Quantitative Economics Key Facts

R Commands

- Variable construction: `dataframe$column1 = dataframe$column2 + dataframe$column3`
- Linear regression: `lm_robust(dep ~ indep1 + indep2, data = dataframe)`
- F test: `linearHypothesis(regression, c("reg1", "reg2"), test = "F")`
- Filter: `dataframe[dataframe$column == 0]`
- Instrumental variable regression: `iv_robust(dep ~ endo + ctrl | ctrl + iv, data = dataframe)`
- Histogram: `hist(dataframe$column)`

Probability

- Kolmogorov Axioms
  - $\forall A \subseteq \Omega : \mathbb{P}A \geq 0$ (i.e. the probability of any event is non-negative),
  - $\mathbb{P}\Omega = 1$ (i.e. the probability of the event Ω is one), and
  - $\mathbb{P}\bigcup_{i=1}^{m} A_i = \sum_{i=1}^{m} \mathbb{P}A_i$ iff $\forall i \neq j : A_i \cap A_j = \emptyset$ (i.e. iff the events $A_1, \ldots, A_m$ are mutually exclusive, then the probability of the event $\bigcup_{i=1}^{m} A_i$ is equal to the sum of the probabilities of the events $A_1, \ldots, A_m$).
- Probability Independence
  - $A \perp\!\!\!\perp B$ iff $\mathbb{P}(A \cap B) = \mathbb{P}A\mathbb{P}B$, $\mathbb{P}(A|B) = \mathbb{P}A$, or $\mathbb{P}(B|A) = \mathbb{P}B$.
- Distribution Functions
  - Single Variable
    - CDF $F_X(x) := \mathbb{P}\{X \leq x\}$. This is fundamental.
    - PMF $f_X(x) := \mathbb{P}\{X = x\}$ for discrete X.
    - PDF $f_X(x) : F_X(x) = \int_{-\infty}^{x} f_X(u)du$ for continuous X.
  - Multiple Variables
    - Joint CDF $F_{X,Y}(x,y) := \mathbb{P}\{X \leq x, Y \leq y\}$. This is fundamental.
  - Discrete Random Variables
    - Joint PMF $f_{X,Y}(x,y) := \mathbb{P}\{X = x, Y = y\}$.
    - Marginal PMF $f_X(x) = \sum_{y \in \mathcal{Y}} f_{X,Y}(x,y)$.
    - Marginal CDF $F_X(x) = \sum_{u \leq x, u \in \mathcal{X}} f_X(u)$.
    - Conditional PMF $f_{X|Y}(x|y) := \mathbb{P}(X = x|Y = y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$.
    - Conditional CDF $F_{X|Y}(x|y) := \mathbb{P}\{X \leq x|Y = y\} = \sum_{u \leq x, u \in \mathcal{X}} f_{X|Y}(u|y)$.
  - Continuous Random Variables
    - Joint PDF $f_{X,Y}(x,y) : F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)dvdu$.
    - Marginal PDF $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$.
    - Marginal CDF $F_X(x) = \int_{-\infty}^{x} f_X(u)du$.
    - Conditional PDF $f_{X|Y}(x|y) := \frac{f_{X,Y}(x,y)}{f_X(x)}$.
    - Conditional CDF $F_{X|Y}(x|y) := \mathbb{P}\{X \leq x|Y = y\} = \int_{-\infty}^{x} f_{X|Y}(u|y)du$.
- Random Variable Independence
  - $X \perp\!\!\!\perp Y$ iff for all x, y, $\mathbb{P}\{X \leq x, Y \leq y\} = \mathbb{P}\{X \leq x\}\mathbb{P}\{Y \leq y\}$.
- Moments
  - $\mathbb{E}X := \begin{cases} \sum_{x \in \mathcal{X}} x f_X(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x f_X(x)dx & \text{if } X \text{ is continuous} \end{cases}$.
  - $var(X) := \mathbb{E}(X - \mathbb{E}X)^2 = \ldots = \mathbb{E}X^2 - (\mathbb{E}X)^2$.
  - $sd(X) := \sqrt{var(X)}$.
  - $cov(X,Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y) = \mathbb{E}XY - (\mathbb{E}X)(\mathbb{E}Y)$.
  - $corr(X,Y) := \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}} = \frac{cov(X,Y)}{sd(X)sd(Y)}$.
  - $\mathbb{E}[X|Y = y] := \begin{cases} \sum_{x \in \mathcal{X}} x f_{X|Y}(x|y) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx & \text{if } X \text{ is continuous} \end{cases}$.
  - $\mathbb{E}[X|Y] := \mu(Y)$, where $\mu(y) := \mathbb{E}[X|Y = y]$.

Statistics

Level of Significance α	Critical Value c_α for Two-Sided Test	Critical Value c_α for One-Sided Test
1%	$2\phi(-c_\alpha) = \alpha \Rightarrow c_\alpha = 2.576$	$\phi(-c_\alpha) = \alpha \Rightarrow c_\alpha = 2.326$
5%	$2\phi(-c_\alpha) = \alpha \Rightarrow c_\alpha = 1.960$	$\phi(-c_\alpha) = \alpha \Rightarrow c_\alpha = 1.645$
10%	$2\phi(-c_\alpha) = \alpha \Rightarrow c_\alpha = 1.645$	$\phi(-c_\alpha) = \alpha \Rightarrow c_\alpha = 1.282$

Parameter of Interest	Estimator	Standard Error Computation	Test Statistic Computation
Mean	$\overline{X}_n$	$se(\overline{X}_n) = \sqrt{\frac{s_{n,X}^2}{n}}$ (The intuition for this is that the sample variance is a consistent estimate of the population variance, the variance of the sample mean is the population variance divided by n.)	$t_n(\mu_0) = \frac{\overline{X}_n - \mu_0}{se(\overline{X}_n)}$
Difference of Means in One Sample	$\overline{D}_n := \hat{\mathbb{E}}(X - Y)$ $= \overline{X}_n - \overline{Y}_n$	$se(\overline{D}_n) = \sqrt{\frac{s_{n,D}^2}{n}}$ (This is exactly analogous to the above. The difference is treated as a variable.)	$t_n(\delta_0) = \frac{\overline{D}_n - \delta_0}{se(\overline{D}_n)}$ (This is again exactly analogous to the above)
Difference of Means in Two Independent Samples	$\hat{\delta} = \overline{X}_{1,n_1} - \overline{X}_{2,n_2}$	$se(\hat{\delta}) = \sqrt{\frac{s_{1,n_1}^2}{n_1} + \frac{s_{2,n_2}^2}{n_2}}$ (This is an estimator of $sd(\hat{\delta}) = \sqrt{var(\hat{\delta})}$ $= \sqrt{var(\overline{X}_{1,n_1}) + var(\overline{X}_{2,n_2})}$, where the second equality follows by the assumption of independent samples.)	$t_n(\delta_0) = \frac{\hat{\delta} - \delta_0}{se(\hat{\delta})}$
Proportion	$\hat{p}_n$	$se(\hat{p}_n) = \sqrt{\frac{\hat{p}_n(1 - \hat{p}_n)}{n}}$ (The numerator within the square root is a consistent estimator of the population variance, then the fraction within the square root is a consistent estimator of the variance of the sample mean.)	$t_n(p_0) = \frac{\hat{p}_n - p_0}{se(\hat{p}_n)}$
Difference of Proportions in One Sample	$\hat{d}_n := \hat{p}_n - \hat{q}_n$ (This is exactly analogous to the case of difference of means in one sample.)	$se(\hat{d}_n) = \sqrt{\frac{s_{n,d}^2}{n}}$	$t_n(\delta_0) = \frac{\hat{d}_n - \delta_0}{se(\hat{d}_n)}$
Difference of Proportions in Two Independent Samples	$\hat{\delta} := \hat{p}_{1,n_1} - \hat{p}_{2,n_2}$	$se(\hat{\delta}) = \sqrt{\frac{\hat{\pi}(1 - \hat{\pi})}{n_1} + \frac{\hat{\pi}(1 - \hat{\pi})}{n_2}}$ where $\hat{\pi} = \frac{n_1}{n_1 + n_2}\hat{p}_{1,n_1} + \frac{n_2}{n_1 + n_2}\hat{p}_{2,n_2}$ (Under the null, the two proportions are equal, and the weighted average of the two proportions is a consistent estimate of the common proportion.)	$t_n(\delta_0) = \frac{\hat{\delta} - \delta_0}{se(\hat{\delta})}$

- Sample Estimators
  - Sample mean $\overline{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$.
  - Sample variance $s_{n,X}^2 \equiv \hat{var}(X) := \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$.
  - Sample covariance $s_{n,XY} \equiv \hat{cov}(X,Y) := \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)(Y_i - \overline{Y}_n) = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)(Y_i - \overline{Y}_n)$.
  - Sample standard deviation $\hat{sd}(X) = \sqrt{\hat{var}(X)}$.
  - Sample correlation $\hat{corr}(X,Y) = \frac{\hat{cov}(X,Y)}{\sqrt{\hat{var}(X)\hat{var}(Y)}}$.
  - Naive analogue estimator of population variance $\hat{\sigma}_{n,X}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$.
  - Naive analogue estimator of population covariance $\hat{\sigma}_{n,XY} = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)(Y_i - \overline{Y}_n)$.
- Estimator Properties
  - $bias(\hat{\theta}_n) := \mathbb{E}\hat{\theta}_n - \theta$. An unbiased estimator has bias 0.
  - Precision is inversely related to the magnitude of the variance of an estimator (due to sampling variation). It is measured by standard error.
    - Variance of sample mean $var(\overline{X}_n) = \frac{\sigma_X^2}{n}$.

- Standard deviation of sample mean $sd(X) = \sqrt{var(X)}$.
- Standard error of sample mean $se(\overline{X}_n) = \frac{s_{n,X}}{\sqrt{n}}$ is the estimable counterpart of the standard deviation of sample mean.
- Mean-squared error $mse(\hat{\theta}_n) := \mathbb{E}(\hat{\theta}_n - \theta)^2$.
- A sequence of estimators $\{\hat{\theta}_i\}_{i=1}^n = \{\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n\}$ is consistent for the population parameter (estimand) θ iff $\hat{\theta}_n \to^p \theta$.
- Law of Large Numbers
  - If the sequence of random variables $\{X_i\}_{i=1}^n$ is independently and identically distributed, and (for all i) $\mathbb{E}X_i = \mu$, where μ is some constant, then the sample mean $\overline{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$ converges in probability to μ (i.e. $\overline{X}_n \to^p \mu$).
- Central Limit Theorem
  - If the sequence of random variables $\{X_i\}_{i=1}^n$ is independently and identically distributed, and (for all i) $\mathbb{E}X_i = \mu$ and $var(X_i) = \sigma^2$, then the standardised sample mean $Z_n := \frac{\overline{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}}$ converges in distribution to $N[0,1]$.
- Hypothesis Tests
  - Null and alternative hypotheses, test statistic, distribution of the test statistic under the null, decision rule, critical value, conclusion.
- Confidence Intervals
  - $\mathcal{C}_n = \{\theta_0 \in \mathbb{R} : |\frac{\hat{\theta}_n - \theta_0}{se(\hat{\theta}_n)}| \le c_\alpha\} == [\hat{\theta}_n - c_\alpha se(\hat{\theta}_n), \hat{\theta}_n + c_\alpha se(\hat{\theta}_n)]$
- p- values
  - State the associated hypothesis test up to the decision rule when computing p-values.

Linear Regression Mechanics

- Descriptive Interpretation
  - "On average, an [object's] having one unit higher [independent variable] is associated with having [coefficient magnitude] [higher/lower] [dependent variable], holding each other regressor constant."
- Population Regression
  - $Y = \rho_0 + \rho_1 X + e$, where (equivalently)
  - $\rho_0, \rho_1 := \arg\min_{b_0, b_1} \mathbb{E}(Y - b_0 - b_1 X)^2$,
  - $\rho_0 = \mathbb{E}Y - \rho_1 \mathbb{E}X, \rho_1 = \frac{cov(X,Y)}{var(X)}$,
  - $\mathbb{E}e = \mathbb{E}Xe = 0$.
  - Where orthogonality holds in the causal model, the population regression coincides with the causal model.
- Sample Regression
  - $Y = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{u}$, where (equivalently)
  - $\hat{\beta}_0, \hat{\beta}_1 := \arg\min_{b_0, b_1} \hat{\mathbb{E}}(Y - b_0 - b_1 X)^2$,
  - $\hat{\beta}_0 = \hat{\mathbb{E}}Y - \hat{\beta}_1 \hat{\mathbb{E}}X, \hat{\beta}_1 = \frac{\hat{cov}(X,Y)}{\hat{var}(X)}$, and
  - $\hat{\mathbb{E}}\hat{u} = \hat{\mathbb{E}}X\hat{u} = 0$.
  - Sample regression coefficients are consistent for population regression coefficients, by the consistency of $\hat{\mathbb{E}}, \hat{cov}, \hat{var}$ for their population counterparts.
- Frisch-Waugh-Lovell Theorem
  - Causal model: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$.
  - Auxiliary population regression: $X_1 = \pi_0 + \pi_2 X_2 + \tilde{X}_1$.
  - By substitution of the auxiliary population regression into the causal model, $Y = (\beta_0 + \beta_1 \pi_0) + \beta_1 \tilde{X}_1 + (\beta_2 + \beta_1 \pi_2)X_2 + u$.
  - $cov(Y, \tilde{X}_1) = \beta_1 var(\tilde{X}_1) \Rightarrow \beta_1 = \frac{cov(Y, \tilde{X}_1)}{var(\tilde{X}_1)}$.
  - Supposing orthogonality holds in the causal model, β_1 can be consistently estimated by the corresponding two-step procedure.
- Conditional Expectation
  - $\mathbb{E}[Y|X] = \arg\min_m \mathbb{E}[Y - m(X)]^2$.
  - If the conditional expectation is linear, then it is the best linear predictor of Y given X, so it coincides with the population linear regression.
  - If errors are mean independent of the regressors, then the conditional expectation is linear.
- Perfect Multicollinearity

- Where some regressors are perfectly multicollinear, (1) the corresponding causal model is not accurate to the reality that it is a model for because it makes no sense to have one of these regressors vary while the others are constant, (2) there is no unique solution to the population regression problem, and the residual of each auxiliary regression is zero and has zero variance, so the population regression parameters are undefined, (3) likewise for the sample regression.
- Perfect multicollinearity could be observed in the data even if the remaining regressors are not, in theory, perfectly multicollinear, because of limitations in the data.

Linear Regression Diagnostics and Inference

- Regression Diagnostics
  - Standard Error of Regression
    - The standard error of regression is an estimate of the standard deviation of the population regression residual. This is an absolute measure of fit.
    - Population variance of population regression residual $\sigma_u^2 := var(u)$.
    - Sample variance of sample regression residual $s_{\hat{u}}^2 := \frac{n}{n-k-1}\hat{\mathbb{E}}(\hat{u} - \hat{\mathbb{E}}\hat{u})^2 = \frac{n}{n-k-1}\hat{\mathbb{E}}\hat{u}^2$.
    - Standard error of regression $s_{\hat{u}} := \sqrt{s_{\hat{u}}^2}$.
  - R^2
    - The R^2 of a regression is the proportion of the total variability in the dependent variable attributable to variation in the regressors. This is a relative measure of fit.
    - $R^2 := \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS}$.
    - $s_{\hat{u}} = \left(\frac{SSR}{n-k-1}\right)^{\frac{1}{2}}$.
  - Adjusted R^2
    - $\overline{R}^2 := 1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS}$.
    - This adjustment can be understood as applying two simultaneous adjustments, one to TSS and the other to SSR. TSS is adjusted by some analogue of Bessel's correction, and SSR is adjusted by some analogue of the degrees of freedom adjustment. Then, the adjusted TSS is an unbiased estimator of $var(Y)$ (in the population) and the adjusted SSR is an unbiased estimator of $var(u)$ (in the population).
- Inference on Regression Parameters
  - Conditions for Valid Inference
    - The causal model satisfies orthogonality.
    - iid random sample.
    - Finite kurtosis.
    - No perfect multicollinearity.
  - t Tests
    - Homoskedastic standard error $se(\hat{\beta}_I) = n^{-\frac{1}{2}}\sqrt{\frac{\hat{var}(\hat{u})}{\hat{var}(\hat{X}_1)}}$. This is $n^{-\frac{1}{2}}$ times the square root of the sample variance of $\hat{u}$ divided by the sample variance of $\hat{\hat{X}}_1$.
    - Heteroskedasticity-robust standard error $se(\hat{\beta}_I) = n^{-\frac{1}{2}}\sqrt{\frac{\hat{var}(\hat{\hat{X}}_I \hat{u})}{[\hat{var}(\hat{X}_I)^2]}}$.
  - F Tests
    - $F = \frac{n-k-1}{q}\frac{SSR_r - SSR_u}{SSR_u}$.
- Model Selection
  - Omission of a control has an ambiguous effect on the precision (measured by standard error) of the remaining regressors because the variance of the sample regression residuals decreases, but so does the variance of the auxiliary regression residuals.
  - An overfit model is too complex or flexible relative to the complexity and volume of data, it will be excessively sensitive to noise. Such a model has low standard error and high R^2, but yields less accurate predictions on out-of-sample data.
  - An underfit model is not complex or flexible enough to capture the underlying dynamics of the statistical process that generated the observations. Such a model will have high standard error and low R^2 and yield less accurate predictions on out-of-sample data.

Endogeneity

- Omitted Variables
  - Long causal model: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$.
  - Short population regression: $Y = \gamma_0 + \gamma_1 X_1 + e$.
  - $\gamma_1 = \frac{cov(Y,X_1)}{var(X_1)} = \beta_1 + \beta_2 \frac{cov(X_1,X_2)}{var(X_1)} = \beta_1 + \beta_2 \pi_1$.
  - Auxiliary population regression: $X_2 = \pi_0 + \pi_1 X_1 + v$.
- Proxying for Omitted Variables
  - Long causal model: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$.
  - Auxiliary population regression: $X_2 = \pi_0 + \pi_1 X_1 + \pi_2 W_2 + v$.
  - By substitution, estimable model: $Y = (\beta_0 + \beta_2 \pi_0) + (\beta_1 + \beta_2 \pi_1)X_1 + \beta_2 \pi_2 W_2 + (u + \beta_2 v)$.
  - Suppose (1) the proxy is correlated with the regressor it proxies for, then $\pi_2 \neq 0$.
  - Suppose (2) the proxy is uncorrelated with unmodelled determinants, then $cov(W_2, u) = 0$ so $cov(W_2, u + \beta_2 v) = 0$.
  - Suppose (3) the prediction of the proxied regressor is not improved by knowledge of the modelled regressors, then $\pi_1 = 0$.
  - Then orthogonality holds in the estimable model, its parameters coincide with population regression parameters, and it can be consistently estimated by OLS.
- Measurement Error in the Dependent Variable
  - Causal model: $Y = \beta_0 + \beta_1 X + u$.
  - $Y^* = Y + e_y$.
  - By substitution, estimable model: $Y^* = \beta_0 + \beta_1 X + (u + e_y)$.
  - Suppose that e_y is uncorrelated with X, then orthogonality holds in the estimable model, its parameters coincide with the population regression parameters, and it can be consistently estimated by OLS.
- Measurement Error in an Independent Variable
  - Causal model: $Y = \beta_0 + \beta_1 X + u$.
  - $X^* = X + e_x$.
  - By substitution, estimable model: $Y = \beta_0 + \beta_1 X^* + \epsilon$, where $\epsilon = u - \beta_1 e_x$.
  - Orthogonality fails in the estimable model because e_x hence ϵ is correlated with X^*.
  -
$$
\begin{aligned}
\beta_1^* &= \frac{cov(Y, X^*)}{var(X^*)} \\
&= \frac{cov(\beta_0 + \beta_1 X^* + \epsilon, X^*)}{var(X^*)} \\
&= \frac{\beta_1 var(X^*) + cov(\epsilon, X^*)}{var(X^*)} \\
&= \frac{\beta_1 var(X^*) + cov(u - \beta_1 e_x, X + e_x)}{var(X^*)}. \\
&= \frac{\beta_1 var(X^*) - \beta_1 var(e_x)}{var(X^*)} \\
&= \beta_1 \frac{var(X^*) - var(e_x)}{var(X^*)} \\
&= \beta_1 \frac{var(X)}{var(X) + var(e_x)}
\end{aligned}
$$
- Simultaneity
  - The failure of orthogonality under simultaneity can be proved by equating the two given causal models.

Instrumental Variables

- Instrument Validity
  - Relevance: $cov(X, Z) \neq 0$, i.e. Z is correlated with X.
  - Exogeneity: $cov(Z, u) = 0$, i.e. Z is uncorrelated with the unmodeled determinants of Y in the causal model above.
  - Exclusion: Z does not appear in the causal model above.
- Indirect Least Squares
  - Structural equation: $Y = \beta_0 + \beta_1 X + u$.
  - First stage regression: $X = \pi_0 + \pi_1 Z + v$.
  - By substitution, reduced form regression: $Y = \gamma_0 + \gamma_1 Z + \epsilon$, where $\gamma_0 = \beta_0 + \beta_1 \pi_0$, $\gamma_1 = \beta_1 \pi_1$, and $\epsilon = 0 + \beta_1 v$.
  - By exogeneity and by construction of the first stage regression, orthogonality holds in the estimable model, its parameters coincide with the population regression parameters, and are consistently estimated by OLS.
  - Then, β_1 is consistently estimated by $\frac{\hat{\gamma}_1}{\hat{\pi}_1}$.

- Where Z is endogenous (i.e. exogeneity fails), the bias of $\hat{\gamma}_1$ is inherited by $\hat{\beta}_1$.
- Two-Stage Least Squares
  - Structural equation: $Y = \beta_0 + \beta_1 X + u$.
  - First stage regression: $X = \pi_0 + \pi_1 Z + v$.
  - By substitution, second stage regression: $Y = \beta_0 + \beta_1 X^* + \epsilon$, where $X^* = \pi_0 + \pi_1 Z$ and $\epsilon = u + \beta_1 v$.
  - By exogeneity and by construction of the first stage regression, orthogonality holds in the estimable model, its parameters coincide with the population regression parameters, and are consistently estimated by OLS.
- Inference on Regression Parameters
  - Conditions for Valid Inference
    - Each control satisfies orthogonality.
    - iid random sample.
    - Finite kurtosis.
    - No perfect multicollinearity (among the controls in the structural equation, and among the instruments in the first stage regression).
    - Instruments are valid.
  - Standard Error
    - Homoskedastic standard error: $se(\hat{\beta}_1) = n^{-\frac{1}{2}} \sqrt{\frac{v\hat{a}r(\hat{u})}{v\hat{a}r(\tilde{X}^*)}}$.
    - Standard errors are exactly analogous to the OLS case, except in replacing $var(X)$ in the denominator with $var(X^*)$ (or, in the case of multiple regression, $var(\tilde{X})$ with $var(\tilde{X}^*)$).
- Test of Relevance
  - Regular F test of
  - $H_0 : \forall I \in \{1, \ldots, m\} : \pi_I = 0$, against
  - $H_1 : \exists I \in \{1, \ldots, m\} : \pi_I \neq 0$, in
  - $X = \pi_0 + \pi_1 Z_1 + \ldots + \pi_m Z_m + \pi_{m+1} W_1 + \ldots + \pi_{m+r} W_r + v$.
  - A rejection of the null hypothesis in this test can be unambiguously interpreted as evidence in favour of Relevance.
  - Relevance is sufficient for consistent estimation of the causal effects of interest but not sufficient for standard inferences. If each π_I is small, the familiar convergence in distribution to the standard normal of the 2SLS estimators does not hold, even in large samples. This is the phenomenon of weak instruments. As a rule of thumb, instruments are weak if the F statistic from the above test is less than 10.
- Test of Exogeneity
  - Adjusted F test of
  - $H_0 : \forall I \in \{1, \ldots, m\} : \delta_I = 0$, against
  - $H_1 : \exists I \in \{1, \ldots, m\} : \delta_I \neq 0$, in
  - $u = \delta_0 + \delta_1 Z_1 + \ldots + \delta_m Z_m + \delta_{m+1} W_1 + \ldots + \delta_{m+r} W_r + \eta$.
  - $F^* = \frac{m}{m-1} F$
  - Compare F^* against critical values drawn from the $F_{m-1,\infty}$ distribution.
  - Estimate the structural equation by 2SLS.
  - Compute the estimated residuals.
  - Estimate the regression of the estimated residuals on the instruments and controls by 2SLS.
  - A failure to reject the null does not constitute unambiguous evidence for Exogeneity because the test of overidentifying restrictions is implicitly a test of the consistency between multiple just-identified models. It is possible that none of the instruments are exogenous but the models they estimate are consistent.

Randomised Control Trials

- Binary Treatment
  - Where treatment is binary, the OLS estimate of the causal effect of treatment coincides with the difference in (sample) mean outcomes.
  - In the population, $\mathbb{E}[Y|D] = \begin{cases} \mathbb{E}[Y|D=0] & \text{if } D = 0 \\ \mathbb{E}[Y|D=1] & \text{if } D = 1 \end{cases}$
    $= \mathbb{E}[Y|D=0] + \{\mathbb{E}[Y|D=1] - \mathbb{E}[Y|D=0]\}D$
  - This is the best predictor of Y given D, hence the best linear predictor, so the population regression parameters coincide with these parameters.
- Endogenous Controls

- An endogenous control is a control that is causally determined by treatment. The practice of including an endogenous control is "over controlling".
  - Orthogonality in the estimable model that includes an endogenous control because the endogenous control is likely correlated with unobserved determinants of the dependent variable.
- Heterogenous Causal Effects
  - Where treatment is randomly assigned, OLS consistently estimates the average causal effect (average treatment effect in the case of binary treatment).
  - In the population,
$$\mathbb{E}[Y_i|X_i] = \mathbb{E}[\beta_0 + \beta_{1i}X_i + u_i|X_i]$$
$$= \beta_0 + X_i\mathbb{E}[\beta_{1i}|X_i] + \mathbb{E}[u_i|X_i].$$
$$= \beta_0 + (\mathbb{E}\beta_{1i})X_i$$
- Instrumental Treatment
  - Structural equation: $Y_i = \beta_0 + \beta_{1i}X_i + u_i$.
  - First stage regression: $X_i = \pi_0 + \pi_{1i}Z_i + v_i$.
  - Exogeneity is guaranteed by successful random assignment.
  - 2SLS estimator of the coefficient on X is consistent for LATE $\mathbb{E}(\beta_{1i} \times \frac{\pi_{1i}}{\mathbb{E}\pi_{1i}})$.
- Internal Validity
  - A study is internally valid iff the estimates it yields are consistent for the causal effects of interest in the study population.
- External Validity
  - A study is externally valid iff its findings can be reliably extrapolated to the real world population or policy of interest. The threats to external validity are heterogeneity between the study population and the population of policy interest, potential spillover effects (which are when the assumption of individualistic treatment response fails), and the failure of surrogate outcomes.

Stationary Time Series

Concept	Series $\{Y_t\}$	Random Variable Y_t
Stationarity	Stationarity is a property of series.	Stationarity is not a property of random variables. If Y_t, Y_s belong to a stationary series $\{Y_t\}$, then the two are in some sense similarly distributed.
Autoregressive Model $AR(p)$	A series is said to follow an autoregressive model $AR(p)$ iff the random variables of that series are distributed according to $AR(p)$. Then, it is also said that the series is a $AR(p)$ process, and that it is generated by an $AR(p)$ model.	An autoregressive model $AR(p)$ describes the distribution of random variables.
Granger Causality	Granger causality is a relation that holds between two (or more) series.	
Unit Root	If the random variables that a series consists in have a unit root (and are distributed according to an $AR(p)$ model) then the series is a unit root autoregressive process, and follows a unit root autoregressive model.	Having a unit root is a property of random variables (that are distributed according to an $AR(p)$ model).
Order of Integration	Order of integration is a property of series.	Number of roots is a property of random variables (and is numerically equivalent to the order of integration of the series that such random variables constitute). $Y_t \sim I(d)$ denotes that Y_t has number of roots d.
Cointegration	Cointegration is not a property of series.	Cointegration is a property of random variables.

- Stationarity
  - Time series $\{Y_t\}$ is weakly stationary iff the mean (of each observation Y_t) $\mathbb{E}Y_t$, variance $var(Y_t)$, and autocovariances $cov(Y_t, Y_{t-h})$ for all $h \in \mathbb{Z}$ are time-invariant, i.e. independent of t.
  - Two time series $\{Y_t\}, \{X_t\}$ are jointly weakly stationary iff their means, variances, autocovariances, and cross covariances $cov(Y_t, X_{t-h})$ and $cov(X_t, Y_{t-h})$ are time-invariant.

- Sample Mean
  - $\overline{Y}_T := \frac{1}{T} \sum_{t=1}^{T} Y_t$.
- Sample Autocovariance
  - $\hat{\gamma}_h := c\hat{o}v(Y_t, Y_{t-h}) := \frac{1}{T} \sum_{t=h+1}^{T} (Y_t - \overline{Y}_T)(Y_{t-h} - \overline{Y}_T)$.
  - Note that the sample autocovariance is computed from $T - h$ observations of (Y_t, Y_{t-h}), but that the denominator in the computation is T rather than $T - h$.
- Sample Variance
  - $\hat{\gamma}_0 = \frac{1}{T} \sum_{t=1}^{T} (Y_t - \overline{Y}_T)^2 =: v\hat{a}r(Y_t)$.
- Sample Autocorrelation
  - $\hat{\rho}_h := \frac{c\hat{o}v(Y_t, Y_{t-h})}{v\hat{a}r(Y_t)} = \frac{\hat{\gamma}_h}{\hat{\gamma}_0}$.
- Autoregressive Models
  - $Y_t = \beta_0 + \sum_{i=1}^{p} \beta_i Y_{t-i} + u_t$, where $\mathbb{E}[u_t | \mathcal{Y}_{t-1}] = 0$.
  - Stationarity
    - $\beta_1 \in (-1, 1)$, and $\mathbb{E}Y_0 = \frac{\beta_0}{1-\beta_1}$ and $var(Y_0) = \frac{\sigma_u^2}{1-\beta_1^2}$ are collectively sufficient for stationarity. The former condition is the substantive condition, whereas the latter two are understood as a mathematical technicality.
  - Forecasting
    - The optimal 1-step ahead forecast of time series $\{Y_t\}$ is $Y_{T+1|T} := \mathbb{E}[Y_{T+1} | \mathcal{Y}_T]$.
    - The optimal 1-step ahead forecast of an $AR(p)$ process $\{Y_t\}$ is
    $$\begin{aligned} Y_{T+1|T} &:= \mathbb{E}[Y_{T+1} | \mathcal{Y}_T] \\ &= \mathbb{E}[\beta_0 + \sum_{i=1}^{p} \beta_i Y_{T+1-i} + u_T | \mathcal{Y}_T] \\ &= \beta_0 + \sum_{i=1}^{p} \beta_i Y_{T+1-i} \end{aligned}.$$
- Estimation of Autoregressive Model
  - Every AR model, by construction, satisfies orthogonality, and so is estimable by OLS regression.
- Estimation of Forecast
  - The estimated optimal 1-step ahead forecast of an $AR(p)$ process $\{Y_t\}$ is $\hat{Y}_{T+1|T} := \hat{\beta}_0 + \sum_{i=1}^{p} \hat{\beta}_i Y_{t+1-i}$, where $\hat{\beta}_0, \ldots, \hat{\beta}_p$ are the OLS estimators for $\beta_0, \ldots, \beta_p$.
  - The error of the estimated optimal 1-step ahead forecast is decomposed into (1) the unforecastable shock component u_{T+1}, and (2) the error introduced by OLS estimation of the model parameters. The two are uncorrelated because the latter is a function of past observations, and the former is uncorrelated with past observations.
  - $$\begin{aligned} \hat{e}_{T+1|T} &:= Y_{T+1} - \hat{Y}_{T+1|T} \\ &= (Y_{T+1} - Y_{T+1|T}) + (Y_{T+1|T} - \hat{Y}_{T+1|T}) \\ &= u_{T+1} + (Y_{T+1|T} - \hat{Y}_{T+1|T}) \\ &= u_{T+1} + [\beta_0 + \sum_{i=1}^{p} \beta_i Y_{T+1-i}] - [\hat{\beta}_0 + \sum_{i=1}^{p} \hat{\beta}_i Y_{T+1-i}] \\ &= u_{T+1} + [(\beta_0 - \hat{\beta}_0) + \sum_{i=1}^{p} (\beta_i - \hat{\beta}_i) Y_{T+1-i}] \end{aligned}$$
  - $$\begin{aligned} MSFE(\hat{Y}_{T+1|T}) &:= \mathbb{E}\hat{e}_{T+1|T}^2 \\ &= \mathbb{E}u_{T+1}^2 + \mathbb{E}(\hat{Y}_{T+1|T} - Y_{T+1|T})^2 \\ &= var(u) + \mathbb{E}(\hat{Y}_{T+1|T} - Y_{T+1|T})^2 \end{aligned}$$
- Forecast Evaluation
  - Suppose that the performance of a forecasting procedure is evaluated by its MSFE. Then, because Y_{T+1} is not known, the MSFE cannot be directly computed and must be estimated.
  - The general difficulty with such estimation is that the residual $\hat{e}_{T+1|T} := Y_{T+1} - \hat{Y}_{T+1|T}$ is an out-of-sample residual and not an in-sample residual, unlike cross-sectional residuals, which are in-sample. Then, the variance of the in-sample residual $\hat{u}_t = Y_t - \hat{Y}_t$, by the above decomposition, systematically underestimates the MSFE.
  - Select some $P \in [0.1T, 0.2T]$. For each $s \in \{T - P, \ldots, T - 1\}$, compute the pseudo out-of-sample forecast error $\hat{e}_{s+1|s} := Y_{s+1} - \hat{Y}_{s+1|s}$, where $\hat{Y}_{s+1|s}$ is the estimated optimal forecast of Y_{s+1}. Then, the estimator for $MSFE(\hat{Y}_{T+1|T})$ is $M\hat{S}FE(\hat{Y}_{T+1|T}) := \frac{1}{P} \sum_{s=T-P}^{T-1} \hat{e}_{s+1|s}^2$, i.e. the average of the squared pseudo out-of-sample forecast errors.
- Model Selection
  - $AIC_m := ln\frac{SSR_m}{T} + m\frac{2}{T}$.

- $BIC_m := ln \frac{SSR_m}{T} + m\frac{lnT}{T}$.
- Where m is the number of parameters, including the constant, so $m = p + 1$ for an $AR(p)$ model. The objective is to minimise the information criterion.
- In general, for small T, both information criteria tend to select models with order lower than p_0. As T grows, both information criteria tend to select larger models.
- As T approaches ∞, the probability that BIC selects $p = p_0$ converges to 1. This is not true of AIC, which even in large samples, selects $p > p_0$ with non-zero probability.
- Autoregressive Distributed Lag Models
  - $Y_t = \beta_0 + \sum_{i=1}^{p} \beta_i Y_{t-i} + \sum_{i=1}^{q} \delta_i X_{t-i} + u_t$, where $\mathbb{E}[u_t|\mathcal{Y}_{t-1}, \mathcal{X}_{t-1}] = 0$
  - Forecasting
    - The optimal 1-step ahead forecast of an $ADL(p,q)$ process $\{Y_t\}$ is
    $$Y_{T+1|T} := \mathbb{E}[Y_{T+1}|\mathcal{Y}_T, \mathcal{X}_T]$$
    $$= \beta_0 + \sum_{i=1}^{p} \beta_i Y_{T+1-i} + \sum_{i=1}^{q} \delta_q X_{T+1-i}.$$
  - Estimation
    - Every ADL model, by construction, satisfies orthogonality, and so is estimable by OLS regression.
- Granger Causality
  - $\{X_t\}$ does not Granger cause $\{Y_t\}$ iff lags of $\{X_t\}$ carry no useful information about $\{Y_t\}$ in addition to that carried by its own lags. Formally, this is iff $\mathbb{E}[Y_{T+1} - \mathbb{E}[Y_{T+1}|\mathcal{Y}_T, \mathcal{X}_T]]^2 = \mathbb{E}[Y_{T+1} - \mathbb{E}[Y_{T+1}|\mathcal{Y}_T]]^2$.
- Test of Granger Causality
  - F test of
  - $H_0 : \delta_1 = \ldots = \delta_p = 0$, against
  - $H_1 : \exists i \in \{1, \ldots, p\} : \delta_i \neq 0$, in
  - $Y_t = \beta_0 + \sum_{i=1}^{p} \beta_i Y_{t-i} + \sum_{i=1}^{p} \delta_i X_{t-i} + u_t$.
  - Supposing that $\{X_t\}$ and $\{Y_t\}$ are jointly stationary, the critical values for this test are drawn from the usual $F_{q,\infty}$ distribution for a test of q restrictions.
  - A rejection of the null can be interpreted as a finding that $\{X_t\}$ Granger causes $\{Y_t\}$.

Non-Stationary Time Series

- Structural Breaks
  - A structural break occurs at time τ iff the parameters of the model change abruptly at τ. A structural break is a source of non-stationarity.
- Chow Breakpoint Test
  - F test of
  - $H_0 : \gamma_0 = \gamma_1 = \ldots = \gamma_p = \theta_1 = \ldots = \theta_q = 0$, against
  - $H_1 : \exists \gamma \in \{\gamma_0, \gamma_1, \ldots, \gamma_p, \theta_1, \ldots, \theta_q\} : \gamma \neq 0$, in
  - $$Y_t = \beta_0 + \sum_{i=1}^{p} \beta_i Y_{t-i} + \sum_{i=1}^{q} \delta_i X_{t-i}$$
    $$+ [t > \tau][\gamma_0 + \sum_{i=1}^{p} \gamma_i Y_{t-i} + \sum_{i=1}^{q} \theta_i X_{t-i}]$$
    $$+ u_t$$
  - Supposing that $\{X_t\}$ and $\{Y_t\}$ are jointly stationary, the critical values for this test are drawn from the usual $F_{q,\infty}$ distribution for a test of q restrictions.
  - A rejection of the null can be interpreted as evidence of a structural break at τ, and that at least one of the parameters changes at τ.
- Quandt Likelihood Ratio Test
  - In practice, the precise time τ of a structural break is not known, hence the hypothesis of interest is not that a structural break occurred at some specific τ, but that a structural break occurred at some τ in some hypothesised interval $\{\tau_0, \tau_0 + 1, \ldots, \tau_1\}$.
  - The naive test that consists in repeating the Chow breakpoint test for each hypothesised time, and rejecting the null that no structural break occurs in the interval if the null of any of the Chow breakpoint tests is rejected is unsatisfactory because the significance of this test is greater than the significance of each individual Chow breakpoint test. When there is in fact no break, this test concludes that some break exists with greater probability than the significance level of each individual Chow test.
  - For each $\tau \in \{\tau_0, \ldots, \tau_1\}$, compute the F statistic $F(\tau)$ of the associated Chow breakpoint test.

- The Quandt likelihood ratio statistic is $QLR := \max_{\tau \in \{\tau_0, \ldots, \tau_1\}} F(\tau)$.
- Compare QLR against a suitable critical value.
  - This critical value depends on (1) the number q of parameters restricted by the null, (2) the number k of slope coefficients in the unrestricted model, and (3) the locations of the endpoints $\frac{\tau_0}{T}$ and $\frac{\tau_1}{T}$ relative to the length of the sample T.
  - The critical values of this test are larger than the critical values of the Chow breakpoint test, drawn from the $F_{q,\infty}$ distribution. This is because QLR is the maximum of several F statistics, so the probability that QLR is larger (in magnitude) than the $F_{q,\infty}$ critical value is greater than the probability that any individual F statistic is greater than that critical value.
  - The difference in magnitude of the Quandt likelihood ratio critical value and the $F_{q,\infty}$ critical value is increasing in the magnitude of $\tau_1 - \tau_0$.
- Unit Roots
  - An $AR(p)$ model $Y_t = \beta_0 + \sum_{i=1}^{p} \beta_i Y_{t-i} + u_t$ has a unit root iff $\sum_{i=1}^{p} \beta_i = 1$.
  - Iff $\{Y_t\}$ is a unit root autoregressive process with order p, then $\{\Delta Y_t\}$ is an autoregressive process with order $p-1$.
- Dickey-Fuller (Constant Only) Test
  - Parameters of a unit root autoregressive process can be consistently estimated by OLS regression. But the CLT does not apply to the OLS estimators of a non-stationary process. And the OLS estimators suffer from non-trivial biases that are particularly pronounced in the vicinity of a unit root.
  - t test of
  - $H_0 : \delta = \delta_0$, where $\delta_0 = 0$, against
  - $H_1 : \delta < \delta_0$, in
  - $\Delta Y_t = \beta_0 + \delta Y_{t-1} + u_t$, where
  - $\delta = \beta_1 - 1$.
  - The Dickey-Fuller test is a one-sided hypothesis test because in general, for economic time series where a unit root is plausible, the relevant competing model is a stationary model with $\beta_1 < 1$. $\beta_1 > 1$ implies an explosive behaviour that is seldom observed in economics, particularly after taking logarithms.
  - Under the null, $t(\delta_0)$ converges in distribution to the DF_{cm} distribution.
  - A rejection of the null can be interpreted as a finding that $\{Y_t\}$ is a stationary (in the substantive sense of being such that $\beta_1 \in (-1, 1)$) autoregressive process and not a unit root autoregressive process.
- Dickey-Fuller (Constant and Trend) Test
  - t test of
  - $H_0 : \delta = \delta_0$, where $\delta_0 = 0$, against
  - $H_1 : \delta < \delta_0$, in
  - $\Delta Y_t = \beta_0 + \alpha t + \delta Y_{t-1} + u_t$, where
  - $\delta = \beta_1 - 1$.
  - Under the null, $t(\delta_0)$ converges in distribution to the DF_{tr} distribution.
  - A rejection of the null can be interpreted as a finding that $\{Y_t\}$ is a (trend) stationary process and does not have a unit root.
- Augmented Dickey-Fuller (Constant Only) Test
  - t test of
  - $H_0 : \delta = \delta_0$, where $\delta_0 = 0$, against
  - $H_1 : \delta < \delta_0$, in
  - $\Delta Y_t = \beta_0 + \delta Y_{t-1} + \sum_{i=1}^{p_\Delta} \gamma_i \Delta Y_{t-i} + u_t$, which is derived from
    - $Y_t = \beta_0 + \sum_{i=1}^{p} \beta_i Y_{t-i} + u_t$, where
    - $\delta = (\sum_{i=1}^{p} \beta_i) - 1$, and
    - $p_\Delta = p - 1$.
  - Under the null, $t(\delta_0)$ converges in distribution to the DF_{cn} distribution.
  - A rejection of the null can be interpreted as a finding that $\{Y_t\}$ is an $AR(p)$ stationary process and does not have a unit root.
- Augmented Dickey-Fuller (Constant and Trend) Test
  - t test of
  - $H_0 : \delta = \delta_0$, where $\delta_0 = 0$, against
  - $H_1 : \delta < \delta_0$, in
  - $\Delta Y_t = \beta_0 + \alpha t + \delta Y_{t-1} + \sum_{i=1}^{p_\Delta} \gamma_i \Delta Y_{t-i} + u_t$, which is derived from

- $Y_t = \beta_0 + \alpha t + \sum_{i=1}^{p} \beta_i Y_{t-i} + u_t$, where
    - $\delta = \left(\sum_{i=1}^{p} \beta_i \right) - 1$,
    - $p_\Delta = p - 1$.
- Under the null, $t(\delta_0)$ converges in distribution to the DF_{cn} distribution.
- A rejection of the null can be interpreted as a finding that $\{Y_t\}$ is an $AR(p)$ stationary process and does not have a unit root.
- Order of Integration
    - The order of integration of a process $\{Y_t\}$ is the smallest $d \in \mathbb{N}_{\geq 0}$ such that $\{\Delta^d Y_t\}$ is stationary.
- Spurious Regression
    - Spurious regression is the tendency to find statistically significant regression relationships between $I(1)$ series, even where they are entirely independent.
    - Spurious regression arises because $I(1)$ series have a stochastic trend, hence a high tendency to exhibit long "swings" which entail that, with surprising regularity, movements in one such series can be "matched" against those in another.
- Cointegration
    - Y_t, X_t are cointegrated iff $Y_t, X_t \sim I(1)$ and there exists $\theta \in \mathbb{R}$ such that $\{\xi_t\} := \{Y_t - \theta X_t\}$ is stationary.
- Cointegration Test
    - Perform an ADF test on each of $\{Y_t\}, \{X_t\}$ to determine the order of integration of each series.
    - If $Y_t \sim I(0)$ or $X_t \sim I(0)$, then there is no possibility of cointegration, but also no possibility of spurious regression.
    - If there is a hypothesised value of θ, compute $\xi_t := Y_t - \theta X_t$, and perform an ADF test (for a suitable range of p_Δ) to determine the order of integration of $\{\xi_t\}$.
        - If the null is rejected, then conclude that $\xi_t \sim I(0)$, then Y_t, X_t are cointegrated with cointegrating coefficient θ.
        - If the null is not rejected, then Y_t, X_t are still potentially cointegrated, but not with the cointegrating coefficient θ.
    - If there is no hypothesised value of θ, estimate θ by OLS regression of Y_t on X_t, then compute the estimated equilibrium error $\hat{\xi}_t = Y_t - \hat{\theta} X_t$. Then, perform an ADF test (for a suitable range of p_Δ) to determine the order of integration of $\hat{\xi}_t$, using adjusted (Engle-Granger) critical values.
        - If the null is rejected, then conclude that Y_t, X_t are cointegrated.
        - If the null is not rejected, then conclude that Y_t, X_t are not cointegrated.