Econometrics Summary

Algebraic and Statistical Preliminaries

Elasticity: The point elasticity of Y with respect to L is given by $\alpha = (\partial Y/\partial L)/(Y/L)$. The arc elasticity is given by $(\Delta Y/\Delta L)/(Y/L)$, when L and Y may be chosen in various ways, including taking the midpoint of the two points in question or using the initial point. We also estimate $\Delta Y/Y \approx \alpha\ (\Delta L/L)$.

Growth Factors: The growth rate of X is given by $g_X = (X_t - X_{t-1})/X_{t-1} = X_t/X_{t-1} - 1$. Notice that $(\ln X_t - \ln X_{t-1}) = \ln (1 + g_X)$. Thus, a constant growth rate would lead to a straight line in a plot of time versus ln X.

- Discrete Time Growth occurs when Pi = (1+r) Pi-1 where i denotes a discrete time period, and therefore takes on only integer values. r is a constant growth factor. Then, $P_i = (1 + r)^i P_0$, and $\ln P_i = \ln P_0 + i \ln (1 + r)$.
- Continuous Time Growth occurs with continuous compounding. Then, we model $P_t = e^g P_{t-1} = e^{gt} P_0$. Then, $\ln P_t = \ln P_0 + gt$.
- As a comparison: For these to be equal, $1 + r = e^g$.

A random variable is a variable that doesn't have a specific value (though it does have a distribution) until an experiment is done. The outcomes of the measurements of the experiment must be numerical. It may also be considered as a function that assigns a numerical value to the outcome of an experiment.

- A discrete random variable has a countable number of possible values, with probability given by a probability density function.
- A continuous random variable can take on values over an interval, given by the cumulative probability function, $P(a < x < b) = \int_a^b p(x)\, dx$, where p(x) is the probability mass function.

Descriptive Statistics:

- Sample mean: x-bar = $(\sum x_i)/n$
- Total variation = $\sum(x_i - \text{x-bar})^2$
- Variance: $s_x^2 = TV/(n-1)$
- Sample Standard Deviation: $s_x = \sqrt{\text{Variance}}$
- Covariance: $s_{xy} = (\sum(x_i - \text{x-bar})(y_i - \text{y-bar}))/(n-1)$
- Correlation Coefficient: $r = s_{xy}/s_x s_y$
- Mean Squared Error: $MSE = E((\beta^* - \beta)^2) = Var(\beta^*) + Bias^2$

- Suppose we have n observations of Y, where Y can take on the values $x_1, \ldots, x_m$. Let $n_k = \{y_i \mid y_i = x_k\}$. Then, $f_k = n_k/n$ is the relative frequency of k. This gives the following new formulas:
  - Sample mean: x-bar = $\sum_{k=1}^m f_k x_k$
  - Sample standard deviation: $s_x = \sqrt{n/(n-1)}\ \sqrt{(\sum(x_k - \text{x-bar})^2 f_k)}$
  - Similar calculations may be done with continuous random variables, by dividing them into m intervals and choosing a class mark (representative value) for each interval.

Sampling: A random sample is a sample such that all units are equally likely to be selected for inclusion in the sample, and each selection is independent of each other selection. This means estimators won't be biased by how the sample was selected. A

sampling distribution of a statistic gives the relative frequency that certain values of the statistic would occur in repeated sampling from the underlying population.
Estimators: An estimator is a rule of calculating estimates based on a set of data. The point estimates are calculated using the rule given by the estimator given the data. (We have distributions for estimators, not estimates).

Elementary Linear Regression

The simple bivariate model is given by $Y_i = \beta_0 + \beta_1 X_i + u_i$

- Y_i is the dependent variable.
- X_i is the explanatory variable.
- β_0 and β_1 are the true population parameters, which are never known.
- u_i are unobservables (differences in variables we are not measuring that affect the dependent variable).

- $Y_i = E(Y_i \mid X_i) + u_i$ (the u_i are the reason that points do not lie on the true regression line).
- The estimated regression line is given by $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$.
  - $\hat{\beta}_0$ and $\hat{\beta}_1$ are the estimates of β_0 and β_1.
  - The residuals (error) are $\hat{u}_i = Y_i - \hat{Y}_i$. (These are only estimates of the unobservables, not the true unobservables… They can't be observed.)

Ordinary Least Squares: We estimate $\hat{\beta}_0$ and $\hat{\beta}_1$ by minimizing the sum of squared residuals: $\sum(\hat{u}_i)^2 = \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$. This involves solving the equations $\sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$ and $\sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)X_i = 0$ (found by taking the partials and setting them to 0). This gives us the solutions:

- $\hat{\beta}_1 = \sum(X_i - \text{X-bar})(Y_i - \text{Y-bar})/\sum(X_i - \text{X-bar})^2$
- $\hat{\beta}_0 = \text{Y-bar} - \hat{\beta}_1 \text{ X-bar}$

- Algebraic Properties:
  - (X-bar, Y-bar) is always on the OLS line.
  - The sum of the residuals, $\sum \hat{u}_i$, is 0.
  - $\text{Cov}(\hat{u}_i, X_i) = 0$, since $\sum \hat{u}_i X_i = 0$.

Measuring the Fit:

- The standard error of regression (also called root mean square error) is $\text{SER} = \sqrt{(\sum (\hat{u}_i)^2 / (n-2))}$.
- The coefficient of variation of a single variable is s_x / x-bar. For regression, this is SER/y-bar.
- The coefficient of determination, is given by $R^2 = 1 - \sum(\hat{u}_i)^2 / \sum(y_i - \text{y-bar})^2 = \sum (\hat{y}_i - \text{y-bar})^2 / \sum(yi - \text{y-bar})^2$.

Interpreting the coefficients of a regression: $\hat{\beta}_0$ is the predicted expectation of Y when X = 0. $\hat{\beta}_1$ is the predicted change in Y given a unit change in X.

- Changing the scales of X and Y do not change the resulting predictions or interpretations.

Alternative Specifications: We may take regressions with functions of X and Y; linear means linear in the estimated parameters.

- log-level specification: $\ln Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$. In this case, $100\hat{\beta}_1\%$ is the approximate percentage change in the dependent variable for a unit change in the independent variable. (The exact value is $g_w = 100(e^{\hat{\beta}_1} - 1)\%$.)

- log-log specification: In this case, $\beta_1^{\wedge}$ is the elasticity of Y with respect to X. (We are implicitly assuming constant elasticity.)

Expected values and variances of OLS Estimators

- The following assumptions are needed for the calculations of bias and variance:
  - The true population model is linear in the parameters.
  - The sample is drawn randomly from the population being modeled.
  - The expected value of the population unobservables is 0, conditional on X. That is, $E(u_i \mid X_i) = 0$.
  - The explanatory variable is not constant.
  - (Homoskedasticity) The variance of the unobservables is constant, conditioning on the explanatory variable. That is, $Var(U_i \mid X_i) = \sigma_u^2$.
- Under the first four assumptions, the OLS estimators are unbiased.
- Under the assumption of homoskedasticity, $Var(\beta_1^{\wedge}) = \sigma_u^2 / \sum (x_i - \text{x-bar})^2$, $Var(\beta_0^{\wedge}) = \sigma_u^2 (\sum x_i^2 / n \sum (x_i - \text{x-bar})^2)$, and $Cov(\beta_0^{\wedge}, \beta_1^{\wedge}) = -(\text{x-bar}) \sigma_u^2 / \sum (x_i - \text{x-bar})^2$.
- We estimate σ_u^2 by $SER^2 = (\sum u_i^{\wedge 2})/(n-2)$. This is unbiased.
- $se(\beta_0^{\wedge}) = SER \sqrt{(\sum x_i^2 / n \sum (x_i - \text{x-bar})^2)}$
- $se(\beta_1^{\wedge}) = SER \sqrt{(1 / \sum (x_i - \text{x-bar})^2)}$
- *Gauss-Markov Theorem*. Under the first four assumptions about, the OLS estimators, in the class of linear unbiased estimators, have minimum variance. That is, the OLS estimators are BLUE (best linear unbiased estimators).

Multiple Regression:

- The model: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i$
  - $E(Y_i \mid X_1, \ldots, X_k) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}$
  - $\beta_j^{\wedge}$ estimates the response in Y if X_j changes by one unit and all other X_l are held constant.
- First Order Conditions (OLS Normal Equations):
  - $\sum (Y_i - \beta_0^{\wedge} - \beta_1^{\wedge} X_{1i} - \ldots - \beta_k^{\wedge} X_{ki}) = 0$
  - $\sum (Y_i - \beta_0^{\wedge} - \beta_1^{\wedge} X_{1i} - \ldots - \beta_k^{\wedge} X_{ki}) X_{ji} = 0$ for $j = 1, \ldots, k$
- Algebraic Properties
  - $\sum u_i^{\wedge} = 0$
  - The fitted regression line passes through the point of means, (X_1-bar, X_2-bar, …, X_k-bar, Y-bar).
  - $Cov(u^{\wedge}, X_j) = 0$ for j = 1, .., k.
- $SER = \sqrt{(\sum u_i^{\wedge 2} / (n-k-1))}$
- $\text{R-bar}^2 = 1 - SER^2/s_y^2$, where $s_y^2 = \sum (Y_i - \text{Y-bar})^2 / (n-1)$.
  - This adjusts R^2 for the number of fitted parameters, penalizing the regression for needing more parameters. (R^2 never decreases with more parameters; R-bar^2 might.)
  - If a model decreases SER^2, then R-bar^2 increases, since s_y^2 is constant.
  - Notice that R-bar^2 approximates $1 - Var(u_i) / Var(Y_i)$.
  - $R^2 = \text{R-bar}^2$ when k = 0.

Multicollinearity: When the independent variables are correlated.

- Under perfect collinearity, we cannot solve for the coefficients on those variables because we cannot separate the effects of two variables that always move together.
    - This requires modification of Assumption 4: In the sample (and in the population), none of the explanatory variables are constant or exact linear functions of each other.
- If there is multicollinearity but not perfect multicollinearity, estimators will be unbiased, but the variance of the estimators will increase:
    - $Var(\hat{\beta_j}) = \sigma_u^2 / (SST_j(1 - R_j^2))$
        - $SST_j = \sum (X_{ji} - X_j\text{-bar})^2$ (total sum of squares)
        - R_j^2 is the R^2 from the regression of X_{ji} on X_{1i}, …, $X_{(j-1)i}$, $X_{(j+1)i}$, …, X_{ki}.
    - This occurs because multicollinearity makes identifying specific influences harder, since variables tend to move together.

Specification Issues:

- Excluding a relevant variable
    - If the excluded variable is correlated with any included variables, the coefficients on the included variables will be biased (omitted variable bias):
        - Suppose we estimate $Y_i = b_0 + b_1 X_{1i} + u_i$' instead of the true $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_{2i} X_{2i} + u_i$. Then, $E(\hat{b_1}) = \beta_1 + \beta_2 b_{21}$, where b_{21} is found from $X_{2i} = b_{20} + b_{21} X_{1i} + u_{2i}$.
- Including an irrelevant variable
    - The results are unbiased if extra variables are included.
    - Irrelevant variables decreases the degrees of freedom, and therefore increases the standard errors.
- We cannot choose the variables to include after running the regression – this violates the theories of hypothesis testing!

Inferences in Regression

For all hypothesis testing, we assume u_i are distributed iid normal, with mean 0 and variance σ_u^2.

- If we consider u_i as a sum of independent, identically distributed influences, u_i is approximately normal by the central limit theorem.
- By a variant of the Central Limit Theorem, there is approximate normality even if there are only a small number of influences which are not quite independent.
- Normality also makes testing and derivations easier.

Assuming normality, since Y_i and all estimators are linear combinations of normally distributed random variables:

- $Y_i \sim N(\beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki}, \sigma_u^2)$
- $\hat{\beta_j} \sim N(\beta_j, \sigma_u^2 / SST_j(1 - R_j^2))$

T-Tests

- The test statistic: $t^* = (\hat{\beta_j} - \beta_j^0) / se(\hat{\beta_j})$
    - Under the null hypothesis, $t^* \sim t_{n-k-1}$

- t measures the "estimation error" – the distance from the estimate to the null hypothesis. Large estimation errors relative to the standard error are evidence against the null hypothesis.

- One-sided (testing for sign, when $\beta_j^0 = 0$)
  - Alternate Hypothesis: $\beta_j > \beta_j^0$ or $\beta_j < \beta_j^0$
  - For the former alternative hypothesis, reject H_0 if $t^* > t^c$, where t^c is determined by $P(t > t^c) = \alpha$, the significance level.
  - Note that rejecting H_0 for extreme values with the wrong sign is incorrect!
- Two-sided
  - Alternative Hypothesis: $\beta_j \neq \beta_j^0$
  - Decision Rule: Reject H_0 if $|t^*| > t^c$, where
    - t^* is calculated as above
    - t^c, the critical value, is determined by $P(|t| > t^c) = \alpha$, where α is the significance level
- Confidence Intervals
  - If we want $P(-t^c \leq (\hat{\beta_j} - \beta_j^*)/se(\hat{\beta_j}) \leq t^c) = 1 - \alpha$, we have a $(1 - \alpha)$% confidence interval of $\hat{\beta_j} \pm t^c \, se(\hat{\beta_j})$.
  - This gives a random interval which, $(1 - \alpha)$% of the time will contain the true population parameter.
- P-Values
  - The p-value gives the probability that a t statistic distributed according to the null hypothesis would be greater than the calculated t*.
    - $p = P(t > t^*)$
  - This allows readers to apply their own significance levels.
  - Note that two-tailed p-values are given by $p = 2\,P(t > |t^*|)$.

F-Tests: To test groups of coefficients.

- We consider two models, the unrestricted model and the restricted model.
  - Restrictions may include removing 1 or more variables or imposing a relationship on the coefficients (for example: $\beta_1 + \beta_2 = 1$; then, we run a new regression with the substitution of $1 - \beta_1$ for β_2 and appropriate algebraic manipulation).
- Test statistic: $F = ((SSR_R - SSR_U)/r) / (SSR_U/(n - k - 1)) = ((R_u^2 - R_R^2)/r)/((1 - R_U^2)/(n - k - 1))$
  - SSR_R is the sum of the squared residuals in the restricted model
  - SSR_U is the sum of squared residuals in the unrestricted model
  - r is the number of restrictions imposed
  - $n - k - 1$ is from the unrestricted model
- Under the null hypothesis, $F \sim F^{r, n-k-1}$. We reject when $F > F^c$.

Types of Errors

- Type I Error: Rejecting the null hypothesis when it is true. The probability of this occurring is α

Prediction

We may use regression lines to predict new values of the dependent variable from other values of the independent variable, assuming the economic conditions hold.

- True Value: $Y_p = \beta_0 + \beta_1 X_p + u_p$
- Predicted Value: $Y_p = \beta_0^{\wedge} + \beta_1^{\wedge} X_p$
- Prediction Error: $u_p^{\wedge} = Y_p^{\wedge} - Y_p = (\beta_0^{\wedge} - \beta_0) + (\beta_1^{\wedge} - \beta_1)X_p - u_p$
- $\text{Var}(u_p^{\wedge}) = \text{Var}(\beta_0^{\wedge}) + X_p^2 \text{Var}(\beta_1^{\wedge}) + \text{Var}(u_p) + 2X_p\text{Cov}(\beta_0^{\wedge}, \beta_1^{\wedge}) = \sigma_u^2(1 + 1/n + (\text{X-bar} - X_p)^2 / \sum (X_i - \text{X-bar})^2)$
- Notice that predictions are less variable for large n and closer to X-bar. However, the variance is never lower than σ^2.

Dummy Variables

Definition. A dummy variable indicates the presence or absence of a quality or attribute for each observation. That is, $D_i = 1$ if the attribute is present, and 0 otherwise.
If there are multiple categories, then there should be one less dummy variable than category (otherwise, perfect collinearity will occur). The category with no dummy variable is called the reference group, and the coefficient on every other variable makes a comparison to the reference group.

Asymptotics

Definition. The probability limit of X is θ if $\lim_{n\to\infty} P(|X - \theta| > \delta) = 0$ for any $\delta > 0$. In this case, we write plim $X = \theta$.

- Properties:
  - plim C = C if C is constant
  - plim $(Z_1 + Z_2)$ = plim Z_1 + plim Z_2
  - plim Z_1Z_2 = (plim Z_1)(plim Z_2)
  - plim Z_1/Z_2 = (plim Z_1)/(plim Z_2)
  - plim g(Z) = g(plim Z)
- Note that this is a generalization of the expectations operator – only the first two properties always hold for expectations.

Definition. $\theta^{\wedge}$ is a consistent estimator of θ if the probability limit of $\theta^{\wedge}$ is θ.
Note that plims hold only for large (infinite) samples. Small samples may have problems.

Heteroskedasticity

We relax the assumption that $E(u_i^2)$ is constant. Instead, $E(u_i^2) = \sigma_i^2$ for each i.
Estimators will still be unbiased with heteroskedasticity; however, the variance estimates will be wrong.

- For example, $\text{Var}(\beta_1^{\wedge}) = (\sum(x_i - \text{x-bar})^2\sigma_i^2)/(\sum(x_i - \text{x-bar})^2)^2$, since we can't pull the σ_i^2 out anymore.
- This means that the variance might not be minimized, and OLS might not be BLUE.

Ideally, we would instead do the regression $Y_i/\sigma_i = \beta_0/\sigma_i + \beta_1 (X_i/\sigma_i) + u_i/\sigma_i$. Then, $u_i/\sigma_i \sim N(0, 1)$, and we would have homoskedasticity. (Note that we are now regressing on two variables – $1/\sigma_i$ and X_i/σ_i – and no error term.) This procedure is called generalized least squares, and yields BLUE estimators.
Detecting heteroskedasticity:

- Might be implied by the nature of the problem.

- Graph the residuals or the squared residuals against independent variables and look for patterns.
- White's Test:
  - Get the residuals from the initial regression
    - $Y_i = \beta_0^{\wedge} + \beta_1^{\wedge} X_{1i} + \beta_2^{\wedge} X_{2i} + u_i^{\wedge}$.
  - Regress $u_i^{\wedge 2}$ on a constant, the independent variables, their squares and crossterms:
    - $u_i^{\wedge 2} = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{1i}^2 + \alpha_4 X_{2i}^2 + \alpha_5 X_{1i} X_{2i}$
  - Under the null hypothesis of homoskedasticity, $nR^2 \sim \chi^2$, where R^2 and the degrees of freedom come from the second equation estimated. Test this hypothesis.

Correcting heteroskedasticity:

- To find the proper standard errors: $se(\beta_1^{\wedge})^2 = (\sum(x_i - \text{x-bar})^2 u_i^{\wedge 2})/(\sum(x_i - \text{x-bar})^2)^2$. (These are called the White standard errors, and are consistent but biased for the true variance.)
- Suppose we have a functional form for the heteroskedasticity: $E(u_i^2 \mid X_i) = \sigma^2 f(X_i)$
  - Transform the model: $Y_i / \sqrt{f(X_i)} = \beta_0 / \sqrt{f(X_i)} + \beta_1 X_i / \sqrt{f(X_i)} + u_i / \sqrt{f(X_i)}$
  - Notice that the error term is now σ^2. Hence, this is heteroskedastic.
  - Now, the model being estimated has two variables ($1 / \sqrt{f(X_i)}$ and $X_i / \sqrt{f(X_i)}$) and no intercept.
  - We may be estimating parameters in choosing f() [say, $u_i^2 = \sigma^2 e^{aXi}$], we cannot ensure that our error and variance estimates will be unbiased. However, they will be consistent.

Time Series Regression

Definition. Suppose we have an infinite sequence of random variables, $\{X_t\}$ for t = …, -2, -1, 0, 1, 2, …. This is called a discrete stochastic process. A realization of a discrete stochastic process for a particular finite time period, {1, 2, …, T}, that is X_1, …, X_T, is called a time series.

This allows us to look at the effect of both lagged (from previous periods) and contemporaneous (determined in the same period) variables on the variable in question – including the dependent variable's own lagged values.

- Often, the result of a variable changing one unit in one period and then returning to its old value is of interest. This is captured (over time) in the coefficients on that variable (in its various lags). This is the short run effect of a change.
- Sometimes, we care more about when the variable changes one unit and stays there. To find this, reparameterize the model in terms of m_t (the current value), Δm_t (the change from t-1 to t), Δm_{t-1}, and so on. The resulting change will be the sum of the coefficients on m_t, m_{t-1}, etc. This is the long run effect of a unit change.

Classical Assumptions for OLS with Time Series Data

- Linearity: the population model is linear in its parameters
- $E(u_t \mid \mathbf{X}) = 0$ for all t
  - Note that **X** includes all past and future values of X_y.

    - This replaces the assumption of random sampling, since we cannot do random sampling in a time series.
- No perfect multicollinearity
- Homoskedasticity
- No autocorrelation: $Corr(u_t, u_s \mid \mathbf{X}) = 0$ for all $t \neq s$.
    - Ensures that disturbances over time periods are not related.
- (for hypothesis testing) $u_t \sim N(0, \sigma^2)$, independently.

Patterns in the Data: Even without economic theory, we may assume that a times series is of the form $Y_t = T_t + S_t + C_t + U_t$, where T_t is a trend, S_t is seasonal, C_t is cyclical, and U_t is random.

- Trends:
    - Deterministic: $T_t = t.$, so that $Y_t = \beta_0 + \beta_1 t + u_t$, and $E(Y_t) = \beta_0 + \beta_1 t$.
    - Semilog Trend: $\ln Y_t = \ln Y_0 + t \ln(1+r) + u_t$, so that $Y_t = Y_0(1+r)^t e^{ut}$.
        - Note that $E(Y_t) = Y_0(1+r)^t E(e^{ut}) > Y_0(1+r)^t$.
        - In particular, $E(e^{ut}) = e^{\sigma\wedge 2/2}$, which we estimate by SER^2
    - Random Walk: $Z_t = Z_{t-1} + e_t$, $e_t \sim [0, \sigma_e^2]$, $Cov(e_t, e_s) = 0$ when $t \neq s$, is a random walk.
        - $Var(Y_t) = Var(\sum e_t) = \sum Var(e_t) = t\sigma_e^2$. This is unbounded.
        - A random walk is said to have a <u>unit root</u>, since the coefficient on Z_{t-1} is 1.
        - Two independent random walks often cause <u>spurious</u> regressions, where the hypothesis that they are unrelated is rejected with probability greater than α.
        - In this regression, $Y_t = \beta_0 + \beta_1 X_t + u_t$. Under the null hypothesis, we have $Y_t = \beta_0 + u_t = u_t$ (since $Y_0 = 0$). But $Cov(u_t, u_{t-1}) = Cov(Y_t, Y_{t-1}) \neq 0$, and $Var(u_t) = Var(Y_t)$ is unbounded.
- Seasonality: When a value may fluctuate depending on the season. To deal with this, create a dummy variable for each season (except for the reference season). Then, the coefficient on each season shows the effect of being in that season relative to the reference season.

Autocorrelation (Serial Correlation)

Definition. Autocorrelation occurs when $Cov(u_t, u_s) \neq 0$ for some $t \neq s$.

- Autocorrelation is more likely to occur with time series, since random sampling (including a random order) removes it.
- This occurs when disturbances are likely to be the same over a few periods.

An example: Suppose $u_t = \rho u_{t-1} + \varepsilon_t$, $|\rho| < 1$ (so that $Var(u_t) < \infty$). (This is <u>first order autocorrelation</u>, written AR(1).) Then, $E(u_t) = 0$ and $Var(u_t) = \sigma_\varepsilon^2/(1 - \rho^2)$.

- The OLS estimators are still unbiased under autocorrelation.
- Notice that $Cov(u_t, u_{t-s}) = Corr(u_t, u_{t-s}) \sqrt{Var(u_t)Var(u_{t-s})} = \rho^s \sigma_u^2$
- $Var(\beta_1\hat{}^A) = Var(\beta_1\hat{}) + \sigma_u^2(2\sum_{s<t}(x_t - \text{x-bar})(x_{t-s} - \text{x-bar})\rho^s)/\sum(x_t - \text{x-bar})^2$.
- Thus, autocorrelation makes the variances of the estimators bigger, if $\rho > 0$ and may make the variance bigger or smaller otherwise.

Detecting autocorrelation: Using the residuals from the standard estimator of the model:

- Look at the residuals graphically to see if they tend to be positive or negative together.
- Regress $u_t^{\wedge}$ on $u_{t-1}^{\wedge}$ and X_t. Test whether the coefficient on the lagged residual is significantly different from 0.

Fixing autocorrelation:

- Assuming ρ were known: $Y_t - \rho Y_{t-1} = (1-\rho)\beta_0 + \beta_1(X_t - \rho X_{t-1}) + (u_t - \rho u_{t-1})$ fits the OLS assumptions.
- Instead, we use feasible GLS to estimate ρ: (This is asymptotically correct.)
  - Estimate $Y_t = \beta_0 + \beta_1 X_t + u_t$ and obtain $u_t^{\wedge}$.
  - Regress $u_t^{\wedge}$ on $u_{t-1}^{\wedge}$. Obtain $r^{\wedge}$, the slope coefficient on $u_{t-1}^{\wedge}$.
  - Use the transformation above: Regress $Y_t - r^{\wedge}Y_{t-1}$ on $X_t - r^{\wedge}X_{t-1}$.

Simultaneous Equations

If a system is described by multiple interdependent equations, then using OLS on each equation independently is biased and inconsistent. This is called simultaneous equations bias (and is like omitted variable bias).

The structural model of a system is the original system of equations (from economic theory). We may solve this system of equations for the endogenous variables in terms of the exogenous variables. This gives the reduced form.

The Identification Problem: Is the reduced form sufficient to make good estimates for the parameters in the structural form?

- *Definition.* An equation is unidentified if there is no way to estimate all its structural parameters from the reduced form. An equation is identified otherwise. An equation is exactly identified if unique parameters values exist and overidentified if more than one value can be obtained for some parameters.
- The Order Condition (a necessary condition for identification): If an equation is identified then the number of exogenous variables excluded from the equation is at least the number of included endogenous variables (on both sides) minus one.
  - It is possible for one equation to be identified when the other is not.
  - Example: In $P_t = \beta_0 + \beta_1 Q_t + u_t$ and $P_t = \alpha_0 + \alpha_1 Q_t + \alpha Y_t + v_t$, the first equation is identified (since Y_t is excluded to make up for P_t and Q_t both being endogenous), but the second is not.

Two Stage Least Squares:

- The Method
  - Find the reduced forms for all endogenous variables.
  - For each endogenous variable on the right-hand-side of the identified equation, estimate the reduced form and obtain the predicted values.
  - Substitute the predicted values for the actual values in the identified equation.
  - Run the regression with these values instead.
- By doing this, we find the part of the endogenous variables that are no correlated with the error term, which removes the bias.
- The exogenous variables that are excluded from the identified equation are called instrumental variables. Thus we assume:
  - They are independent of the error term in the identified equation/

- They are correlated with the endogenous right-hand-side variables in the identified equation.
- They do not appear in the equation being estimated.
- This allows them to substitute for the endogenous variables without having to worry about correlation.

Proxy Variables: When a variable stands in for another (unmeasurable) variable, like IQ for innate ability.

Formulas:

Total variation = $\sum(x_i - \text{x-bar})^2$

$\beta_1^{\wedge} = \sum(X_i - \text{X-bar})(Y_i - \text{Y-bar})/\sum(X_i - \text{X-bar})^2$

$\beta_0^{\wedge} = \text{Y-bar} - \beta_1^{\wedge}\ \text{X-bar}$

CV = SER/y-bar

$R^2 = 1 - \sum(u_i^{\wedge})^2 / \sum(y_i - \text{y-bar})^2 = \sum (y_i^{\wedge} - \text{y-bar})^2/\sum(\text{yi} - \text{y-bar})^2$.

$\text{Var}(\beta_1^{\wedge}) = \sigma_u^2 / \sum (x_i - \text{x-bar})^2$

$\text{Var}(\beta_0^{\wedge}) = \sigma_u^2 (\sum x_i^2 / n \sum (x_i - \text{x-bar})^2)$

$\text{Cov}(\beta_0^{\wedge}, \beta_1^{\wedge}) = - (\text{x-bar})\ \sigma_u^2 / \sum (x_i - \text{x-bar})^2$.

$\text{SER} = \sqrt{} (\sum u_i^{\wedge 2} / (n–k–1))$

$\text{R-bar}^2 = 1 – \text{SER}^2/s_y^2$, where $s_y^2 = \sum (Y_i - \text{Y-bar})^2 / (n-1)$.

$\text{Var}(\beta_j^{\wedge}) = \sigma_u^2 / \text{SST}_j(1 – R_j^2)$

$\text{SST}_j = \sum (X_{ji} - X_j\text{-bar})^2$ (total sum of squares)

R_j^2 is the R^2 from the regression of X_{ji} on X_{1i}, …, $X_{(j-1)i}$, $X_{(j+1)i}$, …, X_{ki}.

$t^* = (\beta_j^{\wedge} - \beta_j^0) / \text{se}(\beta_j^{\wedge})$

$F = ((\text{SSR}_R – \text{SSR}_U)/r) / (\text{SSR}_U/(n – k – 1))$

- Prediction Error: $\text{Var}(u_p^{\wedge}) = \text{Var}(\beta_0^{\wedge}) + X_p^2\ \text{Var}(\beta_1^{\wedge}) + \text{Var}(u_p) + 2X_p\text{Cov}(\beta_0^{\wedge}, \beta_1^{\wedge})$
  $= \sigma_u^2(1 + 1/n + (\text{X-bar} – X_p)^2 / \sum (X_i - \text{X-bar})^2)$

Heteroskedasticity: $\text{Var}(\beta_1^{\wedge}) = (\sum(x_i - \text{x-bar})^2\sigma_i^2)/(\sum(x_i - \text{x-bar})^2)^2$