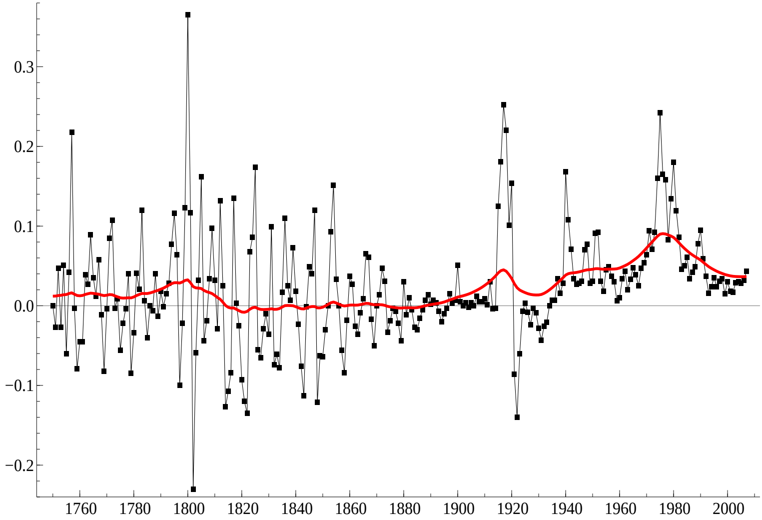
Introduction to Local Level Model and Kalman Filter

S.J. Koopman

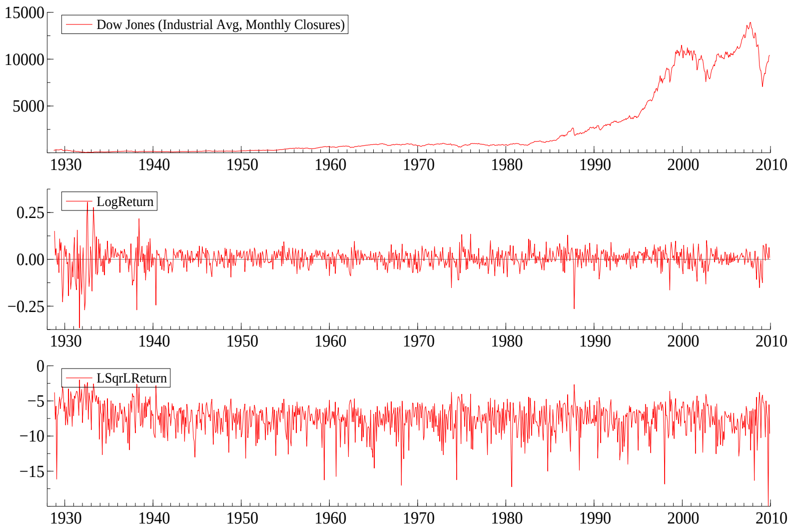
http://staff.feweb.vu.nl/koopman

January 2011

U.K. Yearly Inflation : signal extraction

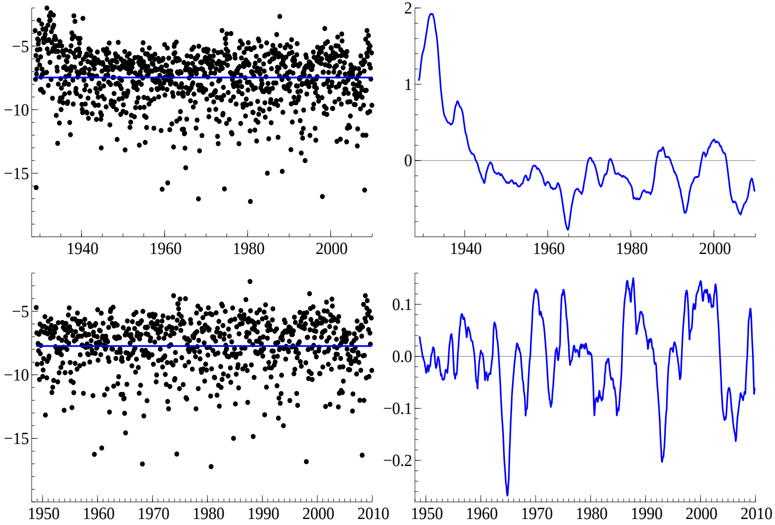


Dow Jones



15000
10000
5000
Dow Jones (Industrial Avg, Monthly Closures)
1930
1940
1950
1960
1970
1980
1990
2000
2010
0.25
0.00
−0.25
LogReturn
1930
1940
1950
1960
1970
1980
1990
2000
2010
0
−5
−10
−15
LSqrLReturn
1930
1940
1950
1960
1970
1980
1990
2000
2010

LogSqr LogReturns Dow Jones : signal extraction



Classical Decomposition

Basic Model

A basic model for representing a time series is the additive model

$$y_t = \mu_t + \gamma_t + \varepsilon_t, \qquad t = 1, \ldots, n,$$

also known as the Classical Decomposition.

$$\begin{aligned} y_t &= \text{observation}, \\ \mu_t &= \text{slowly changing component (trend)}, \\ \gamma_t &= \text{periodic component (seasonal)}, \\ \varepsilon_t &= \text{irregular component (disturbance)}. \end{aligned}$$

Unobserved Components Time Series Model

In a *Structural Time Series Model (STSM)* or *Unobserved Components Model (UCM)*, the various components are modelled explicitly as stochastic processes.

Local Level Model

- Components can be deterministic functions of time (e.g. polynomials), or stochastic processes;
- Deterministic example: $y_t = \mu + \varepsilon_t$ with $\varepsilon_t \sim \mathcal{NID}(0, \sigma_\varepsilon^2)$.
- Stochastic example: the Random Walk plus Noise, or *Local Level* model:

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, \qquad & \varepsilon_t &\sim \mathcal{NID}(0, \sigma_\varepsilon^2) \\ \mu_{t+1} &= \mu_t + \eta_t, \qquad & \eta_t &\sim \mathcal{NID}(0, \sigma_\eta^2), \end{aligned}$$

- The disturbances ε_t, η_s are independent for all s, t;
- The model is incomplete without a specification for μ_1 (note the non-stationarity):

$$\mu_1 \sim \mathcal{N}(a, P)$$

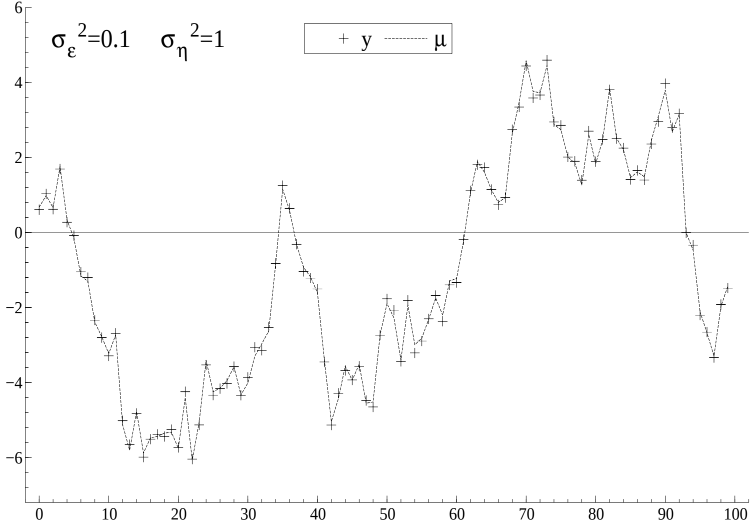
Local Level Model

$$
\begin{aligned}
y_t &= \mu_t + \varepsilon_t, \qquad & \varepsilon_t &\sim \mathcal{NID}(0, \sigma_\varepsilon^2) \\
\mu_{t+1} &= \mu_t + \eta_t, \qquad & \eta_t &\sim \mathcal{NID}(0, \sigma_\eta^2), \\
\mu_1 &\sim \mathcal{N}(a, P)
\end{aligned}
$$

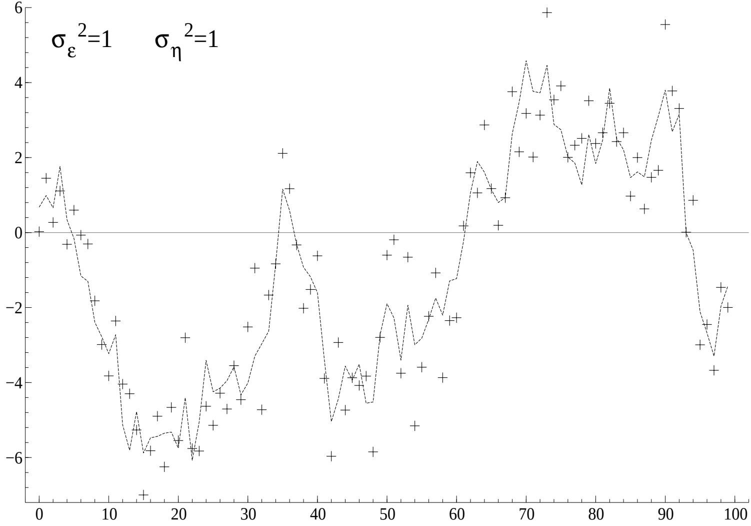
General framework

- The level μ_t and the irregular ε_t are unobservables;
- Parameters: σ_ε^2 and σ_η^2;
- Trivial special cases:
  - $\sigma_\eta^2 = 0 \implies y_t \sim \mathcal{NID}(\mu_1, \sigma_\varepsilon^2)$ (WN with constant level);
  - $\sigma_\varepsilon^2 = 0 \implies y_{t+1} = y_t + \eta_t$ (pure RW);
- Local Level is a model representation for EWMA forecasting.

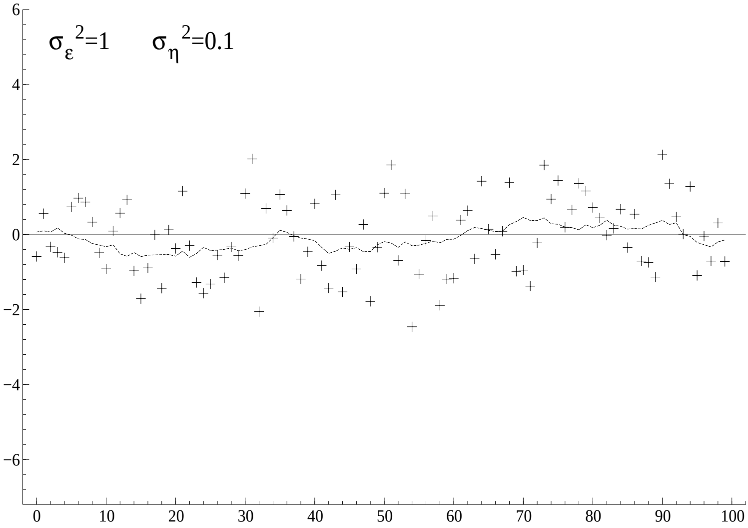
Simulated LL Data



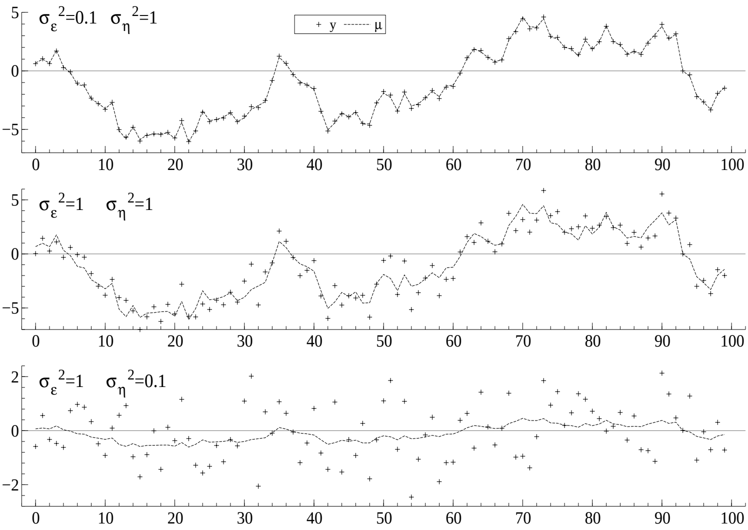
Simulated LL Data



Simulated LL Data



Simulated LL Data



Local Level Model

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, \qquad & \varepsilon_t &\sim \mathcal{NID}(0, \sigma^2_\varepsilon), \\ \mu_{t+1} &= \mu_t + \eta_t, \qquad & \eta_t &\sim \mathcal{NID}(0, \sigma^2_\eta), \end{aligned}$$

Its properties

- First difference is stationary:

$$\Delta y_t = \Delta \mu_t + \Delta \varepsilon_t = \eta_{t-1} + \varepsilon_t - \varepsilon_{t-1}.$$

- Dynamic properties of Δy_t:

$$\begin{aligned} \mathrm{E}(\Delta y_t) &= 0, \\ \gamma_0 &= \mathrm{E}(\Delta y_t \Delta y_t) = \sigma^2_\eta + 2\sigma^2_\varepsilon, \\ \gamma_1 &= \mathrm{E}(\Delta y_t \Delta y_{t-1}) = -\sigma^2_\varepsilon, \\ \gamma_\tau &= \mathrm{E}(\Delta y_t \Delta y_{t-\tau}) = 0 \quad \text{for } \tau \geq 2. \end{aligned}$$

Properties of the LL model

- The ACF of Δy_t is

$$\rho_1 = \frac{-\sigma_\varepsilon^2}{\sigma_\eta^2 + 2\sigma_\varepsilon^2} = -\frac{1}{q+2}, \qquad q = \sigma_\eta^2/\sigma_\varepsilon^2,$$
$$\rho_\tau = 0, \qquad \tau \geq 2.$$

- q is called the *signal-noise ratio*;
- The model for Δy_t is MA(1) with restricted parameters such that

$$-1/2 \leq \rho_1 \leq 0$$

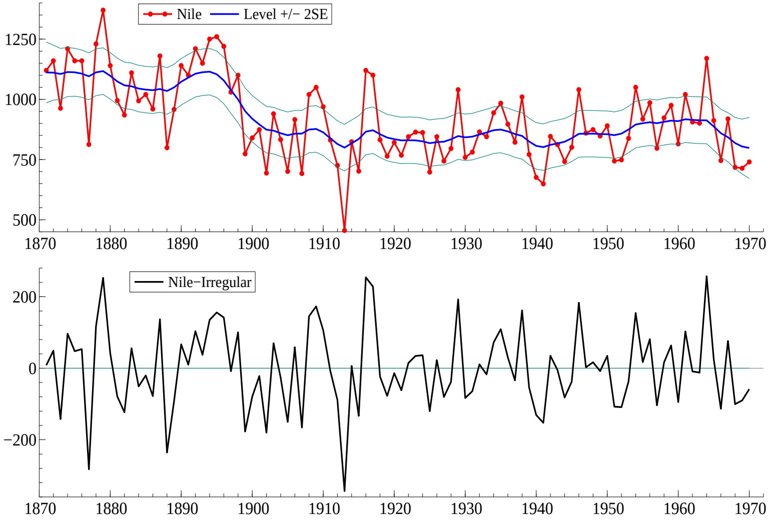
  i.e., y_t is ARIMA(0,1,1);
- Write $\Delta y_t = \xi_t + \theta \xi_{t-1}, \quad \xi_t \sim \mathcal{NID}(0, \sigma^2)$ to solve θ:

$$\theta = \frac{1}{2}\left(\sqrt{q^2 + 4q} - 2 - q\right).$$

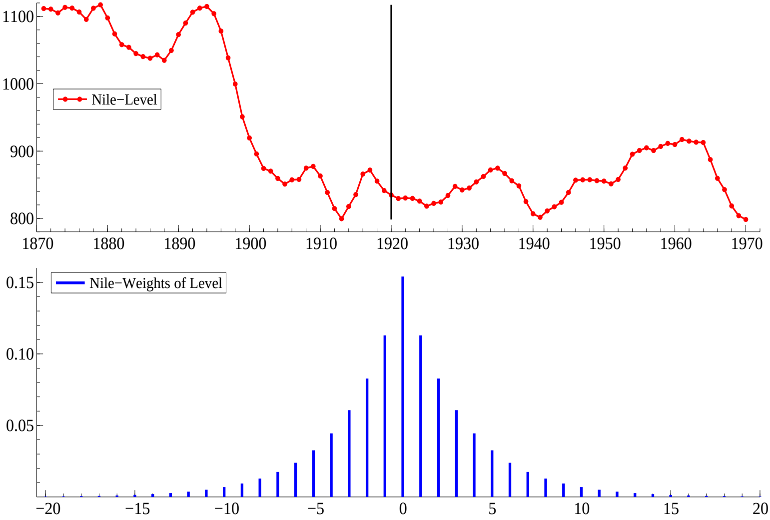
Local Level Model

- The model parameters are estimated by Maximum Likelihood;
- Advantages of model based approach: assumptions can be tested, parameters are estimated...;
- The model with estimated parameters is used for the signal extraction of components;
- The estimated level μ_t is effectively a locally weighted average of the data;
- The distribution of weights can be compared with Kernel functions in nonparametric regressions;
- On basis of model, the methods yield minimum mean square error (MMSE) forecasts and the associated confidence intervals.

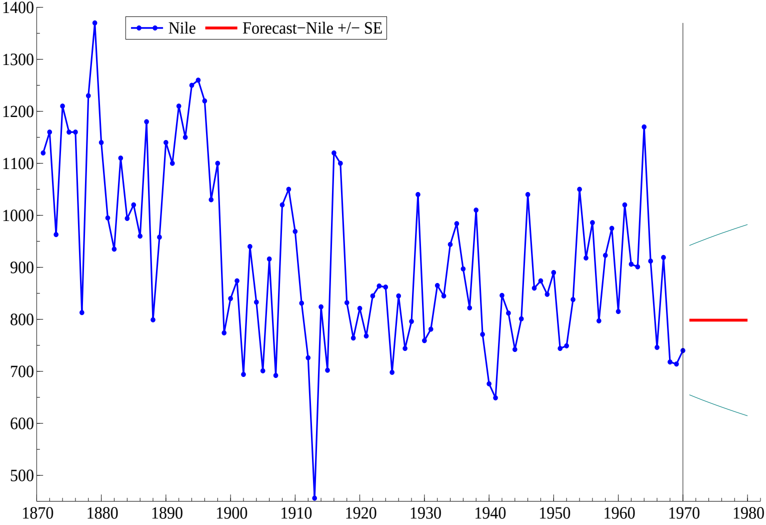
Nile Data: decomposition



Nile Data: decomposition weights



Nile Data: forecasts



Kalman Filter

- The Kalman filter calculates the mean and variance of the unobserved state, given the observations.
- The state is Gaussian: the complete distribution is characterized by the mean and variance.
- The filter is a recursive algorithm; the current best estimate is updated whenever a new observation is obtained.
- To start the recursion, we need a_1 and P_1, which we assumed given.
- There are various ways to initialize when a_1 and P_1 are unknown, which we will not discuss here. See discussion in DK book, Chapter 2.

Kalman Filter

The unobserved variable μ_t can be estimated from the observations with the *Kalman filter*:

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

for $t = 1, \ldots, n$ and starting with given values for a_1 and P_1.

- Writing $Y_t = \{y_1, \ldots, y_t\}$, define

$$a_{t+1} = \mathrm{E}(\mu_{t+1} | Y_t), \qquad P_{t+1} = \mathrm{var}(\mu_{t+1} | Y_t).$$

Kalman Filter

Local level model: $\mu_{t+1} = \mu_t + \eta_t, \qquad y_t = \mu_t + \varepsilon_t.$

- Writing $Y_t = \{y_1, \ldots, y_t\}$, define

$$a_{t+1} = \mathrm{E}(\mu_{t+1}|Y_t), \qquad P_{t+1} = \mathrm{var}(\mu_{t+1}|Y_t);$$

- The prediction error is

$$\begin{aligned} v_t &= y_t - \mathrm{E}(y_t|Y_{t-1}) \\ &= y_t - \mathrm{E}(\mu_t + \varepsilon_t|Y_{t-1}) \\ &= y_t - \mathrm{E}(\mu_t|Y_{t-1}) \\ &= y_t - a_t; \end{aligned}$$

- It follows that $v_t = (\mu_t - a_t) + \varepsilon_t$ and $\mathrm{E}(v_t) = 0$;
- The prediction error variance is $F_t = \mathrm{var}(v_t) = P_t + \sigma_\varepsilon^2$.

Regression theory

The proof of the Kalman filter uses lemmas from the multivariate Normal regression theory.

Lemma 1

Suppose x, y are jointly Normally distributed vectors. Then

$$\begin{aligned} \mathrm{E}(x|y) &= \mathrm{E}(x) + \Sigma_{xy}\Sigma_{y}^{-1}y, \\ \mathrm{var}(x|y) &= \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}'. \end{aligned}$$

Lemma 2

Suppose x, y and z are jointly Normally distributed vectors with $\mathrm{E}(z) = 0$ and $\Sigma_{yz} = 0$. Then

$$\begin{aligned} \mathrm{E}(x|y,z) &= \mathrm{E}(x|y) + \Sigma_{xz}\Sigma_{zz}^{-1}z, \\ \mathrm{var}(x|y,z) &= \mathrm{var}(x|y) - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{xz}'. \end{aligned}$$

Derivation Kalman Filter

Local level model: $\mu_{t+1} = \mu_t + \eta_t, \qquad y_t = \mu_t + \varepsilon_t.$

- We have $Y_t = \{Y_{t-1}, y_t\} = \{Y_{t-1}, v_t\}$ and $\mathrm{E}(v_t y_{t-j}) = 0$ for $j = 1, \ldots, t-1$;
- The lemma is $\mathrm{E}(x|y,z) = \mathrm{E}(x|y) + \Sigma_{xz}\Sigma_{zz}^{-1} z$.
  In our case, take $x = \mu_{t+1}$, $y = Y_{t-1}$ and $z = v_t = (\mu_t - a_t) + \varepsilon_t$;
- $\mathrm{E}(x|y)$ implies that $\mathrm{E}(\mu_{t+1}|Y_{t-1}) = \mathrm{E}(\mu_t|Y_{t-1}) + \mathrm{E}(\eta_t|Y_{t-1}) = a_t$;
- Further, Σ_{xz} provides the expression $\mathrm{E}(\mu_{t+1} v_t) = \mathrm{E}(\mu_t v_t) + \mathrm{E}(\eta_t v_t) = \mathrm{E}[(\mu_t - a_t)(y_t - a_t)] + \mathrm{E}(\eta_t v_t) = \mathrm{E}[(\mu_t - a_t)(\mu_t - a_t)] + \mathrm{E}[(\mu_t - a_t)\varepsilon_t] + \mathrm{E}(\eta_t v_t) = P_t$;
- Since $\Sigma_{zz} = F_t$, we can apply lemma and obtain the state update

$$\begin{aligned} a_{t+1} &= \mathrm{E}(\mu_{t+1}|Y_{t-1}, y_t) \\ &= a_t + P_t F_t^{-1} v_t \\ &= a_t + K_t v_t; \qquad \text{with } K_t = P_t F_t^{-1}. \end{aligned}$$

Kalman Filter Derived

- Our best prediction of y_t based on its past is a_t. When the actual observation arrives, calculate the prediction error $v_t = y_t - a_t$ and its variance $F_t = P_t + \sigma_\varepsilon^2$.
- The best estimate of the state mean for the next period is based on both the current estimate a_t and the new information v_t:

  $$a_{t+1} = a_t + K_t v_t,$$

  similarly for the variance:

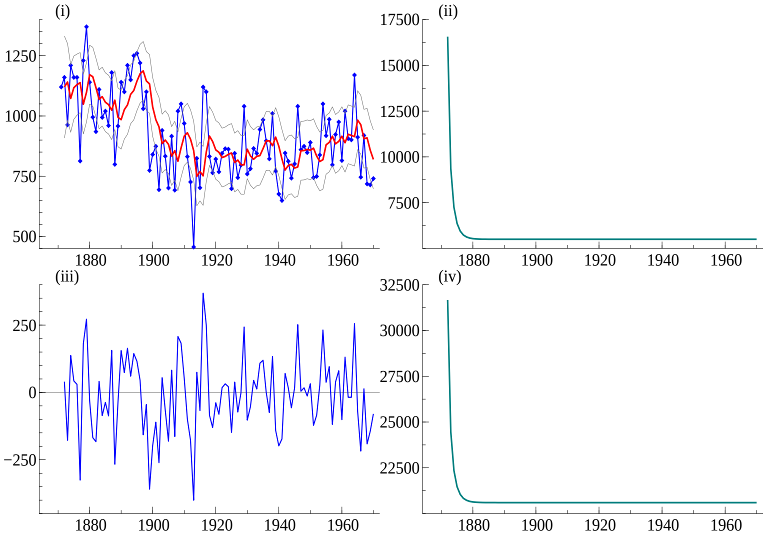
  $$P_{t+1} = P_t + \sigma_\eta^2 - K_t F_t K_t'.$$

- The *Kalman gain*

  $$K_t = P_t F_t^{-1}$$

  is the optimal weighting matrix for the new evidence.
- You should be able to replicate the proof of the Kalman filter for the Local Level Model (DK, Chapter 2).

Kalman filter for Nile Data: (i) a_t; (ii) P_t; (iii) v_t and (iv) F_t.



Steady State Kalman Filter

Kalman filter converges to a positive value, say $P_t \to \bar{P}$. We would then have

$$F_t \to \bar{P} + \sigma_\varepsilon^2, \qquad K_t \to \bar{P}/(\bar{P} + \sigma_\varepsilon^2).$$

The state prediction variance updating leads to

$$\bar{P} = \bar{P}\left(1 - \frac{\bar{P}}{\bar{P} + \sigma_\varepsilon^2}\right) + \sigma_\eta^2,$$

which reduces to the quadratic

$$x^2 - xq - q = 0,$$

where $x = \bar{P}/\sigma_\varepsilon^2$ and $q = \sigma_\eta^2/\sigma_\varepsilon^2$, with solution

$$\bar{P} = \sigma_\varepsilon^2\left(q + \sqrt{q^2 + 4q}\right)/2.$$

Smoothing

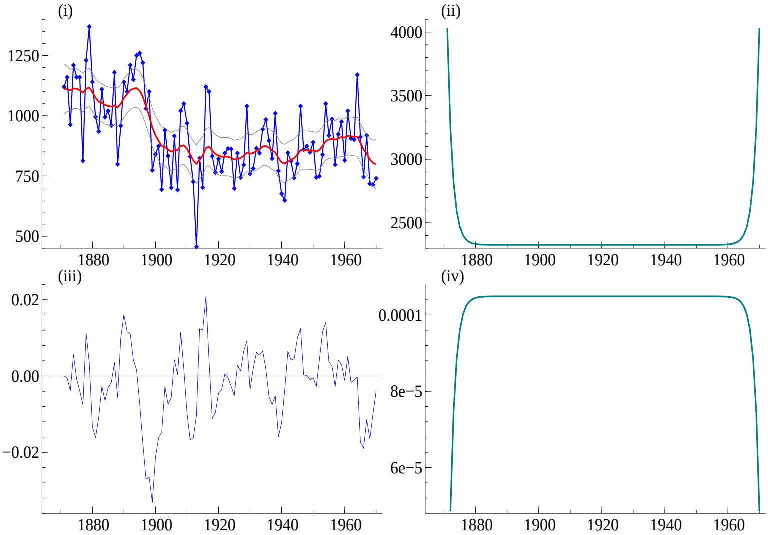
- The filter calculates the mean and variance conditional on Y_t;
- The Kalman smoother calculates the mean and variance conditional on the full set of observations Y_n;
- After the filtered estimates are calculated, the smoothing recursion starts at the last observations and runs until the first.

$$\begin{aligned}
\hat{\mu}_t &= \mathrm{E}(\mu_t | Y_n), \qquad V_t = \mathrm{var}(\mu_t | Y_n), \\
r_t &= \text{weighted sum of future innovations}, \quad N_t = \mathrm{var}(r_t), \\
L_t &= 1 - K_t.
\end{aligned}$$

Starting with $r_n = 0$, $N_n = 0$, the smoothing recursions are given by

$$\begin{aligned}
r_{t-1} &= F_t^{-1} v_t + L_t r_t, \qquad N_{t-1} = F_t^{-1} + L_t^2 N_t, \\
\hat{\mu}_t &= a_t + P_t r_{t-1}, \qquad V_t = P_t - P_t^2 N_{t-1}.
\end{aligned}$$

Kalman smoothing for Nile Data: (i) $\hat{\mu}_t$; (ii) V_t; (iii) r_t and (iv) N_t.



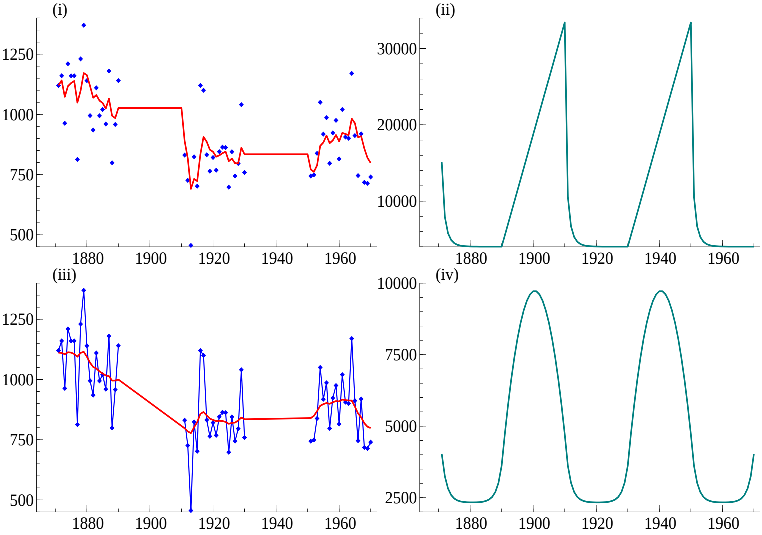
Missing Observations

Missing observations are very easy to handle in Kalman filtering:

- suppose y_j is missing
- put $v_j = 0, K_j = 0$ and $F_j = \infty$ in the algorithm
- proceed further calculations as normal

The filter algorithm extrapolates according to the state equation until a new observation arrives. The smoother interpolates between observations.

Nile Data with missing observations : (i) a_t, (ii) P_t, (iii) $\hat{\mu}_t$ and (iv) V_t.

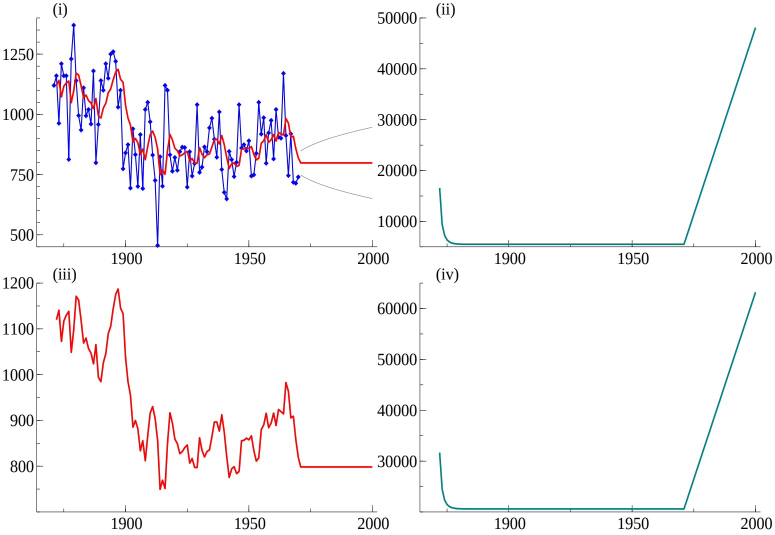


Forecasting

Forecasting requires no extra theory: just treat future observations as missing:

- put $v_j = 0, K_j = 0$ and $F_j = \infty$ for $j = n+1, \ldots, n+k$
- proceed further calculations as normal
- forecast for y_j is a_j

Nile Data: forecasting



Parameters in Local Level Model

We recall the Local Level Model as

$$
\begin{aligned}
y_t &= \mu_t + \varepsilon_t, \qquad && \varepsilon_t \sim \mathcal{NID}(0, \sigma_\varepsilon^2) \\
\mu_{t+1} &= \mu_t + \eta_t, \qquad && \eta_t \sim \mathcal{NID}(0, \sigma_\eta^2), \\
\mu_1 &\sim \mathcal{N}(a, P)
\end{aligned}
$$

General framework

- The unknown μ_t's can be estimated by prediction, filtering and smoothing;
- The other parameters are given by the variances σ_ε^2 and σ_η^2;
- We estimate these parameters by Maximum Likelihood;
- Parameters can be transformed : $\sigma_\varepsilon^2 = \exp(\psi_\varepsilon)$ and $\sigma_\eta^2 = \exp(\psi_\eta)$;
- Parameter vector $\psi = (\psi_\varepsilon\ ,\ \psi_\eta)'$.

Parameter Estimation by ML

The parameters in any state space model can be collected in some vector ψ. When model is linear and Gaussian; we can estimate ψ by Maximum Likelihood.

The loglikelihood af a time series is

$$\log L = \sum_{t=1}^{n} \log p(y_t | Y_{t-1}).$$

In the state space model, $p(y_t | Y_{t-1})$ is a Gaussian density with mean a_t and variance F_t:

$$\log L = -\frac{n}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{n} \left(\log F_t + F_t^{-1} v_t^2 \right),$$

with v_t and F_t from the Kalman filter. This is called the *prediction error decomposition* of the likelihood. Estimation proceeds by numerically maximising $\log L$.

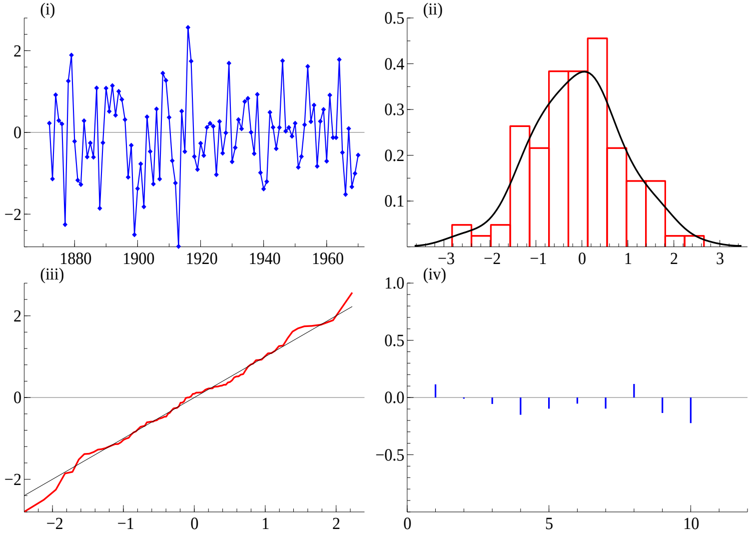
Diagnostics

- Null hypothesis: standardised residuals

$$v_t/\sqrt{F}_t \sim \mathcal{NID}(0, 1)$$

- Apply standard test for Normality, heteroskedasticity, serial correlation;
- A recursive algorithm is available to calculate smoothed disturbances (auxilliary residuals), which can be used to detect breaks and outliers;
- Model comparison and parameter restrictions: use likelihood based procedures (LR test, AIC, BIC).

Nile Data: diagnostics



Three exercises

1. Consider LL model (see slides, see DK chapter 2).
   - Reduced form is ARIMA(0,1,1) process. Derive the relationship between signal-to-noise ratio q of LL model and the θ coefficient of the ARIMA model;
   - Derive the reduced form in the case $\eta_t = \sqrt{q}\varepsilon_t$ and notice the difference in the general case.
   - Give the elements of the mean vector and variance matrix of $y = (y_1, \ldots, y_n)'$ when y_t is generated by a LL model.
   - Show that the forecasts of the Kalman filter (in a steady state) are the same as those generated by the exponentially weighted moving average (EWMA) method of forecasting: $\hat{y}_{t+1} = \hat{y}_t + \lambda(y_t - \hat{y}_t)$ for $t = 1, \ldots, n$. Derive the relationship between λ and the signal-to-noise ratio q ?

Three exercises (cont.)

2. Derive a Kalman filter for the local level model

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \qquad \Delta\mu_{t+1} = \eta_t \sim N(0, \sigma_\eta^2),$$

   with $E(\varepsilon_t \eta_t) = \sigma_{\varepsilon\eta} \neq 0$ and $E(\varepsilon_t \eta_s) = 0$ for all t, s and $t \neq s$. Also discuss the problem of missing obervations in this case.
3. Write `Ox` program(s) that produce all Figures in Ch 2 of DK except Fig. 2.4. Data: `http://www.ssfpack.com/dkbook.html`

Selected references

- A.C.Harvey (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press
- G.Kitagawa & W.Gersch (1996). *Smoothness Priors Analysis of Time Series*. Springer-Verlag
- J.Harrison & M.West (1997). *Bayesian Forecasting and Dynamic Models*. Springer-Verlag
- R.H. Shumway and D.S. Stoffer (2000). *Time Series Analysis and its Applications*. Springer-Verlag.
- J.Durbin & S.J.Koopman (2011). *Time Series Analysis by State Space Methods*. Second Edition, Oxford University Press
- J.Commandeur & S.J.Koopman (2007). *An Introduction to State Space Time Series Analysis*. Oxford University Press