligned}
\alpha_{t+1}^+ &= \alpha_t^+ + \eta_t^+, \\
t &= 1, ..., n \\
\hat{\epsilon}_t^+ &= \mathbb{E}(\epsilon_t|Y_n^+)
\end{aligned}$$

Hence, given a sample $\tilde{\epsilon}_1, ..., \tilde{\epsilon}_n$ we obtain simulated samples for α_t and η_t according to the relations $\tilde{\alpha}_t = y_t - \tilde{\epsilon}_t$, and $\tilde{\eta}_t = \tilde{\alpha}_{t+1} - \tilde{\alpha}_t$. In Figure 9 we have displayed the sample α_t^+, the sample $\tilde{\alpha}_t$, the sample $\tilde{\epsilon}_t$ and the sample $\tilde{\eta}_t$.

In order to account for missing values, we need to adapt our algorithm for the simulation smoothing. We do this by adapting the KFS to work with missing values. That is, in case an observation is missing we let $F_t \to \infty$ such that $K_t \to 0$, see Subsection 1.4. In Figure 10 we show the simulation smoothing for using missing observations.

1.8 Filtering by Importance Sampling

We consider two different methods for filtering by importance sampling: (i) sequential importance sampling (SIS) and (ii) the bootstrap filter. Let the function $x_t(\alpha_{1:t}) \equiv x_t(\alpha_t)$, then the general algorithm for SIS which we repeat for $t = t^*, \ldots, n$ is:

1. Sample α_t: draw N values values $\tilde{\alpha_t}^{(i)}$ from $g\left(\alpha_t|\alpha_{t-1}^{(i)}, Y_t\right)$ and store $\tilde{\alpha}_{t-1:t}^{(i)} = \left\{\alpha_{t-1}^{(i)}, \tilde{\alpha_t}^{(i)}\right\}$.

Figure 9: Simulation: (i) smoothed state $\hat{\alpha}_t$ (solid line) and sample α_t^+ (dots); (ii) smoothed state $\hat{\alpha}_t$ (solid line) and sample $\tilde{\alpha}_t$ (dots); (iii) smoothed observation error $\hat{\epsilon}_t$ (solid line) and sample $\tilde{\epsilon}_t$ (dots); (iv) smoothed state error $\hat{\eta}_t$ (solid line) and sample $\tilde{\eta}_t$ (dots).

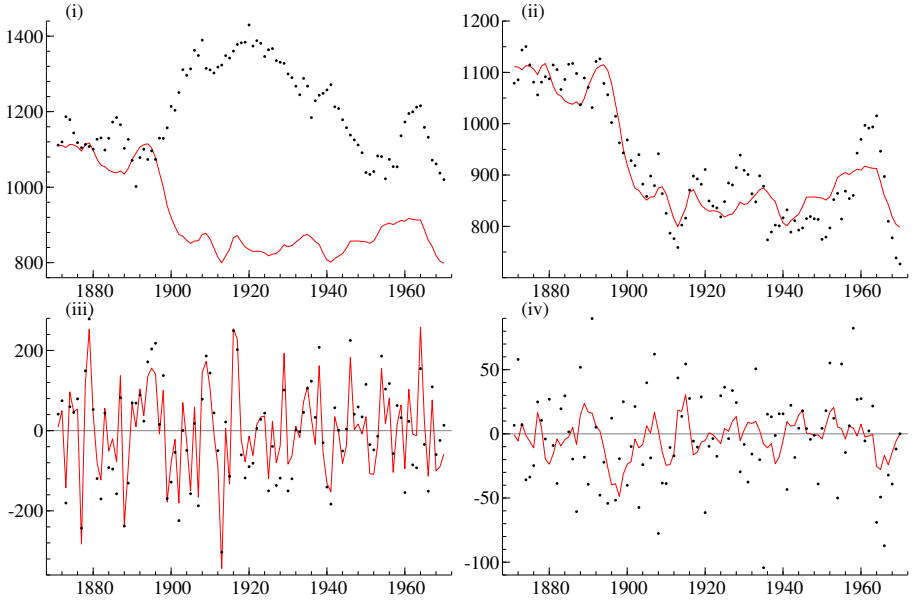
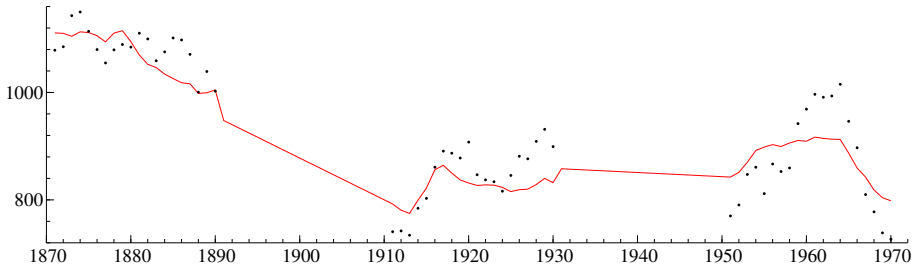


Figure 10: Simulation: smoothed state $\hat{\alpha}_t$ (solid line) and sample $\tilde{\alpha}_t$ (dots) with missing observations.



2. Weights: compute $\tilde{w}_t^{(i)}$ with

$$\tilde{w}_t^{(i)} = \tilde{w}_{t-1}^{(i)} \frac{p\left(\tilde{\alpha}_t^{(i)} | \tilde{\alpha}_{t-1}^{(i)}\right) p\left(y_t | \tilde{\alpha}_t^{(i)}\right)}{g\left(\tilde{\alpha}_t^{(i)} | \tilde{\alpha}_{t-1}^{(i)}, Y_t\right)},$$

and normalise the weights

$$w_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{j=1}^N \tilde{w}_t^{(j)}}.$$

3. Compute the variable of interest x_t: given the set of particles $\left\{\tilde{\alpha}_t^{(1)}, \ldots, \tilde{\alpha}_t^{(N)}\right\}$, compute

$$\hat{x_t} = \sum_{i=1}^N w_t^{(i)} x_t\left(\tilde{\alpha}_t^{(i)}\right).$$

4. Set $\alpha_t^{(i)} = \tilde{\alpha}_t^{(i)}$ for $i = 1, \ldots, N$.

The Bootstrap filter with resampling which is repeated for $t = t^*, \ldots, n$ is given by:

1. Sample α_t: draw N values $\tilde{\alpha}_t^{(i)}$ from $p(\alpha_t|\alpha_{t-1}^{(i)})$.
2. Weights: compute the corresponding weights $\tilde{w}_t^{(i)}$

$$\tilde{w}_t^{(i)} = p\left(y_t|\tilde{\alpha_t}^{(i)}\right), \qquad i = 1, \ldots, N,$$

   and normalise the weights to obtain $w_t^{(i)}$ as in the SIS algorithm.
3. Compute the variable of interest x_t: given the set of particles $\left\{\tilde{\alpha}_t^{(1)}, \ldots, \tilde{\alpha}_t^{(N)}\right\}$, compute

$$\hat{x_t} = \sum_{i=1}^{N} w_t^{(i)} x_t\left(\tilde{\alpha}_t^{(i)}\right).$$

4. Resample: draw N new independent particles $\alpha_t^{(i)}$ from $\left\{\tilde{\alpha}_t^{(1)}, \ldots, \tilde{\alpha}_T^{(N)}\right\}$ with replacement and with corresponding probabilities $\left\{w_t^{(1)}, \ldots, w_t^{(N)}\right\}$.

For the Local Level model we have that $y_t|\alpha_t$ is Gaussian and hence

$$p(y_t|\alpha_t) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left\{-\frac{1}{2}\sigma_\epsilon^{-2}\left(y_t - \tilde{\alpha}_t\right)^2\right\},$$

for $i = 1, \ldots, N$. Moreover, $\alpha_{t+1}|\alpha_t$ is also Gaussian so

$$p(\alpha_{t+1}|\alpha_t) = \frac{1}{\sqrt{2\pi\sigma_\eta^2}} \exp\left\{-\frac{1}{2}\sigma_\eta^{-2}\left(\alpha_{t+1} - \alpha_t\right)^2\right\}.$$

In Figure 15 we have showed the filtered estimate of the Kalman filter along with the estimates from SIS and the Bootstrap filter. Due to the resampling, the Bootstrap filter clearly outperforms SIS and its estimate is rather close to the Kalman filter. This is also what we observe from the efficient sample size, i.e. for SIS it gets close to zero after a few iterations. For the Bootstrap filter we resample in each iteration and hence the ESS is close to $N = 100$. This implies that we have relevant particles in our sample. Note that resampling at every point in time is not necessary, we could also follow a rule in which we only resample if $ESS < k \cdot N$ for $k \in (0, 1)$. We see that the state variance of the Kalman filter is generally higher compared to that of the Bootstrap filter and SIS, but the latter has some very large peaks. This can be solved by implementing resampling, i.e. the SISR algorithm, as the number of relevant particles in the series is very low.

In Figure 16 we show for $N = 15$ particles the associated sample paths from $t = 1$ upto time $t = 5, 10, 15, 20$ from the bootstrap filter together with the filtered state from the Kalman filter. We observe that the sample paths of the particles approximate the filtered of the Kalman filter well.

Figure 11: Kalman filter, sequential importance sampling (SIS) and bootstrap filter with $N = 100$ particles: (i) the filtered state from SIS and the filtered state from the Kalman filter, (ii) filtered state variance from the SIS, Bootstrap filter and the Kalman filter, (iii) the filtered state from Bootstrap filter and the filtered state from the Kalman filter, (iv) the efficient sample size (ESS) of the SIS and Bootstrap filter as a function of time t.

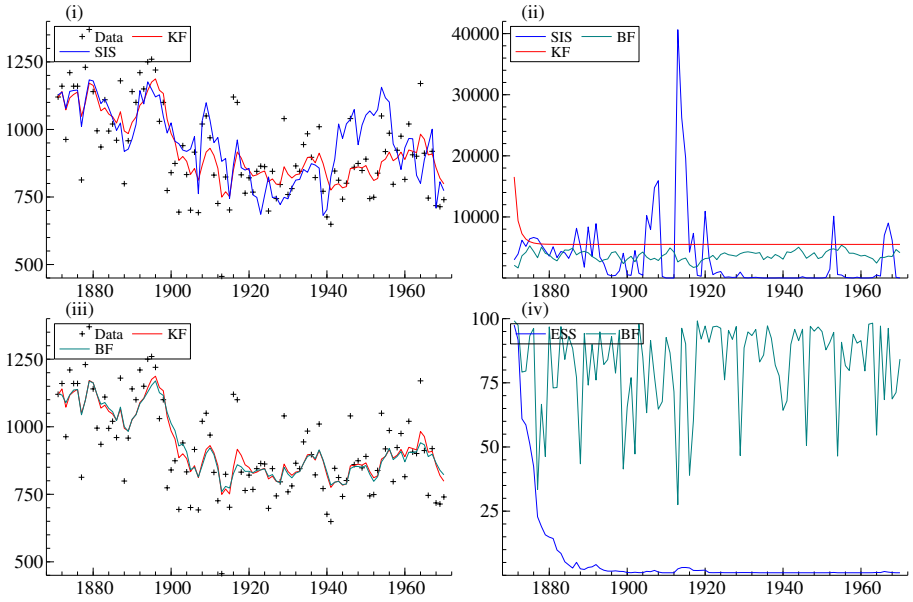
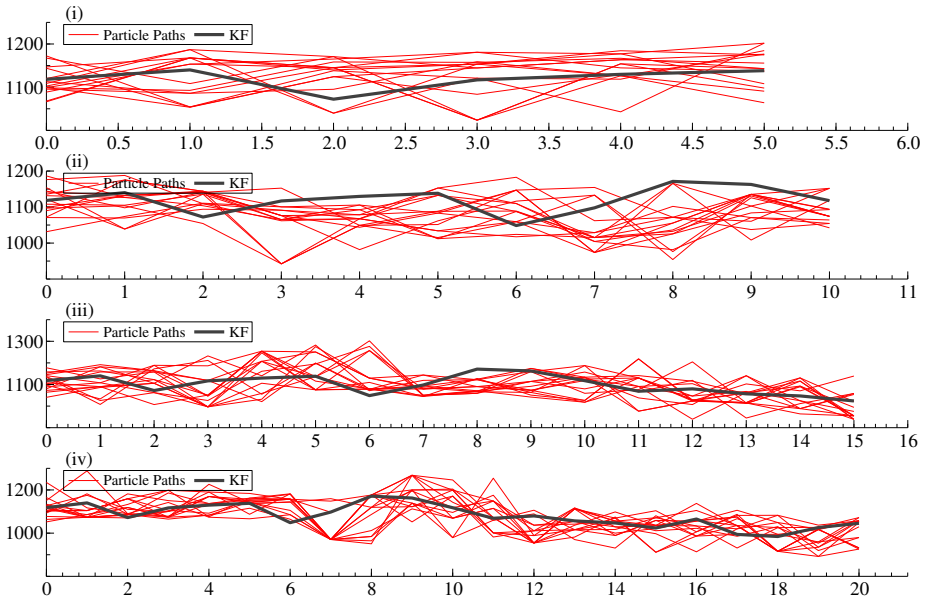


Figure 12: Particle paths for $N = 15$ together with the filtered state from the Kalman filter for time t up to (i) $t = 5$, (ii) $t = 10$, (iii) $t = 15$, (iv) $t = 20$.



2 Formulation of stochastic cycle components

Here we assume that an economic time series can be decomposed into a nonstationary trend μ_t, a stationary cycle ψ_t and an irregular component ϵ_t. The observed time series y_t is then modelled as

$$y_t = \mu_t + \psi_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{NID}(0, \sigma_\epsilon^2),$$

for $t = 1, \ldots, n$. Following Koopman & Lee, we employ a smooth trend specification for the trend-cycle decomposition model defined by

$$\mu_{t+1} = \mu_t + \beta_t, \qquad \beta_{t+1} = \beta_t + \zeta_t, \qquad \zeta_t, \sim \mathcal{NID}(0, \sigma_\zeta^2),$$

for $t = 1, \ldots, n$, where the initial values β_1 and μ_1 are assumed to be diffuse. The resulting model for the trend is also called an integrated random walk. The cyclical component ψ_t is modelled as a stochastic trigonometric process driven by the disturbances κ_t, $\dot{\kappa}_t$, which is discussed in the following subsections.

2.1 Deterministic and stochastic cycles

A deterministic cycle and its partial derivative with amplitude a, phase b and frequency λ can be expressed by

$$\psi_t = a\, \cos(\lambda t - b), \qquad \dot{\psi}_t := d\psi_t/d(\lambda t) = -a\, \sin(\lambda t - b), \tag{1}$$

with $a, b, \lambda, t \in \mathbb{R}$, and $a \neq 0, \lambda \neq 0$. Note that the sign of $\dot{\psi}_t$ indicates whether the cycle is ascending or descending, while its magnitude determines the steepness.

Note that we can obtain a asymmetric cycle be varying the frequency λ for different values of t. For example, the cycle can have different frequencies when ψ_t is ascending or descending, i.e.

$$\psi_t = a\cos(\lambda_t t - b), \qquad \lambda_t = \begin{cases} \lambda^a, & \dot{\psi}_t > 0 \\ \lambda^d, & \dot{\psi}_t \leq 0 \end{cases}, \tag{2}$$

Another example is where we allow the frequency to depend on a continous function of $\dot{\psi}_t$, i.e.

$$\psi_t = a\cos(\lambda_t t - b), \qquad \lambda_t = \lambda + \gamma\dot{\psi}_t. \tag{3}$$

In Figure 13 we have displayed a simulated series of the asymmetric deterministic cycles according to (2) and (3).

By applying the trigonometric identities

$$\cos(x \pm y) = \cos(x)\cos(y) \mp \sin(x)\sin(y), \qquad \sin(x \pm y) = \cos(x)\sin(y) \pm \sin(x)\cos(y),$$

to (1), we obtain

$$\begin{aligned} \psi_t &= a\cos(\lambda t - b) \\ &= a\cos(\lambda t)\cos(b) + \sin(\lambda t)\sin(b) \\ &= \alpha\cos(\lambda t) + \beta\sin(\lambda) \end{aligned} \qquad \begin{aligned} \dot{\psi}_t &= -a\sin(\lambda t - b) \\ &= a\cos(\lambda t)\sin(b) - \sin(\lambda t)\cos(b) \\ &= \beta\cos(\lambda t) - \alpha\sin(\lambda t), \end{aligned}$$

where $\alpha = a\cos(b)$ and $\beta = a\sin(b)$. Suppose $t = t + \delta$, for a $\delta > 0$, then we have that

$$\begin{aligned} \psi_{t+\delta} &= \alpha\cos(\lambda(t+\delta)) + \beta\sin(\lambda(t+\delta)) \\ &= \alpha\cos(\lambda t + \lambda\delta) + \beta\sin(\lambda t + \lambda\delta) \\ &= \alpha\big[\cos(\lambda t)\cos(\lambda\delta) - \sin(\lambda t)\sin(\lambda\delta)\big] + \beta\big[\cos(\lambda t)\sin(\lambda\delta) + \sin(\lambda t)\cos(\lambda\delta)\big] \\ &= \alpha\cos(\lambda t)\cos(\lambda\delta) - \alpha\sin(\lambda t)\sin(\lambda\delta) + \beta\cos(\lambda t)\sin(\lambda\delta) + \beta\sin(\lambda t)\cos(\lambda\delta) \\ &= \alpha\cos(\lambda t)\cos(\lambda\delta) + \beta\sin(\lambda t)\cos(\lambda\delta) + \beta\cos(\lambda t)\sin(\lambda\delta) - \alpha\sin(\lambda t)\sin(\lambda\delta) \\ &= \big[\underbrace{\alpha\cos(\lambda t) + \beta\sin(\lambda t)}_{\psi_t}\big]\cos(\lambda\delta) + \big[\underbrace{\beta\cos(\lambda t) - \alpha\sin(\lambda t)}_{\dot{\psi}_t}\big]\sin(\lambda\delta) \\ &= \psi_t\cos(\lambda\delta) + \dot{\psi}_t\sin(\lambda\delta). \end{aligned}$$

Then it also follows that the partial derivative can be expressed as

$$\begin{aligned}
\dot{\psi}_{t+\delta} &= \beta\cos(\lambda(t+\delta) + \alpha\sin(\lambda(t+\delta) \\
&= \beta\cos(\lambda t + \lambda\delta)) + \alpha\sin(\lambda t + \lambda\delta) \\
&= \beta\big[\cos(\lambda t)\cos(\lambda\delta) - \sin(\lambda t)\sin(\lambda\delta)\big] - \alpha\big[\cos(\lambda t)\sin(\lambda\delta) + \sin(\lambda t)\cos(\lambda\delta)\big] \\
&= \beta\cos(\lambda t)\cos(\lambda\delta) - \beta\sin(\lambda t)\sin(\lambda\delta) - \alpha\cos(\lambda t)\sin(\lambda\delta) + \alpha\sin(\lambda t)\cos(\lambda\delta) \\
&= -\alpha\cos(\lambda t)\sin(\lambda\delta) - \beta\sin(\lambda t)\sin(\lambda\delta) + \beta\cos(\lambda t)\cos(\lambda\delta) + \alpha\sin(\lambda t)\cos(\lambda\delta) \\
&= \big[\underbrace{\alpha\cos(\lambda t) + \beta\sin(\lambda t)}_{\psi_t}\big](-\sin(\lambda\delta)) + \big[\underbrace{\beta\cos(\lambda t) + \alpha\sin(\lambda t)}_{\dot{\psi}_t}\big]\cos(\lambda\delta) \\
&= \psi_t(-\sin(\lambda\delta)) + \dot{\psi}_t\cos(\lambda\delta).
\end{aligned}$$

Hence the cycle ψ_t and its partial derivative $\dot{\psi}_t$ can be expressed via the recursive expression

$$\begin{pmatrix}\psi_{t+\delta}\\ \dot{\psi}_{t+\delta}\end{pmatrix} = \begin{bmatrix}\cos(\lambda\delta) & \sin(\lambda\delta)\\ -\sin(\lambda\delta) & \cos(\lambda\delta)\end{bmatrix}\begin{pmatrix}\psi_t\\ \dot{\psi}_t\end{pmatrix}, \qquad \delta > 0, \qquad t = 0, \delta, 2\delta, \ldots, \tag{4}$$

which follows by repeatingly applying the trigonometric identities.

A stochastic cycle can be based on this recursive expression by including a damping term ϕ and white noise disturbances, which brings us to the following recursive expression

$$\begin{pmatrix}\psi_{t+\delta}\\ \dot{\psi}_{t+\delta}\end{pmatrix} = \phi\begin{bmatrix}\cos(\lambda\delta) & \sin(\lambda\delta)\\ -\sin(\lambda\delta) & \cos(\lambda\delta)\end{bmatrix}\begin{pmatrix}\psi_t\\ \dot{\psi}_t\end{pmatrix} + \begin{pmatrix}\kappa_t\\ \dot{\kappa}_t\end{pmatrix}, \qquad \begin{pmatrix}\kappa_t\\ \dot{\kappa}_t\end{pmatrix} \sim \mathcal{NID}(0, \sigma_\kappa^2 I_2), \qquad \delta > 0, \tag{5}$$

for $t = 0, \delta, 2\delta, \ldots$, where $|\phi| < 1$ is a damping factor, λ is the cycle frequency and the disturbance vectors are Gaussian noise. The damping term ϕ ensures that the stochastic process ψ_t is stationary. The process of ψ_t then follows a nonlinear ARMA(2,1) process where the autoregressive coefficients also depend on the ARMA proces. This can be seen by considering the terms $\phi\cos(\lambda\delta)\psi_t$ and $\phi\sin(\lambda\delta)\dot{\psi}_t$ as the 'autoregressive' terms. We could analyse the ACF, PACF and the spectrum to ensure that the stochastic process is equivalent to an ARMA(2,1) process. Although the interpretation of $\dot{\psi}_t$ as the partial derivative of ψ_t with respect to λt is not strictly valid for the aforementioned stochastic process, it can be taken as a local proxy for the steepness of the cycle ψ_t. In Figure 13 we simulated samples for the deterministic cycle according to (4) and the stochastic cycle according to (5).

2.2 State space form

We can write the symmetric trend-cycle decomposition model as a linear Gaussian linear Gaussian state space form, with $\alpha_t = \begin{pmatrix}\mu_t & \beta_t & \psi_t & \dot{\psi}_t\end{pmatrix}'$ and $\eta = \begin{pmatrix}0 & \zeta_t & \kappa_t & \dot{\kappa}_t\end{pmatrix}'$. The system matrices are given by

$$T_t = \begin{bmatrix}\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix} & \mathbf{O}\\ \mathbf{O} & \phi\begin{bmatrix}\cos(\lambda) & \sin(\lambda)\\ -\sin(\lambda) & \cos(\lambda)\end{bmatrix}\end{bmatrix}, \qquad Z_t = \begin{bmatrix}1 & 0 & 1 & 0\end{bmatrix},$$

$$Q_t = \begin{bmatrix}\begin{bmatrix}0 & 0\\ 0 & \sigma_\zeta^2\end{bmatrix} & \mathbf{O}\\ \mathbf{O} & \begin{bmatrix}\sigma_\kappa^2 & 0\\ 0 & \sigma_\kappa^2\end{bmatrix}\end{bmatrix}, \qquad G_t = \sigma_\varepsilon^2,$$

where $\mathbf{O}$ represents a zero matrix with the appropriate dimensions. Note that the state estimation and likelihood evaluation can be handled by standard Kalman filter. For the initial $m \times 1$ state vector α_1

we consider the following general model

$$\alpha_1 = a + A\delta + R_0\eta_0, \qquad \eta_0 \sim N(0, Q_0),$$

see Durbin and Koopman (2012), where the $m \times 1$ vector a is known, δ is a $q \times 1$ vector of unknown quantities, the $m \times q$ matrix A and the $m \times (m-q)$ matrix R_0 are selection matrices. The vector a will be treated as a zero vector unless some elements of the initial state vector are known constants. The vector δ can be treated as a vector of random variables and we assume that

$$\delta \sim N(0, \kappa I_q),$$

where we let $\kappa \to \infty$. The initial conditions can thus be summarized as

$$\begin{aligned}
a_1 &\equiv \mathbb{E}[\alpha_1] = \mathbb{E}[a + A\delta + R_0\eta_0] \\
&= a + A\mathbb{E}[\delta] + R_0\mathbb{E}[\eta_0] = a \\
P_1 &\equiv \mathbb{V}\text{ar}(\alpha_1) = \mathbb{V}\text{ar}(a + A\delta + R_0\eta_0) \\
&= A\mathbb{V}\text{ar}(\delta)A' + R_0\mathbb{V}\text{ar}(\eta_0)R_0' \\
&= \kappa AA' + R_0Q_0R_0' \\
&= \kappa P_\infty + P_*, \qquad P_\infty \equiv AA', \qquad P_* \equiv R_0Q_0R_0'
\end{aligned}$$

Since the symmetric trend-cycle decomposition model is nonstationary due to the integrated random walk for the trend, we have that

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \qquad R_0 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix},$$

a where Q_0 is the unconditional variance of the stationary VAR(1) model for $\psi_t, \dot{\psi}_t$.

2.3 Illustration

To get a better understanding of the symmetric trend-cycle decomposition model, we simulate four different time-series of length $n = 200$ for different values for the variances of the disturbances of the observation equation, trend and cycle. We set the length of the cycle equal to 20 periods and the damping factor to 0.95. The initial values of μ_1 and β_1 are assumed to be diffuse and for ψ_0 and $\dot{\psi}_0$ we take $\psi_0 = \alpha$ and $\dot{\psi}_0 = \beta$, where $\alpha = a\cos b$ and $\beta = a\sin b$ with $a = 1.5$ and $b = 4$. We obtain ψ_1 and $\dot{\psi}_1$ using the recursive expression

$$\begin{pmatrix} \psi_1 \\ \dot{\psi}_1 \end{pmatrix} = \begin{bmatrix} \cos(\lambda\delta) & \sin(\lambda) \\ -\sin(\lambda\delta) & \cos(\lambda) \end{bmatrix} \begin{pmatrix} \psi_0 \\ \dot{\psi}_0 \end{pmatrix}.$$

The plotted series of the simulate samples are given in Figure 14. We notice that the first (red) and the fourth series (dotted purple) differ due to the variance of the trend, since it is three times as large. We also notice that there is no big difference in the first (red) and the third series (green). This can be explained by the variance of y_t, although it is 50 times as large compared to the first series, it does not have an impact on y_t since it is still small compared to the values of the other components. By making the variance of the cycle component ten times as big, we notice that the series becomes less stable due to the larger shocks.

Figure 13: Simulation of deterministic and stochastic cycles: (i) asymmetric deterministic cycle according to (2) with $\lambda^a = \pi/10$ and $\lambda^d = -\pi/10$, (ii) asymmetric deterministic cylce according to (3) with $\gamma = 0.01$, (iii) deterministic cycle according to (4) where the red line is ψ_t and the blue line is $\dot{\psi}_t$, (iv) stochastic cycle according to (5) with $\rho = 0.95$ and i.i.d. Gaussian innovations. For all simulations we have used $a = 1.5, b = 4, \lambda = \pi/10$.

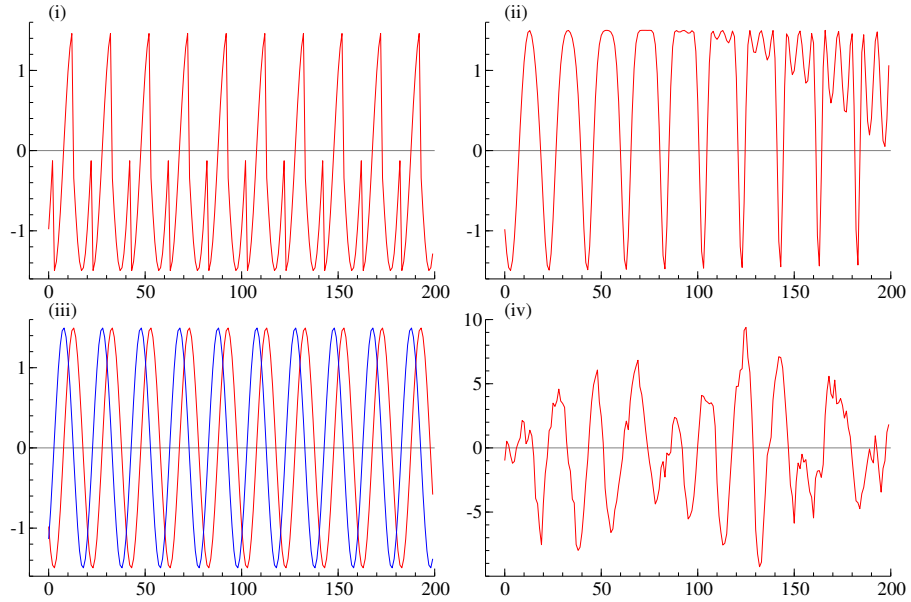
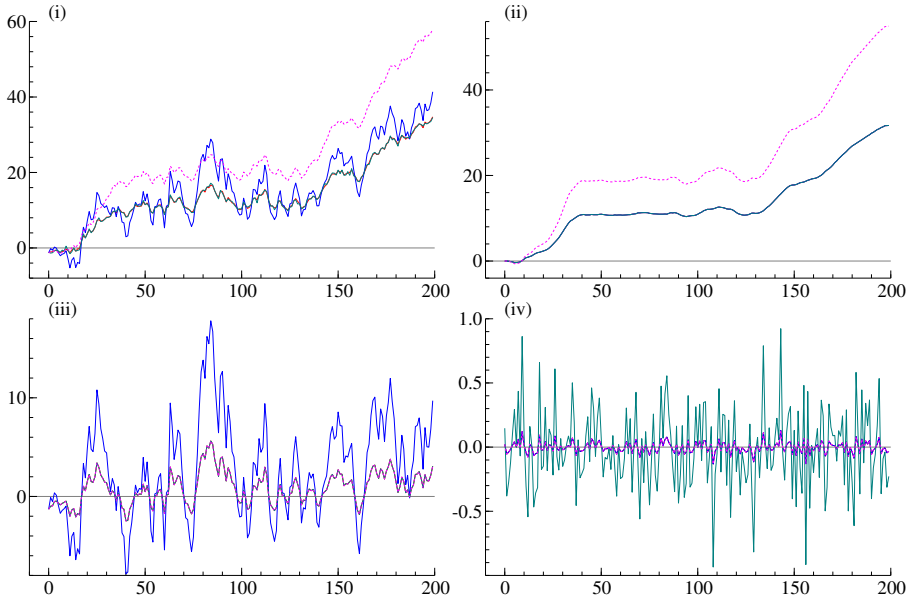


Figure 14: Simulations of the symmetric trend-cycle decomposition model: (i) simulated samples y_t, (ii) simulated trend components μ_t, (iii) simulated cycle components ψ_t, (iv) simulated irregular components ϵ_t. For the red line we have used $\sigma^2_\kappa = 0.5, \sigma^2_\epsilon = 0.002, \sigma^2_\zeta = 0.01$, for the blue line $\sigma^2_\kappa = 5, \sigma^2_\epsilon = 0.002, \sigma^2_\zeta = 0.01$, for the green line $\sigma^2_\kappa = 0.5, \sigma^2_\epsilon = 0.1, \sigma^2_\zeta = 0.01$, and for the dotted purple line $\sigma^2_\kappa = 0.5, \sigma^2_\epsilon = 0.002, \sigma^2_\zeta = 0.03$.



3 Application: Cambridge boat race

The Boat Race is an annual contest between two rowing crews from Oxford and Cambridge universities. The data set contains the results of the boat race contest from the first race in 1829 until 2015. The data is binary, meaning that 0 is a win for Oxford and 1 is a win for Cambridge win. In total, we have 159 observations. The missing observations include, amongst others, the 1st World War and 2nd World War.

Let y_t be the binary variable indicating the winning variable. We have the following specification for the winning probability

$$p(y_t|\theta_t;\psi) \sim \text{Bin}\left(\pi_t = \frac{\exp\theta_t}{1+\exp\theta_t}\right). \tag{6}$$

Note that this binary density with the stochastic time-varying probability π_t can be expressed as

$$\begin{aligned}\log p(y_t|\pi_t) &= \log\left(\pi_t^{y_t}(1-\pi_t)^{1-y_t}\right)\\ &= y_t\log\left(\frac{\pi_t}{1-\pi_t}\right) - \log\left(1+\frac{\pi_t}{1-\pi_t}\right)\\ &= y_t\theta_t - \log(1+\exp\theta_t) = \log p(y_t|\theta_t),\end{aligned} \tag{7}$$

where the signal $\theta_t = \log(\pi_t) - \log(1-\pi_t)$. We consider two different specifications for the signal θ_t, i.e.

$$\theta_t = \mu_t, \qquad \mu_{t+1} = \mu_t + \eta_t, \qquad \eta_t \sim \mathcal{NID}(0,\sigma_\eta^2), \tag{8}$$

$$\theta_t = \mu + u_t, \qquad u_{t+1} = \phi u_t + \eta_t, \qquad \eta_t \sim \mathcal{NID}(0,\sigma_\eta^2). \tag{9}$$

For the random walk, equation (8), we will use $\sigma_\eta^2 = 0.5$ and for the constant plus AR(1), equation (9), we use $\mu = 0.5, \phi = 0.9, \sigma_\eta^2 = 0.05$.

3.1 Mode estimation

For the linear Gaussian model state space model, we have that $p(Y_n|\theta) = \mathcal{N}(\theta, H)$ and $p(\theta) = \mathcal{N}(\mu, \Psi)$ where $\mu = c + Zd$ and $\Psi = Z\Omega Z'$. In the case of the linear Gaussian model we have that the mode of $p(\theta, |Y_n)$ is equal to the mean of $Y_n = \theta + \epsilon$. Hence, the mode $\hat{\theta} = \mathbb{E}(\theta|Y_n)$ can be estimated using the smoothing algorithm. Using the regression lemma, we obtain

$$\hat{\theta} = \mathbb{E}(\theta|Y_n) = \mu + \Psi\Sigma^{-1}(Y_n - \mu) = (H^{-1} + \Psi^{-1})^{-1}(H^{-1}Y_n + \Psi^{-1}\mu),$$

by the Woodbury matrix identity.

We use the above result to estimate the mode for non-Gaussian state space models. In particular, we have that the mode for the non-Gaussian observation density $p(Y_n|\theta)$ an estimate for the mode $\hat{\theta} = \arg_\theta \max p(\theta|Y_n) = \arg_\theta \max \log p(\theta|Y_n)$. The log-density can be rewritten as

$$\log p(\theta|Y_n) = \log\left(\frac{p(Y_n|\theta)p(\theta)}{p(Y_n)}\right) = \log p(Y_n|\theta) + \log p(\theta) - \log p(Y_n).$$

We will adopt the Newton-Raphson method where for a given guess g of mode $\hat{\theta}$, a new guess is given by

$$g^+ = g - \left[\ddot{p}(\theta|Y_n)\big|_{\theta=g}\right]^{-1}\dot{p}(\theta|Y_n)\big|_{\theta=g}, \tag{10}$$

where step-length is one and

$$\dot{p}(\cdot|\cdot) \equiv \frac{\partial \log p(\cdot|\cdot)}{\partial\theta}, \qquad \ddot{p}(\cdot|\cdot) \equiv \frac{\partial^2 \log p(\cdot|\cdot)}{\partial\theta\partial\theta'}.$$

Since it holds that $\log p(\theta) = \mathcal{N}(\mu, \Psi)$, we obtain

$$\dot{p}(\theta|Y_n) = \dot{p}(Y_n|\theta) + \dot{p}(\theta) = \dot{p}(Y_n|\theta) - \Psi^{-1}(\theta - \mu), \quad \ddot{p}(\theta|Y_n) = \ddot{p}(Y_n|\theta) + \ddot{p}(\theta) = \ddot{p}(Y_n|\theta) - \Psi^{-1}. \quad (11)$$

If we substitute (11) in the Newton-Rhapson update (10), we obtain

$$\begin{aligned} g^+ &= g - \left[\ddot{p}(Y_n|\theta)\big|_{\theta=g} - \Psi^{-1}\right]^{-1} \left(\dot{p}(Y_n|\theta)\big|_{\theta=g} - \Psi^{-1}(g-\mu)\right) \\ &= \left[\Psi^{-1} + A^{-1}\right]^{-1} (\Psi^{-1}\mu + A^{-1}z), \end{aligned} \quad (12)$$

where

$$A \equiv -\left[\ddot{p}(Y_n|\theta)\big|_{\theta=g}\right]^{-1}, \qquad z \equiv g + A\dot{p}(Y_n|\theta)\big|_{\theta=g}. \quad (13)$$

We see that we can compute the mode of the non-Gaussian model by considering a Gaussian model and setting $Y_n = z$ and $H = A$. Hence we can use the KFS procedure for the non-Gaussian model.

3.2 Importance sampling

In order to compute the conditional mean and conditional variance of the signal, we apply importance sampling. Let $x(\theta)$ be a function of θ, for an importance density $g(\theta|Y_n)$ we have that

$$\begin{aligned} \bar{x} &= \int x(\theta) p(\theta|Y_n) d\theta \\ &= \int x(\theta) \frac{p(\theta|Y_n)}{g(\theta|Y_n)} g(\theta|Y_n) d\theta \\ &= \mathbb{E}_g \left[x(\theta) \frac{p(\theta|Y_n)}{g(\theta|Y_n)} \right]. \end{aligned}$$

By applying Bayesťheorem and using the fact that $x(\theta) = 1 \Rightarrow \bar{x} = 1$, we obtain

$$\bar{x} = \frac{\mathbb{E}_g\left[x(\theta) w(\theta, Y_n)\right]}{\mathbb{E}_g\left[w(\theta, Y_n)\right]}, \qquad w(\theta, Y_n) = \frac{p(Y_n|\theta)}{g(Y_n|\theta)}.$$

We evaluate these expectation by Monte Carlo estimation, that is

$$\hat{x} = \frac{\sum_{i=1}^N x(\theta^{(i)}) w_i}{\sum_{i=1}^N w_i}, \qquad w_i = \frac{p(Y_n|\theta^{(i)})}{g(Y_n|\theta^{(i)})},$$

where the importance draws $\theta^{(1)}, \ldots, \theta^{(n)}$ are independent and obtained via simulation smoothing from importance density $g(\theta|Y_n)$. Under certain conditions it holds that $\hat{x}$ converges to $\bar{x} = \mathbb{E}(x(\theta)|Y_n)$ in probability as $N \to \infty$.

We use the estimated mode to construct the importance density. More specifically, we take the $g(\theta|Y_n)$ as Gaussian density with a mean (or mode) that equals the mode of $p(\theta|Y_n)$ and a variance that equals the curvature (variance) around the mode of $p(\theta|Y_n)$, which we obtain by KFS.

3.3 Results

For the binary density with the stochastic time-varying probability π_t, we have that $Y_n \sim p(Y_n|\theta)$, where by conditional independence it holds that

$$p(Y_n|\theta) = \prod_{t=1}^n p(y_t|\theta_t) = \prod_{t=1}^n \exp(y_t\theta_t - \log(1 + \exp\theta_t)), \quad (14)$$

and

$$\dot{p}(y_t|\theta_t) = \frac{\partial \log p(y_t|\theta_t)}{\partial \theta_t} = y_t - \frac{\exp\theta_t}{1+\exp\theta_t}, \qquad \ddot{p}(y_t|\theta_t) = \frac{\partial^2 \log p(y_t|\theta_t)}{\partial \theta_t^2} = -\frac{\exp\theta_t}{(1+\exp\theta_t)^2}. \tag{15}$$

We proceed with the mode estimation as described in Subsection 3.1. That is, we take an initial guess g for the mode and consider the linear Gaussian model with

$$Y_n = g + [\ddot{p}(Y_n|\theta)|_{\theta=g}]^{-1}\, \dot{p}(Y_n|\theta)|_{\theta=g}, \qquad H = -\left[\ddot{p}(Y_n|\theta)|_{\theta=g}\right]^{-1}.$$

In particular, we apply the KFS to the following model

$$z_t = \theta_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, H_t), \tag{16}$$

with

$$z_t = g + \ddot{p}(y_t|\theta_t)^{-1}\dot{p}(y_t|\theta_t), \qquad H_t = -[\ddot{p}(y_t|\theta_t)]^{-1},$$

and where we use specification (8), and (9) for the signal θ_t. Note that we modify the KFS to handle the missing observations correctly by letting $P_t \to \infty$ such that $K_t \to 0$. The smooth estimate $\hat{\theta}$ from this model obtained by KFS is set equal to g^+ in the Netwon-Rhapson algorithm (12). We replace g by g^+ until we have reached convergence. We use a stopping criteria of $\max_{1,\dots,t}|g^+ - g| < \epsilon = 0.00001$ and find that for the AR model we have convergence after 6 iterations and for the RW model after 5 iterations. In both cases the initial value for the mode (g) is a vector of zeros.

In Figure 15 we have plotted the mode, the data set with the missing observations and time varying probability π_t. It can be seen that the mode is constantly drawn towards the winning party: winnings for Cambridge let the mode increase while winnings for Oxford let the mode decrease. We observe a similar pattern for the time-varying probability. For example, in the period 1920 to 1935 Cambridge won the race multiple times in a row and hence the probability π_t moves towards 1, whereas in the period 1975 to 1990 the probability moves towards zero as Oxford won for a number of consecutive times. The relationship between the mode and the time-varying probability is that a mode of $\hat{\theta}_t > 0$ implies that $\pi_t > 0.5$.

We see that the random walk specification for θ_t leads to more extreme estimates for the mode compared to the constant plus AR(1) specification. This can be explained by the persistence of the AR(1) model as $|\phi| < 1$ in the AR(1) model. The stationarity causes the process to revert back to its mean of $\mu = 0.5$. Moreover, the AR(1) specification has a much smaller variance σ_η^2. This implies that there is less deviation in the signal. This is also what we observe in the time-varying probability. Namely, it is more stable compared to the random walk case. Consequently, since Cambridge has slightly more wins compared to Oxford (81 against 79), the AR(1) model slightly prefers Cambridge over Oxford.

In Figure 16 we show simulation draws from the density $g(\theta|Y_n)$ by simulation smoothing. For both the constant plus AR(1) and random walk specification of θ_t we show a single draw and the mean of 10.000 draws. We observe that a single draw is close to the estimated mode and that the mean of the 10.000 draws converges to the estimated mode. The latter is because this is the expectation of the importance density

In Figure 17 we show the estimated conditional mean and variance of the signal obtained by importance sampling, see Subsection 3.2. For both specifications of θ_t we see that the conditional mean converges to the estimated mode.

Under the given parameters for the specification of θ_t, we can investigate which model has the best in-sample performance. The model indicates the right outcome if $y_t = 1$ for $\pi_t > 0.5$ and $y_t = 0$ for $\pi_t < 0.5$. We find that the random walk indicates 132 out of 160 times the right outcome, i.e. 82.50%, while the constant plus AR(1) model indicates the correct outcome only 117 times or 73.125%. To see which model is better in terms of forecasting, we recursively compute sixteen 16 one-step ahead forecasts for the years 2001 to 2016 by treating the forecast as a missing value. We find that the random

Figure 15: Data (black plus), missing observations (green cross), mode (red line) and time-varying probability π_t (blue line): (i) constant plus AR(1) specification for θ_t, (ii) random walk specification for θ_t.

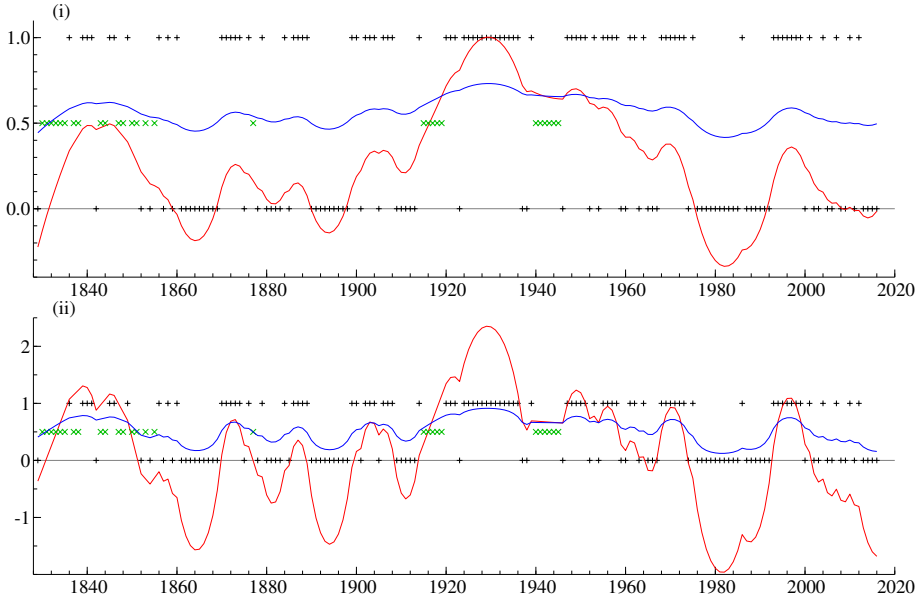
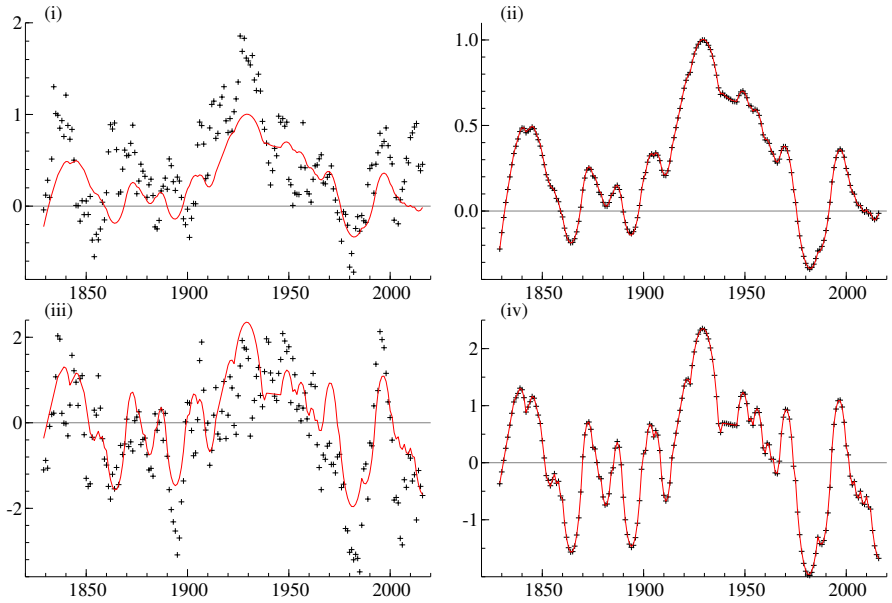
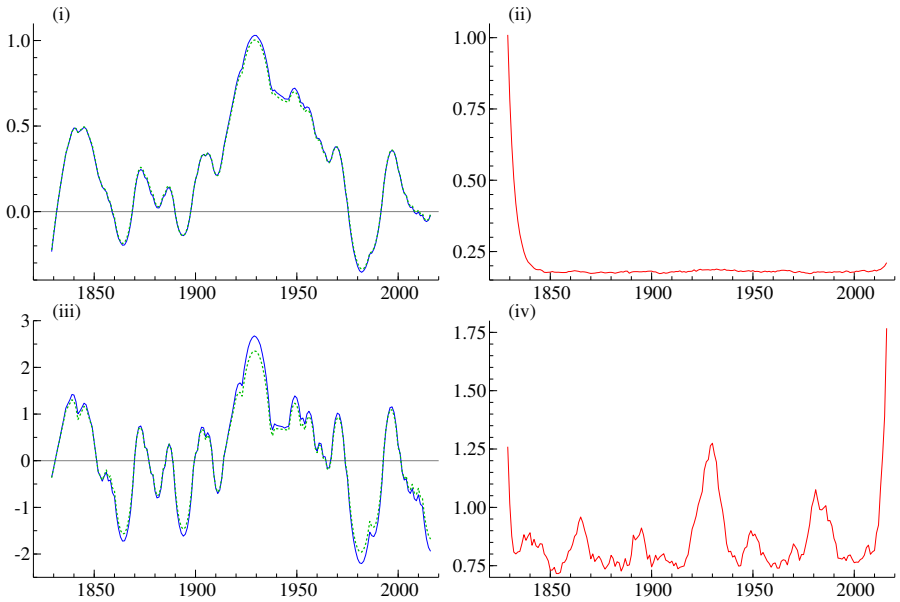


Figure 16: Simulated draw from $g(\theta|Y_n)$ (dots) and estimated mode (line): (i) one draw using constant + AR specification for θ_t, (ii) mean of 10,000 draws using constant + AR specification for θ_t, (iii) one draw using random walk specification for θ_t, (iv) mean of 10,000 draws using random walk specification for θ_t.



walk specification with 8 out of 16 forecasts corrects (50.00%) outperforms the AR(1) specification with

Figure 17: Importance sampling estimated conditional mean (blue), conditional variance (red) and estimated mode (green dotted line): (i) conditional mean and mode for constant + AR(1) specification of θ_t, (ii) conditional variance for constant + AR(1) specification of θ_t, (iii) conditional mean and mode for random walk specification of θ_t, (iv) conditional variance for random walk specification of θ_t.



only 5 out of 16 forecasts correct (31.25%).

Since the constant plus AR(1) specification for θ_t nests the random walk model, i.e. when $\mu = 0, \phi = 1, \sigma_\eta^2 = 0.5$, we experiment with different parameters for this model. As we find that the model for the random walk outperforms the AR(1) specification, we use the fact that high persistence, i.e. ϕ close to 1, is preferred. Note that since in the constant + AR(1) model we have $\mu = 0.5$, we overestimate the odds for a Cambridge win a bit. Since a positive μ suggests Cambridge has in general more chance to win than Oxford, because θ_t has unconditional mean μ and $\theta > 0$ implies $\pi_t > 0.5$. This is visible in Figure 15. Consequently, we would expect that setting $\mu = 0$ is better. Indeed, we find that the in-sample performance increases to 75.625%. If we decrease the parameter of ϕ towards zero, we find that model responds much quicker since there is more randomness in the model. Consequently, one has to make a trade-off between more randomness (high σ_η^2 and low ϕ) that tracks the actual wins more closely but has less predictive power since θ_t is less persistent, or less randomness which has the opposite results.

4 Final Assignment: Stochastic Volatility and NAIS

In the final assignment we employ the Stochastic Volatility (SV) model to analyse the historical (daily) prices of the S&P500 index from 21-03-2005 to 19-03-2015.

4.1 Part (a)

The time series are in levels, but we compute the daily returns for our analysis. In order to avoid numerical complications we take the difference of the logs rather than the ratio of the lags and multiply it by 100 to obtain the percentage growth

$$r_t^* = 100 * (\log(p_t/p_{t-1}) = 100 * (\log(p_t) - \log(p_{t-1})), \qquad t = 1, \ldots, T,$$

where $\{p_t\}$ is the level data and r_t^* the returns. The Stochastic Volatility (SV) model we employ is a mean zero model. In order to avoid taking logs of zero in part (b) we demean the return series, i.e.

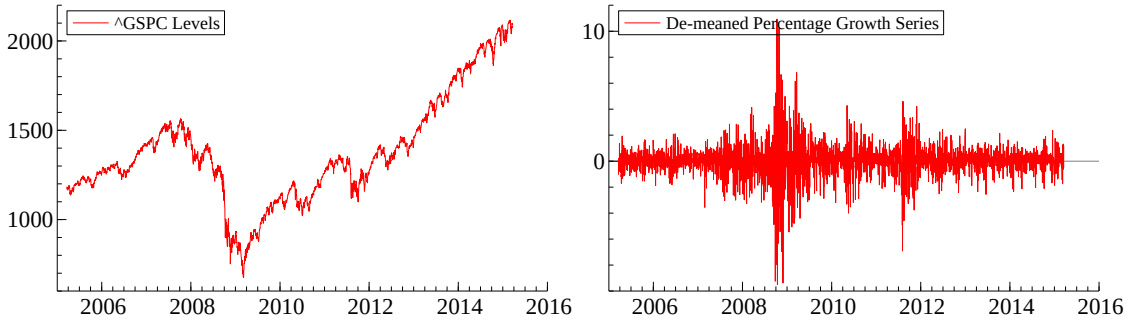
$$r_t = r_t^* - \frac{1}{T}\sum_{t=1}^{T} r_t^*.$$

Table 1: Descriptive statistics of ˆGSPC series.

	Observations	Mean	Max	Min	St. Dev.	Skewness	Kurtosis
ˆGSPC	2516	-4.3942e-16 (≈ 0)	10.935	-9.4921	1.13653	-0.33287	13.8856

In Table 1 we have tabulated the descriptive statistics of the series and in Figure 18 we have displayed the levels series and the demeaned percentage growth series. From the graph the financial crisis is clearly visible, as well as clustering of volatility. From the descriptive statistics we observe that the demeaned percentage growth series indeed has a mean of approximately zero, furthermore the data seems to be skewed with quite some mass in the tails. The standard deviation is clearly not negligible as well. The minimum values and the maximum values are attained in the financial crisis.

Figure 18: The levels and the de-meaned percentage growth series



4.2 Part (b)

We consider the zero mean Stochastic Volatility model

$$r_t^* - \mu = y_t = \sigma_t \varepsilon_t, \qquad \sigma_t^2 = \exp(h_t), \qquad \varepsilon_t \sim \mathcal{NID}(0, 1),$$

where the log-volatility h_t follows a stationary autoregressive process, given by

$$h_{t+1} = \omega + \phi h_t + \sigma_\eta \eta_t, \qquad \eta_t \sim \mathcal{NID}(0,1),$$
$$h_1 \sim \mathcal{N}(\omega/(1-\phi), \sigma_\eta^2/(1-\phi^2)),$$

with ε_t and η_t mutually and serially uncorrelated. The three unknown parameters are ω, ϕ and σ_η. Here ω is the intercept of the AR model, ϕ is the amount of persistence in data, which determines the volatility clustering, and σ_η determines the amount of volatility in the volatility equation.

The SV model can be made linear by transforming the returns $r_t = y_t$ to $x_t = \log y_t^2$, where we avoid taking logs of zeros by demeaning the series, see part (a). Hence, we obtain

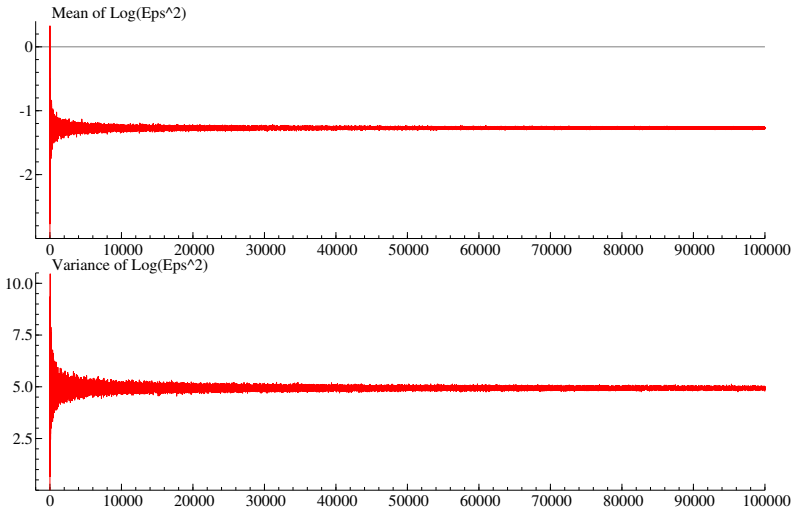
$$\begin{aligned} x_t &= \log y_t^2 = \log\left((\sigma_t \varepsilon_t)^2\right) \\ &= \log\left(\exp(h_t)\varepsilon_t^2\right) \\ &= h_t + \log(\varepsilon_t^2). \end{aligned}$$

Recall that since $\varepsilon_t \sim N(0,1)$, we have that

$$\varepsilon_t^2 \sim \chi^2(1),$$
$$\xi_t = \log(\varepsilon_t^2) \sim \log \chi^2(1).$$

Given that $\log(\varepsilon_t^2)$ follows a $\log \chi^2(1)$, we must determine the mean and the variance of the $\log \chi^2(1)$ distribution, which well do using simulations. In the simulation we transform the standard normally distributed errors by first taking the square and then the log. We generate separate samples with size 1 to 10000 and plot the mean and variance for all samples. The plots are found in figure 19

Figure 19: Top: the convergence of the mean of the log chi-squared distribution with one degree of freedom. Bottom: the convergence of the variance of the log chi-squared distribution with one degree of freedom.

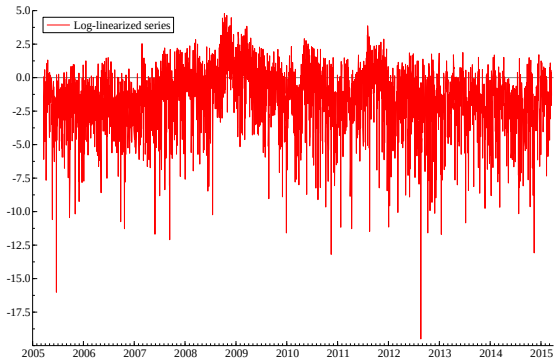


We find the the mean variance converge to -1.272 and 4.960 respectively, which is in line with the theory of Harvey et al. (1994), where they conclude that the mean and variance of $\log \varepsilon_t^2$ are -1.27 and $\pi^2/2 = 4.93$. Thus we can write the model as:

$$x_t = -1.27 + h_t + \xi_t, \qquad \xi_t \sim IID(0, \pi^2/2),$$
$$h_{t+1} = \omega + \phi h_t + \sigma_\eta \eta_t, \qquad \eta_t \sim NID(0,1).$$

In Figure 20 we plot the transformed series.

Figure 20: Plot of series $\log y_t^2$.



4.3 Part (c)

By assuming that $\{\varepsilon_t\}$ is normally distributed we have that $\log(\varepsilon_t^2) \sim IID(-1.27, \pi^2/2)$ is log chi-squared distributed with a mean and variance as found as described in part (b). Now if we let the disturbance term ξ_t be the demeaned log squared transformed disturbances, we obtain zero-mean non-Gaussian innovations. In order to perform Quasi-Maximum Likelikhood Estimation (QMLE) we make the assumption that $\{\xi_t\}$ is normally distributed with zero mean and the variance equal to the one found in part (b). Hence we assume $\xi_t \sim \mathcal{N}(0, \pi^2/2)$ thereby allowing us to estimate the unknown coefficients of our model by QMLE. In essence we estimate a misspecified model using MLE and as long as the quasi-likelihood we are maximizing is not overly simplified, the QMLE is consistent and asymptotically normal, but less efficient. We perform this QMLE with the aid of the Kalman Filter. So we have the model

$$\begin{aligned} x_t &= -1.27 + h_t + \xi_t, & \xi_t &\sim \mathcal{NID}(0, \pi^2/2), \\ h_{t+1} &= \omega + \phi h_t + \sigma_\eta \eta_t, & \eta_t &\sim \mathcal{NID}(0, 1), \end{aligned}$$

which we write in state space form with mean adjustments

$$\begin{aligned} y_t &= c_t + Z_t \alpha_t + \varepsilon_t, & \varepsilon_t &\sim \mathcal{N}(0, H_t), \\ \alpha_{t+1} &= d_t + T_t \alpha_t + R_t \eta_t, & \eta_t &\sim \mathcal{N}(0, Q_t) \\ \alpha_1 &\sim N(a_1, P_1), \end{aligned}$$

where

$$\begin{aligned} y_t &= x_t, & \alpha_t &= h_t \\ c_t &= -1.27, & d_t &= \omega \\ a_1 &= \omega/(1-\phi), & P_1 &= \sigma_\eta^2/(1-\phi^2) \\ Z_t &= 1, & H_t &= \pi^2/2, \\ T_t &= \phi, & R_t &= 1, & Q_t &= \sigma_\eta^2. \end{aligned}$$

From this we obtain the Kalman Filter recursions

$$\begin{aligned} v_t &= y_t - c_t - Z_t \alpha_t, & F_t &= Z_t P_t Z_t' + H_t \\ K_t &= T_t P_t Z_t' F_t^{-1} \\ a_{t+1} &= d_t + T_t a_t + K_t v_t, & P_{t+1} &= T_t P_t T_t' + R_t Q_t R_t' - K_t F_t K_t'. \end{aligned}$$

We estimate the parameter vector $\theta = (\omega, \phi, \sigma_\eta^2, \sigma_\varepsilon^2)$, with the aid of this Kalman Filter. Note that we only use $H_t = \pi^2/2$ as a starting value and estimate $H_t = \sigma_\xi^2$ in the parameter vector. We obtain estimates of F_t and v_t, which are then used to maximize the log-likelihood given by

$$\log P(Y_n) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}(\log|F_t| + v_t' F_t^{-1} v_t).$$

The estimation results are found in Table 2

Table 2: Estimates of θ by QMLE.

	Log Transformed series	
	$\hat{\theta}$	$SE(\hat{\theta})$
ω	-0.0051	0.0038
ϕ	0.9873	0.0047
σ_η^2	0.0282	0.0093
σ_ε^2	5.6419	0.1678

As expected, we find that the estimates show a high level of persistence (visible in $\hat{\phi}$) which explains the volatility clustering we observed in Figure 18.

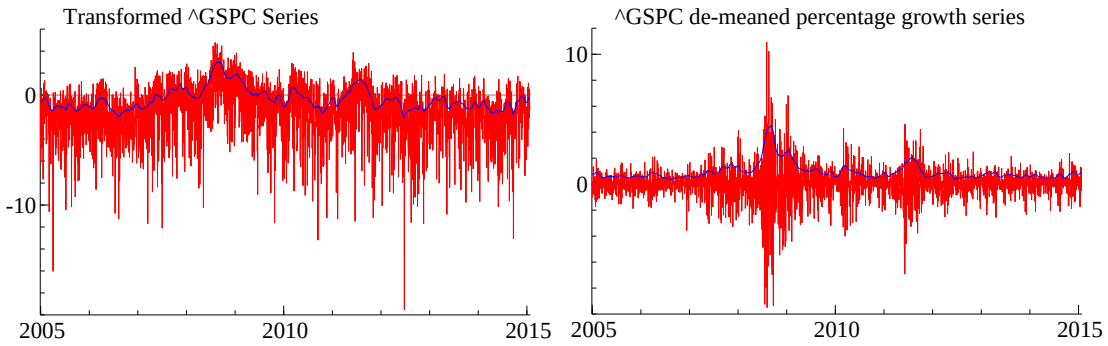
4.4 Part (d)

We use the QMLE estimates to obtain an estimate of h_t by computing $\mathbb{E}[h_t|x_1,\ldots,x_n] = \mathbb{E}[\alpha_t|y_1,\ldots,y_n]$ with the aid of KFS given by the recursions

$$L_t = T_t - K_t Z_t, \qquad K_t = T_t P_t Z_t F_t^{-1},$$
$$r_{t-1} = Z_t' F_t^{-1} v_t + L_t, \qquad N_{t-1} = Z_t' F_t^{-1} Z_t + L_t' N_t L_t,$$
$$\hat{\alpha}_t = a_t + P_t r_{t-1}, \qquad V_t = P_t - P_t N_{t-1} P_t,$$

for $t = 1,\ldots,n$, $r_n = 0$ and $N_n = 0$.

Figure 21: KFS applied to the series where the blue line is smoothed state $\{\hat{\alpha}_t\}$. Left: plot of $\log y_t^2$ against the KFS mean. Right: plot of y_t against $\sqrt{\exp(\hat{\alpha}_t)}$.



In Figure 21 we have graphed the results of the KFS applied to the log transformed series and the percentage growth series.

4.5 Part (e)

Using the same QMLE estimated as before we compute the mode for h_t using the mode estimation method, see Durbin and Koopman (2012). For this purpose we use the original non-linear SV model. From the conditional observation density

$$\log p(y_t|h_t) = -\frac{1}{2}[\log 2\pi\sigma^2 + h_t + y_t^2 \exp(-h_t)].$$

we obtain the first and second order partial derivative

$$\dot{p}_t = -\frac{1}{2}[1 - y_t^2 \exp(-h_t)],$$
$$\ddot{p}_t = -\frac{1}{2} y_t^2 \exp(-h_t).$$

According to Durbin and Koopman (2012) we can expand $\dot{p}_t$ around $\tilde{h}_t$ to get $\dot{p}_t = \dot{p}_t + \ddot{p}_t(h_t - \tilde{h}_t)$ such that the mode can be estimated iteratively by applying the KFS to

$$\tilde{y}_t = h_t - \ddot{p}_t^{-1}\dot{p}_t,$$

where $-\tilde{H}_t = \ddot{p}_t^{-1}$ is the variance matrix. Thus in effect we design a linear approximating model

$$\tilde{y}_t = h_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \tilde{H}_t),$$
$$h_{t+1} = \omega + \phi h_t + \sigma_\eta \eta_t, \qquad \eta_t \sim \mathcal{N}(0,1).,$$

from which we iteratively compute the mode $\tilde{h}_t$ by applying KFS to $\tilde{y}_t$ where we update $\tilde{y}_t$ and $\tilde{H}_t$ in each iteration to get a new guess for $\tilde{h}_t$, defined as:

$$\tilde{y}_t = \tilde{h}_t - \frac{1}{2}\tilde{H}_t + 1,$$
$$\tilde{H}_t = \frac{2\exp(\tilde{h}_t)}{y_t^2}.$$

$\tilde{y}_t$ and $\tilde{H}_t$ are initialized by letting $\tilde{y}_t = \log y_t^2$ and $\tilde{H}_t = 2$. We stop whenever we have $|\tilde{h}_t^{new} - \tilde{h}_t| < \epsilon$ for $\epsilon = 10^{-10}$.

In Figure 22 we have graphed the results of the mode estimation. This graph is quite similar to Figure 21. In general, we expect the mode to react strongly to the jumps in series. However, since the series does not show drastic jumps we do not expect large differences between the mode and KFS mean estimation.

4.6 Part (f)

In this part we estimate $\hat{h_t} = \mathbb{E}[h_t|y_1, \ldots, y_n]$ using by Numerically Accelerated Importance Sampling (NAIS), see Koopman et al. (2015). Note that here we let ψ be the vector of hyper-parameters, which we previously denoted as θ. We let $\theta_t = h_t$ be the signal at time t and write $g(y_t|\theta_t;\psi)$ as:

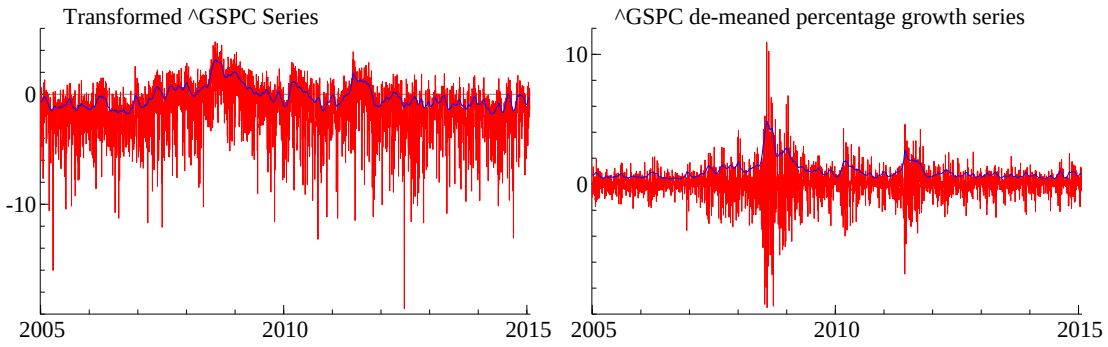
$$g(y_t|\theta_t;\psi) = \exp(a_t + b_t'\theta_t - \frac{1}{2}\theta_t' C_t \theta_t), \tag{17}$$

where a_t is the normalizing constant for $g(y_t, \theta_t; \psi)$, and b_t and C_t importance sampling parameters which we aim to optimize. If we consider the model:

$$y_t^* = \theta_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, C_t^{-1}), \tag{18}$$

where $y_t^* = C_t^{-1}, b_t$ (as with SPDK), then it can be easily verified that $g(y_t|\theta_t;\psi) \equiv g(y_t^*|\theta_t;\psi)$. Hence we have that that $g(\theta_t|y;\psi) = g(\theta_t|y^*;\psi) = \exp\left[-\frac{1}{2}V_t^{-1}(\theta_t - \hat{\theta}_t)^2\right]/\sqrt{2\pi V_t}$ is analytically available where V_t and $\hat{\theta}_t$ are obtained from KFS applied to the linear Gaussian model of Equation 18.

Figure 22: Mode estimation applied to the series where the blue line is the estimated mode $\{\tilde{h}_t\}$. Left: plot of $\log y_t^2$ against the mode. Right: plot of y_t against $\sqrt{\exp(\tilde{h}_t)}$.



Following Richard and Zhang (2007), we have that the minimisation of the variance of the log weights $\omega^*(\theta, y; \psi)$

$$\min_{\chi} \int \lambda^2(\theta, y; \psi)\omega^*(\theta, y; \psi)g(\theta|y; \psi)d\theta,$$

can be approximated by minimizing the parameters $\chi = \{b_1, \ldots, b_n, C_1, \ldots, C_n\}$ separately for each $\chi_t = \{b_t, C_t\}$, thus we have

$$\min_{\chi_t} \int \lambda^2(\theta_t, y_t; \psi)\omega^*(\theta_t, y_t; \psi)g(\theta_t|y; \psi)d\theta, \tag{19}$$

where $\omega^*(\theta_t, y_t; \psi) = \frac{p(y_t|\theta_t;\psi)}{g(y_t^*|\theta_t;\psi)}$ which we can numerically integrate using Gauss-Hermite Quadrature

$$\min_{\chi_t} \sum_{j=1}^{M} \lambda^2(\tilde{\theta_{tj}}, y_t; \psi)\omega^*(\tilde{\theta_{tj}}, y_t; \psi)g(\tilde{\theta_{tj}}|y^*; \psi)h(z_j)e^{z_j^2}. \tag{20}$$

Here $h(z_j)$ are the weights of the nodes and z_j the nodes, i.e. the roots of the Hermite polynomials. For $h(z_j)$ we employ both $h(z_j) = e^{-\frac{1}{2}z_j^2}$ from Monahan (2011) as well as $\frac{2^{n-1}n!\sqrt{\pi}}{n^2[H_{n-1}(x_i)]^2}$ from Abramowitz and Stegun (1964) with $H_{n-1}(x_i)$ the $n-1$th Hermite polynomial and the convergence results are equal. In Equation (20) we have

$$\tilde{\theta_{tj}} = \hat{\theta}_t + V_t^{1/2} z_j,$$

hence it follows that

$$g(\tilde{\theta_{tj}}|y^*; \psi) = \exp\left[-\frac{1}{2}z_j^2\right] / \sqrt{2\pi}, \quad t = 1, \ldots, n.$$

The minimization of Equation (20) for $\tilde{\theta}_{tj} \in \mathbb{R}$ boils down to WLS regression of the linear equation

$$\log p(y_t|\tilde{\theta_{tj}}; \psi) = [1, \tilde{\theta_{tj}}, -\frac{1}{2}\tilde{\theta_{tj}^2}]\beta + v_i,$$

where $w_{t,ii} = \exp(\frac{1}{2}z_j^2)h(z_j)\omega^*(\tilde{\theta_{tj}}, y_t; \psi)\sqrt{2\pi}$ is the diagonal of the weighting matrix and

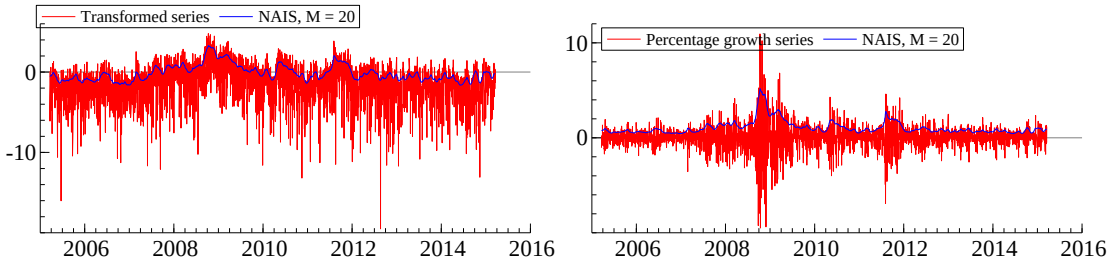
$$\log p(y_t|\tilde{\theta_{tj}}; \psi) = -\frac{1}{2}\left[\log 2\pi + \tilde{\theta_{tj}} + y_t^2 \exp(-\tilde{\theta_{tj}})\right].$$

Note that we can remove the constant independent of z_j, $\sqrt{2\pi}$ and according to Richard and Zhang (2007) removing $\omega^*(\tilde{\theta_{tj}}, y_t; \psi)$ from the regression weights does not results in much numerical efficiency so we adopt:

$$w_{t,ii} = \exp\left(\frac{1}{2}z_j^2\right)h(z_j).$$

Then we have that $\chi_t^{new} = \{b_t, C_t\} = \{\beta_2, \beta_3\}$ repeated for $t = 1, \ldots, n$ and we finally get $\chi^{new} = \{b_t, C_t\}_{t=1}^n$. We repeat the steps for χ^{new} and continue iteratively until convergence. We obtain χ^{final} which we use to construct $y_t^* = C_t^{-1,final} b_t^{final}$ and apply KFS to the model in Equation (18). We can do this because we expect to have $\log g(\theta|y)$ as close as possible to $\log p(\theta|y)$ and $\log g(\theta|y)$ is equal to $\log g(\theta|y^*)$. The result of the smoothed mean estimation using NAIS is graphed in Figure 23. From this figure it looks to be quite similar to the other estimations we have done.

Figure 23: NAIS estimation applied to the series. Left: plot of $\log y_t^2$ against NAIS smoothed mean estimation. Right: plot of y_t against NAIS smoothed mean.



4.7 Part (g)

We graph the three estimations from part (d), (e), and (f) in Figure 24. We see that the QMLE-KFS (KFS applied to the $\log y_t^2$) is the least responsive to shocks, and in fact looks as if it lags behind at certain periods. This can be a results of the model approximation, where the normality assumption violates the tail properties of the series. We see that the NAIS approximation is in general slightly higher than the mode estimation, but the dynamics of the mode estimate and NAIS estimate resemble each other much more. Another thing we notice is that when we increase the nodes M to 100 for instance we see that in Figure 25 the NAIS smoothed mean moves upward.

Figure 24: QMLE smoothed mean, mode estimation applied to the series, and NAIS. Top: plot of the untransformed estimates. Bottom: plot of transformed estimates using $\sqrt{\exp(\tilde{h}_t)}$, M=20

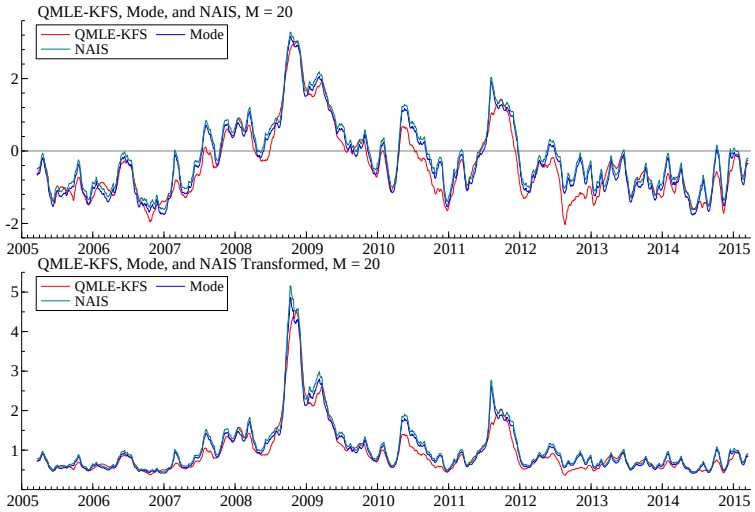
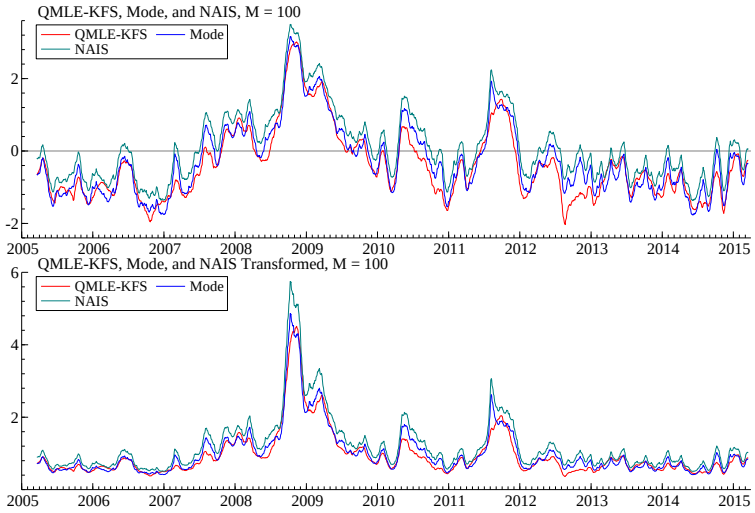


Figure 25: QMLE smoothed mean, mode estimation applied to the series, and NAIS. Top: plot of the untransformed estimates. Bottom: plot of transformed estimates using $\sqrt{\exp(\tilde{h}_t)}$, M=100



4.8 Part (h)

We consider the SV model with the Student's t-distribution

$$\begin{aligned} y_t &= \sigma_t \varepsilon_t, & \sigma_t^2 &= \exp(h_t), & \varepsilon_t &\sim \mathrm{t}(\nu) \\ & & h_{t+1} &= \omega + \phi h_t + \sigma_\eta \eta_t, & \eta_t &\sim \mathcal{NID}(0,1), \end{aligned}$$

where $\mathrm{t}(\nu)$ stands for the standardized student-t distribution with ν degrees of freedom. The AR(1) state equation remains unchanged. From part (b) we have that:

$$\begin{aligned} x_t &= c + h_t + \xi_t, & \xi_t &\sim IID(0, v), \\ h_{t+1} &= \omega + \phi h_t + \sigma_\eta \eta_t, & \eta_t &\sim \mathcal{NID}(0,1), \end{aligned}$$

where if we assumed the ϵ_t to be normally then $c = -1.27, v = \pi^2/2$ and where we used v as the starting value. Now that we have $\varepsilon_t \sim \mathrm{t}(\nu)$ we use the results of Harvey et al. (1994) to define c and v. That is, we let $\epsilon_t = \frac{\zeta_t}{\kappa_t^{1/2}}$ with $\zeta_t \sim \mathcal{N}(0,1)$ and $\kappa_t \sim \chi^2(\nu)$ and both are independent of each other. We obtain

$$\log \epsilon_t^2 = \log \zeta_t^2 - \log \kappa_t,$$

where we use the result that $\mathbb{E}[\kappa_t] = \psi'(\nu/2) - \log(\nu/2)$ and $\mathbb{V}ar[\kappa_t] = \psi'(\nu/2)$ where $\Gamma(\cdot)$ is the usual Gamma function and $\psi'(\cdot)$ is the trigamma function

$$\psi'(x) = \frac{d^2}{dx^2} \log \Gamma(x).$$

Hence we obtain $c = -1.27 - \psi'(\nu/2) + \log(\nu/2)$ and $v = \pi^2/2 - \psi'(\nu/2)$. We now have a approximating model that is linear but non-Gaussian. We can still apply the Kalman filter using the the minimum variance matrix property, see Durbin and Koopman (2012). From the Kalman filter we obtain F_t and v_t which we then use to maximize the log-likelihood

$$\log P(Y_n) = \log \Gamma\Big(\frac{\nu+1}{2}\Big) - \log \Gamma\Big(\frac{\nu}{2}\Big) - \frac{1}{2}\log(\pi(\nu-2)) - \frac{1}{2}\log(F_t) - \frac{(\nu+1)}{2}\log\Big(1 + \frac{v_t^2}{(\nu-2)F_t}\Big).$$

Note that initialisation parameters a_1, P_1 remain unchanged. Using the same state space form as in part (c) we estimate $H_t = \sigma_\xi^2$ along with the other hyper-parameters. From the results in Table 3 we see that indeed our estimates indicate fat tails in the series, hence it is beneficial to model the SV model using Student-t distributed errors.

Table 3: Estimates of θ by QMLE.

	Log Transformed series	
	$\hat{\theta}$	$SE(\hat{\theta})$
ω	0.0005	0.0057
ϕ	0.9866	0.0041
σ_η^2	0.0309	0.0088
σ_ε^2	5.7183	0.2815
ν	5.0032	0.0099

4.9 Part (i)

Now we estimate the hyper-parameters ψ and let $\theta_t = h_t$ be the signal, with $\theta = [\theta_1, \ldots, \theta_n]'$. We maximize the simulated likelihood using importance sampling. We are interested in the likelihood $p(Y_n; \psi)$ which we write as $p(Y_n)$ for convenience

$$p(Y_n) = \int p(\theta, Y_n) d\theta.$$

If we multiply and divide by the importance density $g(\alpha|Y_n)$ we obtain

$$\begin{aligned} p(Y_n) &= \int \frac{p(\theta, Y_n)}{g(\theta|Y_n)} g(\theta|Y_n) d\theta \\ &= g(Y_n) \int \frac{p(\theta, Y_n)}{g(\theta, Y_n)} g(\theta|Y_n) d\theta \\ &= g(Y_n) \int \frac{p(Y_n|\theta)}{g(Y_n|\theta)} g(\theta|Y_n) d\theta \\ &= g(Y_n; \psi) \mathbb{E}_g[\omega(\theta, Y_n)], \end{aligned}$$

because we assume $p(\theta) = g(\theta)$.

With the NAIS procedure we obtain $\chi = \{b_t, C_t\}_{t=1}^n$ from which we also have $y_t^* = C_t^{-1} b_t$, $t = 1, \ldots, n$, which we can model as in Equation (18). Thus we can sample $\{\theta^i\}_{i=1}^N \sim g(\theta|Y_n^*; \psi)$ using the simulation smoother of Koopman et al. (2015). We define $\omega^*(\theta^i, Y_n; \psi) := \dfrac{p(Y_n|\theta^i; \psi)}{g(Y_n^*|\theta^i; \psi)}$ where

$$p(Y_n|\theta^i; \psi) = \prod_{t=1}^n \frac{\Gamma[(\nu+1)/2]}{\sqrt{\pi(\nu-2)}\Gamma(\nu/2)} \exp\left[-\frac{1}{2}\theta_t^i\right] \left(1 + \frac{y_t^2 \exp\left[-\theta_t^i\right]}{\nu - 2}\right)^{\left(-\frac{\nu+1}{2}\right)},$$

and

$$\begin{aligned} g(Y_n|\theta^i; \psi) &= \prod_{t=1}^n (2\pi)^{-\frac{1}{2}} (C_t^{-1})^{-\frac{1}{2}} \exp(-\frac{(y_t^* - \theta_t^i)^2}{2C_t^{-1}}) \\ &= \prod_{t=1}^n \exp\left[-\frac{1}{2}(\log 2\pi - \log(C_t) + (y_t^* - \theta_t^i)^2 C_t)\right]. \end{aligned}$$

Let $\omega_i^* = \omega^*(\theta^i, Y_n; \psi)$, then we estimate $P(Y_n)$ using

$$\hat{p}(Y_n; \psi) = g(Y_n^*; \psi) \frac{1}{N} \sum_{i=1}^{N} \omega_i^*,$$

where we evaluate $g(Y_n^*; \psi)$ by applying KF to the linear model in Equation (18) and maximize

$$\log \hat{p}(Y_n; \psi) = \log g(Y_n^*; \psi) - \log N + \log(\sum_{i=1}^{N} \omega_i^*).$$
$$\propto \log g(Y_n^*; \psi) + \log(\sum_{i=1}^{N} \omega_i^*),$$

for argument ψ.

We could not get this method to work properly however. For instance, we tried using (i) $g(Y_n^*; \psi)$; (ii) $p(Y_n^*; \psi)$ where we evaluate $p(Y_n^*; \psi)$ using the $p(\cdot)$ density with KF applied to Y_n^* from NAIS; (iii) the latter with SPDK; (iv) updating the SPDK as well as NAIS parameters within the maximization. All of this did not work property. Consequently, we implemented $p(X_n; \psi)$ which is the same as the the QMLE density and we applied NAIS for the parameters of the weights of $g(Y_n^*|\theta^i; \psi)$. For the NAIS we used the parameters of the QMLE of part (h). We did not repeat the NAIS within the maximization loop, because this did not work. Therefore, in the end we have a mixture of NAIS and QMLE, where from the results in Table 4 it is clear that the QMLE strongly influences the estimates.

Table 4: Estimates of ψ by Importance Sampling.

	Log Transformed series	
	$\hat{\psi}$	$SE(\hat{\psi})$
ω	0.0160	0.0040
ϕ	0.9866	0.0041
σ_η^2	0.0310	0.0088
σ_ε^2	5.7178	0.2814
ν	5.0041	0.4993

4.10 Part (j)

We now employ the mode estimation, as well as the NAIS estimation for the Student-t SV model. Note that for the mode estimation $\dot{p}_t$ as well as $\ddot{p}_t$ are now different

$$\tilde{H}_t = \frac{1}{2} \left(1 + \frac{(\nu+1)\exp(-h_t) y_t^2}{(\nu + y_t^2 \exp(-h_t))^2} \right)^{-1}$$
$$\tilde{y}_t = h_t - \frac{1}{2} \tilde{H}_t \left(1 - \frac{(\nu+1)\exp(-h_t) y_t^2}{(\nu + y_t^2 \exp(-h_t))} \right)$$

The results of the mode estimation are given in Figure 26.

For applying NAIS to the SV-t model we change $\log p(y_t|\tilde{\theta}_{tj}; \psi)$ to

$$\log p(y_t|\tilde{\theta}_{tj}; \psi) =$$
$$\log \Gamma((\nu+1)/2)) - \log \Gamma(\nu/2)) - \frac{1}{2}\log(\pi(\nu-2)) - \frac{1}{2}\tilde{\theta}_{tj} - \frac{\nu+1}{2} \log\left(1 + \frac{y_t^2 \exp(-\tilde{\theta}_{tj})}{\nu - 2}\right)$$

In Figure 27 we plot the NAIS smoothed mean.

Figure 26: SV-t Mode estimation applied to the series where the blue line is the estimated mode $\{\tilde{h}_t\}$. Left: plot of $\log y_t^2$ against the mode. Right: plot of y_t against $\sqrt{\exp(\tilde{h}_t)}$.

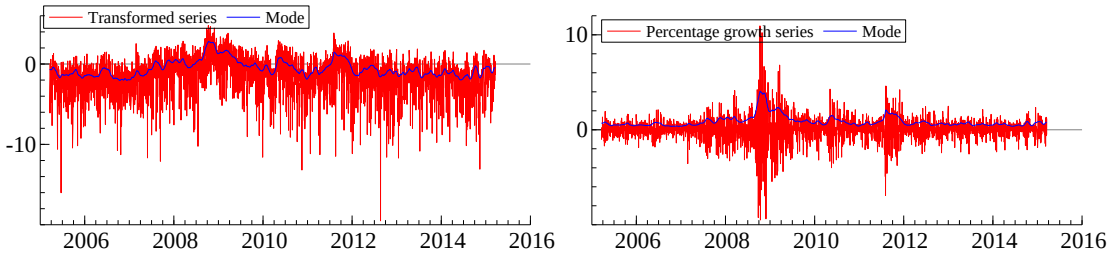
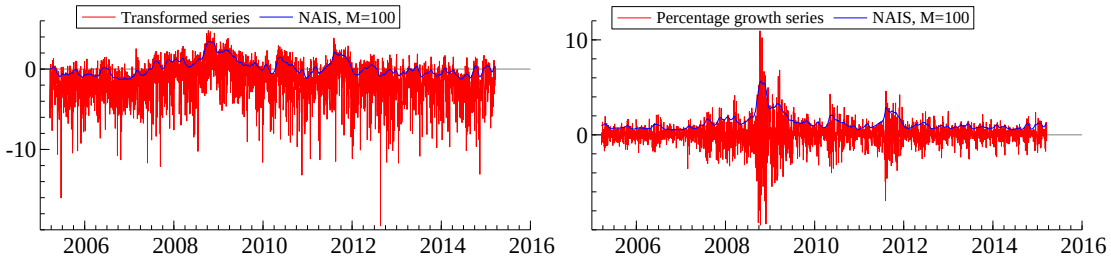


Figure 27: NAIS estimation of SV-t applied to the series. Left: plot of $\log y_t^2$ against NAIS smoothed mean estimation. Right: plot of y_t against NAIS smoothed mean.



4.11 Part (k)

In Figure 28 we have graphed both the mode estimation as well as the NAIS estimation of the SV-t model. We observe first that the mode seems to be lower in the SV-t model compared the Gaussian SV model. Furthermore if we look at the spikes around 2009 for instance we observe that for the SV-t model the spikes are less edged. This can indicate that the SV-t model copes slightly better with outliers than the Gaussian SV model. This is to be expected given the fatter tails of the student-t distribution.

4.12 Part (l)

We implement the bootstrap filter for the SV model with the Student-t disturbances. The details of the bootstrap filter where we perform resampling after each iteration are given in Subsection 1.8. In this application we have that

$$p\left(y_t|\tilde{\theta}_t^{(i)}\right) = \frac{\Gamma[(\nu+1)/2]}{\sqrt{\pi(\nu-2)}\Gamma(\nu/2)} \exp\left[-\frac{1}{2}\theta_t^{(i)}\right]\left(1+\frac{y_t^2\exp\left[-\theta_t^{(i)}\right]}{\nu-2}\right)^{\left(-\frac{\nu+1}{2}\right)},$$

and we draw from $p(\theta_t|\theta_{t-1}^{(i)})$ by using the AR(1) specification for $\theta_t = h_t$

$$h_{t+1} = \omega + \phi h_t + \sigma_\eta \eta_t, \qquad \eta_t \sim \mathcal{NID}(0,1),$$
$$h_1 \sim \mathcal{N}(\omega/(1-\phi), \sigma_\eta^2/(1-\phi^2)).$$

In Figure 29 we have showed the filtered estimate of the Kalman filter along with the estimates from the Bootstrap filter, the ESS and the filtered state variance of the Kalman filter and the variance

obtained from the bootstrap filter. Note that we have applied a transformation $\sqrt{\exp \theta_t}$ to obtain the stochastic volatility. The stochastic volatility obtained by the bootstrap filter is relatively higher compared to that of the Kalman filter, although they react similarly to shocks in the return series. Note however that we cannot benchmark the performance of the bootstrap filter to that of the Kalman filter, because the Kalman filtered states are obtained via an approximate linear model (with normally distributed disturbances), that may be further away from a possible true model than the SV-t.

Figure 28: SV-t, Mode estimation applied to the series, and NAIS. Top: plot of the untransformed estimates. Bottom: plot of transformed estimates using $\sqrt{\exp(\tilde{h}_t)}$, M=100

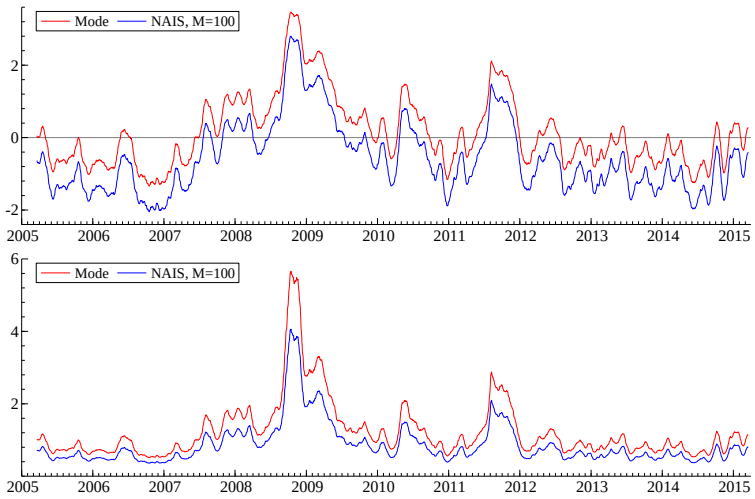
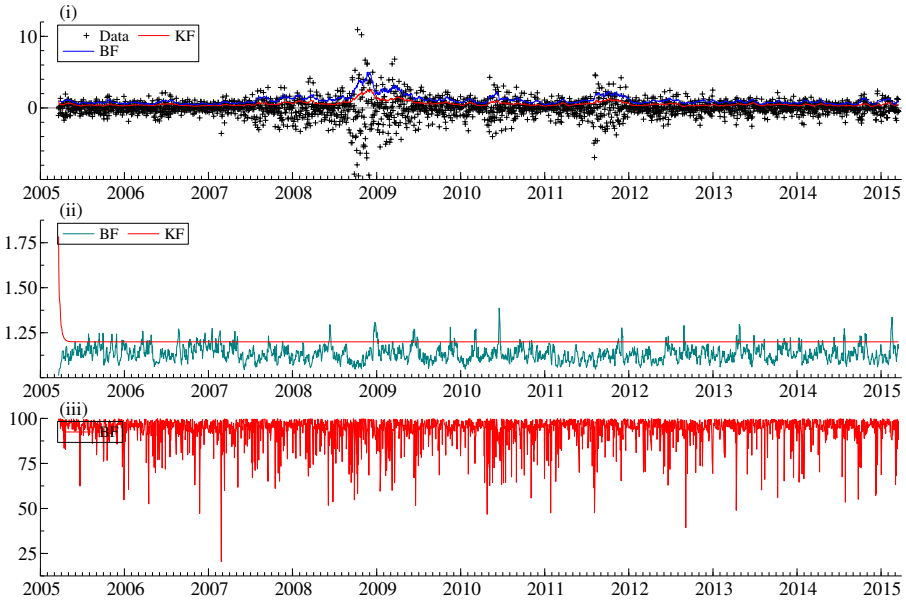


Figure 29: Kalman filter and Bootstrap filter (BF) with $N = 100$ particles: (i) the filtered state from Bootstrap filter and the filtered state from the Kalman filter, (ii) filtered state variance from the Bootstrap filter and the Kalman filter, (iii) the efficient sample size (ESS).



References

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55. Courier Corporation.

Durbin, J. and Koopman, S. (2012). *Time Series Analysis by State Space Methods: Second Edition*. Oxford Statistical Science Series. OUP Oxford.

Harvey, A., Ruiz, E., and Shephard, N. (1994). Multivariate stochastic variance models. *The Review of Economic Studies*, 61(2):247–264.

Koopman, S. J., Lucas, A., and Scharth, M. (2015). Numerically accelerated importance sampling for nonlinear non-gaussian state-space models. *Journal of Business & Economic Statistics*, 33(1):114–127.

Monahan, J. F. (2011). *Numerical methods of statistics*. Cambridge University Press.

Richard, J.-F. and Zhang, W. (2007). Efficient high-dimensional importance sampling. *Journal of Econometrics*, 141(2):1385–1411.