State Space Models and the Kalman Filter

May 25, 2012

State Space Models and the Kalman Filter

What we did last time:

- The scalar filter
  - Combining period t prior and signal is analogous to a simple minimum variance problem with two signals
- Derived the multivariate filter using
  - The projection theorem
  - Projecting onto orthogonal variables
  - The Gram-Schmidt procedure

The basic formulas

The state space system and the Kalman update equation

$$\begin{aligned} X_t &= AX_{t-1} + C\mathbf{u}_t : \mathbf{u}_t \sim N(0, I) \\ Z_t &= DX_t + \mathbf{v}_t : \mathbf{v}_t \sim N(0, \Sigma_{vv}) \end{aligned}$$

$$X_{t|t} = AX_{t-1|t-1} + K_t \left(Z_t - DX_{t|t-1} \right)$$

where K_t is the Kalman gain and $X_{t|t} = E\left[X_t \mid Z^t, X_{0|0} \right]$

- Filter is also the *linear* minimum variance estimator of X_t even if shocks are non-gaussian.

The basic formulas

Most of the time, all you really need to know is how to put these formulas into a computer

$$X_{t|t} = AX_{t-1|t-1} + K_t \left(Z_t - DAX_{t-1|t-1} \right)$$

$$\begin{aligned} K_t &= P_{t|t-1} D' \left(DP_{t|t-1} D' + \Sigma_{vv} \right)^{-1} \\ P_{t+1|t} &= A \left(P_{t|t-1} - P_{t|t-1} D' \left(DP_{t|t-1} D' + \Sigma_{vv} \right)^{-1} DP_{t|t-1} \right) A' \\ &\quad + CC' \end{aligned}$$

This will give you a recursive estimate of X_t

- Remember: $P_{t|t-s} \equiv E(X_t - X_{t|t-s})(X_t - X_{t|t-s})'$

The scalar filter

$$
\begin{aligned}
x_t &= \rho x_{t-1} + u_t : u_t \sim N(0, \sigma_u^2) \text{ (state equation)} \\
z_t &= x_t + v_t : v_t \sim N(0, \sigma_v^2) \text{ (measurement equation)}
\end{aligned}
$$

gives the Kalman update equations

$$
\begin{aligned}
x_{t|t} &= \rho x_{t-1|t-1} + k_t \left(z_1 - \rho x_{t-1|t-1} \right) \\
k_t &= p_{t|t-1}(p_{t|t-1} + \sigma_v^2)^{-1} \\
p_{t|t-1} &= \rho^2 \underbrace{\left(p_{t-1|t-2} - p_{t-1|t-2}^2 (p_{t-1|t-2} + \sigma_v^2)^{-1} \right)}_{p_{t-1|t-1}} + \sigma_u^2
\end{aligned}
$$

Combining information in prior and signal

Kalman filter optimally combine information in prior $\rho x_{t-1|t-1}$ and signal z_t to form posterior estimate $x_{t|t}$ with covariance $p_{t|t}$

$$x_{t|t} = (1 - k_t)\rho x_{t-1|t-1} + k_t z_1$$

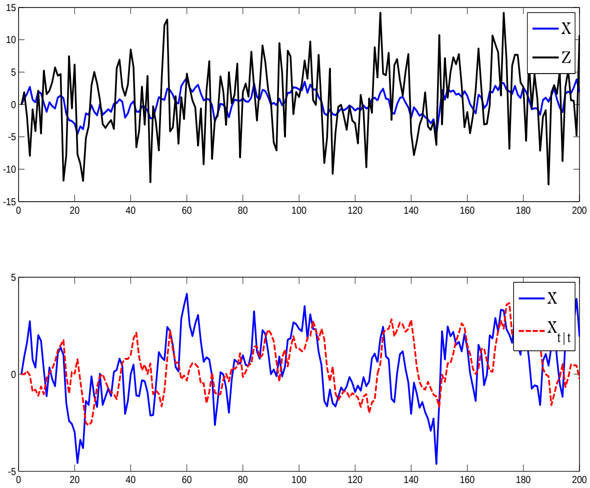
- More weight on signal (large kalman gain k_t) if prior variance is large or if signal is very precise
- Prior variance can be large either because previous state estimate was imprecise (i.e. $p_{t-1|t-1}$ is large) or because state innovations are large (i.e.σ_u^2 is large)

Example 1

Set

- $\rho = 0.9$
- $\sigma_u^2 = 1$
- $\sigma_v^2 = 5$

Example 1

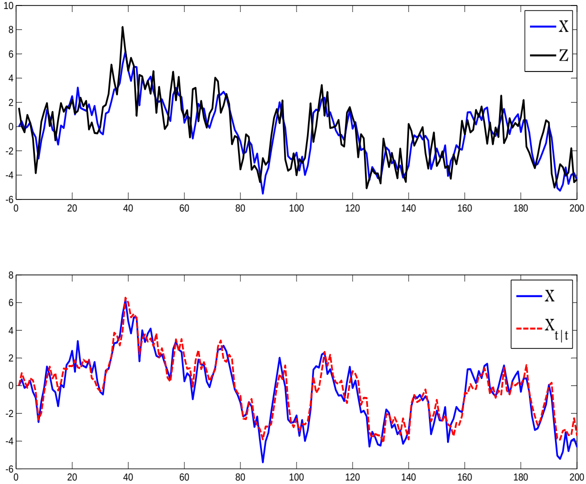


Example 2

Set

- $\rho = 0.9$
- $\sigma_u^2 = 1$
- $\sigma_v^2 = 1$

Example 2: Smaller measurement error variance



Convergence to time invariant filter

If $\rho < 1$ and if ρ, σ_u^2 and σ^2 are constant, the prior variance of the state estimate

$$p_{t|t-1} = \rho^2 \left(p_{t-1|t-2} - p_{t-1|t-2}^2 (p_{t-1|t-2} + \sigma_v^2)^{-1} \right) + \sigma_u^2$$

will converge to

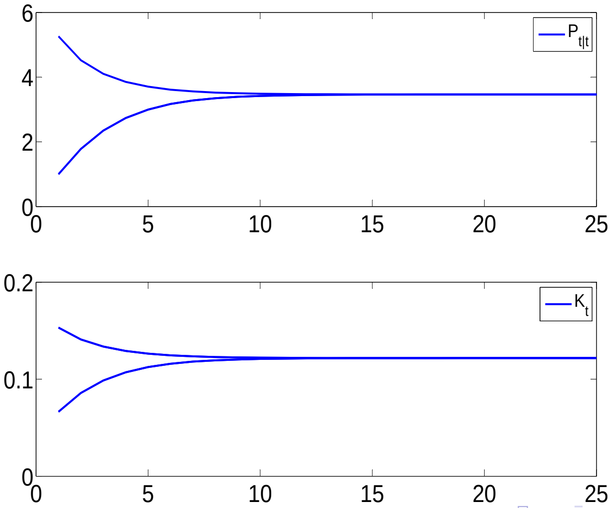
$$p = \rho^2 \left(p - p^2 (p + \sigma_v^2)^{-1} \right) + \sigma_u^2$$

The Kalman gain will then also converge:

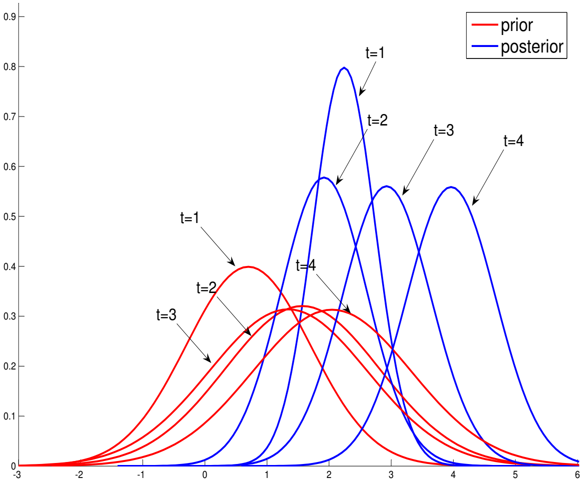
$$k = p(p + \sigma_v^2)^{-1}$$

- We can illustrate this by starting from the boundaries of possible values for $p_{1|0}$
  - Remember: $\sigma_u^2 < p_{t|t-1} < \sigma_u^2 (1 - \rho^2)^{-1}$ (Why?)

Convergence to time invariant filter



Convergence to time invariant filter



State space models are very flexible

We can put all models we have discussed so far in the state space form

$$\begin{aligned} X_t &= AX_{t-1} + C\mathbf{u}_t : \mathbf{u}_t \sim N(0, I) \\ Z_t &= DX_t + \mathbf{v}_t : \mathbf{v}_t \sim N(0, \Sigma_{vv}) \end{aligned}$$

This includes "observable" VAR(p), MA(p) and VARMA(p,q) processes as well as principal components and related models with explicitly latent factor models.

VAR (p) in State Space Form

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-1} + u_t$$

$$\begin{aligned} A &= \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_p \\ I & 0 & & 0 \\ 0 & \ddots & & \ddots \\ 0 & 0 & I & 0 \end{bmatrix}, C = \begin{bmatrix} I \\ 0 \\ 0 \\ 0 \end{bmatrix} u_t \\ D &= \begin{bmatrix} I & 0 & \cdots & 0 \end{bmatrix}, \Sigma_{vv} = 0 \end{aligned}$$

MA(1) in State Space Form

$$y_t = \varepsilon_t + \theta \varepsilon_{t-1}$$

can be written as

$$\begin{aligned} \begin{bmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{bmatrix} &= \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_{t-1} \\ \varepsilon_{t-2} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \varepsilon_t \\ y_t &= \begin{bmatrix} 1 & \theta \end{bmatrix} \end{aligned}$$

Maximum Likelihood Estimation of an MA(1) process

Maximum Likelihood Estimation of an MA(1) process

$$y_t = \varepsilon_t + \theta \varepsilon_{t-1}$$

- No need to assume that $\varepsilon_1 = 0$ and we can estimate exact ML

$$L(Z \mid \theta) = (-T/2)\log(2\pi) - \frac{T}{2}\log|\Omega_t| - \frac{1}{2}\sum_{t=1}^{T} \widetilde{Z}_t' \Omega_t^{-1} \widetilde{Z}_t$$

where

$$\begin{aligned} \widetilde{Z}_t &= Z_t - DAX_{t|t} \\ X_{t|t} &= AX_{t-1|t-1} + K_t \left(Z_t - DAX_{t-1|t-1}\right) \\ \Omega_t &= DP_{t|t-1}D' + \Sigma_{vv} \end{aligned}$$

We can start the Kalman filter recursions from the unconditional mean and variance, i.e. $X_0 = 0$ and $P_{0|0} = \sigma_\varepsilon^2 \times I$.

Alternative state space representations

Sometimes there are more than one state space representation of a given system: But are both

$$\begin{bmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_{t-1} \\ \varepsilon_{t-2} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \varepsilon_t$$
$$y_t = \begin{bmatrix} 1 & \theta \end{bmatrix}$$

and

$$\begin{bmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_{t-1} \\ \varepsilon_{t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{bmatrix}$$
$$y_t = \begin{bmatrix} 1 & \theta \end{bmatrix}$$

valid state space representations of an MA(1) process?

Maximum Likelihood and Unobserved Components Models

Unobserved Component model of Inflation

$$\begin{aligned} \pi_t &= \mu_t + \eta_t \\ \tau_t &= \tau_{t-1} + \varepsilon_t \end{aligned}$$

Decomposes inflation into permanent (τ) and transitory (η) component

- Fits the data well
  - But we may be concerned about having an actual unit root root in inflation on theoretical grounds
- Based on simplified (constant parameters) version of Stock and Watson (JMCB 2007)

The basic formulas

We want to:

1. Estimate the parameters of the system, i.e. estimate σ_η^2 and σ_ε^2
   1.1 Parameter vector is given by $\Theta = \left\{\sigma_\eta^2, \sigma_\varepsilon^2\right\}$
   1.2 $\widehat{\Theta} = \arg\max_{\theta \in \Theta} L(\pi^t \mid \boldsymbol{\Theta})$
2. Find an estimate of the permanent component τ_t at different points in time

The Likelihood function

We have the state space system

$$\begin{aligned}\pi_t &= \mu_t + \eta_t \text{ (measurement equation)} \\ \tau_t &= \tau_{t-1} + \varepsilon_t \text{ (state equation)}\end{aligned}$$

implying that $A = 1, D = 1, C = \sqrt{\sigma_\varepsilon^2}, \Sigma_v = \sigma_\eta^2$. The likelihood function for a state space system is (as always) given by

$$L(Z \mid \mathbf{\Theta}) = -\frac{nT}{2}\log(2\pi) - \frac{T}{2}\log|\Omega_t| - \frac{1}{2}\sum_{t=1}^{T}\widetilde{Z}_t'\Omega_t^{-1}\widetilde{Z}_t$$

where

$$\begin{aligned}\widetilde{Z}_t &= Z_t - DAX_{t-1|t-1} \\ \Omega_t &= DP_{t|t-1}D' + \Sigma_{vv}\end{aligned}$$

and n is the number of observable variables, i.e. the dimension of Z_t.

Starting the Kalman recursions

How can we choose initial values for the Kalman recursions?

- Unconditional variance is infinite because of unit root in permanent component
- A good choice is to choose "neutral" values, i.e. something akin to uninformative priors
  - One such choice is $X_{0|0} = \pi_1$ and $P_{0|0}$ very large (but finite) and constant

$$L(Z \mid \mathbf{\Theta}) = -\frac{nT}{2}\log(2\pi) - \frac{T}{2}\log|\Omega_t| - \frac{1}{2}\sum_{t=1}^{T}\widetilde{Z}_t'\Omega_t^{-1}\widetilde{Z}_t$$

Maximizing the Likelihood function

How can we find $\widehat{\Theta} = \arg\max_{\theta \in \Theta} L(\pi^t \mid \mathbf{\Theta})$?

- The dimension of the parameter vector is low so we can use grid search

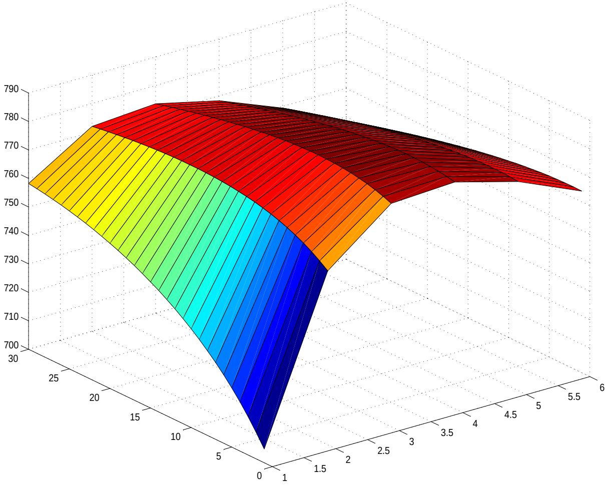
Define grid for variances σ_ε^2 and σ_η^2

$$\begin{aligned} \sigma_\varepsilon^2 &= \{0, 0.001, 0.002, ..., \sigma_{\varepsilon,\max}^2\} \\ \sigma_\eta^2 &= \{0, 0.001, 0.002, ..., \sigma_{\eta,\max}^2\} \end{aligned}$$

and evaluate likelihood function for all combinations.
How do we choose boundaries of grid?

- Variances are non-negative
- Both $\widehat{\sigma}_\varepsilon^2$ and $\widehat{\sigma}_\eta^2$ should be smaller than or equal to the sample variance of inflation so we can set $\sigma_{\varepsilon,\max}^2 = \sigma_{\eta,\max}^2 = \frac{1}{T}\sum \pi_t^2$

Maximizing the Likelihood function



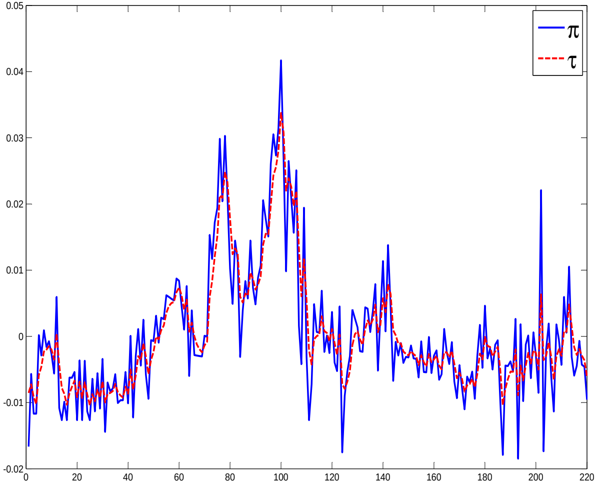
Maximizing the Likelihood function

Estimated parameter values:

- $\widehat{\sigma}_{\varepsilon}^{2} = 0.0028$
- $\widehat{\sigma}_{\eta}^{2} = 0.0051$

We can also estimate the permanent component

Actual Inflation and filtered permanent component



The Estimated path of permanent component and the Kalman Smoother

The standard filter gives an optimal *real time* estimate of the latent state

- Sometimes we are interested in the best estimate given the complete sample, i.e. $X_{t|T}$

$$X_{t|T} = E\left[X_t \mid Z^T, X_{0|0}\right]$$

The *Kalman smoother* can be used to find $X_{t|T}$

The Kalman Smoother: Implementation

Run filter forward, then backward.

$$X_{t|T} = X_{t|t} + J_{t-1}\left(X_{t+1|T} - X_{t+1|t}\right) \tag{1}$$

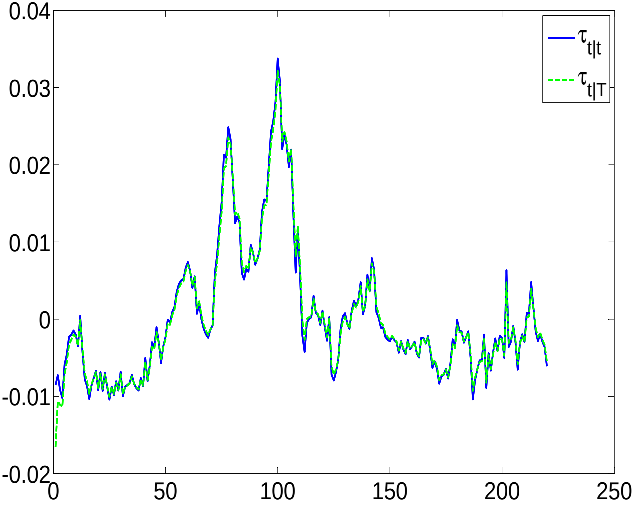
where

$$J_t = P_{t|t} A' P_{t+1|t}^{-1} \tag{2}$$

The covariances of the smoothed state estimation errors can be computed as

$$P_{t|T} = P_{t|t} + J_t\left(P_{t+1|T} - P_{t+1|t}\right) J_t'$$

(for more details, see Hamilton 1994).



0.04
0.03
0.02
0.01
0
-0.01
-0.02
0
50
100
150
200
250
$\tau_{t|t}$
$\tau_{t|T}$

The Kalman Simulation Smoother

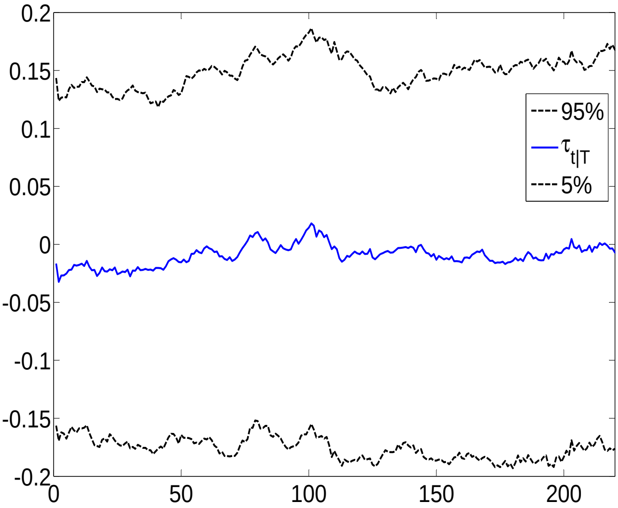
Sometimes we want to know something about the uncertainty of our smoothed estimate.

- One way to illustrate this is to use the *Kalman simulation smoother* to simulate the conditional distribution of X

$$
\begin{aligned}
p\left(X_t \mid Z^T, X_{0|0}\right) &= N(X_{t|T}, P_{t|T}) \\
P_{t|T} &= P_{t|t} + J_t \left(P_{t+1|T} - P_{t+1|t}\right) J_t'
\end{aligned}
$$

See Durbin and Koopman (2002) for more details.

Smoothed Estimate and Prob Intervals



That's it for this morning.