Session 2_2: Distributions

Type	Name	Premise	Parameter(s)	Mean	Variance	PDF	CDF
Discrete	Bernoulli	• Random variable X is equal to 0 or 1 (binary outcomes) $P(X=1)=p,$ $P(X=0)=1-p$ • Is a "one-step" process	• p: Probability of $X=1$	p	$p(1\ -\ p)$	$p^x(1-p)^{1-x}$	**Conditional:** $1-p, \quad x=0$ $1, \quad x=1$
	Binomial	• Shows probability of obtaining exactly n success out of N Bernoulli trials	• N: Total no. of trials • n: No. of successes	Np	$Np(1-p)$	$\binom{N}{n}p^n(1-p)^{N-n}$, where $\binom{N}{n}=\frac{N!}{n!(N-n)!}$	$\sum_{n=0}^{\lfloor N \rfloor}\binom{N}{n}p^n(1-p)^{N-n}$
	Poisson	• **Used to model:** Occurrence of events over time, No. of market crashed, Jumps in jump diffusion models • It is a limiting case of a binomial model, where N is set to tend towards ∞	• n: No. of occurrences of event of which you want the probability of • λ: A constant which denotes the expected occurrence of the event per unit time, where $\lambda=Np$	λ	λ	$\frac{\lambda^n}{n!}e^{-\lambda}$	-
Piecewise Linear	Uniform	• Probability density is constant (horizontal line) between lower and upper bound and zero everywhere else	• a: Lower bound of x values • b: Upper bound of x values	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	**Conditional:** $\frac{1}{b-a}, \quad a<x<b$ $0,$ **Otherwise**	**Conditional:** $0, \quad x<a$ $\frac{x-a}{b-a}, \quad a<x<b$ $1, \quad x\geq b$
	Triangular	• PDF is triangle-shaped • Useful for modelling default rates and recovery rates	• a: Minimum x value • b: Maximum x value • c: Modal x value	-	-	**Conditional:** $\frac{2(x-a)}{(b-a)(c-a)}, \quad a<x<c$ $\frac{2(x-a)}{(b-a)(b-c)}, \quad c<x<b$	-

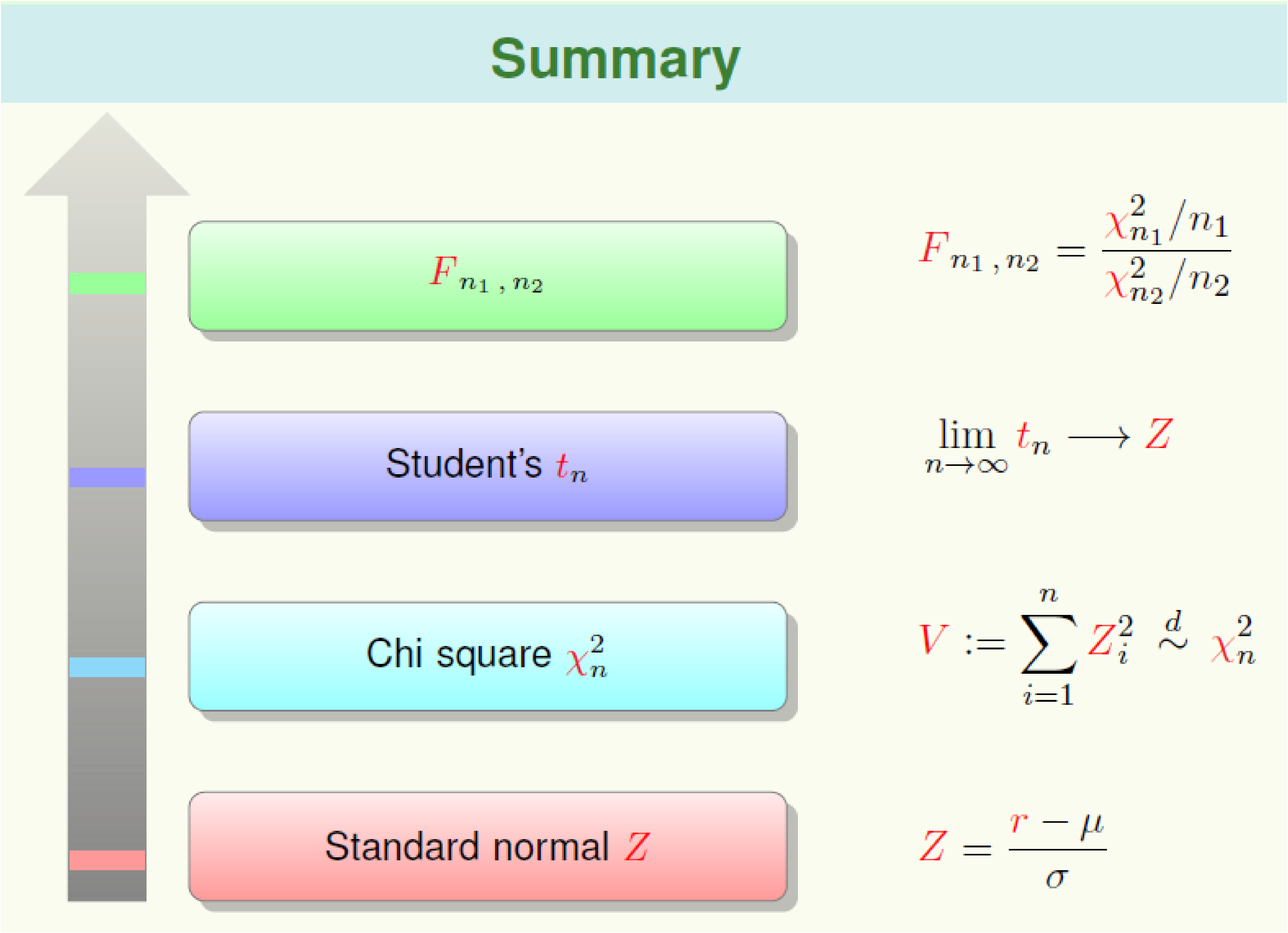
Session 2_2: Distributions

Type	Name	Premise	Parameter(s)	Mean	Variance	PDF
Continuous	**Normal**	• It is a symmetrical distribution • Also known as the Gaussian distribution • Any linear combination of independent normal random variables is also normal • Standard normal distribution is where $\boldsymbol{\mu}$ is 0 and $\boldsymbol{\sigma}$ is 1 • Bell curve has 0 skewness and kurtosis of 3 • To create two correlated normal variables, combine $\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}$: $\boldsymbol{x_1 = \sqrt{p}x_1 + \sqrt{1-p}x_2}$ $\boldsymbol{x_2 = \sqrt{p}x_1 + \sqrt{1-p}x_3}$ • Log returns are normally distributed, and as such log returns are attractive to be used • Easier to work with log returns with normal distribution than standard returns with lognormal distributions	• $\boldsymbol{\mu}$: Mean of $\boldsymbol{x}$ • $\boldsymbol{\sigma}$: Standard deviation of $\boldsymbol{x}$	$\boldsymbol{\mu}$	$\boldsymbol{\sigma^2}$	$\boldsymbol{\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{(x-\mu)^2}{2\sigma^2}}}$
	Lognormal	• It is an asymmetrical distribution • Distribution peaks at $\boldsymbol{e^{(\mu-\sigma^2)}}$ • If a variable has a lognormal distribution, the log of the variable has a normal distribution • If returns are normally distributed, $\boldsymbol{1+r}$ is lognormally distributed • Using lognormal distribution provides easy way to ensure returns of less than -100% are avoided	• $\boldsymbol{\mu}$: Mean of lognormally distributed $\boldsymbol{x}$ • $\boldsymbol{\sigma}$: Standard deviation of lognormally $\boldsymbol{x}$	$\boldsymbol{e^{\mu+\frac{1}{2}\sigma^2}}$	$\boldsymbol{\left(e^{\sigma^2}-1\right)e^{2\mu+\sigma^2}}$	$\boldsymbol{\frac{1}{x\sigma\sqrt{2\,\pi}}e^{-\frac{1}{2}\left(\frac{ln\,(x)-\mu}{\sigma}\right)^2}}$
	Chi-Squared	• If we have $\boldsymbol{k}$ independent standard normal variables, then the respective sum of their squares has a chi-squared distribution • As $\boldsymbol{k}$ increases, chi-squared distribution becomes increasingly symmetrical and converges towards the normal distribution • Widely used in risk management and for hypothesis testing	• $\boldsymbol{k}$: The degrees of freedom	$\boldsymbol{k}$	$\boldsymbol{2k}$	**For positive values of $\boldsymbol{x}$:** $\boldsymbol{\frac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)x^{\frac{k}{2}-1}e^{-\frac{x}{2}}}}$ **where $\boldsymbol{\Gamma(x) = \int_0^\infty x^{n-1}e^{-x}dx}$**

Session 2_2: Distributions

Type	Name	Premise	Parameter(s)	Mean	Variance	PDF
Continuous	Student's t	• If $\boldsymbol{Z}$ is a standard normal variable and U is a chi-squared variable with $\boldsymbol{k}$ degrees of freedom, which is independent of $\boldsymbol{Z}$, then random variable $\boldsymbol{X}$: $$X = \frac{Z}{\sqrt{U/k}}$$ • follows a $\boldsymbol{t}$ distribution with $\boldsymbol{k}$ degrees of freedom • Looks like a standard normal distribution for low values of $\boldsymbol{k}$, but with excess kurtosis and converges towards standard normal as $\boldsymbol{k}$ tends to ∞ • If $\boldsymbol{X}$ is normally distributed with 0 mean and unit variance and $\boldsymbol{X^2}$ is chi-squared distributed with $\boldsymbol{X}$ and $\boldsymbol{V}$ being independent, the following follows a $\boldsymbol{t}$ distribution with $\boldsymbol{m}$ degrees of freedom: $$\frac{X}{\sqrt{V/n}}$$	• $\boldsymbol{k}$: The degrees of freedom	**0**	$\frac{k}{k-2}$ **for $k > 2$**	$\frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\,\Gamma\left(\frac{k}{2}\right)}\left(1+\frac{x^2}{k}\right)^{-\frac{k+1}{2}}$ **where $\Gamma(\mathrm{x}) = \int_0^\infty x^{n-1}e^{-x}dx$**
	F	• If $\boldsymbol{U_1}$ and $\boldsymbol{U_2}$ are independent chi-squared distributions with $\boldsymbol{k_1}$ and $\boldsymbol{k_2}$ degrees of freedom, then the following follows an F-distribution with parameters $\boldsymbol{k_1}$ and $\boldsymbol{k_2}$: $$X = \frac{(U_1/k_1)}{(U_2/k_2)}$$ • As $\boldsymbol{k_1}$ and $\boldsymbol{k_2}$ tend towards , mean and mode converge to 1, F-distribution converges to normal distribution • Square of variable with t distribution will have a F distribution • If $\boldsymbol{X}$ is random variable with t distribution with $\boldsymbol{k}$ degrees of freedom, $\boldsymbol{X^2}$ has F distribution with $\boldsymbol{1}$ and $\boldsymbol{k}$ degrees of freedom, in that order	$\boldsymbol{k}$: The degrees of freedom	$\frac{k_2}{k_2-2}$ **for $k_2 > 2$**	$\frac{2k_2^2(k_1+k_2-2)}{k_1(k_2-2)^2(k_2-4)}$ **for $k_2 > 4$**	$\frac{\sqrt{\frac{(k_1x)^{k_1}k_2^{k_2}}{(k_1x+k_2)^{k_1+k_2}}}}{xB\left(\frac{k_1}{2},\frac{k_2}{2}\right)}$ **where $B = \int_0^1 z^{x-1}(1-z)^{y-1}dz$**

Session 2_2: Distributions



Takeaways

- All the 10 distributions are parametric, being dependent on parameters that can be interpreted intuitively.
- Additionally, Student's t, χ^2, and F distributions need degrees of freedom to determine their shape.
- Probability mass function is the discrete analogue of probability density function.
- χ^2 is the sum of k squared of independent standard normal variables.
- t variable is made of a standard normal random variable and a χ^2 random variable, which are independent.
- The F random variable is the ratio of two independent χ^2 random variables.

Session 3_1: Simple Linear Regression

OLS	
OLS	$\min_{\hat{a},\hat{b}} \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{a} - \hat{b}X_i\right)^2$ $\frac{\partial \sum_{i=1}^{n} e_i^2}{\partial \hat{a}} = -2\sum_{i=1}^{n}\left(Y_i - \hat{a} - \hat{b}X_i\right) = 0$ $\frac{\partial \sum_{i=1}^{n} e_i^2}{\partial \hat{b}} = -2\sum_{i=1}^{n} X_i\left(Y_i - \hat{a} - \hat{b}X_i\right) = 0$
Important Relationships	$n\bar{Y} = n\hat{a} + n\hat{b}\bar{X}$ $\bar{Y} = \hat{a} + \hat{b}\bar{X}$ $\hat{a} = \bar{Y} - \hat{b}\bar{X}$ $\hat{b} = \frac{\sum_{i=1}^{n} X_i(Y_i - \bar{Y})}{\sum_{i=1}^{n} X_i(X_i - \bar{X})}$
Residuals	$\hat{e}_i = Y_i - \hat{Y}_i$ $\hat{\sigma}_e^2 = \frac{1}{n-2}\sum_{i=1}^{n} \hat{e}_i^2$
Variance, Standard Errors and Covariance of Estimators	$V[\hat{a}] = \hat{\sigma}_e^2\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)$ $SE[\hat{a}] = \sqrt{V[\hat{a}]}$ $V\left[\hat{b}\right] = \hat{\sigma}_e^2\left(\frac{1}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)$ $SE\left[\hat{b}\right] = \sqrt{V\left[\hat{b}\right]}$ $C\left[\hat{a},\hat{b}\right] = -\sigma_e^2\left(\frac{\bar{X}}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)$

OLS	
t-Statistics of Estimators	For $\hat{a}$: $t_{n-2} = \frac{\hat{a} - \alpha}{SE[\hat{a}]}$ For $\hat{b}$: $t_{n-2} = \frac{\hat{b} - \beta}{SE\left[\hat{b}\right]}$
Vector-Matrix From	$\hat{\beta} = (X'X)^{-1}X'y$
Sum of Squares	**TSS=ESS+RSS, thus:** $\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2 + \sum_{i=1}^{n} \hat{e}_i^2$ $ESS = \hat{b}^2\sum_{i=1}^{n}(X_i - \bar{X})^2$ $R^2 = \frac{ESS}{TSS}$
Forecasting	**Point Forecast (Unbiased):** $\hat{Y}_{n+1} = \hat{a} + \hat{b}X_{n+1}$ **Confidence Interval of Forecast (95%):** $\hat{Y}_{n+1} \pm t_{n-2,97.5\%} \times \hat{\sigma}_e\sqrt{1 + \frac{1}{n} + \frac{(X_{n+1} - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}$ **t-statistic of Forecast:** $\frac{Y_{n+1} - \hat{Y}_{n+1}}{\hat{\sigma}_e\sqrt{1 + \frac{1}{n} + \frac{(X_{n+1} - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}}$

Session 3_1: Simple Linear Regression

Takeaways

- Scatter plot gives an intuitive view of whether X could explain Y.
- Parameter estimates are obtained by minimizing the sum of squared errors.
- Each residual is the vertical distance from the data point to the OLS fitted line.
- OLS estimators are BLUE.
- Covariance divided by variance of explanatory variable = slope of OLS line.
- Variance decomposition: TSS = ESS + RSS
- R^2 of simple OLS regression = square of correlation coefficient.
- t statistic's degrees of freedom = $n - 2$.
- Many many applications!

Session 4_2: Model k and OLS

$$X'X = \begin{pmatrix} \iota' \\ x' \end{pmatrix} \begin{pmatrix} \iota & x \end{pmatrix} = \begin{pmatrix} \iota'\iota & \iota'x \\ x'\iota & x'x \end{pmatrix} = \begin{pmatrix} n & \sum x_t \\ \sum x_t & \sum x_t^2 \end{pmatrix}$$

$$(X'X)^{-1} = \frac{1}{n\sum x_t^2 - (\sum x_t)^2} \begin{pmatrix} \sum x_t^2 & -\sum x_t \\ -\sum x_t & n \end{pmatrix}$$

$$X'y = \begin{pmatrix} \iota' y \\ x' y \end{pmatrix} = \begin{pmatrix} \sum y_t \\ \sum x_t y_t \end{pmatrix}$$

1. Estimate the model by OLS: $\hat{\beta} = (X'X)^{-1}X'y$
2. Compute the fitted values of y: $\hat{y} = X\hat{\beta}$
3. Compute the residuals or "surprise": $\hat{u} = y - \hat{y}$
4. Compute the residual sum of squares (RSS)

$$\text{SSE} \equiv \text{RSS} = \hat{u}'\hat{u} = \sum_{i=1}^{n} \hat{u}_i^2$$

5. The variance of the residuals is

$$\hat{\sigma}_u^2 = \frac{1}{n-K}\hat{u}'\hat{u}$$

6. Let $\Omega := (X'X)^{-1}$. The variance of $\hat{\beta}_i$ is

$$\mathbb{V}(\hat{\beta}_i) = \hat{\sigma}_u^2 \Omega_{ii}.$$

Akaike Information Criterion (AIC)	$AIC = T\ln\left(\frac{RSS}{T}\right) + 2K$ For small sample sizes $\left(\frac{T}{K} \le 40\right)$, use 2nd order AIC: $AIC_c = AIC + \frac{2K(K+1)}{T-K-1}$ The smaller the AIC, the better the model is in not over-fitting data
Adjusted R^2	Problems with R^2 • If model is reparameterised, R^2 will change • R^2 never falls when more regressors are added Solution: Adjusted R^2 $\bar{R}^2 = 1 - \left[\frac{T-1}{T-K}(1-R^2)\right]$ $\bar{R}^2 = 1 - \frac{\frac{RSS}{n-K}}{\frac{TSS}{n-1}}$ • Takes into account the loss of degrees of freedom • $\bar{R}^2$ will fall if additional regressor does not add sufficient explanatory power • But like R^2, $\bar{R}^2$ also has no distribution

Session 5_2: Serial Correlation or Autocorrelation

Types of Autocorrelation	Description
Positive Autocorrelation	• Indicated by cyclical residual plot over time • Residuals “trend” over time
Negative Autocorrelation	• Indicated by alternating pattern • Residuals cross axis more frequently than if they were distributed randomly

Consequences of Autocorrelation
• Coefficient estimate are still unbiased but they are inefficient, they are not BLUE, even when n is large • Standard error estimates become inappropriate as such, possibility to make wrong inferences • R^2 is likely to be inflated relative to “correct” value for positively correlated residuals

Test for Autocorrelation	Assumptions	Method	Test Statistic
Durbin-Watson Test	• $E[\epsilon_t] = 0$ • $V[\epsilon_t] = \sigma_\epsilon^2$ • $C[\epsilon_t, \epsilon_s] = 0$ • $-1 \leq \phi \leq 1$ • $0 \leq DW \leq 4$	**Hypotheses** • Null Hypothesis (H_0): $\phi = 0$ • Alternate Hypothesis (H_a): $\phi \neq 0$ **Conditions for Valid Test** • Y-intercept must be in regression • Regressor X_i must be non-stochastic • Error term e_i is normally distributed **Interpretation of Results** • No Autocorrelation: $d_U < DW < 4 - d_U$ • Negative Autocorrelation: DW $< 4 - d_L$ • Positive Autocorrelation: $DW > d_L$	Where u_t denotes the residuals of regression at t: $DW = \frac{\sum_{t=2}^{T}(u_t - u_{t-1})^2}{\sum_{t=1}^{T} u_t^2}$ Expansion: $DW = \frac{\sum_{t=2}^{T} u_t^2}{\sum_{t=1}^{T} u_t^2} + \frac{\sum_{t=2}^{T} u_{t-1}^2}{\sum_{t=1}^{T} u_t^2} - 2\frac{\sum_{t=2}^{T} u_t u_{t-1}}{\sum_{t=1}^{T} u_t^2}$ $DW \approx 1 + 1 - 2\,\hat{\phi}$ • 1st and 2nd terms are approximately 1 • 3rd term is 2x estimator of correlation with itself, $2\hat{\phi}$
Breusch-Godfrey Test	-	**Tests for r^{th} order autocorrelation:** $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots + \rho_r u_{t-r} + v_t$ $v_t \sim N(0, \sigma_v^2)$ **Hypotheses** • Null Hypothesis (H_0): $\rho_1 = 0$ and $\rho_2 = 0$ and … $\rho_r = 0$ • Alternate Hypothesis (H_a): $\rho_1 \neq 0$ and $\rho_2 \neq 0$ and … $\rho_r \neq 0$ **Method** • Estimate linear regression using OLS and obtain the residuals $\hat{e}_t$ • Regress $\hat{e}_t$ on X plus $\hat{e}_{t-1}, \hat{e}_{t-2}, \ldots, \hat{e}_{t-r}$ • Obtain R^2 from regression • Get test statistic, if statistic exceeds critical value, reject null hypothesis	$(T - r)R^2 \sim \chi_r^2$

Session 6_1: Diagnostic Tests

<table>
<tr><th>Problem</th><th>Nature of Problem</th><th>Consequences of Problem</th><th>Test(s) for Problem</th><th>Test Statistic</th><th>Remedies for Problem</th></tr>
<tr>
<td rowspan="2">Heteroscedasticity</td>
<td rowspan="2">• When errors of regression do not have a constant variance
• Homoscedasticity: When errors of regression have constant variance</td>
<td rowspan="2">• Coefficient estimates are unbiased but they are no longer BLUE

• Standard errors could be inappropriate, inference we make could be misleading

• Whether standard errors are too big/small depends on form of heteroscedasticity</td>
<td>Goldfeld-Quandt (GQ) Test

• Split total sample of length T to subsamples T_1 and T_2
• Regress both subsamples and calculate the two residual variances

Hypotheses
• Null Hypothesis (H_0): $\sigma_1^2 = \sigma_2^2$

• Alternate Hypothesis (H_a): $\sigma_1^2 \neq \sigma_2^2$

• Problem with the test lies in choice of where to split the sample, which affects outcome of the test</td>
<td>$GQ = \frac{s_1^2}{s_2^2}$
$\sim F(T_1 - K, T_2 - K)$</td>
<td rowspan="2">Generalised Least Squares
• Use generalised least squares method if form of heteroscedasticity is known

• For example, if error variance is related to another variable, divide entire regression equation by that variable to remove heteroscedasticity

Transforming the Data
• Transform variables into logs or reducing by some other measure of size

Use White's Heteroscedasticity Consistent Standard Error Estimates
• Standard errors for slope coefficients are increased relative to usual OLS errors
• Goal is to be more "conservative" in hypothesis testing</td>
</tr>
<tr>
<td>White's Test

• One of the best approaches as it makes few assumptions about the form of heteroscedasticity
• Obtain residuals of regression, run an auxiliary regression (example for 2 regressors shown below):

$\hat{u}_t^2$
$= a_1 + a_2 x_{2,t} + a_3 x_{3,t} + a_4 x_{2,t}^2 + a_5 x_{3,t}^2$
$+ a_6 x_{2,t} x_{3,t} + v_t$

• Obtain R^2 from aux regression and multiply by no. of observations, T
• Reject H_0 of homoscedasticity if χ^2 test stat is greater than critical stat</td>
<td>$T\ x\ R \sim \chi^2(m)$

where m is the no. of regressors in aux regression excl. constant term</td>
</tr>
</table>

Session 6_1: Diagnostic Tests

Problem	Nature of Problem	Consequences of Problem	Test(s) for Problem	Test Statistic	Remedies for Problem
Multicollinearity	Occurs when explanatory variables are very highly correlated with each other	• All coefficients cannot be estimated if there is perfect multicollinearity • R^2 will be very high but individual coefficients will have high standard errors • Regression becomes sensitive to small changes in specification • Confidence intervals of parameters will be very wide • Significance testes may give inappropriate conclusions	• Look at correlation matrix between individual x variables • Correlation matrix method will not work if 3 or more variables are linear	-	• Drop one of the collinear variables • Transform highly correlated variables into a ratio • Collect more data – Longer time series or higher frequency **Other Remarks** • Traditional remedies such as ridge regression and/or principal component usually bring more problems than they solve • Some econometricians argue that if model is otherwise okay, ignore the problem
Specification Errors	Wrong Functional Form	-	**Ramsey's RESET Test** • Add higher order terms of fitted values to auxiliary regression • Regress $\hat{u}_t$ on powers of fitted values: $\hat{u}_t = \beta_0 + \beta_1\hat{y}_t^2 + \beta_2\hat{y}_t^3 + \cdots + \beta_{p-1}\hat{y}_t^p + v_t$ • If test stat is greater than critical stat, reject H_0 that functional form is correct	$TR^2 \sim \chi^2_{p-1}$	• RESET test gives guide as to what better specification (the true model) might be • Transform data into logarithms, which will linearise multiplicative models to additive ones
	Omission of Important Variable	• Also called the Error of Omission • Estimated coefficients on all other variables will be biased and inconsistent unless excl. variable is uncorrelated with all included variables • Even if uncorrelatedness is satisfied, coefficient estimates on constant term will be biased • Standard errors will be biased	-	-	-
	Inclusion of Irrelevant Variable	• Also called the Error of Commission • Coefficient estimates will still be consistent and unbiased • Variance of estimators will be inefficient	-	-	-

Session 6_1: Diagnostic Tests

Problem	Nature of Problem	Consequences of Problem	Test(s) for Problem	Test Statistic	Remedies for Problem
Measurement Errors	• Also known as errors in variables problem • Happens when the explanatory variable is incorrectly measured • Violates assumption that explanatory variables are non-stochastic **Ways Error can Occur** • Macroeconomic variables being estimated quantities • Using proxy variables in place of real data **Measurement Error in Explained Variable** • Much less serious than measurement errors in the explanatory variable(s) • This is so because disturbance term will be a composite of actual term and another source of noise from measurement error • Parameter estimates will still be consistent and unbiased, usual formulas for calculating standard errors will still be appropriate • However, standard errors will be enlarged relative to when there is no measurement error in y	• Composite error term will be correlated with the explanatory variable as measurement error will be incorporated into the original noise term • If coefficient is negative, parameter estimates will be biased positively • If coefficient is positive, parameter estimates will be biased negatively • As such, parameter estimates will always be biased towards zero as result of measurement noise	-	-	-

Session 6_1: Diagnostic Tests

<table>
<tr><th>Problem</th><th>Nature of Problem</th><th>Consequences of Problem</th><th>Test(s) for Problem</th><th>Test Statistic</th><th>Remedies for Problem</th></tr>
<tr><td rowspan="3">Parameter Instability</td><td rowspan="3">Occurs when parameters are not constant over sample period</td><td rowspan="3"></td><td>Split data into sub-periods then estimate up to 3 models, for each sub-part and also for all the data
Then, compare RSS of the models</td><td></td><td rowspan="3"></td></tr>
<tr><td>Chow Test
• Split data into 2 periods
• Estimate regression over whole period and for 2 sub periods
• Restricted Regression: Regression for whole period
• Unrestricted Regression: Has 2 parts, from each sub period

Null Hypothesis (H_0):
$$a_1 = a_2 , \beta_1 = \beta_2$$
• If value of test stat is greater than crit value, reject null hypothesis that parameters are stable over time

Problems
• May not have enough data to do regression on both sub-samples</td><td>$$\frac{RSS - (RSS_1 + RSS_2)}{RSS_1 + RSS_2} \times \frac{T - 2K}{K} \sim F(K, T - 2K)$$
where:
• $RSS = RSS$ for whole sample
• $RSS_1 = RSS$ for sub-sample 1
• $RSS_2 = RSS$ for sub-sample 2
• $K =$ Number of samples, which is 2
• $T =$ Sample period length</td></tr>
<tr><td>Predictive Failure Test
• Estimate regression over a long sub-period then predict values for other period and compare the two
• Restricted Regression: Run regression for whole period
• Unrestricted Regression: Run regression for long sub-period

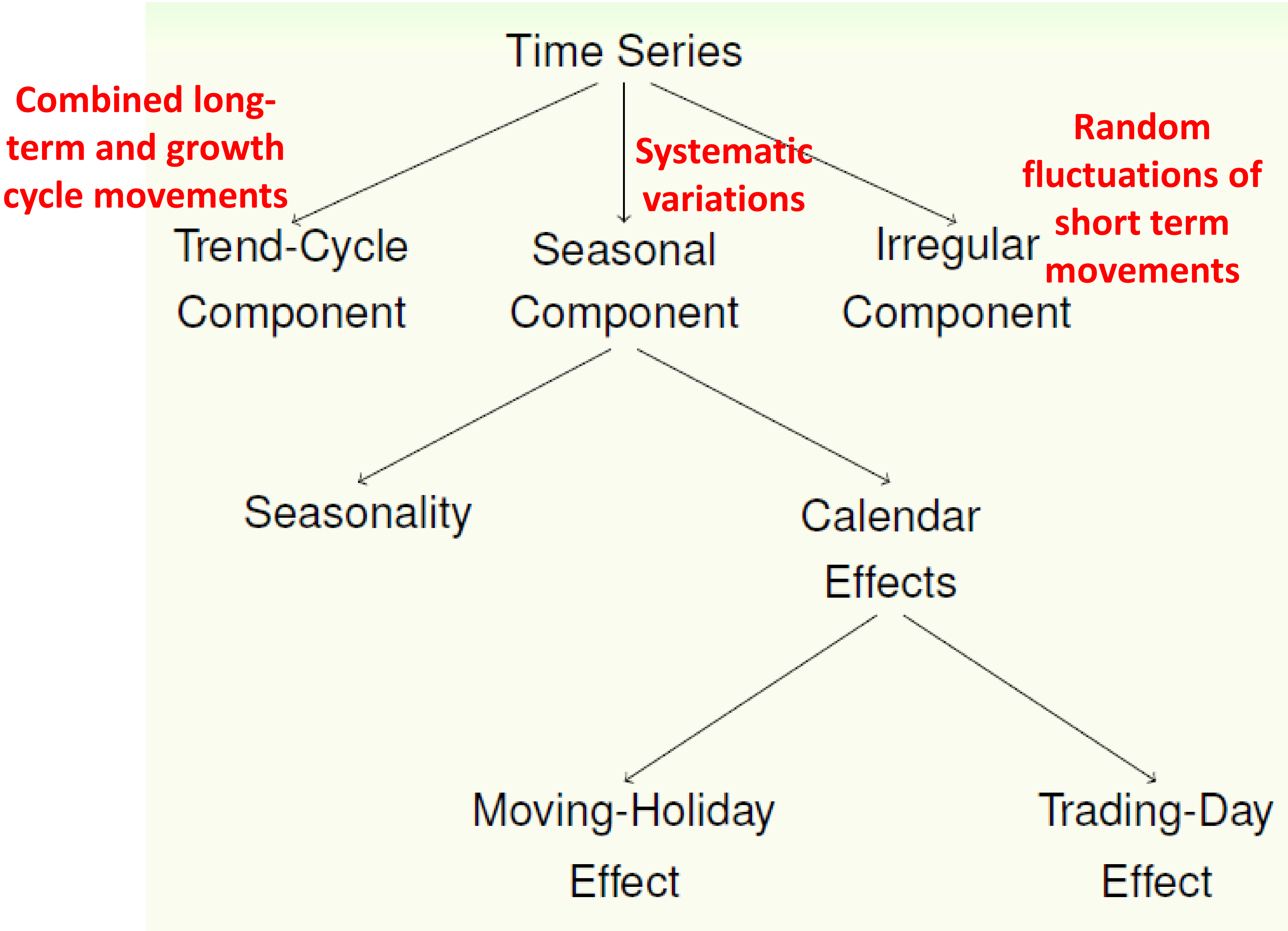
Forward Predictive Failure Test
• Where we keep the last few observations for forecast testing

Backward Predictive Failure Test
• Where we attempt to back-cast the first few observations</td><td>$$\frac{RSS - RSS_1}{RSS_1} \times \frac{T_1 - K}{T_2} \sim F(T_2, T_1 - K)$$
where:
• $RSS = RSS$ for whole sample
• $RSS_1 = RSS$ for long sub-sample
• $K =$ Number of samples, which is 2
• $T_1 =$ Sample period length for long sample
• $T_2 =$ Sample period length for predicted sample</td></tr>
</table>

Session 6_1: Diagnostic Tests

- Many things could go wrong in modeling.
- The most serious ones cause the parameter estimates to be biased and they include
  - Wrong function form
  - Omission of important variables
  - Parameter instability
  - Multicolinearity
  - Measurement errors
- Less serious ones make the variance of the estimate inaccurate and they include
  - Serial correlation
  - Heteroskedasticity
  - Departure from normality
  - Commission error of irrelevant variables
- Nevertheless, multiple linear regression is robust and is the work horse in modeling.

Session 7_1: Stationary Processes



Session 7_1: Stationary Processes

Properties of White Noise
• $E[u_t] = 0$ • $V[u_t] = \sigma_u^2$ • $C[u_t, u_{t+k}] = 0$ • Stronger Definition: u_t is independent of u_{t+k}

Joint Test Statistic	• To test if m autocorrelations are jointly 0 • $H_0: \rho(1) = \rho(2) = \cdots = \rho(m) = 0$
Box and Pierce Q-Statistic	$Q_m = T\sum_{k=1}^{m} r(k)^2 = \sum_{k=1}^{m} z_k^2 \sim \chi_m^2$
Ljung and Box Test Statistic	$Q_m = T(T+2)\sum_{k=1}^{m} \frac{r(k)^2}{T-K} \sim \chi_m^2$

Characteristic Equation	For AR(p) process to be stationary, roots of characteristic equation must lie outside of the unit circle: $\phi(B) = 1 - \sum_{i=1}^{p} \lambda_i B^i = 0$
Yule-Walker Equations	$\rho(k) = \lambda_1 \rho(k-1) + \lambda_2 \rho(k-2) + \cdots + \lambda_p \rho(k-p)$ • where k is the number of lags, p is the order of the AR(p) process • Note that $\rho(0) = 1, \rho(j) = \rho(-j)$

Stationary Process	Characteristics	Process Equation	Mean $E[Y_t]$	Variance $V[Y_t]$	Autocovariance $C[Y_t, Y_{t-k}]$	Autocorrelation $\rho[Y_t, Y_{t-k}]$ or $\rho[Y_{t+k}, Y_t]$	Conditional Mean $E[Y_{t+1} \mid Y_t]$	Conditional Variance $V[Y_{t+1} \mid Y_t]$	Test for $\rho = 0$
AR(1)	• Mean, variance and covariance are stationary • Autocorrelation slowly tapers off towards 0 as lag increases	$Y_t = \theta + \lambda Y_{t-1} + u_t$	$\frac{\theta}{1-\lambda}$	$\frac{(\sigma_u^2)}{1-\lambda^2}$	When k=1: $\lambda \frac{\sigma_u^2}{1-\lambda^2}$ Otherwise: $\gamma(k) = \lambda^k V[Y_t]$	$\lambda^{\lvert k \rvert}$	$\theta + \lambda Y_t$	$\sigma_u^2 < \frac{(\sigma_u^2)}{1-\lambda^2}$	$V[r(k)] \approx \frac{1}{T}$ Test Stat: $z_j = \frac{r(j) - 0}{\sqrt{1/T}} \sim N(0,1)$
MA(1)	• Mean, variance and covariance are stationary • Autocorrelation drops sharply to 0 after the 1st lag	$Y_t = \theta + u_t + a u_{t-1}$	θ	$(1+a^2)\sigma_u^2$	When k=1: $a\sigma_u^2$ Otherwise: 0	When k=1: $\frac{a}{1+a^2}$ Otherwise: 0	$\theta + a u_t$	$\sigma_u^2 < \sigma_u^2(1+a^2)$	$V[r(k)] \approx \frac{1}{T}(1 + 2\sum_{i=1}^{q} \rho(i)^2)$ Test Stat: $z_j = \frac{r(j) - 0}{\sqrt{V[r(k)]}} \sim N(0,1)$
ARMA(1,1)	• Mean, variance and covariance are stationary • Autocorrelation tapers off at speed between AR(1) and MA(1) process	$Y_t = \theta + \lambda Y_{t-1} + u_t + a u_{t-1}$	$\frac{\theta}{1-\lambda}$	$\sigma_u^2(1 + \frac{(\lambda + a)^2}{1-\lambda^2})$	When k=1: $\lambda V[Y_{t-k}] + a\sigma_u^2$ Otherwise: $\lambda^k V[Y_{t-k}]$	-	-	-	-

Duality between AR and MA, ACF and PACF

- While an AR(p) process has a decaying ACF infinite in extent, the PACF cuts off after lag p.
- Recall that an MA(1) process is invertible into AR(∞). In general, this property holds for MA(q) processes.
- So while the ACF of an MA(q) process cuts off after lag q, the PACF is infinite in extent.
- ARMA(p, q)'s ACF follows the same pattern as that of an AR(p) process after $q - p + 1$ initial values $\rho_0, \rho_1, \cdots, \rho_{q-p}$ (if $q - p < 0$, no initial values), while its PACF (for lag $k > p - q$) behaves like that of an MA(q) process.

Session 7_2: GARCH

GARCH	
Monitoring Daily Volatility	$u_i = \frac{S_i - S_{i-1}}{S_{i-1}}, \quad \bar{u} = 0$ $\sigma_n^2 = \frac{1}{m}\sum_{i=1}^{m} u_{n-i}^2$
Weighting Scheme	$\sigma_n^2 = \sum_{i=1}^{m} a_i u_{n-i}^2$ Where: • All weights are positive • More weight is given to more recent data • All weights sum to 1
ARCH Model	$\sigma_n^2 = \omega + \sum_{i=1}^{m} a_i u_{n-i}^2$ Where: • All weights must sum to 1: $\gamma + \sum_{i=1}^{m} a_i = 1$ • $\omega = \gamma V_L$ • V_L is the long-run variance rate
ARCH(1) Model	$V[u_t] = a_0 + a_1 u_{t-1}^2$ $u_t = e_t\sqrt{a_0 + a_1 u_{t-1}^2}$ • Where u_t is unconditionally not a normal distribution but is so when conditional on u_{t-1}
EMWA Model	• It is a special case of ARCH model, where weights a_i decrease exponentially back through time, where: $a_{i+1} = \lambda a_i$ • EMWA Model: $\sigma_n^2 = \lambda\sigma_{n-1}^2 + (1-\lambda)u_{n-1}^2$ • Thus, only current estimate of variance and most recent observation of value of market variable are needed • Relatively little data needs to be stored • Tracks volatility changes • λ is popularly set to 0.94

GARCH	
GARCH (1,1)	$\sigma_n^2 = \omega + a u_{n-1}^2 + \beta\sigma_{n-1}^2$ Where: $\omega = \gamma V_L$ $\gamma + a + \beta = 1$ • For stable GARCH (1,1), $a + \beta < 1, \quad \omega > 0$ • GARCH is unconditionally stationary, but conditionally nonstationary: $\sigma^2 = \frac{\omega}{1 - a - \beta} = V_L$ • GARCH is a mean-reverting process, σ gets pulled towards V_L over time • EWMA has no weight in V_L: $\gamma = 0$ $a = 1 - \lambda$ $\beta = \lambda$
How Good is GARCH?	• If GARCH is working well, it should remove autocorrelation in u_i^2/σ_i^2 • Can test using Ljung-Box statistic for u_i^2/σ_i^2 statistic
Variance Targeting	• Another way of increasing stability of GARCH is through variance targeting • Can set V_L to be sample variance • This way, only 2 other parameters have to be estimated

Volatility	
Volatility Term Structures	**Estimate of instantaneous variance rate in *t* days:** $V(t) = V_L + e^{-at}(V(0) - V_L)$ where: $a = \ln\left(\frac{1}{a+\beta}\right)$ note that $V(t)$ is **variance** **Can be used to calculate maintenance margin:** $x \geq 1.645 \times V(t) \times F_n$ **Volatility per annum for option lasting T days:** $\sigma(T) = \sqrt{252\left(V_L + \frac{1 - e^{-aT}}{aT}(V(0) - V_L)\right)}$ Change in T-day volatility per change in instantaneous volatility: $\Delta\sigma(T) \approx \frac{1 - e^{-aT}}{aT}\frac{\sigma(0)}{\sigma(T)}\Delta\sigma(0)$ note that $V(0) = \sigma(0)^2/252$

Session 7_2: GARCH

<table>
<tr><th colspan="2">Value-at-Risk (VaR)</th></tr>
<tr><td>Value-at-Risk (VaR)</td><td>

• Daily Log Return:

$$\tilde{r} = \ln\left(\frac{\tilde{P}_1}{P_0}\right)$$

• Return-at-Risk:

$$\tilde{r} = -Z \times \sigma$$

where σ is daily volatility, Z is the Z-value corresponding to the % VaR

• Value-at-Risk:

$$P_0 \times (1 - e^{\tilde{r}})$$

• Note that definition of VaR does not invoke assumption that distribution of returns are normal
</td></tr>
<tr><td>Relative VaR</td><td>

• Relative VaR:

$$P_0 \times (1 - e^{r})$$

where $r = \mu - \tilde{r}$
</td></tr>
</table>

Summary

- In the EWMA and the GARCH(1,1) models, the weights assigned to observations decrease exponentially as the observations become older.
- The GARCH(1,1) model differs from the EWMA model in that some weight is also assigned to the long-run average variance rate. It has a structure that enables forecasts of the future level of variance rate to be produced relatively easily.
- Maximum likelihood methods are usually used to estimate parameters from historical data in the EWMA, GARCH(1,1), and similar models.
- Once its parameters have been determined, a GARCH(1,1) model can be judged by how well it removes autocorrelation from the u_i^2.