Math 411 HW1
Section IV 2025
Duke Kunshan University

Problem 1 (Analytical). *Calculate the expectation and variance of a normal random variable.*

Problem 2 (Analytical). *Consider a simple class of model of the form*

$$y_t = r + y_{t-1} + c_0\epsilon_t + c_{-1}\epsilon_{t-1} + \cdots + c_{-p}\epsilon_{t-p}, \tag{1}$$

where $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$. We want to use this model to mimic the stock market. Here, y_t represent the log-price, r represents the expected return, and ϵ_t represents the impact from the news and are independent from each other. We define the residual as $z_t = y_t - r - y_{t-1}$.

1. *Calculate the expectation of log return $\mathbb{E}[y_t - y_{t-1}]$.*
2. *Suppose $y_t = r + y_{t-1} + \epsilon_t$ and $y_0 = 0$. Suppose we are interested in the event when $y_1 > \eta$ for a threshold η to represent the scenario that the company has received good news. Calculate the conditional expectation of $\mathbb{E}[z_2|y_1 > \eta]$.*
3. *Suppose now $y_t = r + y_{t-1} + \epsilon_t + c_{-1}\epsilon_{t-1}$, $y_0 = 0$, and $\epsilon_0 = 0$. Calculate the conditional expectation of $\mathbb{E}[z_2|y_1 > \eta]$ (You can use the $\phi(x)$ to denote the cumulative distribution function of a standard normal distribution in the answer).*
4. *(Optional) Suppose now $y_t = r + y_{t-1} + \epsilon_t + c_{-1}\epsilon_{t-1}$, $y_0 = 0$, and $\epsilon_0 = 0$. Calculate the conditional expectation of $\mathbb{E}[z_3|y_2 > \eta]$ (You can use the $\phi(x)$ to denote the cumulative distribution function of a standard normal distribution in the answer).*

Solution 1.

1. $\mathbb{E}[y_t - y_{t-1}] = \mathbb{E}[r + c_0\epsilon_t + c_{-1}\epsilon_{t-1} + \dots] = r$
2. $\mathbb{E}[z_2|y_1 > \eta] = \mathbb{E}[\epsilon_2|y_1 > \eta] = 0$ *due to independency*
3. *Let $\mathcal{Z}$ be the standard normal, then $\epsilon_1 = \sigma\mathcal{Z}$.*

$$\mathbb{E}[z_2|y_1 > \eta] = \mathbb{E}[\epsilon_2 + c_{-1}\epsilon_1|r + \epsilon_1 > \eta] = \mathbb{E}[\epsilon_2|r + \epsilon_1 > \eta] + c_{-1}\mathbb{E}[\epsilon_1|r + \epsilon_1 > \eta] = c_{-1}\sigma\mathbb{E}\left[\mathcal{Z}\middle|\mathcal{Z} > \frac{\eta - r}{\sigma}\right]$$

$$= c_{-1}\sigma\frac{\int_{\frac{\eta-r}{\sigma}}^{\infty} \frac{x}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\, dx}{\int_{\frac{\eta-r}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\, dx} = c_{-1}\sigma\frac{-\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\Big|_{\frac{\eta-r}{\sigma}}^{\infty}}{1 - \phi\left(\frac{\eta-r}{\sigma}\right)} = \frac{\frac{c_{-1}\sigma}{\sqrt{2\pi}}e^{-\frac{(\eta-r)^2}{2\sigma^2}}}{1 - \phi\left(\frac{\eta-r}{\sigma}\right)}.$$

4. *First, similar to the last problem, we make the expression clear*

$$\begin{aligned}\mathbb{E}[z_3|y_2 > \eta] &= \mathbb{E}[\epsilon_3 + c_{-1}\epsilon_2|2r + \epsilon_1 + \epsilon_2 + c_{-1}\epsilon_1 > \eta] = c_{-1}\mathbb{E}[\epsilon_2|(1 + c_{-1})\epsilon_1 + \epsilon_2 > \eta - 2r] \\ &= c_{-1}\sigma\mathbb{E}\left[\mathcal{Z}_2\middle|(1 + c_{-1})\mathcal{Z}_1 + \mathcal{Z}_2 > \frac{\eta - 2r}{\sigma}\right]\end{aligned}$$

Recall that linear combination of normal is still normal. So we perform the Gram-Schmidt,

$$\begin{aligned}&\widetilde{\mathcal{Z}}_2 = \mathcal{Z}_2 - \frac{Cov((1 + c_{-1})\mathcal{Z}_1 + \mathcal{Z}_2, \mathcal{Z}_2)}{Var((1 + c_{-1})\mathcal{Z}_1 + \mathcal{Z}_2)}((1 + c_{-1})\mathcal{Z}_1 + \mathcal{Z}_2) = \mathcal{Z}_2 - \frac{(1 + c_{-1})\mathcal{Z}_1 + \mathcal{Z}_2}{1 + (1 + c_{-1})^2} \\ &\Rightarrow \mathcal{Z}_2 = \widetilde{\mathcal{Z}}_2 + \frac{(1 + c_{-1})\mathcal{Z}_1 + \mathcal{Z}_2}{1 + (1 + c_{-1})^2},\ \widetilde{\mathcal{Z}}_2 \perp \frac{(1 + c_{-1})\mathcal{Z}_1 + \mathcal{Z}_2}{1 + (1 + c_{-1})^2}\end{aligned}$$

Note

$$\mathbb{E}[\mathcal{Z}_1] = \mathbb{E}[\mathcal{Z}_2] = 0 \Rightarrow \mathbb{E}[\widetilde{\mathcal{Z}}_2] = 0.$$

Therefore, we have

$$
\begin{aligned}
c_{-1}\sigma\mathbb{E}\left[\mathcal{Z}_2\middle|(1+c_{-1})\mathcal{Z}_1+\mathcal{Z}_2>\frac{\eta-2r}{\sigma}\right] &= c_{-1}\sigma\mathbb{E}\left[\widetilde{\mathcal{Z}}_2+\frac{(1+c_{-1})\mathcal{Z}_1+\mathcal{Z}_2}{1+(1+c_{-1})^2}\middle|(1+c_{-1})\mathcal{Z}_1+\mathcal{Z}_2>\frac{\eta-2r}{\sigma}\right] \\
&= c_{-1}\sigma\mathbb{E}\left[\frac{(1+c_{-1})\mathcal{Z}_1+\mathcal{Z}_2}{1+(1+c_{-1})^2}\middle|(1+c_{-1})\mathcal{Z}_1+\mathcal{Z}_2>\frac{\eta-2r}{\sigma}\right] \\
&= \frac{c_{-1}\sigma}{\sqrt{1+(1+c_{-1})^2}}\mathbb{E}\left[\mathcal{Z}\middle|Z>\frac{\eta-2r}{\sigma\sqrt{1+(1+c_{-1})^2}}\right] \\
&= \frac{\frac{c_{-1}\sigma}{\sqrt{2\pi}}}{\sqrt{1+(1+c_{-1})^2}}\frac{e^{-\frac{(\eta-2r)^2}{2\sigma^2(1+(1+c_{-1})^2)}}}{1-\phi\left(\frac{\eta-2r}{\sigma\sqrt{1+(1+c_{-1})^2}}\right)}.
\end{aligned}
$$

Problem 3 (Coding)**.** *Now consider we have a collection of log prices for different companies*

$$y_{i,t}=r_i+y_{i,t-1}+c_0\epsilon_{i,t}+c_{-1}\epsilon_{i,t}+\cdots+c_{-p}\epsilon_{i,t-p},\ i=1,\ldots,M, \tag{2}$$

whereas $y_{i,t}$ represent the log-price, $r_{i,t}$ represents the expected return, and $\epsilon_{i,t}\sim\mathcal{N}(0,\sigma_i^2)$ represents the impact from the news for the ith company and are independent from each other. We define the residual as $z_{i,t}=y_{i,t}-r_i-y_{i,t-1}$. For r_i and σ_i, you can randomly select values as long as they are positive. (As an example of demonstration, I used $r_i\sim\frac{0.05}{252}(1+\mathcal{N}(0,1))$ and $\sigma_i\sim\frac{0.05}{252}\cdot|\mathcal{N}(0,1)|$). Use the Monte Carlo simulation to generate a sequence of N (e.g., 10^4) log prices for M (e.g., 10^3) companies. Consider the event that the company keep receives the good news, that is given y_i, we find τ_i s.t. $y_{\tau_i}-y_0>\eta$, whereas η is the threshold (e.g., $\log(2)$). We wish to estimate the impact of the news by calculating $w_k=\frac{\sum_{j|y_{j,\tau_i}>\eta}z_{j,k}}{\sum_{j|y_{j,\tau_i}>\eta}1}$. Then we look at $v_l=\sum_{k=1}^{l}w_k$ to test the Markovian. For convenience, we can plot v_l from $\tau-m$ to $\tau+m$ (e.g., $m=10$).

1. *Use this model together with Monte Carlo simulation to generate a result that indicates the Markovian behavior. .*
2. *Use this model together with Monte Carlo simulation to generate a result that indicates the Non-Markovian behavior.*

Here is an example of what we should observe