Lecture 6: State Space Models: part I

Geert Mesters (Universitat Pompeu Fabra, Barcelona GSE and VU Amsterdam)

Motivation

Motivation

- In these slides we introduce state space models
- This class of models includes many of the previously discussed models and includes many new models such as time-varying parameter models and dynamic factor models
- State space models are used in a variety of fields including economics, statistics and engineering

Local level model

These slides concentrate on the local level model

$$\begin{aligned} y_t &= \alpha_t + \epsilon_t \\ \alpha_{t+1} &= \alpha_t + \eta_t \end{aligned}$$

where

- y_t observed variable
- α_t unobserved trend
- ϵ_t, η_t disturbances

Linear Gaussian state space model

The local level model is the simplest example of the general linear Gaussian state space model:

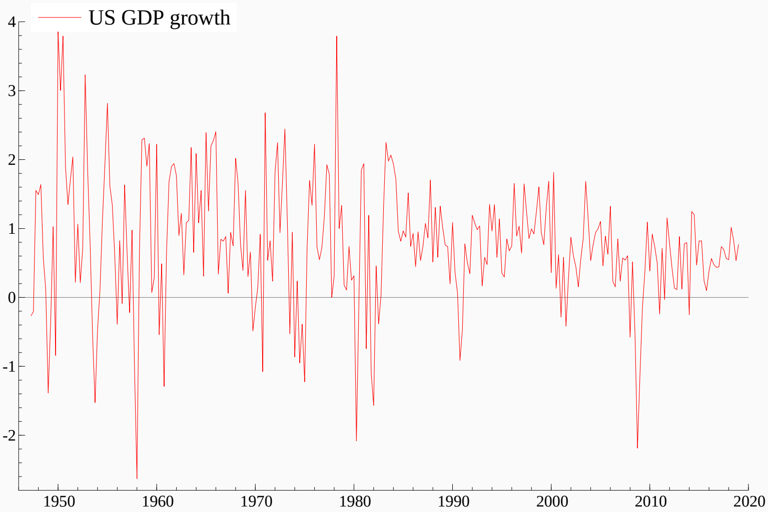
$$\begin{aligned} \boldsymbol{Y}_t &= \boldsymbol{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\epsilon}_t \\ \boldsymbol{\alpha}_{t+1} &= \boldsymbol{T}_t \boldsymbol{\alpha}_t + \boldsymbol{R}_t \boldsymbol{\eta}_t \end{aligned}$$

which will be discussed in the next lecture in detail.

- All intuition of local level model applies for general model

Running example

US GDP growth



Some observations

- Reason to believe that there is an underlying trend that moves slowly over time
- Additionally there are some deviations from the trend
- We are interested in extracting the trend from the time series

Methodology

Some comments

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

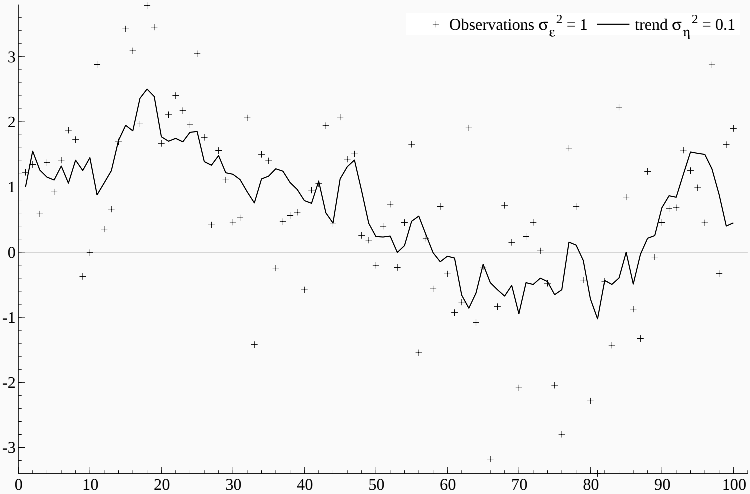
- The equation for y_t is the observation equation
- The equation for α_t is the state equation

Local level model

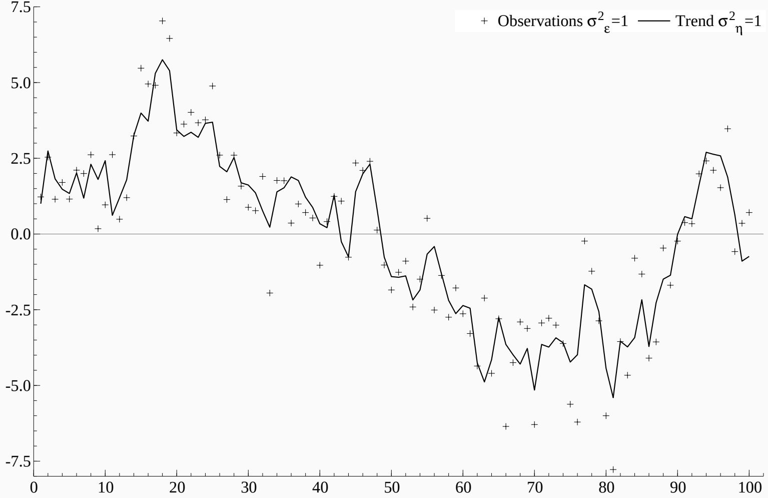
$$
\begin{aligned}
y_t &= \alpha_t + \epsilon_t & \epsilon_t &\sim N(0, \sigma_\epsilon^2) \\
\alpha_{t+1} &= \alpha_t + \eta_t & \eta_t &\sim N(0, \sigma_\eta^2)
\end{aligned}
$$

- Normality is imposed for convenience only; otherwise all could be stated in terms of best linear prediction (see lecture 3)
- We assume for now that the variances σ_ϵ^2 and σ_η^2 are known in the derivations.

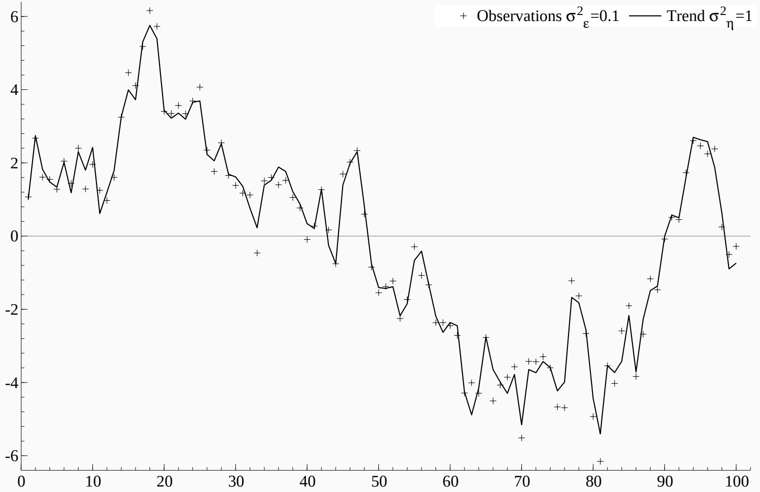
Simulated data local level



Simulated data local level



Simulated data local level



Econometrics of the local level model

We discuss

- Filtering
- Smoothing
- Missing observations
- Forecasting
- Initialization
- Parameter estimation
- Diagnostic checking

Filtering

Prediction and Filtering

The objective of filtering is to update our knowledge of the state when a new observation becomes available. To establish some convenient notation let $Y_{s:t} = (y_s, y_{s+1}, \ldots, y_t)'$ and define

Predictive estimates:
$$\begin{aligned} a_{t+1|t} &= \mathrm{E}(\alpha_{t+1}|Y_{1:t}) \\ P_{t+1|t} &= \mathrm{Var}(\alpha_{t+1}|Y_{1:t}) \end{aligned}$$

Filtered estimates:
$$\begin{aligned} a_{t|t} &= \mathrm{E}(\alpha_{t}|Y_{1:t}) \\ P_{t|t} &= \mathrm{Var}(\alpha_{t}|Y_{1:t}) \end{aligned}$$

The objective is to obtain these estimates by passing through the data from $t = 1$ until $t = T$ and updating the estimates recursively.

- This is done by an algorithm called the Kalman filter

Multivariate regression lemma

Theorem

Let x and y be jointly normally distributed with

$$\mathrm{E}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} \qquad \mathrm{Var}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}$$

then

$$\begin{aligned} \mathrm{E}(x|y) &= \mathrm{E}(x) + \mathrm{Cov}(x,y)\mathrm{Var}(y)^{-1}(y - \mathrm{E}(y)) \\ &= \mu_x + \tfrac{\sigma_{xy}}{\sigma_x^2}(y - \mu_y) \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(x|y) &= \mathrm{Var}(x) - \mathrm{Cov}(x,y)\mathrm{Var}(y)^{-1}\mathrm{Cov}(y,x) \\ &= \sigma_x^2 - \tfrac{(\sigma_{xy})^2}{\sigma_x^2} \end{aligned}$$

Predictions

We derive the prediction

$$\begin{aligned} a_{t+1|t} &= \mathrm{E}(\alpha_{t+1}|Y_{1:t}) \\ &= \mathrm{E}(\alpha_t + \eta_t|Y_{1:t}) \\ &= \mathrm{E}(\alpha_t|Y_{1:t}) = a_{t|t} \end{aligned}$$

Further, define the prediction variance

$$\begin{aligned} P_{t+1|t} &= \mathrm{Var}(\alpha_{t+1}|Y_{1:t}) \\ &= \mathrm{Var}(\alpha_t + \eta_t|Y_{1:t}) \\ &= \mathrm{Var}(\alpha_t|Y_{1:t}) + \sigma_\eta^2 \\ &= P_{t|t} + \sigma_\eta^2 \end{aligned}$$

Prediction errors

Define the prediction errors

$$\begin{aligned} v_t &= y_t - \mathrm{E}(y_t | Y_{1:t-1}) \\ &= y_t - \mathrm{E}(\alpha_t + \epsilon_t | Y_{1:t-1}) \\ &= y_t - \mathrm{E}(\alpha_t | Y_{1:t-1}) \\ &= y_t - a_{t|t-1} \end{aligned}$$

and note that $\mathrm{E}(v_t | Y_{1:t-1}) = 0$.

Further, define the prediction error variance

$$\begin{aligned} F_t &= \mathrm{Var}(v_t | Y_{1:t-1}) \\ &= \mathrm{E}\left((y_t - a_{t|t-1})^2\right) \\ &= \mathrm{E}\left((\alpha_t - a_{t|t-1} + \epsilon_t)^2\right) \\ &= \mathrm{E}\left((\alpha_t - a_{t|t-1})^2\right) + \sigma_\epsilon^2 \\ &= P_{t|t-1} + \sigma_\epsilon^2 \end{aligned}$$

Filtering

We apply the regression lemma while holding $Y_{1:t-1}$ fixed. We have

$$\begin{aligned} a_{t|t} &= \mathrm{E}(\alpha_t | Y_{1:t-1}, v_t) \\ &= \mathrm{E}(\alpha_t | Y_{1:t-1}) \\ &\quad + \mathrm{Cov}(\alpha_t, v_t | Y_{1:t-1}) \mathrm{Var}(v_t | Y_{1:t-1})^{-1} (v_t - \mathrm{E}(v_t | Y_{t:t-1})) \\ &= a_{t|t-1} + \mathrm{Cov}(\alpha_t, v_t | Y_{1:t-1}) F_t^{-1} v_t \end{aligned}$$

which follows from the $\mathrm{E}(v_t | Y_{1:t-1}) = 0$ and $\mathrm{Var}(v_t | Y_{1:t-1}) = F_t$

Filtering

The covariance can be computed as follows

$$\begin{aligned}\mathrm{Cov}(\alpha_t, v_t | Y_{1:t-1}) &= \mathrm{E}(\alpha_t v_t | Y_{1:t-1}) \\ &= \mathrm{E}(\alpha_t(\alpha_t + \epsilon_t - a_{t|t-1}) | Y_{1:t-1}) \\ &= \mathrm{E}(\alpha_t^2 | Y_{1:t-1}) - a_{t|t-1}^2 \\ &= \mathrm{Var}(\alpha_t | Y_{1:t-1}) = P_{t|t-1}\end{aligned}$$

We obtain the filtered estimate

$$a_{t|t} = a_{t|t-1} + P_{t|t-1} F_t^{-1} v_t$$

Prediction and Filtering

Similarly from the regression lemma we get for the filtered variance

$$
\begin{aligned}
P_{t|t} &= \mathrm{Var}(\alpha_t | Y_{1:t-1}, v_t) \\
&= \mathrm{Var}(\alpha_t | Y_{1:t-1}) \\
&\quad -\mathrm{Cov}(\alpha_t, v_t | Y_{1:t-1}) \mathrm{Var}(v_t | Y_{1:t-1})^{-1} \mathrm{Cov}(v_t, \alpha_t | Y_{1:t-1}) \\
&= P_{t|t-1} - P_{t|t-1} F_t^{-1} P_{t|t-1}
\end{aligned}
$$

Kalman filter

Theorem

When collecting the equations we obtain the Kalman filter for the local level model:

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

for $t = 1, \ldots, T$ *and* $K_t = P_{t|t-1}/F_t$.

Prediction and Filtering: some comments

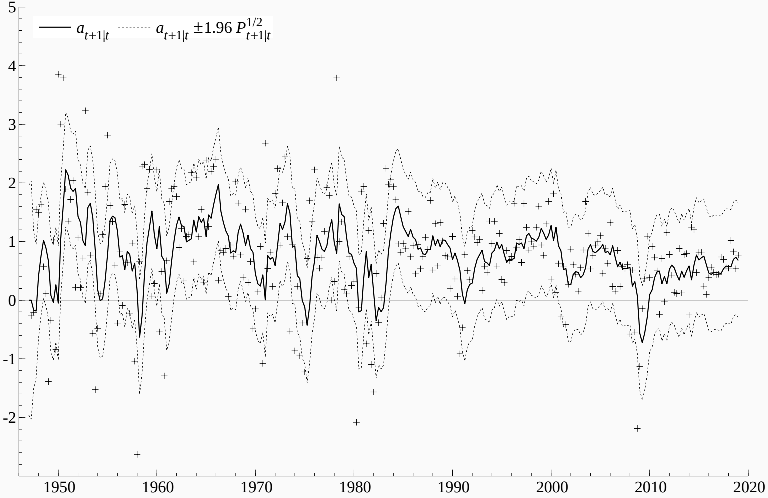
- K_t is known as the Kalman gain

$$K_t = P_{t|t-1}/F_t = \frac{P_{t|t-1}}{P_{t|t-1} + \sigma_\epsilon^2}$$

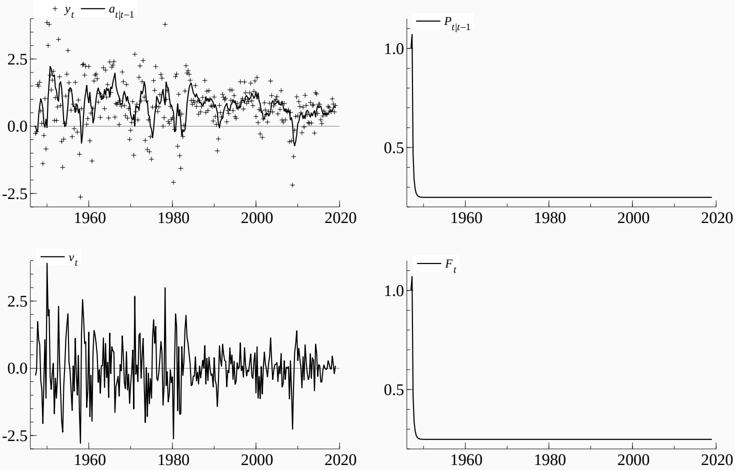
  measures how much one should update the state given the relative uncertainty in the state.

- The Kalman filter recursions can be implemented in Matlab/R by writing a for-loop for $t = 1, \dots, n$; initializing with $a_{1|0}$, $P_{1|0}$; compute v_1, F_1, compute $a_{1|1}$, $P_{1|1}$; compute $a_{2|1}$, $P_{2|1}$; compute v_2, F_2 and so on

Filtered US GDP



Filtered US GDP



Error recursions

The state prediction error is defined as

$$x_t = \alpha_t - a_{t|t-1}, \qquad \mathrm{Var}(x_t | Y_{1:t-1}) = P_{t|t-1}$$

The state prediction error x_t and the observation prediction error v_t are linear functions of the initial error x_1 and the disturbances ϵ_t and η_t. To see this

$$\begin{aligned} v_t &= y_t - a_{t|t-1} \\ &= \alpha_t + \epsilon_t - a_{t|t-1} \\ &= x_t + \epsilon_t \end{aligned}$$

$$\begin{aligned} x_{t+1} &= \alpha_{t+1} - a_{t+1|t} \\ &= \alpha_t + \eta_t - a_{t|t-1} - K_t v_t \\ &= x_t + \eta_t - K_t v_t \\ &= x_t + \eta_t - K_t(x_t + \epsilon_t) \\ &= L_t x_t + \eta_t - K_t \epsilon_t \end{aligned}$$

where $L_t = 1 - K_t$

Error recursions

Thus analog to the local level model recursions

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

we have the error recursions

$$v_t = x_t + \epsilon_t \qquad x_{t+1} = L_t x_t + \eta_t - K_t \epsilon_t$$

Smoothing

Smoothing

The objective of smoothing is to update our knowledge of the state using all available observations. We define

$$\text{Smoothed estimates:} \qquad \begin{aligned} a_{t|T} &= \mathrm{E}(\alpha_t | Y_{1:T}) \\ P_{t|T} &= \mathrm{Var}(\alpha_t | Y_{1:T}) \end{aligned}$$

The objective is to obtain these estimates by passing backwards through the data from $t = T$ until $t = 1$ and updating the filtered estimates recursively.

- This is done by an algorithm called the backward smoothing

Smoothing

Two initial comments that help to derive the smoothing recursions

- First, notice that the prediction errors are uncorrelated

$$\mathrm{E}(v_t v_{t-j}) = \mathrm{E}(\mathrm{E}(v_t | Y_{1:t-1}) v_{t-j}) = 0$$

  and since they are Gaussian we may conclude they are independent.

- Second, notice that transformation from $v_t = y_t - \mathrm{E}(y_t | Y_{1:t-1})$ to y_t has a Jacobian of 1

Smoothing

We have

$$\begin{aligned} a_{t|T} &= \mathrm{E}(\alpha_t|Y_{1:T}) \\ &= \mathrm{E}(\alpha_t|y_1, \ldots, y_{t-1}, v_t, \ldots, v_T) \\ &= \mathrm{E}(\alpha_t|Y_{1:t-1}) + \textstyle\sum_{j=t}^{T} \mathrm{Cov}(\alpha_t, v_j|Y_{1:t-1}) F_j^{-1} v_j \end{aligned}$$

And we need to work out the covariance terms. Consider $j = t$

$$\begin{aligned} \mathrm{Cov}(\alpha_t, v_t|Y_{1:t-1}) &= \mathrm{Cov}(x_t, v_t|Y_{1:t-1}) \\ &= \mathrm{E}(x_t(x_t + \epsilon_t)|Y_{1:t-1}) \\ &= \mathrm{E}(x_t^2|Y_{1:t-1}) = P_{t|t-1} \end{aligned}$$

Smoothing

Consider $j = t + 1$

$$\begin{aligned} \mathrm{Cov}(\alpha_t, v_{t+1}|Y_{1:t-1}) &= \mathrm{Cov}(x_t, v_{t+1}|Y_{1:t-1}) \\ &= \mathrm{E}(x_t(x_{t+1} + \epsilon_{t+1})|Y_{1:t-1}) \\ &= \mathrm{E}(x_t(L_t x_t + \eta_t - K_t \epsilon_t)|Y_{1:t-1}) = P_t L_t \end{aligned}$$

Similarly we can work out

$$\begin{aligned} \mathrm{Cov}(\alpha_t, v_{t+2}|Y_{1:t-1}) &= P_t L_t L_{t+1} \\ &\vdots \\ \mathrm{Cov}(\alpha_t, v_T|Y_{1:t-1}) &= P_t L_t L_{t+1} \dots L_{T-1} \end{aligned}$$

The covariances admit a recursive structure

Smoothing

When plugging in the covariances we can write the smoothing recursions as

$$a_{t|T} = a_{t|t-1} + P_t r_{t-1} \qquad r_{t-1} = F_t^{-1} v_t + L_t r_t$$

which is initialized by $r_T = 0$.

We can do similar steps for the smoothed variances $P_{t|T}$ to obtain

$$P_{t|T} = P_t - P_t^2 N_{t-1} \qquad N_{t-1} = F_t^{-1} + L_t^2 N_t$$

which is initialized by $N_T = 0$.

Backward smoothing recursions

Theorem

When collecting the equations we obtain the backward smoothing recursions for the local level model:

$$a_{t|T} = a_{t|t-1} + P_t r_{t-1} \qquad r_{t-1} = F_t^{-1} v_t + L_t r_t$$

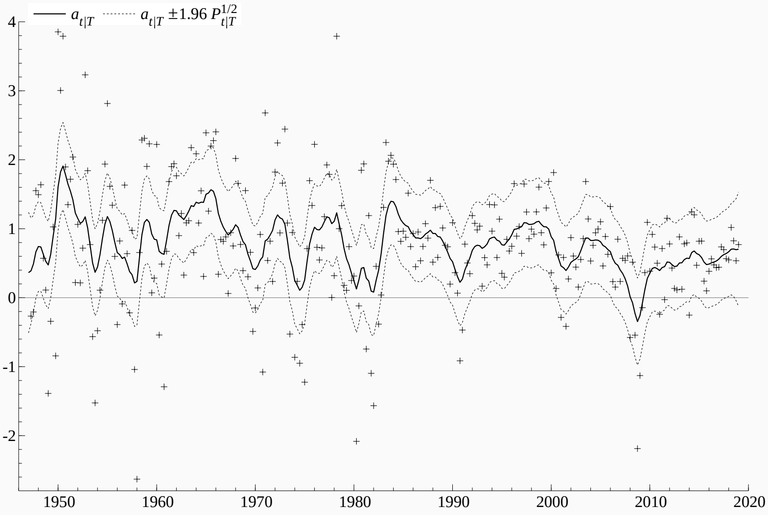
$$P_{t|T} = P_t - P_t^2 N_{t-1} \qquad N_{t-1} = F_t^{-1} + L_t^2 N_t$$

for $t = T, \ldots, 1$ *and* $L_t = 1 - K_t$ *and* $K_t = P_{t|t-1}/F_t$.

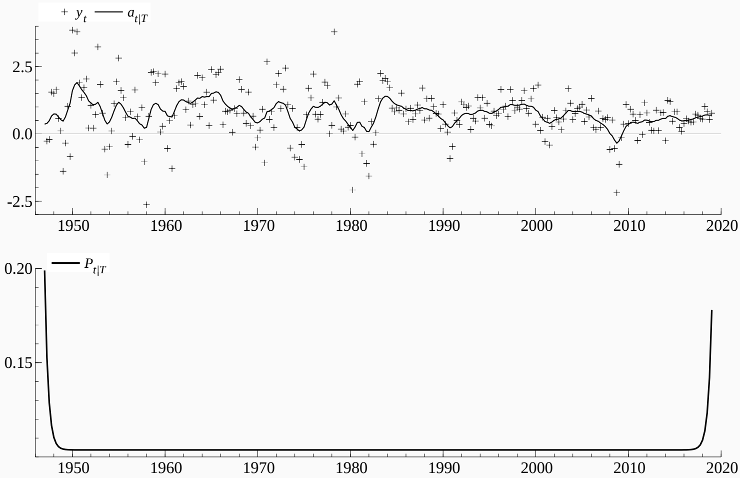
Smoothing: some comments

- After running the Kalman filter for $t = 1, \ldots, T$ we obtain the smoothing recursions as a function of the output of the Kalman filter for $t = T, \ldots, 1$
- The smoothing recursions can be implemented in Matlab/R by writing a for-loop for $t = T, \ldots, 1$; initializing with $r_T = 0$, $N_T = 0$; compute r_{T-1}, N_{T-1}, compute $a_{T|T}$, $P_{T|T}$; compute r_{T-2}, N_{T-2}; compute $a_{T-1|T}$, $P_{T-1|T}$ and so on

Smoothed US GDP



Smoothed US GDP



Missing observations

Missing observations

- A big advantage of the state space approach is the ease by which missing values can be dealt with
- Suppose y_j is missing for $j = \tau, \ldots, \tau^* - 1$
- We discuss how the Kalman filter can be logically adjusted to deal with the missing values

Missing observations

We have for any $t = \tau, \ldots, \tau^* - 1$ that

$$\begin{aligned} \mathrm{E}(\alpha_t | Y_t) &= \mathrm{E}(\alpha_t | Y_{\tau-1}) \\ &= \mathrm{E}(\alpha_\tau + \textstyle\sum_{j=\tau}^{t} \eta_j | Y_{\tau-1}) \\ &= a_{\tau|\tau-1} \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(\alpha_t | Y_t) &= \mathrm{Var}(\alpha_t | Y_{\tau-1}) \\ &= \mathrm{Var}(\alpha_\tau + \textstyle\sum_{j=\tau}^{t} \eta_j | Y_{\tau-1}) \\ &= P_{\tau|\tau-1} + (t - \tau)\sigma_\eta^2 \end{aligned}$$

which is equivalent to setting the Kalman gain $K_t = 0$ in the original Kalman filter

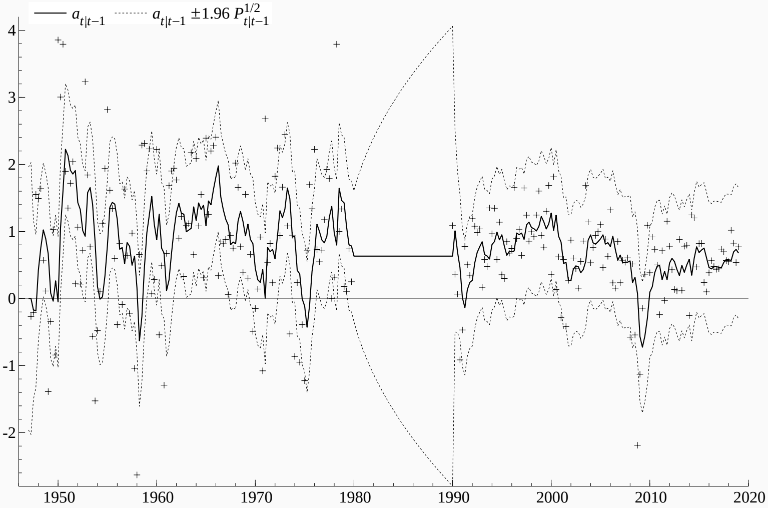
Missing observations

Recall

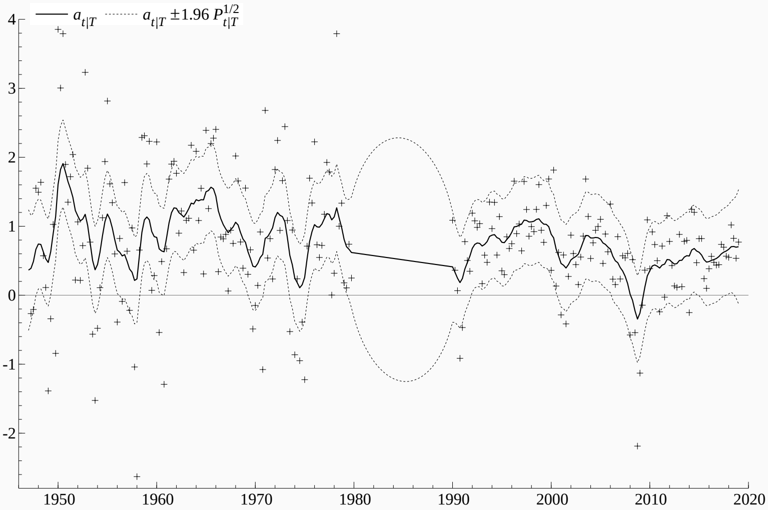
$$K_t = \frac{P_{t|t-1}}{P_{t|t-1} + \sigma_\epsilon^2}$$

- Now when an observation is missing we can think about this as its variance being infinite, e.g. $\sigma_\epsilon^2 \to \infty$, which makes the Kalman gain zero
- In the code this is easy to adjust with an if statement for missing observations
- Once the Kalman filter is adjusted the smoothing recursions require no further adjustment

Filtered US GDP: missing 80s



Smoothed US GDP: missing 80s



Forecasting

Missing observations

- Forecasting is equally easy as handling missing values
- Future values can be treated as missing values

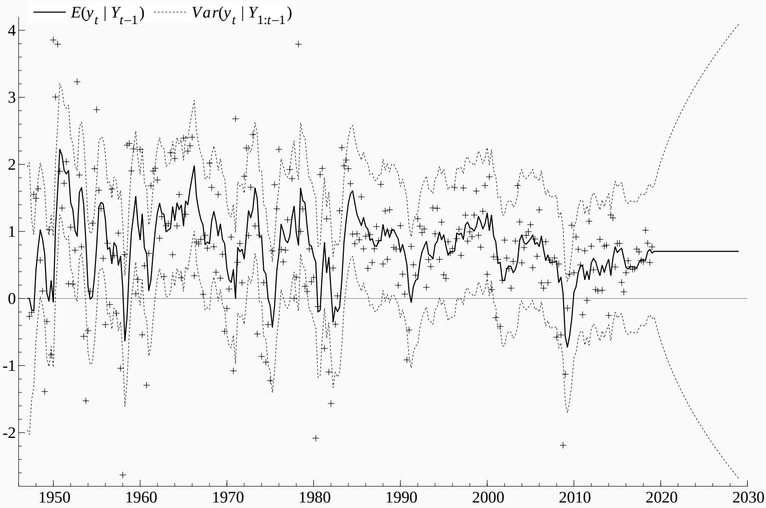
Missing observations

Using similar arguments as before

$$\begin{aligned}
\mathrm{E}(y_{T+h}|Y_{1:T}) &= \mathrm{E}(\alpha_{T+h} + \epsilon_{T+h}|Y_{1:T}) \\
&= \mathrm{E}(\alpha_{T+h}|Y_{1:T}) \\
&= \mathrm{E}(\alpha_T + \textstyle\sum_{j=T}^{T+h-1} \eta_j|Y_{1:T}) \\
&= \mathrm{E}(\alpha_T|Y_{1:T}) \\
&= a_{T|T}
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(y_{T+h}|Y_{1:T}) &= \mathrm{Var}(\alpha_{T+h} + \epsilon_{T+h}|Y_{1:T}) \\
&= \mathrm{Var}(\alpha_{T+h}|Y_{1:T}) + \sigma_\epsilon^2 \\
&= \mathrm{Var}(\alpha_T + \textstyle\sum_{j=T}^{T+h-1} \eta_j|Y_{1:T}) + \sigma_\epsilon^2 \\
&= \mathrm{Var}(\alpha_T|Y_{1:T}) + h\sigma_\eta^2 + \sigma_\epsilon^2 \\
&= P_{T|T} + h\sigma_\eta^2 + \sigma_\epsilon^2
\end{aligned}$$

Forecast US GDP



Initialization

Initialization

Recall the Kalman filter

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

- So far we did not discuss initialization, e.g. what to do with $a_{1|0}$ and $P_{1|0}$?

Initialization

Two cases

- Suppose the state is non-stationary like in the local level model

  $$\alpha_{t+1} = \alpha_t + \eta_t$$

  then there is no way of knowing its location if you have not looked at the data! To mimic this we consider diffuse initialization

  $$a_{1|0} = 0 \qquad P_{1|0} = 10^5$$

  infinite variance and arbitrary mean

Initialization

- Suppose the state is stationary causal like an AR(1) with $|\phi| < 1$

$$\alpha_{t+1} = \phi\alpha_t + \eta_t$$

then the initialization can be taken unconditionally

$$a_{1|0} = \mathrm{E}(\alpha_t) \qquad P_{1|0} = \mathrm{Var}(\alpha_t)$$

Hence we distinguish between stationary (unconditional initialization) and non-stationary (diffuse initialization)

Parameter estimation

Parameter estimation

The parameters of the local level model $\{\sigma_\epsilon^2, \sigma_\eta^2\}$ can be estimated by maximum likelihood

$$
\begin{aligned}
\ell(Y_{1:T}; \sigma_\epsilon^2, \sigma_\eta^2) &= \log p(Y_{1:T}) \\
&= \log p(y_1) + \textstyle\sum_{t=2}^{T} \log p(y_t | Y_{1:t-1}) \\
&= \log p(v_1) + \textstyle\sum_{t=2}^{T} \log p(v_t | Y_{1:t-1})
\end{aligned}
$$

where the last equality follows as the Jacobian of the transformation from y_t to v_t is one. Also note that

- v_t is Gaussian
- $\mathrm{E}(v_t | Y_{1:t-1}) = 0$
- $\mathrm{Var}(v_t | Y_{1:t-1}) = F_t$

Parameter estimation

We can write the log likelihood as

$$\ell(Y_{1:T}; \sigma_\epsilon^2, \sigma_\eta^2) = -\frac{1}{2} \sum_{t=1}^{T} \left(\log F_t + \frac{v_t^2}{F_t} \right)$$

In words

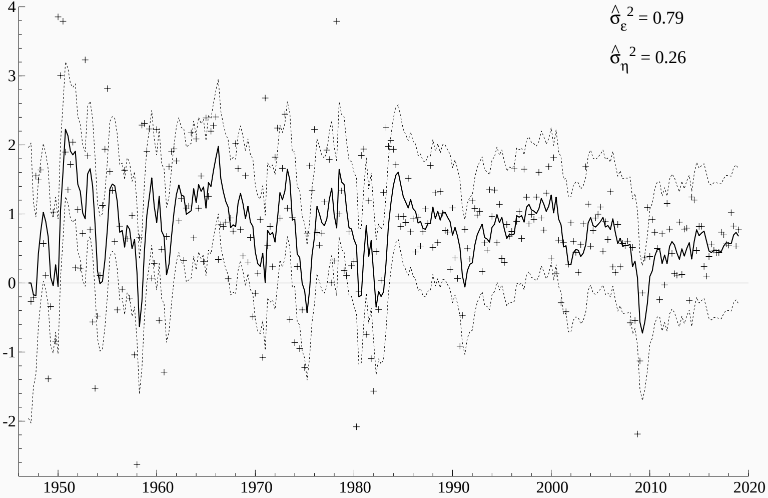
- The log likelihood is a direct function of the output of the Kalman filter

We define the ML estimates

$$\{\hat{\sigma}_\epsilon^2, \hat{\sigma}_\eta^2\} = \arg\max_{\sigma_\epsilon^2, \sigma_\eta^2} \ell(Y_{1:T}; \sigma_\epsilon^2, \sigma_\eta^2)$$

which can be obtained by numerically maximizing the likelihood wrt to the parameters.

Parameter estimates US GDP



Diagnostic checking

Diagnostic checking

A key assumption underlying the local level model is that the errors satisfy

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

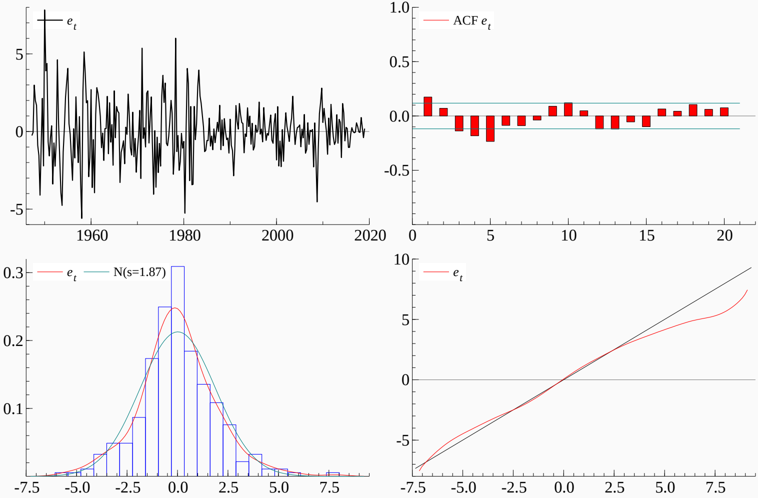
It can be shown that under this assumption we should have

$$e_t = \frac{v_t}{\sqrt{F_t}} \sim NID(0, 1)$$

We can test $\{e_t\}$ for

- Normality
- Heteroskedasticity
- Serial correlation

Diagnostics US GDP



References & Material

Reference & Material

- References:
  'Time Series Analysis by State Space Methods' by Jim Durbin and Siem Jan Koopman, Oxford University Press, 2012.