Introduction to
Local Level Model and Kalman Filter

S.J. Koopman

http://staff.feweb.vu.nl/koopman
January 2011
Classical Decomposition

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

\[ y_t = \mu_t + \gamma_t + \varepsilon_t, \quad t = 1, \ldots, n, \]

also known as the Classical Decomposition.

- \( y_t \) = observation,
- \( \mu_t \) = slowly changing component (trend),
- \( \gamma_t \) = periodic component (seasonal),
- \( \varepsilon_t \) = irregular component (disturbance).

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 \( \varepsilon_t \sim \mathcal{N}(0, \sigma_{\varepsilon}^2) \).
- Stochastic example: the Random Walk plus Noise, or *Local Level* model:

\[
\begin{align*}
\ y_t & = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_{\varepsilon}^2) \\
\mu_t+1 & = \mu_t + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \sigma_{\eta}^2),
\end{align*}
\]

- The disturbances \( \varepsilon_t, \eta_t \) are independent for all \( s, t \);
- The model is incomplete without a specification for \( \mu_1 \) (note the non-stationarity):

\[ \mu_1 \sim \mathcal{N}(a, P) \]
Local Level Model

\[ y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2_\varepsilon) \]
\[ \mu_{t+1} = \mu_t + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \sigma^2_\eta) \]
\[ \mu_1 \sim \mathcal{N}(a, P) \]

General framework

- The level \( \mu_t \) and the irregular \( \varepsilon_t \) are unobservables;
- Parameters: \( \sigma^2_\varepsilon \) and \( \sigma^2_\eta \);
- Trivial special cases:
  - \( \sigma^2_\eta = 0 \implies y_t \sim \mathcal{N}(\mu_1, \sigma^2_\varepsilon) \) (WN with constant level);
  - \( \sigma^2_\varepsilon = 0 \implies y_{t+1} = y_t + \eta_t \) (pure RW);
- Local Level is a model representation for EWMA forecasting.
Simulated LL Data

$\sigma_\varepsilon^2 = 0.1 \quad \sigma_\eta^2 = 1$

$+$ $y$ $-$ $\mu$
Simulated LL Data

\[ \sigma^2_\varepsilon = 1 \quad \sigma^2_\eta = 1 \]
Simulated LL Data

\[ \sigma_\varepsilon^2 = 1 \quad \sigma_\eta^2 = 0.1 \]
Simulated LL Data

\[ \sigma^2_\varepsilon = 0.1 \quad \sigma^2_\eta = 1 \]

\[ \sigma^2_\varepsilon = 1 \quad \sigma^2_\eta = 1 \]

\[ \sigma^2_\varepsilon = 1 \quad \sigma^2_\eta = 0.1 \]
Local Level Model

\[ y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{NID}(0, \sigma_\varepsilon^2), \]
\[ \mu_{t+1} = \mu_t + \eta_t, \quad \eta_t \sim \mathcal{NID}(0, \sigma_\eta^2), \]

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 \( \Delta y_t \):
  \[ E(\Delta y_t) = 0, \]
  \[ \gamma_0 = E(\Delta y_t \Delta y_t) = \sigma_\eta^2 + 2\sigma_\varepsilon^2, \]
  \[ \gamma_1 = E(\Delta y_t \Delta y_{t-1}) = -\sigma_\varepsilon^2, \]
  \[ \gamma_\tau = E(\Delta y_t \Delta y_{t-\tau}) = 0 \quad \text{for } \tau \geq 2. \]
Properties of the LL model

- The ACF of $\Delta y_t$ is

  $$\rho_1 = \frac{-\sigma_\varepsilon^2}{\sigma_\eta^2 + 2\sigma_\varepsilon^2} = -\frac{1}{q + 2}, \quad q = \frac{\sigma_\eta^2}{\sigma_\varepsilon^2},$$

  $$\rho_\tau = 0, \quad \tau \geq 2.$$

- $q$ is called the signal-noise ratio;

- The model for $\Delta y_t$ 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 $\theta$:

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

- Nile-Level
- Nile-Weights of Level

Graph showing Nile-Weights of Level against Nile-Level with x-axis from 1870 to 1970 and y-axis values ranging from -20 to 20.
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 $\mu_t$ can be estimated from the observations with the *Kalman filter*:

\[ v_t = y_t - a_t, \]
\[ F_t = P_t + \sigma^2_{\varepsilon}, \]
\[ K_t = P_t F_t^{-1}, \]
\[ a_{t+1} = a_t + K_t v_t, \]
\[ P_{t+1} = P_t + \sigma^2_\eta - K_t^2 F_t, \]

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

  \[ a_{t+1} = \mathbb{E}(\mu_{t+1} | Y_t), \quad P_{t+1} = \text{var}(\mu_{t+1} | Y_t). \]
Kalman Filter

Local level model: \( \mu_{t+1} = \mu_t + \eta_t, \quad y_t = \mu_t + \varepsilon_t. \)

- Writing \( Y_t = \{y_1, \ldots, y_t\} \), define
  \[
  a_{t+1} = \mathbb{E}(\mu_{t+1} | Y_t), \quad P_{t+1} = \text{var}(\mu_{t+1} | Y_t);
  \]

- The prediction error is
  \[
  v_t = y_t - \mathbb{E}(y_t | Y_{t-1})
  = y_t - \mathbb{E}(\mu_t + \varepsilon_t | Y_{t-1})
  = y_t - \mathbb{E}(\mu_t | Y_{t-1})
  = y_t - a_t;
  \]

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

\[
E(x|y) = E(x) + \Sigma_{xy} \Sigma_{y}^{-1} y,
\]
\[
\text{var}(x|y) = \Sigma_{xx} - \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma'_{xy}.
\]

Lemma 2
Suppose \( x, y \) and \( z \) are jointly Normally distributed vectors with \( E(z) = 0 \) and \( \Sigma_{yz} = 0 \). Then

\[
E(x|y, z) = E(x|y) + \Sigma_{xz} \Sigma_{zz}^{-1} z,
\]
\[
\text{var}(x|y, z) = \text{var}(x|y) - \Sigma_{xz} \Sigma_{zz}^{-1} \Sigma'_{xz}.
\]
Derivation Kalman Filter

Local level model: \( \mu_{t+1} = \mu_t + \eta_t, \quad y_t = \mu_t + \varepsilon_t. \)

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

\[
a_{t+1} = \text{E}(\mu_{t+1}|Y_{t-1}, y_t) \\
= a_t + P_t F_t^{-1} v_t \\
= a_t + K_t v_t; \quad \text{with} \ K_t = P_t F_t^{-1}.
\]
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 \( \nu_t = y_t - a_t \) 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 \( \nu_t \):

\[
a_{t+1} = a_t + K_t \nu_t,
\]

similarly for the variance:

\[
P_{t+1} = P_t + \sigma_\eta^2 - K_t F_t K'_t.
\]

- The Kalman gain

\[
K_t = P_tF_t^{-1}
\]

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

$$F_t \to \bar{P} + \sigma^2_\epsilon, \quad K_t \to \frac{\bar{P}}{(\bar{P} + \sigma^2_\epsilon)}.$$

The state prediction variance updating leads to

$$\bar{P} = \bar{P} \left(1 - \frac{\bar{P}}{\bar{P} + \sigma^2_\epsilon}\right) + \sigma^2_\eta,$$

which reduces to the quadratic

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

where $x = \bar{P}/\sigma^2_\epsilon$ and $q = \sigma^2_\eta/\sigma^2_\epsilon$, with solution

$$\bar{P} = \sigma^2_\epsilon \left(q + \sqrt{q^2 + 4q}\right)/2.$$
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.

$$\hat{\mu}_t = \mathbb{E}(\mu_t | Y_n), \quad V_t = \text{var}(\mu_t | Y_n),$$

$$r_t = \text{weighted sum of future innovations}, \quad N_t = \text{var}(r_t),$$

$$L_t = 1 - K_t.$$ 

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

$$r_{t-1} = F_{t}^{-1} v_t + L_t r_t, \quad \ N_{t-1} = F_{t}^{-1} + L_t L_t N_t,$$

$$\hat{\mu}_t = a_t + P_t r_{t-1}, \quad V_t = P_t - P_t^2 N_{t-1}.$$
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) $V_t$. 

(i) 

(ii) 

(iii) 

(iv)
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 $a_j$
Parameters in Local Level Model

We recall the Local Level Model as

\[ y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\mathcal{I}\mathcal{D}(0, \sigma^2_\varepsilon) \]

\[ \mu_{t+1} = \mu_t + \eta_t, \quad \eta_t \sim \mathcal{N}\mathcal{I}\mathcal{D}(0, \sigma^2_\eta), \]

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

General framework

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

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

The loglikelihood of 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
  \[ \frac{\nu_t}{\sqrt{F_t}} \sim \mathcal{N}(0, 1) \]

- 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 $\theta$ 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 $\lambda$ and the signal-to-noise ratio $q$?
Three exercises (cont.)

2. Derive a Kalman filter for the local level model

\[ y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2), \quad \Delta \mu_{t+1} = \eta_t \sim N(0, \sigma^2), \]

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 observations 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

- J. Harrison & M. West (1997). *Bayesian Forecasting and Dynamic Models*. Springer-Verlag