Construct a stationary portfolio by augmented Dickey-Fuller(ADF) test and Kalman filter(KF)

Introduction

Kálmán filter has been widely used in a lot of areas, especially, in engineering and finance. It is named by Rudolf E. Kálmán. The original idea of Kálmán filter is extremely simple. This situation doesn't reduce its power; furthermore, Kálmán filter becomes popular in numerous places because of its naturity.

E. Chan mentioned that Kálmán filter is a good tool in algorithmic trading in an interview which can be found on [3]. However, most people neglect its potential. The main content in this project is to find an example in finance, more precisely, in trading, which Kálmán filter can be applied. We organize the article in the following way. Most of the material comes from [1] and [2]. First, we introduce augmented Dickey-Fuller test. The model for this test fits a concept in algorithmic trading, which is called mean reversion strategy. Second, we would explain cointegration. Roughly speaking, it is trying to combine several assets into a portfolio which satisfies stationarity. However, the portion of each asset is unknown. Therefore, our main topic of this project is to demonstrate how Kálmán filter can play an important role on deciding and updating the portion by using the data of two stock prices from Taiwanese banks in 368 days. Conclusion and discussion will be provided in the end of the article.

Background

Augmented Dickey-Fuller Test

Suppose Y_t is a time series. Since Y_t stands for the price of an asset at time t, we also call it as price series. The mathematical description for mean-reversion price series is that the change of price Y_t at next period which is proportional to the difference of mean price $\bar{Y}_{t-1}$(it can be weighted mean) and current price Y_{t-1}. To be clear, if the price Y_{t-1} at time $t-1$ is higher than $\bar{Y}_{t-1}$, then the difference of price $Y_t - Y_{t-1}$ at time t has larger possibility to be negative. (If this situation must happen, then the price series become predictable.) Therefore, the best thing for this idea is that it can be interpreted as an equation.

$$Y_t - Y_{t-1} = \gamma_t Y_{t-1} + [\beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_r Y_{t-r}] \tag{1}$$

where $\beta_1, \cdots, \beta_r$ are all given nonnegative numbers and $\sum_{i=1}^{r} \beta_i = 1$. The last part can be viewed as the mean of the price at time $\bar{Y}_{t-1}$ and γ is the proportion of price difference and mean price. For the situation we consider, we hope γ is negative. It is worth to mention the character of γ here since it plays an important role for the time series of the our model.

Assume a time series is of the form

$$Y_t = \gamma_1 Y_{t-1} + \gamma_2 Y_{t-2} + \cdots + \gamma_r Y_{t-r} + \epsilon_t. \tag{2}$$

The characteristic polynomial of (2) is

$$x^r - \gamma_1 x^{r-1} - \gamma_2 x^{r-2} - \cdots - \gamma_r = 0 \tag{3}$$

and we consider the roots of (3). A well-known result is that if all the roots are within unit circle, then the time series is stationary. Roughly speaking, stationarity means the variation of the time series is not too wide; hence, some control can be done for the time series. For explicit definition, [2] is a good reference. Moreover, if 1 is a root of (3), then the time series (2) is non-stationary. One thing needs to be mention here, the term "stationary"" in [1] means "trend-stationary"; approximately, the variance increases slower than normal diffusion.

Based on above ideas, extremely luckily, we can rewrite it and create a linear model as following,

$$\Delta Y_t = \gamma Y_{t-1} + [\beta_1 \Delta Y_{t-1} + \beta_2 \Delta Y_{t-2} + \cdots + \beta_r \Delta Y_{t-r}] + \epsilon_t \tag{4}$$

where ϵ_t's are i.i.d white noises and $\Delta Y_j := Y_j - Y_{j-1}$. β_j's which compare to the ones in (1) are different. We abuse the notation here. A more general case can be considered,

$$\Delta Y_t = \gamma Y_{t-1} + \mu + \alpha t + [\beta_1 \Delta Y_{t-1} + \beta_2 \Delta Y_{t-2} + \cdots + \beta_r \Delta Y_{t-r}] + \epsilon_t \tag{5}$$

where μ and α stand for the drift and diffusion of the price series. (4) and (5) are for stationarity and trend-stationarity, respectively.

Now, lets move to ADF(short for Augmented Dickey-Fuller) test. Dickey-Fuller test is a test for whether there is a unit root for an autoregressive model. ADF test is for the same issue, but has something more. For our model, either (4) or (5), 1 will be a root for the characteristic polynomial if and only if $\gamma = 0$. The null and alternative hypotheses for ADF test are $\gamma = 0$ and $\gamma < 0$, respective. Surprisingly, $\gamma < 0$ is the desired result. That's the reason why ADF test is mentioned here.

Cointegration

In reality, to find a stationary asset is hard. Stationarity plays an important role for mean-reversion strategy. It seems this kind of methods faces difficulty on putting it into practice. However, it is not the end of whole story. Smart people figured out another way to make it. Most of time, we care about a portfolio instead of just a specific asset. Hence, we can combine two or several assets into one portfolio such that stationarity appears for this portfolio. This method is called **cointegration**.

Now, there are infinitely many choices of the combination of assets depending on the weights. Negative and positive weights represent sell and buy the assets, respectively. Sometimes, assets are hard to sell due to plenty of issues, such as liquidity and constraint on the laws. However, in this article, all assets we consider can be traded on both directions freely. To simplify, only two asset M_t and N_t will be consider, the choices of portfolio are

$$\{Y_t = M_t + \delta_t^1 N_t + \delta_t^2 : \delta_t^1, \delta_t^2 \in \mathbb{R}\}.$$

Choices of δ_t^1 and δ_t^2 are another issue on cointegration and it is changing with respect to t. We will use Kálmán filter to determine these coefficients in application.

Kálmán Filter

Suppose we'd like to develop an auto-driving system. To know the exact position X_t is hard since the position Y_t from GPS is not as same with X_t. Y_t is called **observable variable** while X_t is called **latent variable** or **hidden variable**. Scientists try to have estimate on X_t by using two equations, first the **system equation** or **hidden system**

$$X_t = A_t X_{t-1} + B_t u_{t-1} + \epsilon_t^1 \tag{6}$$

and another equation, **observation equation**

$$Y_t = C_t X_t + \epsilon_t^2. \tag{7}$$

ϵ_t^1 and ϵ_t^2 are independent white noises with means m_i's and variances Σ_i. For $i = 1$ or 2, ϵ_t^i are i.i.d.. u_t is the predictor of Y_t. A_t, B_t and C_t are known matrices which do or don't depend on time t. The idea for the model is clear and clever. We do not know the desired quantities immediately, and using the hidden layer to display the interactions of Y_t and X_t. Sometimes, A_t, B_t and C_t are determined from truth or differential equations. Hence, these crucial matrices are hard to know for lots of problems. Chapter 11 in [2] decribes a way to estimate them; however, for this case, the method doesn't involve Bayesian in this chapter. There should be at least some previous work which consider this problem with Bayesian method. Although, it is still not clear yet if someone has worked on this before.

Now, we describe the one with Bayesian in the following for later use. The content is mainly from [2]. Let $\bar{\mathcal{Y}}_t = \{\mathcal{Y}_1, \cdots, \mathcal{Y}_t\}$ be the data by time t. Let $X_{t|t-1}$ and $X_{t|t}$ be the estimates for X_t conditioned on the data $\bar{\mathcal{Y}}_{t-1}$ and $\bar{\mathcal{Y}}_t$, respectively. Hence, $X_{t|t-1}$ is the prior of X_t by using (6) and $X_{t|t}$ is the posterior of X_t updating by $\bar{\mathcal{Y}}_t$.

Assume that

$$X_{t-1|t-1} \sim N(\hat{X}_{t-1|t-1}, \Sigma^{xx}_{t-1|t-1}).$$

Using (6), the prior distribution

$$X_{t|t-1} \sim N(A_t\hat{X}_{t-1|t-1} + B_t u_{t-1}, A_t\Sigma^{xx}_{t-1|t-1}A_t^T + \Sigma_1).$$

We choose the prediction $\hat{X}_{t|t-1}$ to be the mean of $X_{t|t-1}$ which is $A_t\hat{X}_{t-1|t-1} + B_t u_{t-1}$. By (7), $Y_{t|t-1}$ has the distribution

$$N\left(C_t(A_t\hat{X}_{t-1|t-1} + B_t u_{t-1}), C_t(A\Sigma^{xx}_{t-1|t-1}A^T + \Sigma_1)C_t^T + \Sigma_2\right).$$

Therefore, the prediction of $\hat{Y}_{t|t-1} = C_t(A_t\hat{X}_{t-1|t-1} + B_t u_{t-1})$. Moreover,

$$P[X_t|Y_t, \mathcal{Y}_{t-1}] \propto P[\tilde{Y}_{t|t-1}|X_t, \mathcal{Y}_{t-1}]P[X_t|\mathcal{Y}_t]$$

where $\tilde{Y}_t := Y_t - \hat{Y}_t$. Via some arguments, we have the results which show as formulas (10.100) and (10.101) in [2]. Here, we display the final result of $X_{t|t}$. $X_{t|t}$ has mean

$$A_t\hat{X}_{t-1|t-1} + B_t u_{t-1} + R_t C_t^T(\Sigma_2 + C_t R_t C_t^T)^{-1}\tilde{Y}_{t|t-1}$$

and covariance

$$\Sigma^{xx}_{t|t} = R_t - R_t C_t^T(\Sigma_2 + C_t R_t C_t^T)^{-1}C_t R_t$$

where $R_t = A_t\Sigma^{xx}_{t-1|t-1}A_t^T + \Sigma_1$. Let's restate above result as a theorem which is same as the one in [2].

{Theorem}

(The Kálmán Filter) The optimal linear reconstruction $\hat{X}_{t|t}$ and the prediction of the system (4) and (5) is obtained in terms of a reconstruction(updating)

$$\hat{X}_{t|t} = \hat{X}_{t|t-1} + K_t(Y_t - C_t\hat{X}_{t|t-1}), \tag{8}$$

$$\Sigma^{xx}_{t|t} = \Sigma^{xx}_{t|t-1} - K_t\Sigma^{yy}_{t|t-1}K_t^T \tag{9}$$

where the Kálmán gain is

$$K_t = \Sigma^{xx}_{t|t-1}C_t^T(\Sigma^{yy}_{t|t-1})^{-1},$$

and the prediction is

$$\hat{X}_{t+1|t} = A_{t+1}\hat{X}_{t|t} + B_{t+1}u_t, \tag{10}$$

$$\Sigma^{xx}_{t+1|t} = A_{t+1}\Sigma^{xx}_{t|t}A_{t+1}^T \tag{11}$$

$$\Sigma^{yy}_{t+1|t} = C_{t+1}\Sigma^{xx}_{t+1|t}C_{t+1}^T + \Sigma_2. \tag{12}$$

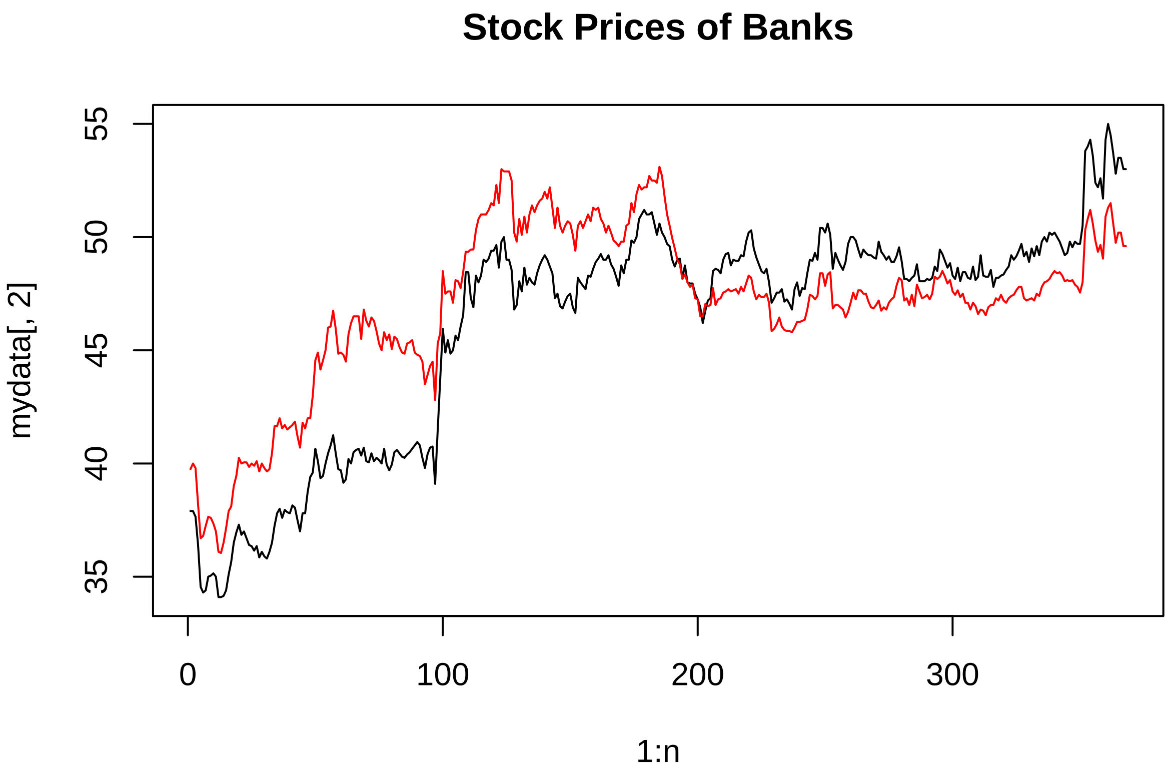
with initial condition,

$$\hat{X}_{1|0} = E[X_1] = \mu_0, \tag{13}$$

$$\Sigma^{xx}_{1|0} = Var[X] = V_0. \tag{14}$$

Application

For the application, X, Y are two stock prices of Taiwanese banks in 368 days. First, we plot the prices of these banks. It shows that they are correlated and is possible to be combined as a stationary portfolio.



Assume Y_t can be expressed as $\alpha_t^1 + \alpha_t^2 X_t$. Let $\alpha_t = (\alpha_t^1, \alpha_t^2)$ be the unknown quantity. $Z_t = (1, X_t)$ and Y_t are observable variables. We set the hidden equation as

$$\alpha_t = \alpha_{t-1} + \epsilon_t^1$$

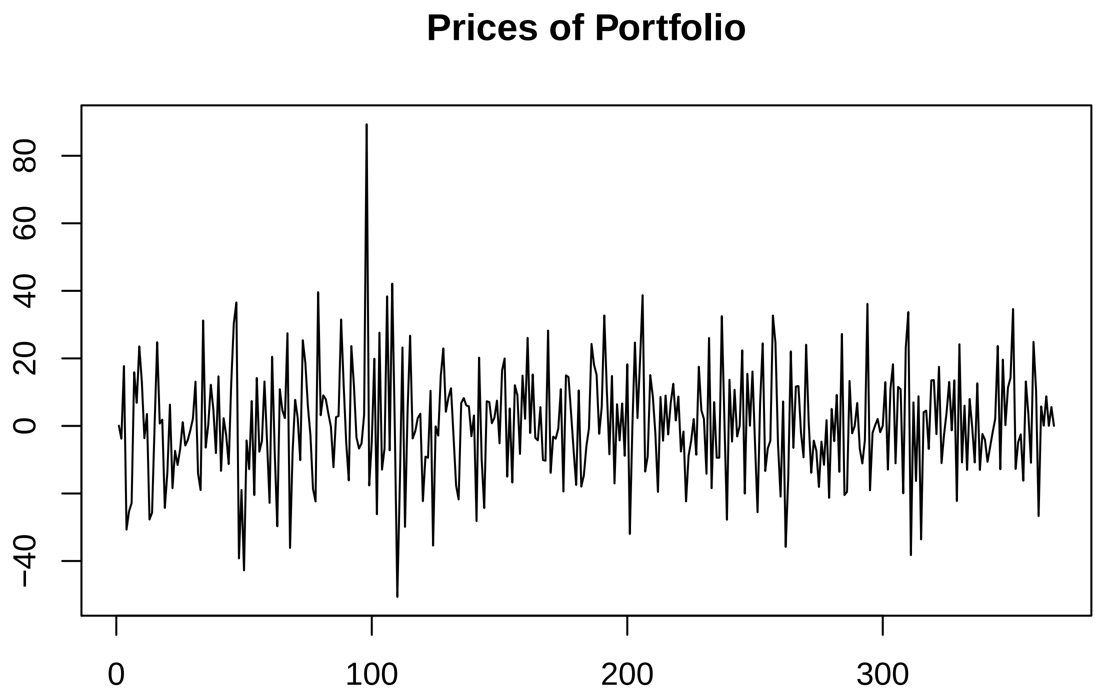
where $\epsilon_t^1 \sim N(0, \Sigma_1)$ and the observation equation is

$$Y_t = \alpha_t Z_t + \epsilon_t^2$$

where $\epsilon_t^2 \sim N(0, \sigma^2)$. In [1], the auther mentions that we can set up the initial conditions

$$\beta_{1|0} = 0,\ \Sigma_{0|0}^{\beta\beta} = 0,\ \Sigma_1 = \frac{\delta}{1-\delta} I_2,\ \delta = 0.0001,\ \Sigma_2 = 0.001.$$

Apply Kálmán filter we have the graph for the price portfolio at each time t. We multiply it by 10000 to make the plot easier to read.



It shows that after time 100, it reaches stationarity. Let's also test it by ADF. The library "tseries" is used here. Finally, the following result shows that the portifolio created by Kálmán filter is stationary.

```
##
##  Augmented Dickey-Fuller Test
##
## data:  new.portfolio
## Dickey-Fuller = -8.7845, Lag order = 7, p-value = 0.01
## alternative hypothesis: stationary
```

Discussion

In [1], the auther metions that all strategy in trading can be roughly separated into two main categories, mean reversion strategies and momentum strategies. The key issue on using mean reversion strategies is to find a stationary portfolio. Usually, an asset doesn't hold this property. However, by combining two or more assets can help us to create plenty of stationary portfolios. Moreover, Kálmán filter provides a way to merge non-stationary assets.

It is still unclear that how to use the nonstationary portifolio to make a profit. The time of holding positions is usually a tough question in trading and it is not clearly stated in [1].

It is also vague that why Kálmán filter is used here instead of other methods. We don't know which methods should be compared with Kálmán filter. It may be interested to have a Bayesian version of Kálmán filter with varied coefficients. However, it may be hard on simulation since the posterior distributions are not well-known at each time step and it becomes worse when t is large.

References

[1] E. Chan *Algorithmic Trading: Winning Strategy and their Rationale*. Wiley, 2013.

[2] H. Madsen *Time Series Analysis*. Chapman & Hall.

[3] Interview with E. Chan, *https://www.youtube.com/watch?v=DW9zQ8Hz6gQ*

[4] Augmented Dickey-Fuller Test in Wikipedia,
https://en.wikipedia.org/wiki/Augmented_Dickey%E2%80%93Fuller_test