4 Exploratory Data Analysis on Returns

Text: pages 41-62.

4.1 What to do

1. Plot your data. Look for
   (a) Outliers? Mistakes?
   (b) Does it trend? How to make it stationary?
   (c) Structual break?
   (d) Volatility clustering?
2. Ask if observations are normal.
3. Graph the kernel density against the implied normal denstity.
4. Relation between sample moments and theoretical moments
5. Jarque-Bera test for normality
   (a) The Jarque-Bera statistic measures the difference between skewness and kurtosis in the data and the normal distribution.
   (b) Let sk_T be sample skewness, and $kurt_T$ be sample kurtosis. Jarque and Bera showed that their statistic JB

$$\mathrm{JB} = \frac{T}{6}\left(sk_T^2 + \frac{(kurt_T - 3)^2}{4}\right) \sim \chi_2^2$$

   under the **null** hypothesis of normality.
   (c) Eviews produces JB test and p-values when asking for descriptive statistics.

5 Autoregressive Moving Average (ARMA) models. Specification and Estimation

Text: pp. 246-276.
Overview. These are simple parametric models of univariate time-series. They are purely statistical models with no economic content, to model simple types of observational dependence over time. We use the dependence to forecast future values.

Motivate with graphs of STicky price CPI from FRED and graph of stock returns. Illustrates different types of dependence over time.

Weak Stationarity. The time series $\{y_t\}_{t=1}^{T}$ is weakly stationary if the mean, variance, and autocovariances of the process are constant.

$$\begin{aligned} E\left(y_t\right) &= \mu_y \\ E\left(y_t - \mu_y\right)^2 &= \sigma^2 \\ E\left(y_t - \mu_y\right)\left(y_{t-k} - \mu_y\right) &= \gamma_k \end{aligned}$$

Important result: Conditional expectation is **minimum mean-square error predictor**. Let I_t be the observable information set. This can include current and past values of y_t and other variables. If we ask, what function minimizes the mean square prediction error,

$$E\left[y_{t+1} - P\left(y_{t+1}|I_t\right)\right]^2$$

answer is

$$E\left(y_{t+1}|I_t\right) = \int y_{t+1} f\left(y_{t+1}|I_t\right) dy_{t+1}$$

where $f\left(y_{t+1}|I_t\right)$ is conditional pdf of y_{t+1}.

- Think of **fitted value** of regression as conditional expectation. **Systematic** part of regression also called **projection**.
- **Notational Convention** $E_t(y_{t+k}) \equiv E(y_{t+k}|I_t)$

5.1 The white noise process

The white noise process is the basic building block of all time series

$$y_t = \epsilon_t, \quad \epsilon_t \overset{iid}{\sim} (0, \sigma_\epsilon^2)$$

1. It reflects the stochastic (random) nature of the world. View these as **exogenous** shocks. These shocks are random and have no dependence over time, representing purely unpredictable events. It's a model of news.

2. Notice we didn't say they are normally distributed. In time-series, it doesn't matter because all inference is asymptotic

3. By itself, it is uninteresting, because there is no dependence over time.

5.2 The Moving Average (MA) Model

The moving averge model strings together a finite (usually small) number of current and past white noise shocks.

An MA(k) process. y_t is correlated with y_{t-k} and possibly those observations in between, $y_{t-1}, ..., y_{t-k+1}$. k is the 'order' of the moving average process.

1. MA(1). Example might be daily returns with slow moving capital. News occurs today. High frequency traders pounce, institutional investors, move later in the day. Retail investors don't know until they see the nightly Bloomberg report.

5.2.1 MA(1) model

1. Let y_t be the stock return.
$$y_t = \mu + \epsilon_t + \theta\epsilon_{t-1}$$
where ϵ_t is white noise with mean 0 and variance σ_ϵ^2. Now see, if we shift time index back one period,
$$y_{t-1} = \mu + \epsilon_{t-1} + \theta\epsilon_{t-2}$$
Since ϵ_{t-1} is common to both y_t and y_{t-1}, they are correlated.

2. Calculate the interesting moments (mean, variance, autocovariance, autocorrelation).
$$E\left(y_t\right) = E\left(\mu + \epsilon_t + \theta\epsilon_{t-1}\right) = \mu_y$$
$$Var\left(y_t\right) = E\left(y_t - \mu_y\right)^2 = E\left(\tilde{y}_t\right) = \sigma_y^2 = \left(1 + \theta^2\right)\sigma_\epsilon^2$$
$$Cov\left(y_t, y_{t-1}\right) = \gamma_1 = E\left(\tilde{y}_t, \tilde{y}_{t-1}\right) = E\left(\epsilon_t + \theta\epsilon_{t-1}\right)\left(\epsilon_{t-1} + \theta\epsilon_{t-2}\right) = \theta\sigma_\epsilon^2$$
$$Corr\left(y_t, y_{t-1}\right) = \frac{\gamma_1}{\left(1+\theta^2\right)\sigma_\epsilon^2} = \frac{\theta}{\left(1+\theta^2\right)}$$
and for any $k > 1$,
$$Cov\left(y_t, y_{t-k}\right) = 0.$$

3. **Forecasting** formula. Conditional expectation is projection, is fitted value of model (regression), is the optimal forecast. Forecast formula is the conditional expectation. Hence, find the conditional expectation.

   Let's forecast **deviation from mean** since mean is constant. Let $\tilde{y}_t \equiv y_t - \mu_y$

$$E_t\left(\tilde{y}_{t+1}\right) = E_t\left(\epsilon_{t+1} + \theta\epsilon_t\right) = \theta\epsilon_t$$

$$E_t\left(\tilde{y}_{t+2}\right) = 0$$

   And for any $k \geq 2$

$$E_t(\tilde{y}_{t+k}) = 0$$

4. **Impulse response** function. How does y_t respond to a one-time shock? MA(1) model has **memory** of only one period. Hence there is only a one period response.

   Set all $\epsilon_{t-s} = 0$ for $s \neq 0$, $\epsilon_t = \sigma_\epsilon$, which we'll assume is $\sigma_\epsilon = 1$.

$$\begin{aligned}\tilde{y}_{t-1} &= 0\\ \tilde{y}_t &= 1\\ \tilde{y}_{t+1} &= \theta_1\\ \tilde{y}_{t+2} &= 0\end{aligned}$$

5. Wait! How do you estimate an MA(1) model? There are no independent variables, so you can't run least squares regression. We do something called **maximum likelihood estimation**.

   - The ϵ_t are random variables. Let's assume they are drawn from a normal distribution, $N\left(0, \sigma_\epsilon^2\right)$. The marginal probability density function (pdf) for ϵ_t is

$$f_1\left(\epsilon_t\right) = \frac{1}{\sigma_\epsilon\sqrt{2\pi}} e^{-\frac{\epsilon_t^2}{2\sigma_\epsilon^2}}$$

     The joint pdf for $\epsilon_1, \epsilon_2, ..., \epsilon_t, \epsilon_{t+1}, ..., \epsilon_T$ is the product of the $f_1\left(\right)$, because the $\epsilon's$ are independent.

$$f\left(\epsilon_T, \epsilon_{T-1}, ..., \epsilon_1\right) = f_1\left(\epsilon_1\right) f_1\left(\epsilon_2\right) \cdots f_1\left(\epsilon_T\right) = \left(\frac{2}{\sigma_\epsilon\sqrt{2\pi}}\right)^T e^{-\frac{1}{2\sigma_\epsilon^2}\sum_{t=1}^T \epsilon_t^2}$$

   - Notice that $\epsilon_t = y_t - \mu - \theta\epsilon_{t-1}$, $\epsilon_{t-1} = y_{t-1} - \mu - \theta\epsilon_{t-2}, \epsilon_{t-2} = y_{t-2} - \mu - \theta\epsilon_{t-3}, ...$ This