Kalman Filtering and Model Estimation

Steven Lillywhite

Introduction

We aim to do the following.

- Explain the basics of the Kalman Filter .
- Explain the relationship with MLE estimation.
- Show some real applications.

Overview

1. Some Applications
2. Some History
3. Minimum Variance Estimation
4. Kalman Filter
   - State-Space Form
   - Kalman Filter Algorithm
   - Initial State Conditions
   - Stability
5. Maximum Likelihood Estimation
6. Estimating Commodities Models

Applications

- Keywords: estimation, control theory, signal processing, filtering, linear stochastic systems.
- Navigational systems. Used in the Apollo lunar landing. From the NASA Ames' website:
  "The Kalman-Schmidt filter was embedded in the Apollo navigation computer and ultimately into all air navigation systems, and laid the foundation for Ames' future leadership in flight and air traffic research."
- Satellite tracking. Missile tracking. Radar tracking.
- Computer vision. Robotics. Speech enhancement.
- Economics, Math Finance.

Gauss was no dummy!

- Late 1700s. Problem: estimate planet and comet motion using data from telescopes.
- 1795. Gauss first uses least-squares method at age of 18.
- 1912. Fisher introduces the method of maximum likelihood.
- 1940s. Wiener-Kolmogorov linear minimum variance estimation technique. Signal processing. Unwieldly for large data sets.
- 1960. Kalman introduces Kalman filter.

Minimum Variance Estimators

Let $(\Omega, \mathcal{P})$ be a probability space.

Definition (Estimator)

Let X, $Y_1, Y_2, \dots Y_n \in \mathcal{L}^2(\mathcal{P})$ be random variables. Let $\mathbf{Y} \overset{\text{def}}{=} (Y_1, Y_2, \dots, Y_n)$. By an *estimator* $\hat{X}$ for X given $\mathbf{Y}$ we mean a random variable of the form $\hat{X} = g\mathbf{Y}$, where $g : \mathbb{R}^n \to \mathbb{R}$ is a given Borel-measurable function.

Definition (Minimum Variance Estimator)

An estimator $\hat{X}$ of X given $\mathbf{Y}$ is called a *minimum variance estimator* if

$$\|\hat{X} - X\| \leq \|h\mathbf{Y} - X\| \tag{1}$$

for all Borel-measurable h. Let us denote $MVE(X|\mathbf{Y}) \overset{\text{def}}{=} \hat{X}$.

Minimum Variance Estimators

- $MVE(X) = E(X|\mathbf{Y})$
- Let $\mathcal{M}(\mathbf{Y}) \overset{\text{def}}{=} \{g\mathbf{Y}|g \text{ Borel-measurable, } g\mathbf{Y} \in \mathcal{L}^2(\mathcal{P})\}$. Then $\mathcal{M}(\mathbf{Y})$ is a closed subspace of $\mathcal{L}^2(\mathcal{P})$, and $MVE(X|\mathbf{Y})$ is the projection of X onto $\mathcal{M}(\mathbf{Y})$.
- As a corollary, $MVE(X|\mathbf{Y})$ exists, is unique, and is characterized by the condition:
$$(MVE(X|\mathbf{Y}) - X) \perp \mathcal{M}(\mathbf{Y}) \tag{2}$$

Linear Minimum Variance Estimators

Definition (Estimator)

Let X, $Y_1, Y_2, \ldots Y_n \in \mathcal{L}^2(\mathcal{P})$ be random variables. Let $\mathbf{Y} \overset{\text{def}}{=} (Y_1, Y_2, \ldots, Y_n)$. By an *linear estimator* $\hat{X}$ for X given $\mathbf{Y}$ we mean a random variable of the form $\hat{X} = g\mathbf{Y}$, where $g : \mathbb{R}^n \to \mathbb{R}$ is a given linear function.

Let's ramp it up a bit by letting X be multi-dimensional.

Definition (Best Linear Minimum Variance Estimator)

Let $\mathbf{X} \in \mathcal{L}^2(\mathcal{P})^n$,$\mathbf{Y} \in \mathcal{L}^2(\mathcal{P})^m$. An linear estimator $\hat{\mathbf{X}}$ of $\mathbf{X}$ given $\mathbf{Y}$ is called a *best linear minimum variance estimator* if

$$\|\hat{\mathbf{X}} - \mathbf{X}\| \leq \|h\mathbf{Y} - \mathbf{X}\| \tag{3}$$

for all linear h. Let us denote $BLMVE(\mathbf{X}|\mathbf{Y}) \overset{\text{def}}{=} \hat{X}$. Here h is given by an $n \times m$ matrix.

Linear Minimum Variance Estimators

- If $\mathbf{X}$ and $\mathbf{Y}$ are multivariate normal, then $MVE(\mathbf{X}|\mathbf{Y}) = BLMVE(\mathbf{X}|\mathbf{Y})$ (up to a constant term).

State-Space Form

Definition (State-Space Form)

The state-space form is defined by the following pair of equations:

$$x_{i+1} = J_i x_i + g_i + u_i \qquad \text{(state)}$$

$$z_i = H_i x_i + b_i + w_i \qquad \text{(observation)}$$

- Here x_i, z_i are vectors representing a discrete random variables.
- In general the elements of x_i are not observable.
- We assume that the elements of z_i are observable.
- u_i and w_i are white noise processes.
- We assume that all vectors and matrices take values in Euclidean space and can vary with i, but apart from x_i and z_i, that they only vary in a deterministic manner.

State-Space Form

Furthermore, we denote $Q_i \overset{\text{def}}{=} E(u_i u_i^T)$ and $R_i \overset{\text{def}}{=} E(w_i w_i^T)$, and assume that the following hold:

$$E(u_i x_0^T) = 0 \tag{4}$$

$$E(w_i x_0^T) = 0 \tag{5}$$

$$E(u_i w_j^T) = 0 \; for \; all \; i, j \tag{6}$$

Kalman Filter Notation

Definition

Denote $\mathcal{Y}_j \overset{\text{def}}{=} (z_0, z_1, \ldots z_j)$. By $\hat{x}_{i|j}$, (resp. $\hat{z}_{i|j}$) we shall mean the best linear minimum variance estimate(BLMVE) of x_i (resp. z_i) based on $\mathcal{Y}_j$. We also define

$$P_{i|j} \overset{\text{def}}{=} E\{(x_i - \hat{x}_{i|j})(x_i - \hat{x}_{i|j})^T\} \tag{7}$$

and call this the error matrix. When $i = j$, the estimate is called a *filtered* estimate, when $i > j$, the estimate is called a *predicted* estimate, and when $i < j$, the estimate is called a *smoothed* estimate

Discrete Kalman Filter

Theorem (Kalman, 1960)

The BLMVE $\hat{x}_{i|i}$ may be generated recursively by

$$
\begin{aligned}
\hat{x}_{i+1|i} &= J_i \hat{x}_{i|i} + g_i && \text{(predicted state)}\\
P_{i+1|i} &= J_i P_{i|i} J_i^T + Q_i && \text{(predicted state error matrix)}\\
\hat{z}_{i+1|i} &= H_{i+1} \hat{x}_{i+1|i} + b_{i+1} && \text{(predicted observation)}\\
r_{i+1} &\overset{def}{=} z_{i+1} - \hat{z}_{i+1|i} && \text{(predicted obs error)}\\
\Sigma_{i+1} &\overset{def}{=} H_{i+1} P_{i+1|i} H_{i+1}^T + R_{i+1} && \text{(predicted obs error matrix)}\\
K_{i+1} &= P_{i+1|i} H_{i+1}^T \Sigma_{i+1}^{-1} && \text{(Kalman gain)}\\
\hat{x}_{i+1|i+1} &= \hat{x}_{i+1|i} + K_{i+1} r_{i+1} && \text{(next filtered state)}\\
P_{i+1|i+1} &= [I - K_{i+1} H_{i+1}] P_{i+1|i} && \text{(next filtered state error matrix)}
\end{aligned}
$$

Discrete Kalman Filter

- If the initial state x_0 and the innovations u_i, w_i are multivariate Gaussian, then the forecasts $\hat{x}_{i|j}$, (resp. $\hat{z}_{i|j}$) are minimum variance estimators(MVE).
- Note that the updated filtered state estimate is a sum of the predicted state estimate and the predicted observation error weighted by the gain matrix.
- Observe that the gain matrix is proportional to the predicted state error covariance matrix, and inversely proportional to the predicted observation error covariance matrix. Thus, in updating the state estimator, more weight is given to the observation error when the error in the predicted state estimate is large, and less when the observation error is large.

Kalman Filter Initial State Conditions

To run the Kalman filter, we begin with the pair $\hat{x}_{0|0}$, $P_{0|0}$ (alternatively, one may also use $\hat{x}_{1|0}$, $P_{1|0}$). A difficuly with the Kalman filter is the determination of these initial conditions. In many real applications, the distribution for x_0 is unknown. Several approaches are possible.

- For stationary state series, we can compute $\hat{x}_{0|0}$, $P_{0|0}$ directly.
- Prior information.
- Diffuse prior: $\hat{x}_{0|0} = 0$, and $P_{0|0} = kI, \; k \gg 0$. The details are more involved.
- Or one may treat x_0 as a fixed vector, taking $\hat{x}_{0|0} = x_0$, and $P_{0|0} = 0$, and estimate its components by treating them as extra parameters in the model. The details are more involved.
- General rule of thumb is that for long time series, the initial state conditions will have little impact.

Kalman Filter Stability

- Under certain conditions, the err matrices $P_{i+1|i}$(equivalently $P_{i|i}$) will stabilize

$$\lim_{i\to\infty} P_{i+1|i} = \bar{P} \tag{8}$$

  with $\bar{P}$ independent of $P_{1|0}$. Convergence is often exponentially fast.
- This means that for stable filters, the initial state conditions won't have much impact so long as we have enough data to get to a stable state. Need to be more concerned with initial state conditions in small samples.
- We gain significant computational advantage exploiting convergence in the filter. Especially when the matrices are time-invariant, the the predicted observation err matrix and the Kalman gain stabilize, too. See next slide.

Kalman Filter Stability

This part is independent of the data

$$P_{i+1|i} = J_i P_{i|i} J_i^T + Q_i \qquad \text{(predicted state error matrix)}$$

$$\Sigma_{i+1} \overset{\text{def}}{=} H_{i+1} P_{i+1|i} H_{i+1}^T + R_{i+1} \qquad \text{(predicted obs error matrix)}$$

$$K_{i+1} = P_{i+1|i} H_{i+1}^T \Sigma_{i+1}^{-1} \qquad \text{(Kalman gain)}$$

$$P_{i+1|i+1} = [I - K_{i+1} H_{i+1}] P_{i+1|i} \qquad \text{(next filtered state error matrix)}$$

$$\hat{x}_{i+1|i} = J_i \hat{x}_{i|i} + g_i \qquad \text{(predicted state)}$$

$$\hat{z}_{i+1|i} = H_{i+1} \hat{x}_{i+1|i} + b_{i+1} \qquad \text{(predicted observation)}$$

$$r_{i+1} \overset{\text{def}}{=} z_{i+1} - \hat{z}_{i+1|i} \qquad \text{(predicted obs error)}$$

$$\hat{x}_{i+1|i+1} = \hat{x}_{i+1|i} + K_{i+1} r_{i+1} \qquad \text{(next filtered state)}$$

Kalman Filter Divergence

- Numerical instability in the algorithm, round-off errors, etc., can cause divergence in the filter.
- Model fit. If the underlying state model does not fit the real-world process well, then the filter can diverge.
- Observability. If we cannot observe some of the state variables(or linear combinations), then we can get divergence in the filter.

Kalman Filter Other Items

- Kalman advantage: real-time updating. No need to store past data to update current data.
- Can handle missing data, since the matrices in the algorithm can vary over time.
- Smoothing. The filter algorithm above gives BLMVE at time t based on data up to time t. However, once all data is in, we can make better estimates of the state variables at time t using also data after time t.
- Alternative forms for the filter algorithm based on algebraic manipulation of the equations. Information filter computes $P_{i|i}^{-1}$. Depending on the situation, this can be more(or less) useful. Square-Root filter uses square roots of $P_{i|i}^{-1}$. It is more computationally burdensome, but can improve numerical instability problems.

Kalman Filter Other Items

- Non-linear state-space filters. This is called the *Extended Kalman Filter*. Here, we allow arbitrary functions in the state-space formulation, rather than the linear functions above.

$$x_{i+1} = f(x_i, g_i, u_i) \qquad \text{(state)}$$

$$z_i = h(x_i, b_i, w_i) \qquad \text{(observation)}$$

  One proceeds by linearizing the functions about the estimates at each step, and thereby obtain an analogous filter algorithm.

- There is a continuous version of the filter due to Kalman and Bucy.

Maximum Likelihood Estimation

If the initial state x_0 and the innovations u_i, w_i are multivariate Gaussian, then the distribution of z_i conditional on the set $\mathcal{Y}_{i-1}$ is also Gaussian, and the error matrices above are covariance matrices of the error random variables.

$$z_i | \mathcal{Y}_{i-1} \sim N(\hat{z}_{i|i-1}, \Sigma_i) \tag{9}$$

Now let us suppose that the state-space vectors and matrices depend on certain unknown parameters. Let us denote by θ the vector of these parameters. We may form the likelihood function by taking the joint probability density function(pdf):

$$L(z; \theta) = \prod_{i=1}^{n} pdf(z_i | \mathcal{Y}_{i-1}) \tag{10}$$

Maximum Likelihood Estimation

Then the log of the likelihood function is

$$\log(L(z;\theta)) = -\tfrac{mn}{2}\log(2\pi) - \tfrac{1}{2}\sum_{i=1}^{n}\{\log(\det(\Sigma_i)) + {r_i}^T \Sigma_i^{-1} r_i\} \qquad (11)$$

By maximizing $\log(L(z;\theta))$ with respect to θ for a particular realization of z, we obtain the maximum likelihood estimates for the parameters. The main point of this section is that $\log(L(z;\theta))$ may be computed via the Kalman filter, since the algorithm naturally computes both r_i and Σ_i.

Estimating Commodities Models

- Kalman filtering with maximum likelihood can be used to estimate parameters in various models in financial engineering applications.
- One such use is for the estimation of parameters in commodities models. Here the state system would model the spot price, and the observation system would be futures prices.

  Kalman filtering can be useful here for the following reasons
- It is not uncommon that there is no true spot price process in the real world.
- Even if there is a spot price process, it can be be highly illiquid, error prone, and unreliable for modelling.
- In multi-factor models, we may have the spot price divided into unobservable components.

A Two-Factor Model of Schwartz-Smith for Oil prices

Combining geometric brownian motion with mean-reversion. This is the short-long model of Schwartz-Smith. The idea is that short-term variations revert back to an equilibrium level. But the equilibrium level is uncertain, and follows a Brownian motion process with drift.

$$ln(S_t) = \xi_t + \chi_t \tag{12}$$

$$d\xi_t = \mu_\xi dt + \sigma_\xi dW_\chi \tag{13}$$

$$d\chi_t = -\kappa\chi_t dt + \sigma_\chi dW_\xi \tag{14}$$

$$dW_\chi dW_\xi = \rho_{\chi\xi} dt \tag{15}$$

A Two-Factor Model of Schwartz-Smith for Oil prices

We shall use the risk-neutral framework to price derivatives. We shall assume that the market price of risk is constant, and we write:

$$d\xi_t = (\mu_\xi - \lambda_\xi)dt + \sigma_\xi d\tilde{W}_\chi \tag{16}$$

$$d\chi_t = (-\kappa\chi_t - \lambda_\chi)dt + \sigma_\chi d\tilde{W}_\xi \tag{17}$$

Here $(\lambda_\xi\sigma_\xi, \lambda_\chi\sigma_\chi)$ denotes the change of measure according to the Girsanov theorem. We shall denote $\mu_\xi^* = \mu_\xi - \lambda_\xi$. Under these assumptions, we obtain that the distribution of $ln(S_t)$ is normal, and

$$ln(F(t,T)) = e^{-\kappa(T-t)}\chi_t + \xi_t + A(T-t) \tag{18}$$

$$\begin{aligned} A(T-t) \quad = \quad & \mu_\xi^*(T-t) - (1 - e^{(-\kappa(T-t))}\lambda_\chi/\kappa \\ & + \tfrac{1}{2}(1 - e^{-2\kappa(T-t)}) + \tfrac{1}{2}\sigma_\xi^2(T-t) \\ & + (1 - e^{-\kappa(T-t)})\rho_{\chi\xi}\sigma_\chi\sigma_\xi/\kappa \end{aligned}$$

Parameter Estimation Via Kalman filter MLE

- State Equation: discretize the model for the spot price
- Measurement Equation: discretize the formula for the futures price in terms of the spot. Add noise term.
- We estimate the parameters in the model using maximum likelihood with the Kalman filter.

State-Space Formulation

$$x_t = Jx_{t-1} + g + u_{t-1} \tag{19}$$

where

$$x_t = \begin{bmatrix} \chi_t \\ \xi_t \end{bmatrix} \qquad g = \begin{bmatrix} 0 \\ \mu \Delta t \end{bmatrix} \qquad J = \begin{bmatrix} e^{-\kappa \Delta t} & 0 \\ 0 & 1 \end{bmatrix} \tag{20}$$

and $u_t \sim N(0, Q)$, with

$$Q = \begin{bmatrix} (1 - e^{-2\kappa \Delta t})\sigma_\chi^2 / 2\kappa & (1 - e^{-\kappa \Delta t})\rho\sigma_\chi\sigma_\xi/\kappa \\ (1 - e^{-\kappa \Delta t})\rho\sigma_\chi\sigma_\xi/\kappa & \sigma_\xi^2 \Delta t \end{bmatrix} \tag{21}$$

Here, Δt represents the data frequency, or time between observations. Note that J, g, and Q do not vary with time.

State-Space Formulation

For the observation equation, we have

$$z_t = H_t x_t + b_t + w_t \tag{22}$$

where

$$z_t = [\log F_{T_1}(t) \ \ \log F_{T_2}(t) \ \ \ldots \ \ \log F_{T_m}(t)]^T \tag{23}$$

$$b_t = [A(\phi(t, T_1) - t) \ A(\phi(t, T_2) - t) \ \ \ldots \ \ A(\phi(t, T_m) - t)]^T \tag{24}$$

$$H_t = \begin{bmatrix} e^{-\kappa(\phi(t,T_1)-t)} & 1 \\ e^{-\kappa(\phi(t,T_2)-t)} & 1 \\ \ldots & \\ e^{-\kappa(\phi(t,T_m)-t)} & 1 \end{bmatrix} \tag{25}$$

State-Space Formulation

Here $w_t \sim N(0, R)$ represents measurement error, which could come about via error in price reporting, or alternatively can represent errors in fitting the model. To simplify the problem, it is common practice to take the matrix R to be diagonal, or even a constant times the identity. This corresponds to assuming that the measurement errors are not correlated, resp. that measurement errors are uncorrelated *and* equal for all maturities. This assumption has the effect of introducing m, resp. 1, extra parameter(s) to be estimated with the model.