Dynamical Systems: Kalman Filtering

We can extend the linear regression models by allowing the beta coefficients to evolve slowly with time. Development of this idea led to the concept of a *Kalman filter*. Kalman filtering is often used in statistical trading (evolving beta) and volatility estimations (evolving volatility regimes). In a Kalman filter, the beta coefficient varies continuously over a certain range and is estimated recursively. If we discretize this range to a finite set of values, then we can derive a *Hidden Markov Model* (HMM) which will be discussed in the section on Classification methods.

The Kalman Filter (Kalman, 1960) is a technique that combines a series of observations, in the presence of uncertainty, in order to estimate and forecast parameters of an evolving system. The algorithm usually works in two-steps. In the first step, one comes with an estimate of the current state and an estimated error. The next observation (and its error) is then incorporated to obtain a new forecast (by properly weighting the previous estimate and error, and new observation and error).

The dynamic system is described by a State Space model (or, Dynamic Linear Models, DLMs), where there are 2 components:

1) State evolution:

The unobserved variables of interest are labelled as the state of the system, and we know how the state evolves over time via a linear expression with Gaussian noise:

$$x_t = Fx_{t-1} + w_t, \qquad w_t \sim N(0, Q)$$

This is the first piece of information: Given the previous state x_(t-1), we have some knowledge of the current state x_t, but there are uncertainties due to external randomn influence.

2) Measurement:

Although we cannot directly observe the state x_t, we can make some measurements z_t that are related to the state x_t, but again, our measurements come with Gaussian noise:

$$z_t = Hx_t + v_t, \qquad v_t \sim N(0, R)$$

This is the second piece of information: We have the measurements that can infer the state, but there are uncertainties. The Kalman Filter combines the above two pieces of information and gives the optimal state variable as a Gaussian distribution[37] $N(x_{t|t}, P_{t|t})$, where the mean and covariance are

$$x_{t|t} = x_{t|t-1} + K_t\left(z_t - Hx_{t|t-1}\right)$$
$$P_{t|t} = P_{t|t-1} - KHP_{t|t-1}$$

Here K is the Kalman gain and is given by $K = P_{t|t-1}H^T\left(HP_{t|t-1}H^T + R\right)^{-1}$ Despite the complicated expressions, the derivation of the formulae relies only on conditional Gaussian distributions:

$$\begin{pmatrix} X \\ Z \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_X \\ \mu_Z \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$$

$$X|(Z = z) \sim N(\hat{\mu}, \hat{\Sigma})$$

[37] This utilizes the property in which the product of two Gaussian distributions is also a Gaussian distribution.

Marko Kolanovic, PhD
(1-212) 272-1438
marko.kolanovic@jpmorgan.com

Where

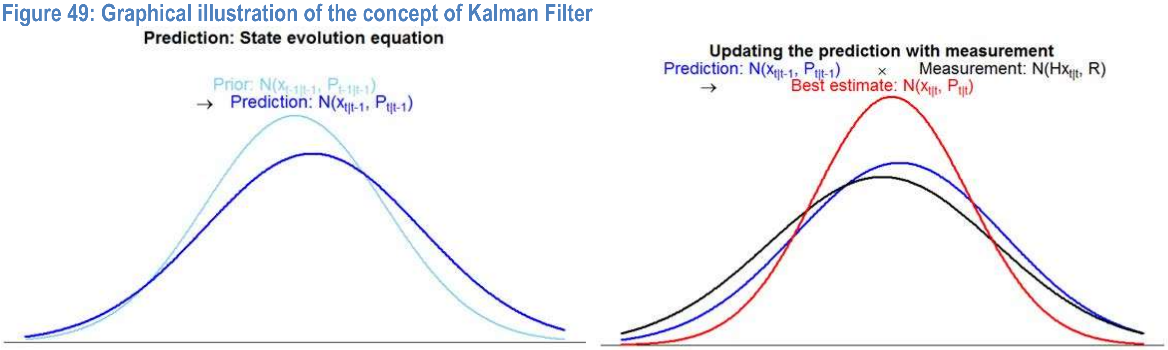
$$\hat{\mu} = \mu_X + \sum_{12}\sum_{22}^{-1}(z - \mu_Z)$$

$$\hat{\Sigma} = \sum_{11} - \sum_{12}\sum_{22}^{-1}\sum_{21}$$

Figure 49 gives a graphical illustration of the workings of a Kalman Filter, which in essence is a Bayesian prediction conditioned on measurements.

Readers who want to have a step-by-step derivations of the formulae can refer to the tutorial by Faragher (2012), and "How a Kalman filter works, in pictures". The most famous application of the Kalman Filter was in spacecraft navigation system in the Apollo mission. Nowadays non-linear versions of the Kalman Filter are also applied in self-driving cars. In finance, Kalman Filters can be used to estimate trends and de-noise signals, to estimate unobserved economic activities, or to describe the dynamic relationships between assets and the market[38] (Petris et al (2009)).

Figure 49: Graphical illustration of the concept of Kalman Filter



Source: J.P.Morgan Macro QDS. Procedure: in the first step, we predict the state based on the evolution, and because of noise the variance increases. In the second step, we combine the predicted distribution and the measurement to update the state distribution, which is more precise than any single piece of information

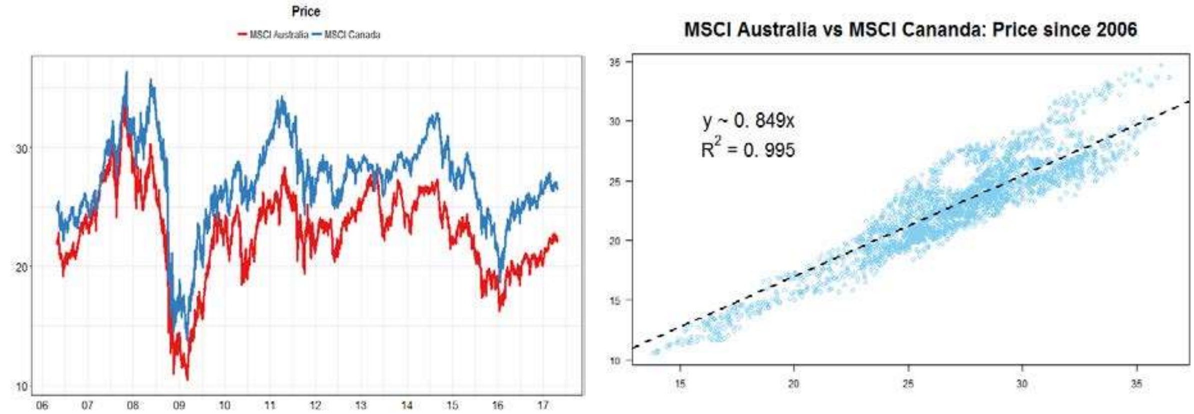
Financial Example:

To illustrate how one can apply Kalman Filter in finance, let us consider an example on pair trading. In general, one has to determine a prominent pair of assets before applying any pair trading strategies. Let us consider some widely known potential pairs of ETFs[39]. We choose the iShares MSCI Australia ETF and Canada ETF as a pair, given strong correlation between the two.

[38] There is an example on the dynamic CAPM model in Petris et al (2009).
[39] Please refer to http://etfdb.com/etf-trading-strategies/how-to-use-a-pairs-trading-strategy-with-etfs/

Marko Kolanovic, PhD
(1-212) 272-1438
marko.kolanovic@jpmorgan.com

Figure 50: MSCI Australia and MSCI Canada indices used in Kalman Filtering example



Source: J.P.Morgan Macro QDS

As such, let us determine the beta that relates the two asset prices S_t, in other words, the co-integration factor:

$$S_{t,AU} = \beta_t S_{t,CA} + v_t, \qquad v_t \sim N(0, \sigma_v^2)$$

We may further assume that beta is time-dependent instead of being a constant. This has the advantage of capturing the dynamics of the system, and to make the residuals v_t to be stationary. For simplicity, let us assume that the evolution of beta simply follows a random walk:

$$\beta_t = \beta_{t-1} + w_t, \qquad w_t \sim N(0, \sigma_w^2)$$

This is a dynamic linear regression, and in the State Space notations, β_t is the state variable. In State Space notations, our system is:

State equation:

$$\beta_t = F_t \beta_{t-1} + w_t, \qquad w_t \sim N(0, Q_t)$$

where $F_t = 1$, $Q_t = \sigma_w^2$

Measurement equation:

$$z_t = H_t x_t + v_t, \qquad v_t \sim N(0, R_t)$$

where $z_t = S_{t,AU}$, $H_t = S_{t,CA}$ and $R_t = \sigma_v^2$.

Hence, we can use the Kalman Filter to estimate the optimal values of beta when we update our observations (in this case, the prices of the ETFs).With some algebra, the Kalman gain in our univariate example is given by

$$K_t = \frac{S_{t,CA}}{S_{t,CA}^2 + \gamma^{-1}}$$

where

$$\gamma = \frac{\hat{P}_{t|t-1}}{\sigma_v^2}$$

is the Signal-to-Noise Ratio (SNR): The ratio of the state variance to the measurement error. If SNR is small, the measurement is noisy and not informative, hence a larger weight is put on the prior knowledge $\hat{\beta}_{t|t-1}$. If SNR is large, more weight is put on the observation. Substituting the variables into $\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t\left(z_t - H_t \hat{x}_{t|t-1}\right)$, we have

$$\hat{\beta}_{t|t} = \hat{\beta}_{t|t-1} + \frac{S_{t,CA}}{S_{t,CA}^2 + \gamma^{-1}}\left(S_{t,AU} - \hat{\beta}_{t|t-1} S_{t,CA}\right)$$

$$= \hat{\beta}_{t|t-1}\left(1 - \frac{S_{t,CA}^2}{S_{t,CA}^2 + \gamma^{-1}}\right) + \frac{S_{t,CA}S_{t,AU}}{S_{t,CA}^2 + \gamma^{-1}}$$

As a result, if SNR γ is very large, we basically just use the latest observation to estimate beta, without using any prior:

$$\hat{\beta}_{t|t} \approx \frac{S_{t,AU}}{S_{t,CA}}$$

If SNR γ is very small, we basically ignore the observation (because it is noisy), and we just use the information from the prior:

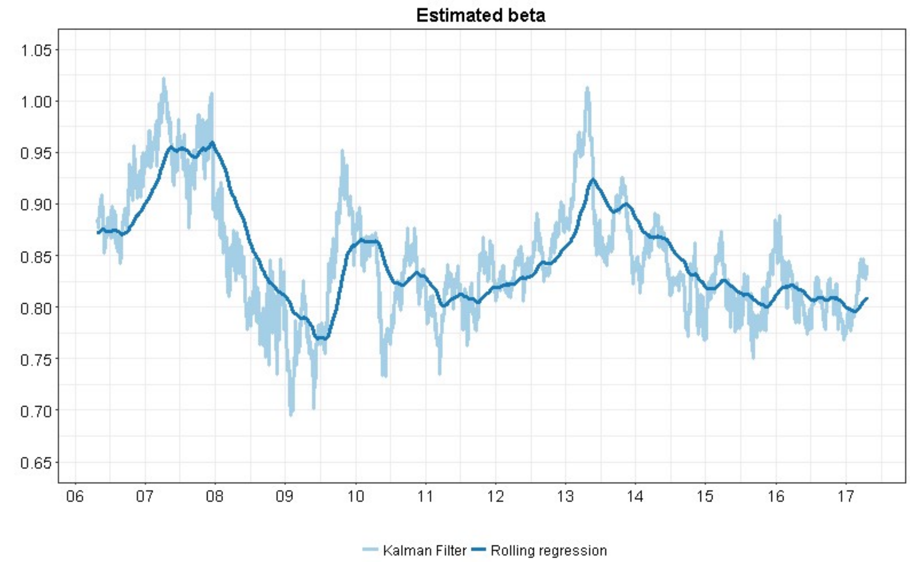
$$\hat{\beta}_{t|t} \approx \hat{\beta}_{t|t-1}$$

Instead of using a Kalman Filter, of course we can also estimate β using standard linear regression, without imposing any dynamics on it:

$$S_{t,AU} = \beta S_{t,CA} + v_t, \qquad v_t \sim N(0, \sigma_v^2)$$

We compare the estimates of beta using the Kalman Filter, and the one using rolling linear regressions with a lookback window of 6 months. Note that Kalman Filter is more reactive. In fact, the Kalman Filter is closely related to exponential smoothing where it puts more weights on recent observations, and it can adjust the weights depending on how 'noisy' are the measurements[40].

Figure 51: Estimate of beta using Kalman Filter, and rolling linear regression – note that Kalman filter Is more reactive to price movements



Source: J.P.Morgan Macro QDS

Our pairs trading signal simply depends on the residuals v_t, which is supposed to fluctuate around mean zero. At the end of each trading day, we use the closing price of the ETFs to update our estimate of the beta β_t, and then calculate the residuals:

$$v_t = S_{t,AU} - \beta_t S_{t,CA}$$

[40] Meinhold and Singpurwalla, Understanding the Kalman Filter. Available at http://www-stat.wharton.upenn.edu/~steele/Resources/FTSResources/StateSpaceModels/KFExposition/MeinSing83.pdf

We also record the uncertainty of the residuals σ_t, so that we can use it to determine whether the magnitude of the residual is large enough to trigger our strategy (otherwise no position is taken):

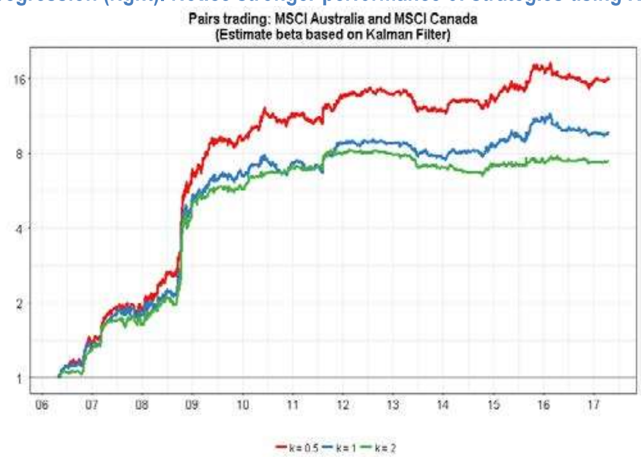
If $v_t \geq k\sigma_t$, we long β_t unit of MSCI Canada and short one unit of MSCI Australia

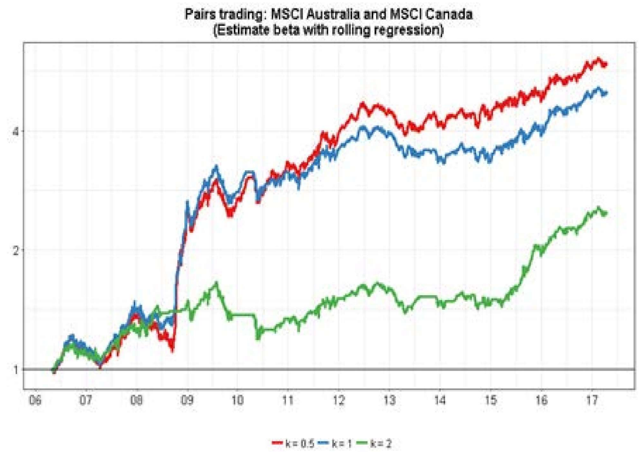
If $v_t \leq -k\sigma_t$, we short β_t unit of MSCI Canada and long one unit of MSCI Australia

As an example, we take $k = 0.5, 1, 2$. This scaling parameter can actually be tuned in a more optimal way by modelling the mean-reversion process, but we will not investigate further here.

The performance of the strategy since 2006 is given in Figure 52. Next to it, we show the performance of the same strategy when using a simple rolling regression beta

Figure 52: MSCI Australia vs. MSCI Canada pair trading strategy. Beta is estimated from the Kalman Filter (left), and from rolling 6 months regression (right). Notice stronger performance of strategies using Kalman Filter





Source: J.P.Morgan Macro QDS, Bloomberg

To compare the 2 strategies, we choose $k = 0.5$. We see that using a Kalman Filter that has a faster response helped deliver higher returns during time of market dislocations in 2008/2009.

Figure 53: Comparing the pairs trading strategy - Kalman Filter vs. rolling regressions

	CAGR (%)	Vol (%)	Sharpe	Max DD (%)
Since 2006				
Kalman Filter	28.6	18.1	1.58	21.1
Regression	17.4	17.9	0.97	22.5
Since 2009				
Kalman Filter	10.9	14.5	0.75	21.1
Regression	10.4	15.3	0.68	19.9

Source: J.P. Morgan Macro QDS, Bloomberg

Kalman filter gives good results (closed form solution) when applied to linear systems with Gaussian noise. For non linear system or systems with non-Gaussian noise, a numerical technique called Particle Filtering is used. For technical details and an example of a Particle Filtering application, see the Appendix.