Time Varying Beta:
Intro to Kalman Filters

Financial Modeling
Professor Doug Blackburn

CAPM

- You have discussed the CAPM extensively from a theoretical perspective in Finance Theory.
- To estimate: we need expected market return, expected asset return, and a risk-free rate.
  - Expected returns are not observable
  - Typically estimated using historic averages.
- This is enough for us to estimate beta for a stock.
  - Beta is not observable

The linear regression

- This is covered extensively in Financial Econometrics
- The idea
  - Use past data to determine the single beta that minimizes the distance between the estimated CAPM return and actual observed returns.

Linear Regression

We start with the empirical form of the CAPM:

$$r_{i,t} - r_f = \alpha_i + \beta_i \left[r_{m,t} - r_f \right] + \varepsilon_t$$

such that $\mathrm{E}[\varepsilon_t] = 0, \mathrm{E}[\varepsilon_t^2] = \sigma_\varepsilon^2, \mathrm{cov}(\varepsilon_t, r_{m,t} - r_f) = 0,$

$\mathrm{cov}(\varepsilon_t, \varepsilon_{t-j}) = 0$ for all $\mathrm{j} \neq 0$.

We want to minimize the sum of squared residuals

$$Min \sum \varepsilon_t^2$$

Linear Regression

Consider the simple regression model: $y_t = \alpha + \beta x_t + \varepsilon_t \quad = \quad \vec{\beta}^T \vec{x}_t + \varepsilon_t$

where $\vec{\beta}^T = (\alpha, \beta)$, and $\vec{x}_t^{\,T} = (1, x_t)$

Then we can write $\varepsilon_t = y_t - \vec{\beta}^T \vec{x}_t$

Then, $\sum \varepsilon_t^2 = \sum (y_t - \vec{\beta}^T \vec{x}_t)^2$

Rewriting in matrix notation: $\vec{\varepsilon}^T \vec{\varepsilon} = \left(\vec{y} - \mathbf{X}\vec{\beta}\right)^T \left(\vec{y} - \mathbf{X}\vec{\beta}\right)$

Linear Regression

Multiply: $\vec{\varepsilon}^T \vec{\varepsilon} = \left(\vec{y} - \mathbf{X}\vec{\beta}\right)^T \left(\vec{y} - \mathbf{X}\vec{\beta}\right)$

$$\vec{\varepsilon}^T \vec{\varepsilon} = \vec{y}^T \vec{y} - \vec{\beta}^T \mathbf{X} \vec{y} - \vec{y}^T \mathbf{X} \vec{\beta} + \vec{\beta}^T \mathbf{X}^T \mathbf{X} \vec{\beta}$$

$$s(\vec{\beta}) \; = \vec{y}^T \vec{y} - 2\vec{y}^T \mathbf{X} \vec{\beta} + \vec{\beta}^T \mathbf{X}^T \mathbf{X} \vec{\beta}$$

$$\frac{\partial s}{\partial \vec{\beta}} = -2\mathbf{X}^T \vec{y} + 2\mathbf{X}^T \mathbf{X} \vec{\beta} = 0 \quad \Rightarrow \quad \mathbf{X}^T \mathbf{X} \vec{\beta} = \mathbf{X}^T \vec{y}$$

$$\boxed{\vec{\beta} = \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \vec{y}}$$

The optimal linear regression coefficients

Linear Regression

Simple Example: Suppose we are given the following data

1/2010	$r_j - r_f = 5.0\%$	$r_m - r_f = 3.5\%$
2/2010	$r_j - r_f = 6.2\%$	$r_m - r_f = 4.1\%$
3/2010	$r_j - r_f = 5.5\%$	$r_m - r_f = 3.3\%$
4/2010	$r_j - r_f = 3.2\%$	$r_m - r_f = 2.8\%$
5/2010	$r_j - r_f = -1.0\%$	$r_m - r_f = 1.1\%$

$$\text{Then we have } \vec{y} = \begin{pmatrix} 5.0 \\ 6.2 \\ 5.5 \\ 3.2 \\ -1.0 \end{pmatrix} \quad \text{and} \quad \mathbf{X} = \begin{pmatrix} 1 & 3.5 \\ 1 & 4.1 \\ 1 & 3.3 \\ 1 & 2.8 \\ 1 & 1.1 \end{pmatrix}$$

Linear Regression

$$\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \begin{pmatrix} 1.88 & -0.57 \\ -0.57 & 0.19 \end{pmatrix}$$

$$\mathbf{X}^T\vec{y} = \begin{pmatrix} 18.9 \\ 68.93 \end{pmatrix}$$

$$\vec{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\vec{y} = \begin{pmatrix} -3.62 \\ 2.50 \end{pmatrix}$$

Constant

Beta

Alternative: Estimating a Single Beta

We know from theory that $\beta = \dfrac{\operatorname{cov}(r_m, r_i)}{\sigma_m^2}$.

This gives us the same estimate for beta.

From the previous example, $\operatorname{cov} = 2.5972$ and $\sigma_m^2 = 1.0384 \Rightarrow \beta = 2.50$.

Note: When computing variance in Excel using VAR( ), Excel calculates a sample variance $\sum \dfrac{(\mathrm{x} - \bar{\mathrm{x}})^2}{N-1}$. You should use the command VARP( ) to estimate the population variance $\sum \dfrac{(\mathrm{x} - \bar{\mathrm{x}})^2}{N}$

2nd Alternative: Estimating Beta and Alpha

- Use the worksheet functions
  - Slope(y_variable, x_variable)
  - Intercept(y_variable, x_variable)
- This will work for the simple case of only one factor … but not the multi-factor case.

$$y_t = \alpha + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{3,t} + \varepsilon_t$$

  - Returns may depend on many different variables
    - Market return, unemployment, GDP, interest rates, etc …

What is beta?

- Theoretically
  - Two types of risk: systematic and idiosyncratic
  - Beta measures systematic risk
- Practically
  - Measures sensitivity to a particular factor
  - You can have:
    - Market betas, Interest rate betas, liquidity betas, etc …

What is alpha?

- The CAPM states:

$$E[r_i - r_f] = \beta\left(E[r_m - r_f]\right)$$

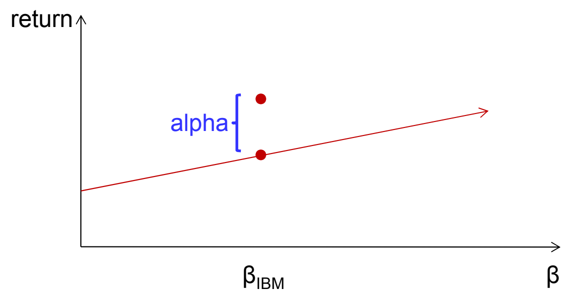
- But, the empirical model looks like

$$r_{i,t} - r_f = \alpha_i + \beta_i \left[r_{m,t} - r_f\right] + \varepsilon_t$$

What is alpha?

- Alpha is the difference between the observed returns and the CAPM expected returns.

$$\alpha_i = \left(r_{i,t} - r_f\right) - \left(\beta_i \left[r_{m,t} - r_f\right] + \varepsilon_t\right)$$



Problems using Linear Regression

- There are many assumptions behind the linear regression.
  - You will talk about these in other classes.
  - Warning: do not naively use linear regressions. First understand the assumptions!!
- Fundamentally, we need to use about 60 months of return data to estimate beta.
  - Is beta for a stock constant over 60 months?
  - Is variance and covariance constant over 60 months?
  - What if beta is time-varying?

CAPM with Time-Varying Beta

- Beta changes with time
  - Time-varying variances (GARCH, Stochastic volatility models)
  - Time-varying correlations (multivariate GARCH)

- CAPM Model

$$E[r_{j,t}] = r_{f,t} + \beta_t \left(E[r_{m,t}] - r_{f,t} \right)$$

or

$$E[R_{j,t}] = \beta_t E[R_{m,t}]$$

Rolling Linear Regression

- This is the commonly used method of estimating time-varying beta.

- Similar to a moving average.

- The idea is to use the past N data points and estimate beta using linear regression.

- Use t=1 to t=N to estimate beta at t=N+1
- Use t=2 to t=N+1 to estimate beta at t=N+2
- Use t=3 to t=N+2 to estimate beta at t=N+3
- And so on …

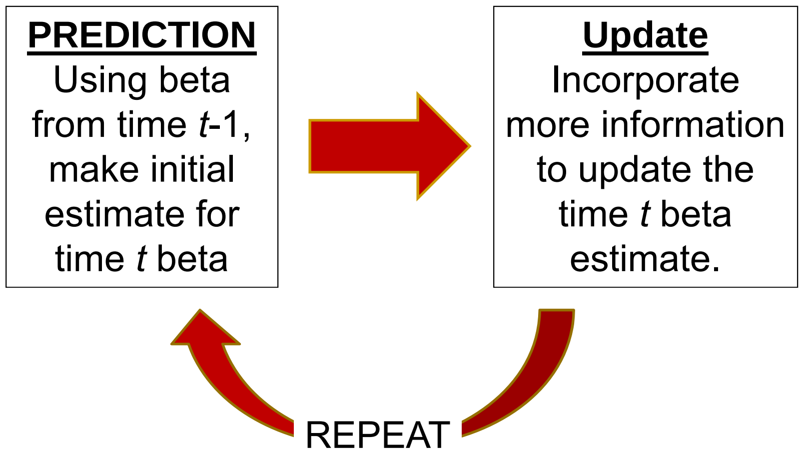
Kalman Filter

- Purpose
  - Estimate the state of a system through time from measurements which contain random errors.
- Example:
  - Your car has a GPS that tracks your position.
  - You enter into a tunnel losing the connection to the satellite.
  - A Kalman Filter can be used to estimate your position as you travel through the tunnel based on the last position, the direction you were driving, and the velocity of your car.

Kalman Filter

- The Kalman filter was first used in the 1960s.
- Since, it has been used extensively
  - In signal processing
  - Tracking moving objects (space crafts, satellites, etc.)
  - And, more recently, in finance to forecast movements in financial and economic variables.

Kalman Filter



Two Stages

- Stage 1 – Prediction (Time Update)
  - *A state equation is assumed*. This describes how the state variable changes over time. We use this assumption to create a prior estimate for beta.
  - Of course, *the state equation contains error*. With each updating of the state, we need to describe the error in our prediction.
  - To estimate beta, we can assume that the beta follows an AR(1) process.
    - AR(1)=First order AutoRegression $\beta_t = A\beta_{t-1}$

Two Stages

- Stage 2 – Correction (Measurement Update)
  - We assume a measurement equation. We use this equation to update our prior estimate.
  - Again, this will be measured with error so we must also update the error measurement.
  - For our example, we will assume that the CAPM is true and use it to update our prior estimate of beta.

Detailed Example with Derivations

- Suppose the CAPM beta is time-varying
- We believe that beta follows the dynamic system.

$$\beta_{t+1} = A\beta_t + \omega_t$$

The error is uncorrelated with beta.

$$\omega_t \sim N(0, Q)$$

- We assume here that A=1 and Q=0.002.

Initial Guess

- We start with an initial guess of beta
  - Beta at t=0 is 0.96.

- We know that this is measured with error.
  - We assume that our initial guess of beta has variance of P_0=0.0025.
  - That is, we believe with 68% probability that the true beta is between 0.91 and 1.01.

Stage 1 - Prediction

We make a first guess of beta at time 1.
Use the state equation:

$\hat{\beta}_1^- = 1(0.96) = 0.96$ with error:

$$\hat{P}_1^- = E\left[\left(\beta_1 - \hat{\beta}_1^-\right)^2\right] = E\left[\left(A\beta_0 + \omega_0 - A\hat{\beta}_0^-\right)^2\right]$$

Is zero because of zero correlation between estimate and error

$$= E\left[A^2\left(\beta_0 - \hat{\beta}_0^-\right)^2\right] + E\left[\omega_0^2\right] + 2AE\left[\left(\beta_0 - \hat{\beta}_0^-\right)\omega_0\right] = A^2E\left[\left(\beta_0 - \hat{\beta}_0^-\right)^2\right] + Q$$

$$\boxed{\hat{P}_1^- = A^2P_0 + Q}$$

Stage 1

- For our problem:

$$\hat{\beta}_1^- = 0.96$$

$$\hat{P}_1^- = 1^2(0.0025) + 0.002 = 0.0045$$

- This is our best initial guess (prior estimate) for beta at t=1.
- We now need to use more information to update our estimate of beta.

Stage 2

We observe a market return of $R_m = 0.016$ and asset j's return $R_j = 0.053$.
The CAPM relation is (note: R_m and R_j are excess of R_f):

$$R_j = \beta_t R_m + \varepsilon_{j,t} \quad \text{with} \quad \varepsilon_{j,t} \sim N(0,R) \quad \text{and} \quad Cov(\varepsilon_{j,t}, R_m) = 0.$$

Assume $R = 0.003$.

Plugging our initial guess $\hat{\beta}_1^-$ into the CAPM, we find that $\hat{\varepsilon}_{j,1} = 0.03764$.

This error can be used to correct our initial guess $\hat{\beta}_1^-$.

Stage 2

We assume that $\hat{\beta}_t = \hat{\beta}_t^- + K\,\hat{\varepsilon}_t$ for some choice of K (the Kalman Gain).

Note that the we have two sources of error in the CAPM.

1) We have error in our first estimate of beta.

2) We have error caused by the random noise ω.

If all errors were of the first type, we could simply adjust our beta by $\hat{\varepsilon}_1$ $(K = 1)$.

Stage 2 – What is K?

We assume $\hat{\beta}_1 = \hat{\beta}_1^- + K_1\hat{\varepsilon}_1$, and compute the updated error $\hat{P}_1$.

$$\hat{P}_1 = E\left[\left(\beta_1 - \hat{\beta}_1\right)^2\right] = E\left[\left(\beta_1 - \hat{\beta}_1^- - K_1\hat{\varepsilon}_1\right)^2\right]$$

$$= E\left[\left(\beta_1 - \hat{\beta}_1^- - K_1(R_{j,1} - \hat{\beta}_1^- R_{m,1})\right)^2\right] \quad \text{using the CAPM relation}$$

$$= E\left[\left(\beta_1 - \hat{\beta}_1^- - K_1(\beta_1 R_{m,1} + \varepsilon_1 - \hat{\beta}_1^- R_{m,1})\right)^2\right] \quad \text{using the CAPM with true } \beta$$

$$= E\left[(1 - K_1 R_{m,1})^2\left(\beta_1 - \hat{\beta}_1^-\right)^2 + (K_1\varepsilon_1)^2\right] \quad \text{terms drop out due to zero correlation}$$

Stage 2 – What is K?

We assume $\hat{\beta}_1 = \hat{\beta}_1^- + K_1\hat{\varepsilon}_1$. We compute the updated error $\hat{P}_1$.

$$\hat{P}_1 = E\left[\left(1 - K_1 R_{m,1}\right)^2\left(\beta_1 - \hat{\beta}_1^-\right)^2 + \left(K_1\varepsilon_1\right)^2\right]$$

$$= \left(1 - K_1 R_{m,1}\right)^2 E\left[\left(\beta_1 - \hat{\beta}_1^-\right)^2\right] + K_1^2 E\left[\varepsilon_1^2\right] = \left(1 - K_1 R_{m,1}\right)^2 \hat{P}_1^- + K_1^2 R$$

We choose K_1 to minimize the variance $\hat{P}_1$. Take a derivative and solve for K_1.

$$\boxed{K_1 = \frac{\hat{P}_1^- R_{m,1}}{\hat{P}_1^- R_{m,1}^2 + R}}$$

Stage 2 - Correction

We use our derived form of K_1 to determine the corrected β.

$$K_1 = \frac{0.0045(0.016)}{0.0045(0.016)^2 + 0.003} = 0.023991$$

$$\hat{P}_1 = (1 - 0.023991(0.016))^2(0.0045) + 0.023991^2(0.003)$$
$$= 0.004498$$

$$\hat{\beta}_1 = \hat{\beta}_1^- + K_1\varepsilon_1 = 0.96 + 0.023991(0.03764)$$
$$= 0.960903$$

Our best guess of beta at $t = 1$ is $\hat{\beta}_1 = 0.960903$ with error $\hat{P}_1 = 0.004498$

Repeat the process for t=2

$$\hat{\beta}_2^- = A\hat{\beta}_1 = 1(0.960903) = 0.960903$$

$$\hat{P}_2^- = A^2\hat{P}_1 + Q = 1^2(0.004498) + 0.002 = 0.006498$$

We observe $R_{m,2} = 0.014$ and $R_{j,2} = -0.009$.

$$\hat{\varepsilon}_2 = -0.009 - 0.960903(0.014) = -0.02245$$

$$K_2 = \frac{\hat{P}_2^- R_{m,2}}{\hat{P}_2^- R_{m,2}^2 + R} = \frac{0.006498(0.014)}{0.006498(0.014)^2 + 0.003} = 0.030312$$

Repeat the Process for t=2

$$\hat{\beta}_2 = \hat{\beta}_2^- + K_2\varepsilon_2 = 0.960903 + 0.030312(-0.02245)$$
$$= 0.960222$$

$$\hat{P}_2 = (1 - K_2 R_{m,2})^2 \hat{P}_2^- + K_2^2 R$$
$$= 0.006496$$

And repeat for time t = 3, ...

In General

Prediction Stage:

$$\hat{\beta}_t^- = A\hat{\beta}_{t-1} \qquad \omega_t \sim N(0,Q) \quad Cov(\omega_t, \hat{\beta}_t) = 0 \text{ for all } t$$

$$\hat{P}_t^- = A^2\hat{P}_{t-1} + Q$$

Correction Stage:

$$R_{j,t} = \hat{\beta}_t^-\left(R_{m,t}\right) + \varepsilon_t \qquad \varepsilon_t \sim N(0,R) \quad Cov(\varepsilon_t, R_{m,t}) = 0 \text{ for all } t$$

$$\hat{\beta}_t = \hat{\beta}_t^- + K_t\varepsilon_t$$

$$K_t = \frac{\hat{P}_t^- R_{m,t}}{\hat{P}_t^- R_{m,t}^2 + R}$$

$$\hat{P}_t = \left(1 - K_t R_{m,t}\right)^2 \hat{P}_t^- + K_t^2 R$$

An Example - methodology

- To see how well the three methods perform, I do the following experiment:

1. Download returns for a large index of stocks (market returns) and returns for a 3-month bond (risk-free rate) … 1950-2008 monthly data
2. Choose a time-series for beta.
3. Randomly determine an error term $\varepsilon_t \sim N(0,\sigma)$
4. Use the CAPM to determine the return for a “fake” stock that uses our chosen beta and the random error.

Simulated Stock Returns

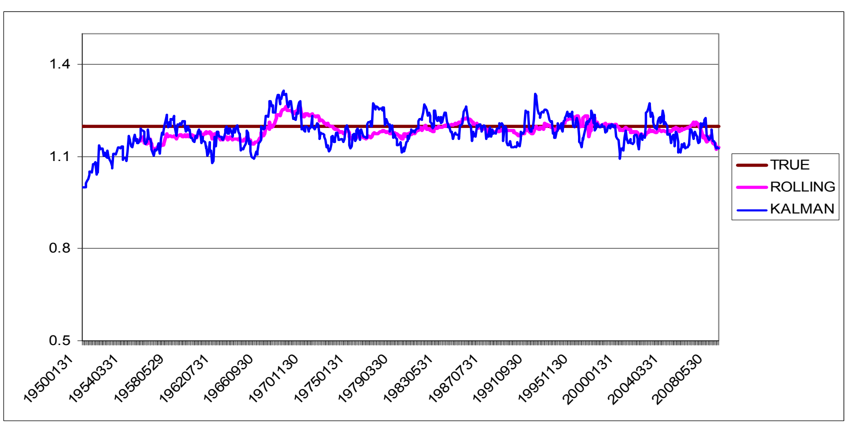
- I start with a time-varying beta (not random).
  - I know beta … I want to see of the method is able to estimate betas similar to the time series of true betas.
- For the random errors, I chose $\varepsilon_t \sim N(0, 0.01)$
- Then use the CAPM to determine the return.
- Suppose at t=1, $\beta=1$, $r_m - r_f$ = 1.2% and the randomly picked error is 0.008. Then,

$$r_j - r_f = \beta(r_m - r_f) + \varepsilon = 1(0.012)+0.008 = 0.02$$

Kalman Filter vs. Rolling vs. True Beta

example 1: Constant Beta = 1.2

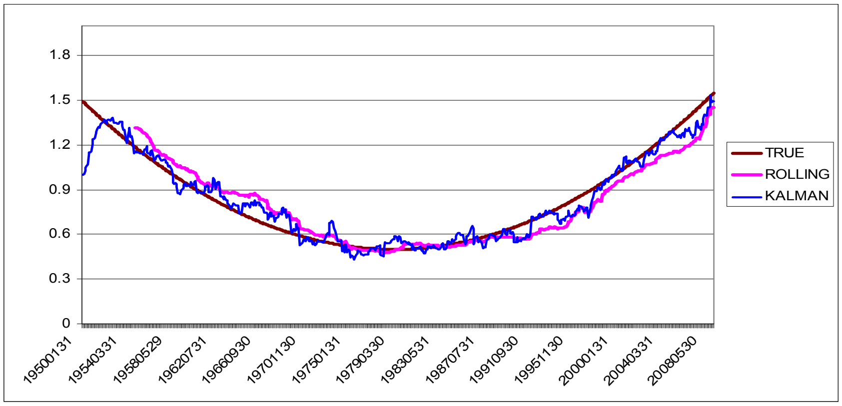
Linear regression finds β=1.18.



Kalman Filter Beta vs. True Beta

example 2: Smooth Beta

Linear regression finds β=0.810.



Kalman Filter Beta vs. True Beta

example 3: Step Function Beta

Linear regression finds β=1.013.

Other uses of Kalman Filters

- Following the evolution of short rates using interest rate / term structure models
- Modeling stochastic volatility
- Pairs trading
  - The formulation is more complicated than the simple algorithm we discussed.
  - Using Kalman filters, we can forecast the evolution of the price ratio.

Kalman Filter and Pairs Trading

Consider the time grid given by: $t_k = k\tau$ where τ is one day.

We begin by taking the difference between two prices $y_k = P_{1k} - aP_{2k}$.
It is assumed that the observed spread y_k is a noisy observation of some mean-reverting state process x_k.

The mean reverting state process is modeled using an Ornstein-Uhlenbeck process:

$$x_{k+1} - x_k = (a - bx_k)\tau + \sigma\sqrt{\tau}\varepsilon_{k+1} \qquad \sigma > 0,\ b > 0,\ a \in \Re,\ \varepsilon \sim N(0,1)$$

Note that b is the speed of mean reversion and $\frac{a}{b}$ is the long term mean.

Kalman Filter and Pairs Trading

$$x_{k+1} - x_k = (a - bx_k)\tau + \sigma\sqrt{\tau}\,\varepsilon_{k+1} \qquad \sigma > 0,\ b > 0,\ a \in \Re,\ \varepsilon \sim N(0,1)$$

The O - U state process can be rewritten as

$$x_{k+1} = A + Bx_k + C\varepsilon_{k+1} \qquad \Leftarrow \text{This is the state equation}$$

The observation process will simply compare the true observed spread y_k with the state variable x_k :

$$y_k = x_k + D\omega_k \qquad \text{where } \omega_k \sim N(0,1),\ D > 0$$

Kalman Filter and Pairs Trading

Given A, B, C, and D, we can derive all equations to implement the Kalman Filter :

$$x_{k+1}^{-} = A + Bx_k^{+}$$
$$P_{k+1}^{-} = B^2 P_k^{+} + C^2$$
$$K_{k+1} = P_{k+1}^{-} / \left(P_{k+1}^{-} + D^2\right)$$
$$x_{k+1}^{+} = x_{k+1}^{-} + K_{k+1}\left(y_{k+1} - x_{k+1}^{-}\right)$$
$$P_{k+1}^{+} = D^2 K_{k+1} = P_{k+1}^{-} - K_{k+1} P_{k+1}^{-}$$

Implementing the Strategy

To implement the strategy, it needs to be determined when to initiate the trade and when to unwind the trade.

Long when $y_k > \bar{x_k} + \Delta_L$

Short when $y_k < \bar{x_k} + \Delta_S$

Δ is some constant that reflects a threshold that must be past. This threshold should take into consideration such things as cost of trading, any risk premiums, etc.

Unwind the trade when y_k crosses $\bar{x_k}$ or when y_k crosses the long - term mean.