Empirical Methods in Finance Spring 2008

- Professor Rossen Valkanov, Rady School of Mgmt, UCSD.
  Email: rvalkanov@ucsd.edu
  Tel: (858) 534-0898, Cell: (310) 351-5107
- Class:
  - Tuesdays (11:00-1:00 and 2:00-4:00)
  - Wednesdays (8:00-10:00)
- Attendance is mandatory.
- Homework each week.
- At the end of week 4, project proposal is due.
- At the end of course, project is due.

- TA:

Class Policy

- There is no midterm
- But there is:
  - Homework
  - Final Project
  - Final Exam
- Work in groups of no more than 4 people.
- Homework, assigned at the end of each class, is due before Tue next week, 12am, in my and the TA's email. Absolutely no delays allowed!
- Project Proposal and Project: From beginning of class, start thinking of possible projects. I have a list of about 5 VERY INTERESTING projects, on topics that are of current research interest.
- A finished paper is expected by the end of class. The paper must contain original work and be of quality publishable in a good empirical journal.

Empirical Methods in Finance

- Empirical Finance/Economics has always been at the forefront of research

- The empirical tools are very useful, since:

  - We want to capture features of the data (descriptive)

  - We have to test hypotheses/models

  - Models are approximations of reality

  - We cannot conduct controlled experiments

  - We want to conduct exploratory work (data mining)

- The empirical methods in finance have evolved to capture the "peculiarities" of financial data. If you have done work in probability, statistics, or engineering, you will be familiar with some of the notions, but their use in finance might be different (e.g. estimation, forecasting, Kalman filtering).

- There is a huge demand for *capable* econometricians on Wall Street. The ability to synthesize information into a few, easy to interpret statistics is an invaluable analytical tool.

- We will emphasize solid foundations and important applications.

In this class:

1. Basics of Probability and Statistics

2. Basics of Econometrics

3. Intro to Time Series

4. Time Series Linear Regression Models

5. Elements of Linear Algebra

1   Basics of Probability and Statistics

- The collection of all possible outcomes (ω) of a random experiment is called a sample space and is denoted by Ω.
  – EX: Tossing a die, or $\Omega = \{$one spot, two spots, three spots

- Let event A be a subset of Ω, and let F be a collection of all such subsets, such that F is a σ-algebra
  – EX: Let A be the event that we have at least three spots, then $A = \{$three spots, ..., six spots$\}$

- Let's define $P(A)$, the probability that A will occur, such that
  – $P(A) \geq 0$ for every $A \in F$
  – $P(\Omega) = 1$
  – If A_1, A_2, ... is a countable sequence of events from F and $A_i \cap A_j$ is the null set for all $i \neq j$, then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

- Note the triple (Ω, F, P) is a probability space

- We will work with random variables instead of events.

- A random variable X is a real valued function defined on Ω such that the set $\{\omega : X(\omega) \leq x\}$ is a member of F for every real number x.

- Note: A random variable is *discrete* if the set of outcomes is either finite or is countably infinite (EX: number of trades in a given time period)

- Note: A random variable is *continuous* if the set of outcomes is infinitely divisible.

- Example: The simple return R_t of an asset (any asset) is defined as:
$$R_t = \frac{P_t + D_t - P_{t-1}}{P_{t-1}}$$
$$P_{t-1}(1 + R_t) = P_t + D_t$$
where P_t is the price at time t, P_{t-1} is the price at time $t-1$, D_t is the dividend at time t.
  - R_t, P_t, and D_t are random variables.
  - Q: Continuous or discrete?

- If we denote $f_X(x) = P_X(X = x)$ where X is a discrete variable, then it can be shown that:
  - $0 \leq f_X(x) \leq 1$
  - $\sum_x f(x_i) = 1$
- For a continuous variable Y, $f_Y(y) = 0$, but $P_Y(a < y < b) = \int_a^b f_Y(y)\, dy \geq 0$. Also $P_Y(-\infty < y < \infty) = \int_{-\infty}^{\infty} f_Y(y)\, dy = 1$.
- Cumulative distribution function (cdf) in discrete case: $F_X(x) = P(X \leq x) = \sum_{X \leq x} f(x)$
- Cumulative distribution function (cdf) in continuous case: $F_Y(y) = P(Y \leq y) = \int_{-\infty}^{x} f(t)dt$
- It can be shown that:
  - $0 \leq F_X(x) \leq 1$
  - If $x > y$, then $F_X(x) > F_X(y)$.
  - $F_X(+\infty) = 1$ and $F_X(-\infty) = 0$
  - $P_X(a < x \leq b) = F_X(b) - F_X(a)$

- Expectation:
  - The expected value of a discrete random variable X is $E(X) = \sum_x x f_X(x)$
  - The expected value of a continuous random variable Y is $E(Y) = \int_y y f_Y(y) dy$
  - We usually denote the expectation of a variable, also called its first moment, by μ. It is a "weighted average" of the values taken by X, where the weights are the respective probabilities.

- Let $g(.)$ be a function of x. Then
  - $E(g(X)) = \sum_x g(x) f_X(x)$, if X is a discrete r.v.
  - $E(g(X)) = \int_x g(x) f_X(x) dx$, if X is a continuous r.v.

- Using the above results, it is easy to see that if $g(x) = a + bx$, then $E(g(x)) = a + bE(x)$.

- Examples:
  (a) Difference between **expected return**, $E(R_t)$, and **realized return**, R_t.
    - $*$ Never to confuse one for the other!
    - $*$ We model $E(R_t)$
    - $*$ R_t is usually hard to forecast
  (b) Portfolio Returns: Value-Weighted and Equally-Weighted Portfolio Returns.
    - $*$ Suppose we have two stocks with returns, R_t^1 and R_t^2, respectively. A portfolio return is $R_t^p = w^1 R_t^1 + w^2 R_t^2$, s.t. $w^1 + w^2 = 1$.
    - $*$ $E\left(R_t^p\right) = w^1 E\left(R_t^1\right) + w^2 E\left(R_t^2\right)$.
    - $*$ Suppose that $w^1 = w^2 = 1/2$. Both stocks are equally weighted. This portfolio is called equally weighted!
    - $*$ Suppose that the market value of company 1 is 1 and that of company 2 is 3. Suppose we define $w^1 = 1/4$ and $w^2 = 3/4$. This portfolio is called value-weighted.

* Ex: The market portfolio is a value-weighted portfolio.

- The variance of a discrete random variable X is
$Var\left(x\right) = E\left[\left(x - \mu_x\right)^2\right] = \sum_x \left(x - \mu_x\right)^2 f_x\left(x\right)$

- The variance of a continuous random variable Y is
$Var\left(y\right) = E\left[\left(y - \mu_y\right)^2\right] = \int_y \left(y - \mu_y\right)^2 f_y\left(y\right) dy$

- Note: The variance of a random variable X, usually denoted by σ_x^2 or $Var(x)$, is positive.

- Note: $Var(x) = E\left[\left(x - \mu_x\right)^2\right] = E\left[x^2 - 2x\mu_x + \mu_x^2\right] = E\left[x^2\right] - \left(E\left[x\right]\right)^2$

- Note: Recall that $E\left[x\right]$ was called the first moment of X. Similarly, $E\left[x^2\right]$ is called the second moment of X.

- Note: $Var(x) = E\left[x^2\right] - \left(E\left[x\right]\right)^2$ is called the second centered moment of X.

- Let $g()$ be a function of x. Then
  - $Var(g(x)) = \sum_x \left(g(x) - E\left[g\left(x\right)\right]\right)^2 f_X(x)$, if X is a discrete r.v.
  - $Var(g(x)) = \int_x \left(g(x) - E\left[g\left(x\right)\right]\right)^2 f_X(x)dx$, if X is a continuous r.v.

- Let $g(x) = a + bx$. Using the above results, we can show that $Var\left(g(x)\right) = b^2 Var(x)$.

- Two other useful moments:
  - Skewness: $E\left[(x - \mu_x)^3\right]$ is used as a measure of the asymmetry of a distribution. Note that for a distribution that is symmetric around the mean, we have $f_X(\mu - x) = f_X(\mu + x)$. Therefore, is such a case, we can show that $E\left[(x - \mu_x)^3\right] = 0$. Skewness it the third centered moment of X.
  - Kurtosis: $E\left[(x - \mu_x)^4\right]$ is used as a measure of the "thickness" of the tails of a distribution.

- Note: It is not true that $E(g(x)) = g(E(x))$ for any function $g()$.

- Useful heuristics: $g(x) \approx g(\tilde{x}) + g'(\tilde{x})(x - \tilde{x})$ (Taylor's expansion). But here we are working we random variables. Let $\tilde{x} = \mu_x$. Then $g(x) \approx g(\mu_x) + g'(\mu_x)(x - \mu_x)$. Then,
  - $E(g(x)) \approx g(\mu_x)$
  - $Var(g(x)) \approx [g'(\mu_x)]^2 Var(x)$

- Tremendously useful densities
  - Normal density: $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu_x)^2}{2\sigma^2}}$. The usual notation is $X \sim N(\mu_x, \sigma^2)$.
    * It is important to note that if X has a normal distribution, it is characterized by its first two moments. I.e. if we know μ_x and σ^2, we know the entire distribution of X.
    * If $X \sim N(\mu_x, \sigma^2)$, and $Y = a + bX$, then $Y \sim N(a + b\mu_x, b^2\sigma^2)$.
    * Let $Z = a + bX$ where $a = -\mu_x/\sigma$ and $b = 1/\sigma$. Then $Z \sim N(0, 1)$. We say that Z has a standard normal distribution, denoted by $N(0, 1)$. Note that the standard normal distribution does not depend on any parameters. In other words, $f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$ can be plotted.

– Chi-squared (χ^2) distribution: If $Z \sim N(0, 1)$ and $X = Z^2$, then $X \sim \chi^2(1)$, i.e. X has a chi-squared distribution with 1 degree of freedom.

– If X_1, X_2, ... X_n are n independent $\chi^2(1)$ variables and $Y = \sum_{i=1}^{n} X_i$, then $Y \sim \chi^2(n)$, i.e. Y has a chi-squared distribution with n degrees of freedom.

* Note: The χ^2 distribution is characterized by its degree of freedom (so, if we want to plot the chi-squared distribution, we have to do so for each value of n).

* Note: The mean and variance of a chi-squared variable with n degrees of freedom are n and $2n$, respecively. In other words, $E(Y) = n$ and $Var(Y) = 2n$.

* Note: If Z_1, Z_2, ... Z_n are n independent standard normal r.v's, then $Y = \sum_{i=1}^{n} Z_i^2 \sim \chi^2(n)$

* If Y_1 and Y_2 are two independent chi-squared r.v's with n_1 and n_2 degrees of freedom, then $Y_1 + Y_2 \sim \chi^2(n_1 + n_2)$

* Note: Many statistical tests will have a chi-squared distribution, because they will be constructed by taking squares of standard normal variables.

– F distribution: If X_1 and X_2 are two independent chi-squared variables with degree of freedom parameters n_1 and n_2, respectively. Then, the ratio

$$Y = \frac{X_1/n_1}{X_2/n_2}$$

has an F distribution with (n_1, n_2) degrees of freedom.

* Note: The F distribution is characterized by its two degrees of freedom (so, if we want to plot the F distribution, we have to do so for each pair (n_1, n_2)).

* Note: Many statistical tests will have a F distribution, because they will be constructed by taking ratios of chi-squared variables.

* Note: The independence assumption is crucial.

– t distribution: If $Z \sim N(0,1)$ and $Y \sim \chi^2(n)$, then
$$T = \frac{Z}{\sqrt{Y/n}}$$
has a t distribution with n degrees of freedom.

* Note: The t distribution is characterized by its degree of freedom (so, if we want to plot the chi-squared distribution, we have to do so for each value of n).

* The t-distribution "looks" like the normal distribution but has thicker tails. It is used in empirical finance to model the distribution of stock returns.

* Note: $T^2 = \frac{Z^2/1}{Y/n}$ has what distribution?

* Note: $T(n \rightarrow \infty)$ has a standard normal distribution.

– Lognormal distribution
  * Suppose that $X \sim N(\mu, \sigma^2)$ and $Y = e^X$, then Y has a lognormal distribution if $\ln(Y) \sim N(\mu, \sigma^2)$.
  * $E(Y) = e^{\mu + \sigma^2/2}$
  * $Var(Y) = \left(e^{\sigma^2} - 1\right) e^{2\mu + \sigma^2}.$
  * In the above definition, we started with a standard normal variable and derived the moments of the lognormal variable. But the other way around also works. If Y is a lognormal distribution with mean θ and variance λ^2, then $\ln(Y) \sim N(\mu, \sigma^2)$, where $\mu = \ln \theta^2 - \frac{1}{2} \ln \left(\theta^2 + \lambda^2\right)$ and $\sigma^2 = \ln \left(1 + \lambda^2/\theta^2\right)$.

– Use of lognormality in finance: Let
$$1 + R_t = \frac{P_t + D_t}{P_{t-1}}$$
where R_t is the simple return between periods $t - 1$ and t, P_t is the price at time t, D_t is the dividend at time t. $(1 + R_t)$ is called the gross simple return between periods $t - 1$ and t. Let r_t be the continuously compounded one-period return (between $t - 1$ and t). We define r_t as:
$$P_{t-1} (1 + r_t \Delta)^{1/\Delta} = P_t + D_t$$
as $\Delta \to 0$. But $\lim_{\Delta \to 0} (1 + r_t \Delta)^{1/\Delta} = e^{r_t}$. Assume $D_t = 0$ for all t. Therefore
$$P_{t-1} e^{r_t} = P_t$$
Similarly, the 2-period continuously compounded return (between $t - 2$ and t) is
$$P_{t-2} e^{r_{t-2,t}} = P_t$$

– It is even easier in logs:
$$\ln P_{t-1} + r_t = \ln P_t$$
$$\ln P_t - \ln P_{t-1} = r_t$$
The log (or continuously compounded) return is the first difference of log prices.

– This is convenient, because:

$$P_{t-2}e^{r_{t-1}} = P_{t-1}$$
$$P_{t-1}e^{r_t} = P_t$$

and then

$$P_{t-2}e^{r_{t-1}}e^{r_t} = P_{t-2}e^{r_{t-1}+r_t} = P_{t-2}e^{r_{t-2,t}} = P_t$$

– Similarly in logs:

$$r_{t-2,t} = \ln P_t - \ln P_{t-2} = (\ln P_t - \ln P_{t-1}) + (\ln P_{t-1} - \ln P_{t-2}) = r$$

– Punchline: The two-period continuously compounded return is equal to the sum of the one-period continuously compounded returns, or $r_{t-2,t} = r_t + r_{t-1}$.

– The same argument would carry through for a k-period continuously compounded return $r_{t-k,t} = r_{t-k+1} + ... + r_t$

– Q: What does this have to do with lognormality?

$$P_{t-1}(1 + R_t) = P_t$$
$$P_{t-1}e^{r_t} = P_t$$

Then,

$$R_t = e^{r_t} - 1$$

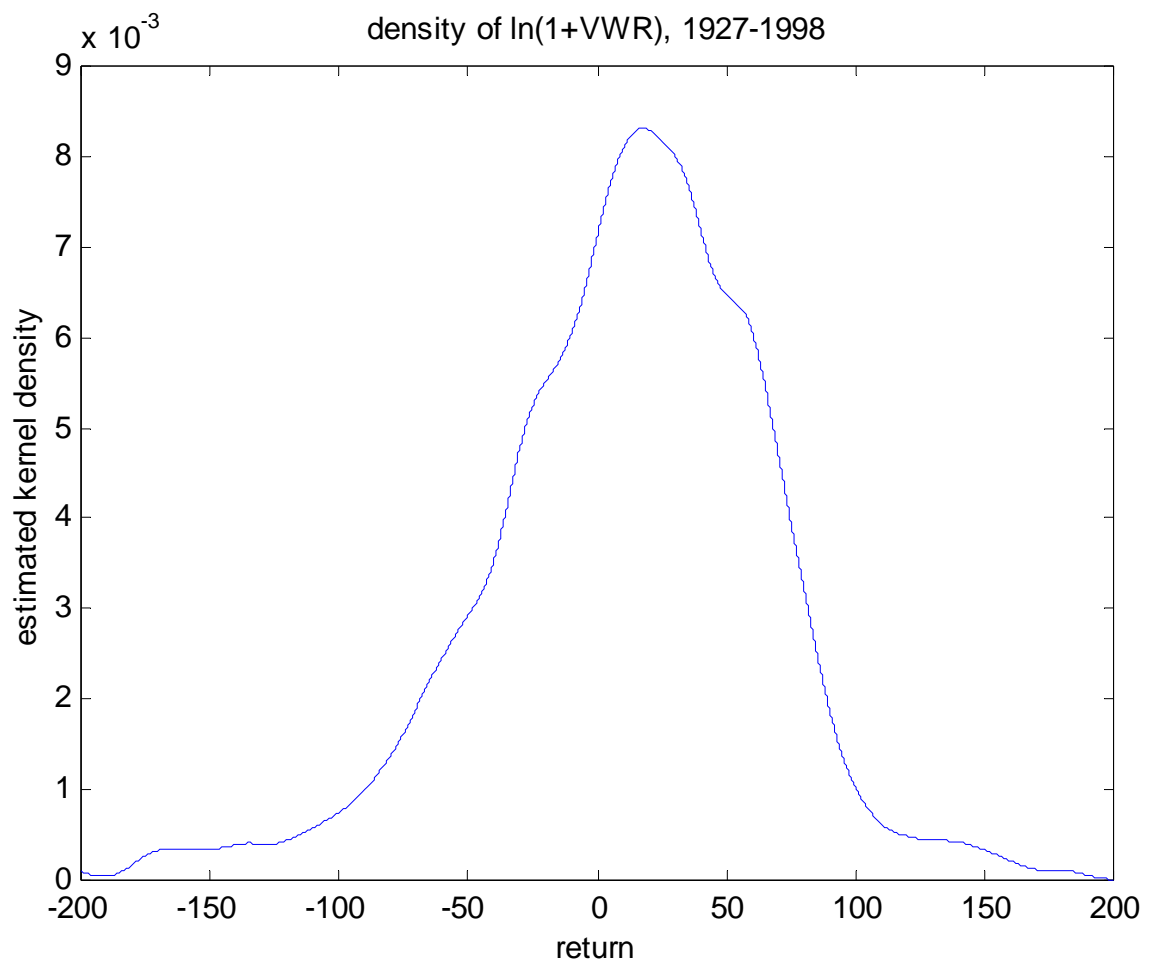* Assume that $r_t \sim N(\mu, \sigma^2)$, then R_t has a lognormal distribution.

$$E(R_t) = e^{\mu + \sigma^2/2} - 1$$
$$Var(R_t) = e^{2\mu + \sigma^2}\left(e^{\sigma^2} - 1\right)$$

* Alternatively, if we assume that the mean and variance of the simple return R_t are θ and λ^2, then

$$E(r_t) = ...$$
$$Var(r_t) = ...$$

density of ln(1+VWR), 1927-1998

1.

– Plot of r_t for the market portfolio.

– Multivariate: The joint distribution of X and Y, denoted by $f_{X,Y}(x,y)$ is defined so that

– $P(a < x < b, c < y < d) = \int_a^b \int_c^d f_{X,Y}(x,y)\,dydx$, for continuous X and Y.

– $P(a < x < b, c < y < d) = \sum_{a<x<b} \sum_{c<y<d} f_{X,Y}(x,y)\,dydx$, for discrete X and Y.

– The cumulative distribution function $F_{X,Y}(x,y) = P(X \leq x, Y \leq y)$

– Also, $F_{X,Y}(-\infty, -\infty) = 0$ and $F_{X,Y}(\infty, \infty) = 1$, etc.

– Marginal density:
  * The marginal density of X, $f_X(x) = \int_y f_{X,Y}(x,y)\,dy$, if y is continuous
  * The marginal density of X, $f_X(x) = \sum_y f_{X,Y}(x,y)$, if y is discrete
  * Similarly for $f_Y(y)$.

– Independence: Two random variables X and Y are statistically independent if and only if
$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

– Alternatively, two random variables X and Y are statistically independent if and only if
$$F_{X,Y}(x,y) = F_X(x) F_Y(y)$$

– Note that the above calculations and definitions are still true:
  * EX: $E(X) = \int_x x f_X(x)\, dx = \int_x \int_y x f_{X,Y}(x,y)\, dy\, dx$
  * Similarly for variances, etc.

– For any function $g(.,.)$, $E(g(.,.)) = \sum_x \sum_y g(x,y) f_{X,Y}(x,$

– The covariance is a special case, for $g(x,y) = (x - \mu_x)(y - \mu_y)$.

$$Cov(X,Y) = \sigma_{x,y} = E\left((x - \mu_x)(y - \mu_y)\right)$$
$$= E(xy) - \mu_x \mu_y$$

– If X and Y are independent, then

$$\sigma_{x,y} = \int_x \int_y (x - \mu_x)(y - \mu_y) f_{X,Y}(x,y)\, dy\, dx$$
$$= \int_x \int_y (x - \mu_x)(y - \mu_y) f_X(x) f_Y(y)\, dy\, dx$$
$$= \int_x (x - \mu_x) f_X(x)\, dx \int_y (y - \mu_y) f_Y(y)\, dy$$
$$= E(x - \mu_x) E(y - \mu_y) = 0$$

– IMPORTANT: The other way is not true.
  * Example: corr_sim.m

– Correlation:
$$\rho_{x,y} = \frac{\sigma_{X,Y}}{\sqrt{\sigma_X^2}\sqrt{\sigma_Y^2}}$$

– Conclusion: Covariances and correlations are appropriate measures of DEPENDENCE only when we have LINEAR relationships.

– Conditioning: We can ask, what is the distribution of Y, given that X is a certain number. In other words, we want

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

Similarly,

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

– If X and Y are statistically independent, then $f_{Y|X}(y|x) = f_Y(y)$ and $f_{X|Y}(x|y) = f_X(x)$.

2 Basics of Econometrics

- Conditional Mean–Regression (regression function)
  - Suppose that X and Y have a joint distribution $f_{X,Y}(x,y)$. Then $E(y|x) = \int_y y f_{Y|X}(y|x)\, dy$ is called the **regression** of y on x.
  - Why is this a useful relationship?
    * Any variable Y can be written as:
    $$\begin{aligned} y &= E(y|x) + \{y - E(y|x)\} \\ &= E(y|x) + \varepsilon \end{aligned}$$
    * The first term is called the systematic part of y.
    * Note that the regression $E(y|x)$ is a function of x.

- Important: In finance, some of the most important models are linear
  - Example: the CAPM
    * The return of asset i, R_i is a random variable. The risk free rate is denoted by r^f.
    * R_M is the return of the entire market and it is also a random variable
    * Suppose we can write
      $$R_i = r^f + \beta_i(R_M - r^f) + \varepsilon_i$$
    * Suppose that $E(\varepsilon_i | R_M) = R_i - E(R_i | R_M) = 0$
    * Then, the conditional mean is:
      $$E(R_i | R_M) = r^f + \beta_i \left(R_M - r^f \right)$$
    * Also, note that the unconditional mean is (assuming $E(\varepsilon_i) = 0$):
      $$E(R_i) = r^f + \beta_i E \left(R_M - r^f \right)$$
    * Can you see the difference between the conditional and the unconditional mean?
    * The CAPM is usually written in terms of excess returns, i.e.
      $$E \left(R_i - r^f \right) = \alpha_i + \beta_i E \left(R_M - r^f \right)$$

- Conditional Variance–Scedasticity (scedastic function)
  - Suppose that X and Y have a joint distribution $f_{X,Y}(x, y)$. Then $Var(y|x) = \int_y (y - E(Y|X))^2 f_{Y|X}(y|x)$ ⊘ is called the **conditional variance** of y.
  - We can simply write $Var(y|x) = E(y^2|x) - (E(y|x))^2$
  - Before we can compute $Var(y|x)$, we have to know $E(y|x)$
  - Note: Scedasticity is a function of x. People are used to assume $Var(y|x) = Var(y)$. In such a case, y is homoskedastic. Otherwise, it is heteroskedastic. In finance, the homoskedasticity is not a tenable assumption (for returns).
  - Example: GARCH (Generalized Autoregressive Conditional Heteroskedasticity)
    * I.e., The variance today depends on the variance yesterday (the "auto-regressive" part) and it changes over time (the "conditional heteroskedastic" part).

– Relationship between Marginal and Conditional Moments: Law of Iterated Expectations

$$E\left(y\right) = E_x\left(E\left(y|x\right)\right)$$

* Note: The first expectation is the conditional expectation, taken with respect to $f_{Y|X}\left(y|x\right)$. It produces the regression, a function of x.

* Note: The second expectation is taken with respect to $f_X\left(x\right)$. The result is a number.

* This result is very useful and will be used repeatedly in this class.

- Law of Iterated Expectations: Precise formulation of the Efficient Markets Hypothesis
  - Suppose that I_t denotes the information at time t. Similarly I_{t+1} denotes the information at time $t+1$.
  - Let $I_t \subset I_{t+1}$ (we don't forget, and we learn extra stuff)
  - We value a certain asset with fundamental value V. Then, we can write the price at time t (today):
    $$P_t = E(V|I_t) = E_t(V)$$
  - At time $t+1$ (tomorrow), similarly
    $$P_{t+1} = E(V|I_{t+1}) = E_{t+1}(V)$$
- Q: Can we forecast changes in prices? Well,
  $$\begin{aligned} E_t(P_{t+1} - P_t) &= E_t(E(V|I_{t+1})) - E_t(P_t) \\ &= E(E(V|I_{t+1})|I_t) - E(V|I_t) \\ &= E(V|I_t) - E(V|I_t) \\ &= 0 \end{aligned}$$
- In other words, $E_t P_{t+1} = P_t.$ Such a process is called m_____le.
- If we define a process $e_{t+1} = P_{t+1} - P_t$, then e_t is called a martingale difference sequence if $E(e_t|e_{t-1}, e_{t-2}, ...) = 0$

- – Example of Conditional Expectations: Linear regression.
$$y = E\left(y|x\right) + \varepsilon$$
Thus far, we have said nothing about $E\left(y|x\right)$ except that it is a function of x.
  * Suppose that $E\left(y|x\right) = \alpha + \beta x$.
  * Then, using the law of iterated expectations
$$\alpha = E\left(y\right) - \beta E\left(x\right)$$
  * To get β, we note that
$$\begin{aligned} Cov\left(x, y\right) &= cov\left(x, E\left(y|x\right)\right) \\ &= cov\left(x, \alpha + \beta x\right) \\ &= \beta Var(x) \end{aligned}$$
  * Therefore,
$$\beta = \frac{Cov\left(x, y\right)}{Var(x)}$$

– Decomposition of Variance (ANOVA):
  * Another useful trick:
  $$Var(y) = Var_x \left(E\left(y|x\right) \right) + E_x \left(Var\left(y|x\right) \right)$$
  In other words, the variance of y can be decomposed into
    · Variance of the conditional expectation (variance of the regression)
    · Expected value of the variance around the conditional mean (residual variance)
  * We can write:
  $$1 = \frac{Var_x \left(E\left(y|x\right) \right)}{Var(y)} + \frac{E_x \left(Var\left(y|x\right) \right)}{Var(y)}$$

- Sample: Collection of observations.

- Random sample: A sample of n observations denoted by $\{x_i\}_{i=1}^{n}$ is a random sample if the n observations are drawn independently from the same population, or density $f_\theta(x)$, where θ are the parameters of the distribution

- We will estimate the population parameters, such as μ and σ^2 using their sample analogues.

- The estimate of μ is: $\overline{x} = \sum_{i=1}^{n} x_i \frac{1}{n}$.

- The estimate of σ^2 is $s^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2 \frac{1}{n-1}$

- Note: If we have a truly random sample, then everything is easy, but we don't....

– We will distinguish between three (non-random) samples:

 * a cross section is a sample drawn at the same point in time (prices of stocks at the close of market today).

 * a time series is a sample drawn from the same observational unit at various points in time (price of the stock market everyday for the last 30 years).

 * a panel data is a sample of many observations, followed at various points in time. (daily prices of all stocks for the past 30 years)

– This distinction is necessary because the samples present different issues to deal with.

– Note that if we know the exact distribution $f_\theta(x)$ (functional form and θ), then we know the properties of x. However, we don't know $f_\theta(x)$.

– However, we observe a sample of a finite number of observations and want to infer the properties of X (its mean, variance, and entire distribution).

- A time series will be denoted by $\{r_t\}$, $t = 1, ..., T$

- A cross section will be denoted by $\{q_i\}$, $i = 1, ..., N$

- Note: In finance, we don't have random samples and herein lies the problem.

- Note: The samples of stock and bond returns were examples of two time series

- Together, they constitute a (small) panel dataset.

- We will devote most of our time at studying time series processes.

3 Intro to Time Series

- Suppose we have a sample of size T of independent and identically distributed (iid) variables $\{\varepsilon_t\}_{t=1}^{T}$. Then ε_t is called a white noise.

- If we add the assumption the the distribution of (all ε_t) is Normal, then ε_t is called a Gaussian (or Normal) white noise.

- Suppose we have observed a sample of size T of some random variable, Y_t: $\{y_t\}_{t=1}^{T}$

- Supoose we observe another sample of Y_t, denoted by $\{y_t^1\}_{t=1}^{T}$.

- Then we observe another, and another and another sample of Y_t, denoted by $\{y_t^i\}_{t=1}^{T}$, $i = 1, ..., I$.

- Let's suppose that the samples are independent.

- The variable Y_t has some density, $f_{Y_t}(y_t)$ and expectation $E(Y_t) = \int y_t \, f_{Y_t}(y_t) \, dy_t = \mu_t$

- We approximate $E(Y_t) \approx \frac{1}{I} \sum_{i=1}^{I} y_t^i = \overline{y}_t$

- Similarly, $Var(Y_t) = \int (y_t - \mu_t)^2 \, f_{Y_t}(y_t) \, dy_t = \gamma_t(0)$

- We approximate $Var(Y_t) \approx \frac{1}{I} \sum_{i=1}^{I} \left(y_t^i - \overline{y}_t\right)^2 = \hat{\gamma}_t(0)$

- Define: $Cov(Y_t, Y_{t-j}) = \gamma_t(j) = \int \int ... \int (y_t - \mu_t) \left(y_{t-j} - \mu_{t-j}\right) f_{Y_t, Y_{t-1}, ... Y_{t-j}}(y_t, y_{t-1}, ..., y_{t-j}) \, dy_t dy_{t-1} ... dy_{t-j}$

- Note: $Cov(\varepsilon_t \varepsilon_{t-j}) = 0$ for $j \neq 0$.

- Note: We can compute $\gamma_t(j)$ for all j. This is called the autocovariance function of Y_t. It captures the time-dependence between Y_t and Y_{t-j} for various $j's$.

- Covariance stationarity: If neither μ_t nor $\gamma_t(j)$ depend on t, then the process Y_t is said to be covariance stationary (or weakly stationary). In other words
  - $E(Y_t) = \mu$, for all t.
  - $E(Y_t - \mu)(Y_{t-j} - \mu) = \gamma(j)$, for all t and any j.
- Example. Suppose that $Y_t = \beta t + \varepsilon_t$, where ε_t is gaussian white noise. Is Y_t covariance stationary? Why?
- Example: Suppose that $Y_t = \varepsilon_t$, where ε_t is gaussian white noise. Is Y_t covariance stationary? Why?
- Note: For a covariance stationary process
  $$E(Y_t - \mu)(Y_{t-j} - \mu) = E(Y_t - \mu)(Y_{t+j} - \mu) = \gamma(j)$$
- Strict stationarity: The joint distribution of $(Y_{t+j1}, Y_{t+j2,...,}Y_{t+jk})$ depends only on the intervals separating the dates $(j1, j2, ...jk)$ and not on the date t itself.
- We will work mostly with covariance stationarity.

- So far so good, but we usually have only one observation of a series. Think of stock market. Or, we do not have many independent observations of a time series.

- Q: What to do? How can we estimate $E(Y_t)$?

- In practice, we have $\left(y_t^1\right)_{t=1}^T$

- It is tempting to estimate $\overline{y} = \frac{1}{T}\sum_{t=1}^T y_t^1$, but is it true that $\overline{y} \approx E(Y_t) = \mu$.

- Definition: A covariance stationary process is said to be ergodic for the mean if

$$p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^T y_t^1 = \mu$$

- If Y_t is covariance stationary, it is sufficient that $\Sigma_{j=0}^{\infty}\left|\gamma(j)\right| < \infty$ for the process to be ergodic for the mean. The above condition is called absolute summability.

- Definition: A covariance stationary process is said to be ergodic for the second moments if

$$p \lim_{T \to \infty} \frac{1}{T-k} \sum_{t=1}^{T} \left(y_t^1 - \mu \right) \left(y_{t-k}^1 - \mu \right) = \gamma \left(k \right)$$

- In the special case where Y_t is Gaussian (stationary) process, then absolute summability is enough to insure ergodicity for all moments.

An interesting time series process: Autoregressive process of order 1 (AR(1))

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t$$

where $\{\varepsilon_t\}$ is a white noise sequence (uncorrelated, with mean zero and variance σ^2).

• Note: The parameter ϕ is very important.

   – If $\phi = 0$, $Y_t = \varepsilon_t$.

   – If $\phi = 1$, Y_t is non-stationary.

• Suppose $|\phi| < 1$. Then $E(Y_t) = \frac{c}{1-\phi}$

• Similarly, $\gamma(0) = \frac{\sigma^2}{1-\phi^2}$

• Also, $\gamma(j) = \frac{\sigma^2}{1-\phi^2} \phi^j$

• The autocorrelation function $\rho(j) = \gamma(j)/\gamma(0) = \phi^j$.

- Note that, $P_t = P_{t-1} + \varepsilon_t$ is an example of an autoregressive process with $\phi = 1$.

- Q: Is P_t stationary? Why? (Hint: Think second moments)

- Q: Suppose that r_t the (continuously compounded) return. Suppose that $r_t = c + \phi r_{t-1} + \varepsilon_t$, where ε_t is white noise. Under the efficient market hypothesis, what is the magnitude of ϕ? Should c be equal to zero?

- Q: What is $E(r_t) = ?$

- Q: What is $E_{t-1}(r_t) = ?$

- Q: What is the difference?

The process
$$Y_t = c + \phi Y_{t-1} + \varepsilon_t$$
is governed by the parameter ϕ. How do we estimate ϕ? Are there other parameters to estimate? How do we estimate those parameters?

- Idea: Treat Y_t as the left hand side variable in a regression and Y_{t-1} as the right hand side variable.

- Q: Is this legitimate?

4   Time Series Linear Regression Models

$$y_t = x_t \beta + \varepsilon_t$$

- Given the observations (y_t, x_t), the ordinary least squares estimate of β, denoted by $\hat{\beta}^{ols}$, or simply $\hat{\beta}$, is the value of β that minimized the residual sum of squares, or

$$\hat{\beta}^{ols} = \arg\min_{b} \Sigma_{t=1}^{T} \left(y_t - x_t\beta\right)^2$$

- The solution is

$$\hat{\beta} = \left(\Sigma_{t=1}^{T} x_t x_t\right)^{-1} \left(\Sigma_{t=1}^{T} x_t y_t\right)$$

- Recall $(\beta = cov/var)$

- Note that

$$
\begin{aligned}
\hat{\beta} &= \left(\Sigma_{t=1}^{T} x_t x_t\right)^{-1} \left(\Sigma_{t=1}^{T} x_t y_t\right) \\
&= \left(\Sigma_{t=1}^{T} x_t x_t\right)^{-1} \left(\Sigma_{t=1}^{T} x_t \left(x_t\beta + \varepsilon_t\right)\right) \\
&= \left(\Sigma_{t=1}^{T} x_t x_t\right)^{-1} \left(\Sigma_{t=1}^{T} x_t x_t\right) \beta + \left(\Sigma_{t=1}^{T} x_t x_t\right)^{-1} \Sigma_{t=1}^{T} x_t \varepsilon_t \\
&= \beta + \left(\Sigma_{t=1}^{T} x_t x_t\right)^{-1} \Sigma_{t=1}^{T} x_t \varepsilon_t
\end{aligned}
$$

- Note:
$$\hat{\beta} - \beta = \left(\Sigma_{t=1}^{T} x_t x_t\right)^{-1} \Sigma_{t=1}^{T} x_t \varepsilon_t$$

- If $\hat{\beta}$ is a good estimate of β, it better be the case that $\left(\Sigma_{t=1}^{T} x_t x_t\right)^{-1} \Sigma_{t=1}^{T} x_t \varepsilon_t$ is small. In fact, we can argue that this piece is "small" and "decreases" as $T \to \infty$. For that, we need that $E\left(x_t \varepsilon_t\right) = 0$. Then, we can argue that $plim_{T \to \infty} \frac{1}{T} \Sigma_{t=1}^{T} x_t \varepsilon_t = E\left(x_t \varepsilon_t\right) = 0$ and that $\frac{1}{T} \Sigma_{t=1}^{T} x_t x_t$ converges to a finite number.

- Next class, we will discuss the properties of $\hat{\beta}$ in details.

- Now, we want to build some intuition: simpleregression.m and simpleregression2.m

- But we have not answered whether it is legitimate to run the regression
$$Y_t = c + \phi Y_{t-1} + \varepsilon_t$$

- When I say "legitimate," I mean would we get consistent estimates of the parameters of interest.

- From the above discussion we need that $E\left(Y_{t-1}\varepsilon_t\right) = 0$.

- Note that we can write
$$
\begin{aligned}
Y_{t-1} &= c + \phi Y_{t-2} + \varepsilon_{t-1} \\
&= c + \phi\left[c + \phi Y_{t-3} + \varepsilon_{t-2}\right] + \varepsilon_{t-1} \\
&= c + \phi c + \phi^2 Y_{t-3} + \varepsilon_{t-1} + \phi\varepsilon_{t-2} \\
&= \frac{c}{1-\phi} + \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-1-i}
\end{aligned}
$$

- Hence, we need
$$
E\left(\left[\sum_{i=0}^{\infty} \phi^i \varepsilon_{t-1-i}\right] \varepsilon_t\right) = 0
$$

- In other words, we need the $\varepsilon_t's$ to be serially uncorrelated.

- If we have serially correlated $\varepsilon's$, we have to take that correlation into account.

- Let's look at the bivariate forecasting relation:

$$y_t = x_{t-1}\beta + \varepsilon_t$$

- Note, x_{t-1} forecasts (lags) y_t.

- This is not a contemporaneous relation.

- We need that $E\left(x_{t-1}\varepsilon_t\right) = 0$.

- Suppose that

$$x_t = \phi x_{t-1} + u_t$$

- In other words, x_t is an autoregressive process.

- We will need that $E\left(\varepsilon_t u_{t-j}\right) = 0$ all $j's$.

- If this is not the case, we will have problems!

- The problems will be more severe for more persistent x_t.

- Aside: How can we define persistence?

- All this will become intuitive.

5   Elements of Linear Algebra

- – We are almost there! Now we have to sprint through linear algebra.

  – A k-dimentional vector μ is defined as $\mu = \begin{bmatrix} \mu_1 \\ \mu_1 \\ \vdots \\ \mu_k \end{bmatrix}$.

  – A k by k matrix Σ is defined as $\sum = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \sigma_{1k}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \\ & & \\ \sigma_{11}^2 & & \sigma_{kk}^2 \end{bmatrix}$

  – We all know what is a symmetric, diagonal, triangular and an identity matrix.

  – Idempotent matrix: If M is idempotent, then $MM = M$.

  – Idempotent symmetric matrix: $M'M = M$.

– A matrix M is invertible if there exist a matrix M^{-1} such that $MM^{-1} = M^{-1}M = I$

– Linear form: let r and w be two k dimentional vectors. Then $R = r'w$ is a linear form.

– Quadratic form: If A is a symmetric matrix, then $Q = w'Qw$ is a quadratic form.
  * If $Q > (\geq)$ for all non-zero w, then A is positive definite (semidefinite)
  * If $Q < (\leq)$ for all non-zero w, then A is negative definite (semidefinite)

- Suppose that the return on stocks is r_1 and the return on bonds is r_2.

- Let $E(r_1) = \mu_1$ and $E(r_2) = \mu_2$.

- Let $Var(r_1) = \sigma_1^2$, $Var(r_2) = \sigma_2^2$, and $Cov(r_1, r_2) = \sigma_{12}$.

- Supose that we invest w of our wealth in stocks and the rest, $1 - w$, in bonds. The return from such a portfolio is $r_p = wr_1 + (1 - w) r_2$

- The expected return from the portfolio is: $E(r_p) = [w \; (1-w)] \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = w\mu_1 + (1-w)\mu_2.$

- The variance of such a portfolio is: $Var(r_p) = [w \; (1-w)] \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \begin{bmatrix} w \\ 1 - w \end{bmatrix} = \mathbf{w}'\Sigma\mathbf{w}$

- – In finance, matrix algebra is very useful when we work with portfolios.

- – Portfolio choice problems: What fraction of your wealth shoud be invested in stocks and how much in bonds, given that you want a certain level of return from the portfolio?

- – Mean Variance: Portfolio p is a minimum variance portfolio of all portfolios with mean return μ_p if its portfolio weights vector is the solution to the following problem:

$$\min_{w} w'\Sigma w$$

  subject to

$$w'\mu = \mu_p$$
$$w'e = 1$$

- – You should know by now what is the solution to this problem.

- – One of the most unrealistic assumptions of the above setup is the homoskedasticity assumption. There is ample evidence that variances vary over time. Modelling such variations is one of the goals of this class.

– Before we can solve the minimization problem above, we have to know what Σ and μ are, i.e. we have to know the mean and the variance-covariance of the return vector.

– Since we don't know what the true mean is, we have to estimate it.

– Hence, we are faced with two problems, or rather one problem that must be approached in two steps:

  * Step1: Estimate Σ and μ

  * Step2: Given the estimates of Σ and μ, say $\hat{\Sigma}$ and $\hat{\mu}$, we solve the minimization problem.

– In practice, step 2 is very heavily influenced by small errors in the estimation of Σ and μ. Therefore, one must have good estimators of those parameters. But this is not always possible if we have, say 10,000 returns.