QUANTITATIVE RESEARCH METHODS II

VERA E. TROEGER
DEPARTMENT OF ECONOMICS

E-MAIL: V.E.TROEGER@WARWICK.AC.UK
OFFICE HOURS: BY APPOINTMENT
COURSE OUTLINE

1. Introduction to the Course: the OLS model, Gauss-Markov Assumptions and Violations

2. Heteroskedasticity, cross-sectional correlation, multicollinearity, omitted variable bias: tests and common solutions.

3. Specification issues in Linear Models: Non-Linearities and Interaction Effects

4. Dynamics, serial correlation and dependence over time

5. Pooling Data 1: Heterogeneity: How to choose the right model – how good are the available tests.
6. Pooling Data 2: Slope and Parameter Heterogeneity: Seemingly Unrelated Regression and Random Coefficient Models, modelling parameter instability

7. Specification Issues: Endogeneity and spatial econometrics: instrumental variable approaches and simultaneous equation models

8. Limited Dependent Variable Models: Binary and Ordered Choice Models

9. Limited Dependent Variable Models with Pooled and Panel Data

10. Specification Issues in Limited Dependent Variable Models: Dynamics and Heterogeneity

11. Wrap Up, Q&A
THE SIMPLE LINEAR REGRESSION MODEL

Regression models help investigating bivariate and multivariate relationships between variables, where we can hypothesize that 1 variable depends on another variable or a combination of other variables.

Relationships between variables in political science and economics are not exact – unless true by definition, but relationships include most often a non-structural or random component, due to the probabilistic nature of theories and hypotheses in PolSci, measurement errors etc.

Regression analysis enables to find average relationships that may not be obvious by just „eye-balling“ the data – explicit formulation of structural and random components of a hypothesized relationship between variables.

Example: positive relationship between unemployment and government spending
SIMPLE LINEAR REGRESSION ANALYSIS

Linear relationship between $x$ (explanatory variable) and $y$ (dependent variable)

Epsilon describes the random component of the linear relationship between $x$ and $y$

$$y_i = \alpha + \beta \cdot x_i + \varepsilon_i$$
\[ y_i = \alpha + \beta \cdot x_i + \varepsilon_i \]

**Y** is the value of the dependent variable (spending) in observation i (e.g. in the UK)

**Y is determined by 2 components:**

1. the non-random/ structural component \( \alpha + \beta \cdot x_i \) - where \( x \) is the independent/ explanatory variable (unemployment) in observation i (UK) and \( \alpha \) and \( \beta \) are fixed quantities, the parameters of the model; \( \alpha \) is called constant or intercept and measures the value where the regression line crosses the y-axis; \( \beta \) is called coefficient/ slope, and measures the steepness of the regression line.

2. the random component called disturbance or error term \( \varepsilon_i \) in observation i
A simple example:

x has 10 observations: 0,1,2,3,4,5,6,7,8,9

The true relationship between y and x is: y = 5 + 1*x, thus, the true y takes on the values: 5,6,7,8,9,10,11,12,13,14

There is some disturbance e.g. a measurement error, which is standard normally distributed: thus the y we can measure takes on the values: 6.95,5.22,6.36,7.03,9.71,9.67,10.69,13.85, 13.21,14.82 – which are close to the true values, but for any given observation the observed values are a little larger or smaller than the true values.

the relationship between x and y should hold on average true but is not exact

When we do our analysis, we don’t know the true relationship and the true y, we just have the observed x and y.

We know that the relationship between x and y should have the following form: y = alpha + beta * x + epsilon (we hypothesize a linear relationship)

The regression analysis „estimates“ the parameters alpha and beta by using the given observations for x and y.

The simplest form of estimating alpha and beta is called ordinary least squares (OLS) regression
OLS-Regression:

Draw a line through the scatter plot in a way to minimize the deviations of the single observations from the line:

\[
\hat{y}_i = \hat{\alpha} + \hat{\beta} \cdot x_i + \varepsilon_i \quad \Rightarrow \quad \varepsilon_i = y_i - \left( \hat{\alpha} + \hat{\beta} \cdot x_i \right)
\]

Minimize the sum of all squared deviations from the line (squared residuals)

This is done mathematically by the statistical program at hand

the values of the dependent variable (values on the line) are called predicted values of the regression (yhat): 4.97, 6.03, 7.10, 8.16, 9.22, 10.28, 11.34, 12.41, 13.47, 14.53 – these are very close to the „true values“; the estimated alpha = 4.97 and beta = 1.06
**OLS REGRESSION**

Ordinary least squares regression: minimizes the squared residuals

\[ y_i = \alpha + \beta \cdot x_i + \varepsilon_i \quad \Rightarrow \quad \varepsilon_i = y_i - \alpha - \beta \cdot x_i \]

\[ \hat{y}_i = \hat{\alpha} + \hat{\beta} \cdot x_i + \hat{\varepsilon}_i \]

\[ \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} (\hat{\varepsilon}_i)^2 = \text{min} \]

Components:

**DY:** \( y \); at least 1 **IV:** \( x \)

Constant or

intercept term: alpha

Regression coefficient,

slope: beta

Error term, residuals: epsilon
DERIVATION

Linear relationship between x (explanatory variable) and y (dependent variable)

\[ y_i = \alpha + \beta \cdot x_i + \varepsilon_i \]

\[ \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} (\varepsilon_i)^2 = \text{min} \]

\[ \Rightarrow \hat{\alpha} = \sum_{i=1}^{n} (y_i - \hat{\beta} \cdot x_i) \Rightarrow \hat{\alpha} = \bar{y} - \hat{\beta} \cdot \bar{x} \]

\[ \Rightarrow \hat{\beta} = \frac{\sum_{i=1}^{n} (y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})} \]

\[ \Rightarrow \hat{\beta} = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\text{Cov}(x, y)}{\text{Var}(x)} = X'X^{-1}X'y \]
Naturally we still have to verify whether $\hat{\alpha}$ and $\hat{\beta}$ really minimize the sum of squared residuals and satisfy the second order conditions of the minimizing problem. Thus we need the second derivatives of the two functions with respect to alpha and beta which are given by the so called Hessian matrix (matrix of second derivatives). (I spare the mathematical derivation)

The Hessian matrix has to be positive definite (the determinant must be larger than 0) so that $\hat{\alpha}$ and $\hat{\beta}$ globally minimize the sum of squared residuals. Only in this case alpha and beta are optimal estimates for the relationship between the dependent variable y and the independent variable x.
Regression coefficient:

\[ \hat{\beta}_{yx} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \]

Beta equals the covariance between y and x divided by the variance of x.
Interpretation: example

\( \alpha = 4.97, \beta = 1.06 \)

Education and earnings: no education gives you a minimal hourly wage of around 5 pounds. Each additional year of education increases the hourly wage by app. 1 pound:
Properties of the OLS estimator:

Since alpha and beta are estimates of the unknown parameters, 
\[ \hat{y}_i = \hat{\alpha} + \hat{\beta} \cdot x_i \] estimates the mean function or the systematic part of the regression equation. Since a random variable can be predicted best by the mean function (under the mean squared error criterion), \( \hat{y} \) can be interpreted as the best prediction of \( y \). The difference between the dependent variable \( y \) and its least squares prediction is the least squares residual: 
\[ e = y - \hat{y} = y - (\alpha + \beta \cdot x). \]

A large residual \( e \) can either be due to a poor estimation of the parameters of the model or to a large unsystematic part of the regression equation.

For the OLS model to be the best estimator of the relationship between \( x \) and \( y \) several conditions (full ideal conditions, Gauss-Markov conditions) have to be met.

If the „full ideal conditions“ are met one can argue that the OLS-estimator imitates the properties of the unknown model of the population. This means e.g. that the explanatory variables and the error term are uncorrelated.
GAUSS-MARKOV ASSUMPTIONS, FULL IDEAL CONDITIONS OF OLS

The full ideal conditions consist of a collection of assumptions about the true regression model and the data generating process and can be thought of as a description of an ideal data set. Ideal conditions have to be met in order for OLS to be a good estimate (BLUE, unbiased and efficient).

Most real data do not satisfy these conditions, since they are not generated by an ideal experiment. However, the linear regression model under full ideal conditions can be thought of as being the benchmark case with which other models assuming a more realistic DGP should be compared.

One has to be aware of ideal conditions and their violation to be able to control for deviations from these conditions and render results unbiased or at least consistent:

1. Linearity in parameters alpha and beta: the DV is a linear function of a set of IV and a random error component
   → Problems: non-linearity, wrong determinants, wrong estimates; a relationship that is actually there can not be detected with a linear model

2. The expected value of the error term is zero for all observations
   \[ E(\varepsilon_i) = 0 \]
   → Problem: intercept is biased
3. Homoskedasticity: The conditional variance of the error term is constant in all x and over time: the error variance is a measure of model uncertainty. Homoskedasticity implies that the model uncertainty is identical across observations.

\[ V(\varepsilon_i) = E(\varepsilon_i^2) = \sigma_{\varepsilon}^2 = \text{constant} \]

→ Problem: heteroskedasticity – variance of error term is different across observations – model uncertainty varies from observation to observation – often a problem in cross-sectional data, omitted variables bias

4. Error term is independently distributed and not correlated, no correlation between observations of the DV.

\[ \text{Cov}(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i \varepsilon_j) = 0, \ i \neq j \]

→ Problem: spatial correlation (panel and cross-sectional data), serial correlation/autocorrelation (panel and time-series data)
5. **Xi is deterministic:** $x$ is uncorrelated with the error term since $x_i$ is deterministic:

\[
\text{Cov}(X_i, \varepsilon_i) = E(X_i \varepsilon_i) - E(X_i) * E(\varepsilon_i) \\
= X_i E(\varepsilon_i) - X_i E(\varepsilon_i) \rightarrow \text{since } X_i \text{ is det} \\
= 0
\]

$\rightarrow$ **Problems:** omitted variable bias, endogeneity and simultaneity

6. **Other problems:** measurement errors, multicollinearity

If all Gauss-Markov assumptions are met than the OLS estimators alpha and beta are BLUE – best linear unbiased estimators:

- **best:** variance of the OLS estimator is minimal, smaller than the variance of any other estimator
- **linear:** if the relationship is not linear – OLS is not applicable.
- **unbiased:** the expected values of the estimated beta and alpha equal the true values describing the relationship between $x$ and $y$. 
Is it possible to generalize the regression results for the sample under observation to the universe of cases (the population)?

Can you draw conclusions for individuals, countries, time-points beyond those observations in your data-set?

Significance tests are designed to answer exactly these questions.

If a coefficient is significant (p-value<0.10, 0.05, 0.01) then you can draw conclusions for observations beyond the sample under observation.

But…

Only in case the samples matches the characteristics of the population

This is normally the case if all (Gauss-Markov) assumptions of OLS regressions are met by the data under observation.

If this is not the case the standard errors of the coefficients might be biased and therefore the result of the significance test might be wrong as well leading to false conclusions.
SIGNIFICANCE TEST: THE T-TEST

- Test whether the coefficient beta is significantly different from zero: t-test follows the student t-distribution – this is the simplest test for significance.
- Inference from a sample to a population
- To be able to calculate the t-statistic we first need an estimate of the precision of the regression coefficient beta:
- **Standard error of beta**: measures the precision of the regression coefficient, it equals the standard deviation of beta if we could estimate beta an infinite number of times (or at least a very large number)
- Since in reality we have only 1 estimated beta we have to approximate the variation of beta.
- This is done by using the variation of x and the overall error-term of the regression:

\[
\sigma^2_\beta = \frac{\sigma_e^2}{N \cdot \text{Var}(X)} \quad , \quad \text{s.e.}(\beta) = \sqrt{\frac{SSR}{N \cdot \text{Var}(X)}}
\]
The t-test:

T-test for significance: testing the H0 (Null-Hypothesis) that beta equals zero: H0: beta=0; HA: beta≠0

The test statistic follows a student t distribution under the Null

\[
\frac{\hat{\beta} - r}{SE(\hat{\beta})} = \frac{\hat{\beta} - r}{\sqrt{\frac{SSR}{N \cdot Var(X)}}} \sim t_{(n-2)}
\]

\[
\frac{\hat{\beta}}{SE(\hat{\beta})} = \frac{\hat{\beta}}{\sqrt{\frac{SSR}{N \cdot Var(X)}}} \sim t_{(n-2)}
\]

t is the critical value of a t – distribution for a specific number of observations and a specific level of significance: convention in statistics is a significance level of 5% (2.5% on each side of the t-distribution for a 2-sided test) – this is also called the p-value.
Significance test – rule of thumb:

If the regression-coefficient (beta) is at least twice as large as the corresponding standard error of beta the result is statistically significant at the 5% level.
POWER OF A TEST

For a given test statistic and a critical region of a given significance level we define the probability of rejecting the null hypothesis as the power of a test.

The power would be optimal if the probability of rejecting the null would be 0 if there is a relationship and 1 otherwise.

This is, however, not the case in reality. There is always a positive probability to draw the wrong conclusions from the results:

One can reject the null-hypothesis even though it is true (type 1 error, alpha error):

\[ \alpha = \Pr[\text{Type I Error}] = \Pr[\text{rejecting the } H_0 \mid H_0 \text{ is true}] \]

Or not reject the null-hypothesis even though it is wrong (type 2 error, beta error):

\[ \beta = \Pr[\text{Type II Error}] = \Pr[\text{accepting the } H_0 \mid H_a \text{ is true}] \]
Confidence intervals give us a range of numbers that are plausible and are compatible with the hypothesis.

As for significance test the researcher has to choose the level of confidence (95% is convention)

Using the same example again: estimated beta is 1 and the SE(beta) is 0.4; the critical value of the two-sided t-distribution are 1.96 and -1.96
Calculation of the confidence interval:

The question is how far can a hypothetical value differ from the estimated result before they become incompatible with the estimated value?

The regression coefficient $b$ and the hypothetical value $\beta$ are incompatible if either

$$\frac{b - \beta}{SE(b)} > t_{crit} \quad \text{or} \quad \frac{b - \beta}{SE(b)} < -t_{crit}$$

That is if $\beta$ satisfies the double inequality:

$$b - SE(b) * t_{crit} \leq \beta \leq b + SE(b) * t_{crit}$$

Any hypothetical value of $\beta$ that satisfies this inequality will therefore automatically be compatible with the estimate $b$, that is will not be rejected. The set of all such values, given by the interval between the lower and upper limits of the inequality, is known as the confidence interval for $b$. The centre of the confidence interval is the estimated $b$.

If the 5% significance level is adopted the corresponding confidence interval is known as the 95% confidence interval (1% - 99%).
DEFINITIONS

Total Sum of Squares (SST):

\[ SST = \sum_{i=1}^{n} (y_i - \bar{y})^2 \]

Explained (Estimation) Sum of Squares (SSE):

\[ SSE = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \]

Residual Sum of Squares or Sum of Squares Residuals (SSR):

\[ SSR = \sum_{i=1}^{n} \hat{e}_i^2 = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 \]
GOODNESS OF FIT

How well does the explanatory variable explain the dependent variable?

How well does the regression line fit the data?

The R-squared (coefficient of determination) measures how much variation of the dependent variable can be explained by the explanatory variables.

The $R^2$ is the ratio of the explained variation compared to the total variation: it is interpreted as the fraction of the sample variation in $y$ that is explained by $x$.

Explained variation of $y$ / total variation of $y$:

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{Y} - \bar{Y})^2}{\sum_{i=1}^{n} (Y - \bar{Y})^2} = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$
Properties of $R^2$:

$0 \leq R^2 \leq 1$, often the $R^2$ is multiplied by 100 to get the percentage of the sample variation in y that is explained by $x$.

If the data points all lie on the same line, OLS provides a perfect fit to the data. In this case the $R^2$ equals 1 or 100%.

A value of $R^2$ that is nearly equal to zero indicates a poor fit of the OLS line: very little of the variation in the $y$ is captured by the variation in the $y_{\text{hat}}$ (which all lie on the regression line).

$R^2 = (\text{corr}(y, y_{\text{hat}}))^2$

The $R^2$ follows a complex distribution which depends on the explanatory variable.

Adding further explanatory variables leads to an increase the $R^2$.

The $R^2$ can have a reasonable size in spurious regressions if the regressors are non-stationary.

Linear transformations of the regression model change the value of the $R^2$ coefficient.

The $R^2$ is not bounded between 0 and 1 in models without intercept.
PROPERTIES OF AN ESTIMATOR
1. FINITE SAMPLE PROPERTIES

There are often more than 1 possible estimators to estimate a relationship between x and y (e.g. OLS or Maximum Likelihood).

How do we choose between two estimators: the 2 mostly used selection criteria are bias and efficiency.

Bias and efficiency are finite sample properties, because they describe how an estimator behaves when we only have a finite sample (even though the sample might be large).

In comparison so called “asymptotic properties” of an estimator have to do with the behaviour of estimators as the sample size grows without bound.

Since we always deal with finite samples and it is hard to say whether asymptotic properties translate to finite samples, examining the behaviour of estimators in finite samples seems to be more important.
**UNBIASEDNESS**

UnBiasedness: the estimated coefficient is on average true:

That is: in repeated samples of size n the mean outcome of the estimate equals the true – but unknown – value of the parameter to be estimated.

\[ E(\hat{\beta} - \beta) = 0 \]

If an estimator is unbiased, then its probability distribution has an expected value equal to the parameter it is supposed to be estimating. Unbiasedness does not mean that the estimate we get with any particular sample is equal to the true parameter or even close. Rather the mean of all estimates from infinitely drawn random samples equals the true parameter.
SAMPLING VARIANCE OF AN ESTIMATOR

Efficiency: is a relative measure between two estimators – measures the sampling variance of an estimator: $V(\beta)$

Let $\hat{\beta}$ and $\tilde{\beta}$ be two unbiased estimator of the true parameter $\beta$. With variances $V[\hat{\beta}]$ and $V[\tilde{\beta}]$. Then $\hat{\beta}$ is called to be relative more efficient than $\tilde{\beta}$ if $V[\hat{\beta}]$ is smaller than $V[\tilde{\beta}]$.

The property of relative efficiency only helps us to rank two unbiased estimators.
TRADE-OFF BETWEEN BIAS AND EFFICIENCY

With real world data and the related problems we sometimes have only the choice between a biased but efficient and an unbiased but inefficient estimator. Then another criterion can be used to choose between the two estimators, the root mean squared error (RMSE). The RMSE is a combination of bias and efficiency and gives us a measure of overall performance of an estimator.

RMSE:

\[
RMSE(\hat{\beta}) = \frac{1}{K} \sum_{k=1}^{K} \sqrt{(\hat{\beta} - \beta_{true})^2}
\]

\[
MSE = E\left[(\hat{\beta} - \beta_{true})^2\right]
\]

\[
MSE = Var(\hat{\beta}) + \left[\text{Bias}\left(\hat{\beta}, \beta_{true}\right)^2\right]
\]

\(k\) measures the number of experiments, trials or simulations.
ASYMPTOTIC PROPERTIES OF ESTIMATORS

We can rule out certain silly estimators by studying the asymptotic or large sample properties of estimators. We can say something about estimators that are biased and whose variances are not easily found. Asymptotic analysis involves approximating the features of the sampling distribution of an estimator.

Consistency: how far is the estimator likely to be from the true parameter as we let the sample size increase indefinitely. If \( N \to \infty \) the estimated beta equals the true beta:

\[
\lim_{n \to \infty} \Pr \left( \left| \hat{\beta}_n - \beta \right| > \varepsilon \right) = 0, \quad \text{plim} \hat{\beta}_n = \beta,
\]

\[
\lim_{n \to \infty} E \left[ \hat{\beta}_n \right] = \beta
\]

Unlike unbiasedness, consistency involves that the variance of the estimator collapses around the true value as \( N \) approaches infinity. Thus unbiased estimators are not necessarily consistent, but those whose variance shrink to zero as the sample size grows are consistent.
MULTIPLE REGRESSIONS

In most cases the dependent variable $y$ is not just a function of a single explanatory variable but a combination of several explanatory variables.

THUS: drawback of binary regression: impossible to draw ceteris paribus conclusions about how $x$ affects $y$ (omitted variable bias).

Models with $k$-independent variables:

$$y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \varepsilon_i$$

Control for omitted variable bias
But: increases inefficiency of the regression since explanatory variables might be collinear.
OBTAINING OLS ESTIMATES IN MULTIPLE REGRESSIONS

\[ y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i \]

\[
\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_{i2} - \bar{x}_2)^2 \sum_{i=1}^{n} (x_{i1} - \bar{x}_1)(y_i - \bar{y}) - \sum_{i=1}^{n} (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2) \sum_{i=1}^{n} (x_{i2} - \bar{x}_2)(y_i - \bar{y})}{\sum_{i=1}^{n} (x_{i1} - \bar{x}_1)^2 \sum_{i=1}^{n} (x_{i2} - \bar{x}_2)^2 - \left(\sum_{i=1}^{n} (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)\right)^2}
\]

\[
\hat{\beta} = X' X^{-1} X' y
\]

The intercept is the predicted value of \( y \) when all explanatory variables equal zero.

The estimated betas have partial effect or ceteris paribus interpretations.

We can obtain the predicted change in \( y \) given the changes in each \( x \). When \( x_2 \) is held fixed then \( \beta_1 \) gives the change in \( y \) if \( x_1 \) changes by one unit.
“HOLDING OTHER FACTORS FIXED”

The power of multiple regression analysis is that it provides a ceteris paribus interpretation even though the data have not been collected in a ceteris paribus fashion.

Example: multiple regression coefficients tell us what effect an additional year of education has on personal income if we hold social background, intelligence, sex, number of children, marital status and all other factors constant that also influence personal income.
STANDARD ERROR AND SIGNIFICANCE IN MULTIPLE REGRESSIONS

\[ \text{Var} \left( \hat{\beta}_1 \right) = \frac{\sigma^2}{SST_1 \left( 1 - R_1^2 \right)} \implies \hat{\sigma}^2 = \frac{1}{n - (k + 1)} \sum_{i=1}^{n} \varepsilon_i^2 \]

\[ SST_1 = \sum_{i=1}^{n} (x_{i1} - \bar{x}_1)^2 \]

\[ R_1^2 \text{ for the regression of } x_{i1} \text{ on } x_{i2} : R_1^2 = \frac{SSE}{SST} = \frac{\sum_{i=1}^{n} (\hat{x}_{i1} - \bar{x}_1)^2}{\sum_{i=1}^{n} (x_{i1} - \bar{x}_1)^2} \]

\[ SD \left( \hat{\beta}_1 \right) = SE \left( \hat{\beta}_1 \right) = \frac{\sigma}{\sqrt{SST_1 \left( 1 - R_1^2 \right)}} \]
F – TEST: TESTING MULTIPLE LINEAR RESTRICTIONS

t-test (as significance test) is associated with any OLS coefficient. We also want to test multiple hypotheses about the underlying parameters beta_0…beta_k.

The F-test, tests multiple restriction: e.g. all coefficients jointly equal zero:
H0: beta_0=beta_1=…=beta_k=0
Ha: H0 is not true, thus at least one beta differs from zero

The F-statistic (or F-ratio) is defined as:
$$ F = \frac{(SSR_r - SSR_w) / q}{SSR_w / (n - k - 1)} $$

The F-test for overall significance of a regression: H0: all coefficients are jointly zero – in this case we can also compute the F-statistic by using the $R^2$ of the Regression:
$$ F = \frac{R^2 / k}{(1 - R^2) / (n - k - 1)} $$
SSR_r: Sum of Squared Residuals of the restricted model (constant only)

SSR_ur: Sum of Squared Residuals of the unrestricted model (all regressors)

SSR_r can be never smaller than SSR_ur → F is always non-negative

k – number of explanatory variables (regressors), n – number of observations, q – number of exclusion restrictions (q of the variables have zero coefficients): q = df_r – df_ur (difference in degrees of freedom between the restricted and unrestricted models; df_r > df_ur)

The F-test is a one sided test, since the F-statistic is always non-negative

We reject the Null at a given significance level if F>F_critical for this significance level.

If H0 is rejected than we say that all explanatory variables are jointly statistically significant at the chosen significance level.

THUS: The F-test only allows to not reject H0 if all t-tests for all single variables are insignificant too.
Goodness of Fit in multiple Regressions:

As with simple binary regressions we can define SST, SSE and SSR. And we can calculate the $R^2$ in the same way.

BUT: $R^2$ never decreases but tends to increase with the number of explanatory variables.

THUS, $R^2$ is a poor tool for deciding whether one variable of several variables should be added to a model.

We want to know whether a variable has a nonzero partial effect on $y$ in the population.

Adjusted $R^2$: takes the number of explanatory variables into account since the $R^2$ increases with the number of regressors:

$$R^2_{adj} = 1 - \frac{n-1}{n-k}(1 - R^2)$$

$k$ is the number of explanatory variables and $n$ the number of observations.
The size of the slope parameters depends on the scaling of the variables (on which scale a variable is measured), e.g. population in thousands or in millions etc.

To be able to compare the size effects of different explanatory variables in a multiple regression we can use standardized coefficients:

\[
\hat{b}_j = \frac{\hat{\sigma}_{x_j} \hat{\beta}_j}{\hat{\sigma}_y}
\]

for \( j = 1, ..., k \)

Standardized coefficients take the standard deviation of the dependent and explanatory variables into account. So they describe how much y changes if x changes by one standard deviation instead of one unit. If x changes by 1 SD – y changes by \( \hat{b}_\text{hat} \) SD. This makes the scale of the regressors irrelevant and we can compare the magnitude of the effects of different explanatory variables (the variables with the largest standardized coefficient is most important in explaining changes in the dependent variable).
SESSION 2

Testing for Violations of Gauss-Markov Assumptions
Solutions to deal with most common violations

Specification Issues:
Non-linear Relationships
Interaction Effects and Dummy Variables
PROBLEMS IN MULTIPLE REGRESSIONS: 1. MULTICOLLINEARITY

Perfect multicollinearity leads to drop out of one of the variables: if x1 and x2 are perfectly correlated (correlation of 1) – the statistical program at hand does the job.

The higher the correlation the larger the population variance of the coefficients, the less efficient the estimation and the higher the probability to get erratic point estimates. Multicollinearity can result in numerically unstable estimates of the regression coefficients (small changes in X can result in large changes to the estimated regression coefficients).

Trade off between omitted variable bias and inefficiency due to multicollinearity.
TESTING FOR MULTICOLLINEARITY

Correlation between explanatory variables: Pairwise colinearity can be determined from viewing a correlation matrix of the independent variables. However, correlation matrices will not reveal higher order colinearity.

Variance Inflation Factor (vif): measures the impact of collinearity among the x in a regression model on the precision of estimation. vif detects higher order multicolinearity: one or more x is/are close to a linear combination of the other x.

Variance inflation factors are a scaled version of the multiple correlation coefficient between variable j and the rest of the independent variables. Specifically,

\[ VIF_j = \frac{1}{1 - R_j^2} \]

where \( R_j \) is the multiple correlation coefficient.

Variance inflation factors are often given as the reciprocal of the above formula. In this case, they are referred to as the tolerances.

If \( R_j \) equals zero (i.e., no correlation between \( X_j \) and the remaining independent variables), then \( VIF_j \) equals 1. This is the minimum value. Neter, Wasserman, and Kutner (1990) recommend looking at the largest VIF value. A value greater than 10 is an indication of potential multicolinearity problems.
POSSIBLE SOLUTIONS

Reduce the overall error: by including explanatory variables not correlated with other variables but explaining the dependent variable

Drop variables which are highly multi-collinear

Increase the variance by increasing the number of observations

Increase the variance of the explanatory variables

If variables are conceptually similar – combine them into a single index, e.g. by factor or principal component analysis
2. OUTLIERS

Problem:

The OLS principle implies the minimization of squared residuals. From this follows that extreme cases can have a strong impact on the regression line. Inclusion/exclusion of extreme cases might change the results significantly.

The slope and intercept of the least squares line is very sensitive to data points which lie far from the true regression line. These points are called outliers, i.e. extreme values of observed variables that can distort estimates of regression coefficients.
TEST FOR OUTLIERS

symmetry (symplot) and normality (dotplot) of dependent variable gives first indication for outlier cases.

Residual-vs.-fitted plots (rvfplot) indicate which observations of the DV are far away from the predicted values.

Ivr2plot is the leverage against residual squared plot. The upper left corner of the plot will be points that are high in leverage and the lower right corner will be points that are high in the absolute of residuals. The upper right portion will be those points that are both high in leverage and in the absolute of residuals.

DFBETA: how much would the coefficient of an explanatory variable change if we omitted one observation?

DFBETA measures how much impact each observation has on a particular coefficient is. The DFBETA for an explanatory variable and for a particular observation is the difference between the regression coefficient calculated for all of the data and the regression coefficient calculated with the observation deleted, scaled by the standard error calculated with the observation deleted. The cut-off value for DFBETAs is 2/sqrt(n), where n is the number of observations.
SOLUTIONS: OUTLIERS

Include or exclude obvious outlier cases and check their impact on the regression coefficients.

Logarithmize the dependent variable and possibly the explanatory variables as well – this reduces the impact of larger values.

jacknife, bootstrap:

Are both tests and solutions at the same time: they show whether single observations have an impact on the results. If so, one can use the jacknifed and bootstrapped coefficients and standard errors which are more robust to outliers than normal OLS results.

Jacknife: takes the original dataset, runs the same regression N-1 times, leaving one observation out at a time.

Example command in STATA: „jacknife _b _se, eclass: reg spend unem growthpc depratio left cdem trade lowwage fdi skand“

Bootstrapping is a re-sampling technique: for the specified number of repetitions, the same regression is run for a different sample randomly drawn from the original dataset.

Example command: „bootstrap _b _se, reps(1000): reg spend unem growthpc depratio left cdem trade lowwage fdi skand“
3. OMITTED VARIABLE BIAS

The effect of omitted variables that ought to be included:

Suppose the dependent variable $y$ depends on two explanatory variables:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$$

But you are unaware of the importance of $x_2$ and only include $x$

$$y_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i$$

If $x_2$ is omitted from the regression equation, $x_1$ will have a “double” effect on $y$ (a direct effect and one mimicking $x_2$)

The mimicking effect depends on the ability of $x_1$ to mimic $x_2$ (the correlation) and how much $x_2$ would explain $y$

Beta1 in the second equation is biased upwards in case $x_1$ and $x_2$ are positively correlated and downward biased otherwise

Beta1 is only unbiased if $x_1$ and $x_2$ are not related ($\text{corr}(x_1,x_2)=0$

However: including variable that is unnecessary, because it does not explain an variation in $y$ the regression becomes inefficient and the reliability of point estimates decreases.
TESTING FOR OMITTED VARIABLES

Heteroskedasticity of the error term with respect to the observation of a specific independent variable is a good indication for omitted variable bias:

Plot the error term against all explanatory variables

Ramsey RESET F-test for omitted variables in the whole model: tests for wrong functional form (if e.g. an interaction term is omitted):

• Regress $Y$ on the $X$'s and keep the fitted value $Y_{\hat{}}$;
• Regress $Y$ on the $X$'s, and $Y_{\hat{}}^2$ and $Y_{\hat{}}^3$.
• Test the significance of the fitted value terms using an F test.

Szroeter test for monotonic variance of the error term in the explanatory variables

Solutions:

Include variables that are theoretically important and have a high probability of being correlated with one or more variables in the model and explaining significant parts of the variance in the DV.

Fixed unit effects for unobserved unit heterogeneity (time invariant unmeasurable characteristics of e.g. countries – culture, institutions)
4. HETEROSKEDASTICITY

The variance of the error term is not constant in each observation but dependent on unobserved effects, not controlling for this problem violates one of the basic assumptions of linear regressions and renders the estimation results inefficient.

Possible causes:
Omitted variables, for example: spending might vary with the economic size of a country, but size is not included in the model.

Test:
plot the error term against all independent variables
White test if the form of Heteroskedasticity is unknown
Breusch-Pagan Lagrange Multiplier test if the form is known

Solutions:
Robust Huber-White sandwich estimator (GLS)
White Heteroskedasticity consistent VC estimate: manipulates the variance-covariance matrix of the error term.
More substantially: include omitted variables
Dummies for groups of individuals or countries that are assumed to behave more similar than others
Tests for Heteroskedasticity:

a. Breusch-Pagan LM test for known form of Heteroskedasticity: groupwise

\[ LM = \frac{T}{2} \sum_{i=1}^{n} \left( \frac{s_i^2}{s^2} - 1 \right)^2 \]

- \( s_i^2 \) = sum of group-specific squared residuals
- \( s^2 \) = OLS residuals

H0: homoskedasticity \( \sim \) \( \chi^2 \) with \( n-1 \) degrees of freedom

LM-test assumes normality of residuals, not appropriate if assumption not met.

b. Likelihood Ratio Statistic

Residuals are computed using MLE (e.g. iterated FGLS, OLS loss of power)

\[ -2 \ln (\lambda) = (NT) \ln (\sigma^2) - \Sigma \left( T \ln (\sigma_i^2) \right) \sim \chi^2 (dF = n - 1) \]
c. White test if form of Heteroskedasticity is unknown:

H0: \[ V[\varepsilon_i | x_i] = \sigma^2 \]
Ha: \[ V[\varepsilon_i | x_i] = \sigma_i^2 \]

1. Estimate the model under H0 \( \hat{e}_i^2 \)
2. Compute squared residuals:
3. Use squared residuals as dependent variable of auxiliary regression: RHS: all regressors, their quadratic forms and interaction terms
   \[ e_i^2 = \delta_0 + \delta_1 x_{i2} + ... + \delta_{k-1} x_{ik} + \delta_{k-1} x_{i2}^2 + \delta_{k+1} x_{i2} x_{i3} + ... + \delta_q x_{ik}^2 + \xi_i \]
4. Compute White statistic from \( R^2 \) of auxiliary regression:
   \[ n^* R^2 \xrightarrow{a} \chi^2_{(q)} \]
5. Use one-sided test and check if \( n^* R^2 \) is larger than 95\% quantile of \( \chi^2 \)-distribution
Robust White Heteroskedasticity Consistent Variance-Covariance Estimator:

Normal Variance of beta:

\[ \sigma^2_\beta = \frac{\sigma^2_\varepsilon}{N \cdot \text{Var}(X)} \]

Robust White VC matrix:

\[
\hat{V} \left[ \hat{\beta} \right] = \frac{1}{n} \hat{\Omega} \\
\hat{\Omega} = n \left( X'X \right)^{-1} X' \hat{D} X \left( X'X \right)^{-1} \\
\hat{D} = \text{diag} \left[ e_i^2 \right] 
\]

D is a n*n matrix with off-diagonals=0 and diagonal the squared residuals. The normal variance covariance matrix is weighted by the non-constant error variance. Robust Standard errors therefore tend to be larger.
GENERALIZED LEAST SQUARES APPROACHES

The structure of the variance covariance matrix Omega is used not just to adjust the standard errors but also the estimated coefficient.

GLS can be an econometric solution to many violations of the G-M conditions (Autocorrelation, Heteroskedasticity, Spatial Correlation…), since the Omega Matrix can be flexibly specified.

Since the Omega matrix is not known, it has to be estimated and GLS becomes FGLS (Feasible Generalized Least Squares).

All FGLS approaches are problematic if number of observations is limited – very inefficient, since the Omega matrix has to be estimated.

Beta:

\[
\hat{\beta} = \left( \sum_{i=1}^{N} X_i' \Omega^{-1} X_i \right)^{-1} \left( \sum_{i=1}^{N} X_i' \Omega^{-1} y_i \right)
\]

\[
\hat{\beta} = \left( \sum_{i=1}^{N} X_i' \hat{\Omega}^{-1} X_i \right)^{-1} \left( \sum_{i=1}^{N} X_i' \hat{\Omega}^{-1} y_i \right)
\]

Estimated covariance matrix:

\[
\left( X' \Omega^{-1} X \right)^{-1}
\]
Omega matrix with heteroscedastic error structure and contemporaneously correlated errors, but in principle FGLS can handle all different correlation structures…:

\[
\begin{pmatrix}
\varepsilon_1^2 & \varepsilon_{21} & \varepsilon_{31} & \ldots & \varepsilon_{n1} \\
\varepsilon_{12} & \varepsilon_2 & \varepsilon_{32} & \ldots & \varepsilon_{n2} \\
\varepsilon_{13} & \varepsilon_{23} & \varepsilon_3^2 & \ldots & \varepsilon_{n3} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\varepsilon_{1n} & \varepsilon_{2n} & \varepsilon_{3n} & \ldots & \varepsilon_n^2
\end{pmatrix}
\]
HETEROSKEDASTICITY IN POOLED DATA

Is a bigger problem than in pure cross-sections since the error term of one unit i can be correlated with the error term of unit j at the same point in time and different points in time!

Test: Breusch-Pagan test, Cook-Weisberg test (estat hettest after reg) – for unspecific heteroskedasticity, Ramsey test for omitted variables (estat ovtest after reg), Szroeter’s test of heteroskedasticity for specific right-hand side variables (estat szroeter).

Solutions:
• White Heteroskedasticity consistent VC estimate
• More substantially: include omitted variables
• Parks-Kmenta approach: FGLS
• Beck/ Katz: Panel Corrected Standard Errors
• More theoretically: Dummies for groups of individuals or countries that are assumed to behave more similar than others
THE PARKS-KMENTA APPROACH

Basically a Feasible Generalized Least Squares Estimator which takes the two-dimensional nature of the pooled data into account.

As all FGLS – problematic if number of observations is limited – very inefficient, since the Omega matrix has to be estimated

\[
\left( \sum_{i=1}^{N} X_i \Omega^{-1} X_i \right)^{-1} \left( \sum_{i=1}^{N} X_i \Omega^{-1} y_i \right) \Rightarrow \left( \sum_{i=1}^{N} X_i \hat{\Omega}^{-1} X_i \right)^{-1} \left( \sum_{i=1}^{N} X_i \hat{\Omega}^{-1} y_i \right)
\]

The estimated covariance matrix is

\[
\left( X \Omega^{-1} X \right)^{-1}
\]
Parks (1967) and Kmenta suggested an Omega matrix with panel specific AR1 error structure and contemporaneously correlated errors, but in principle FGLS can handle all different correlation structures...

\[
\begin{bmatrix}
\varepsilon_{11} & \varepsilon_{21} & \varepsilon_{31} & \cdots & \varepsilon_{n1} \\
\varepsilon_{12} & \varepsilon_{22} & \varepsilon_{32} & \cdots & \varepsilon_{n2} \\
\varepsilon_{13} & \varepsilon_{23} & \varepsilon_{33} & \cdots & \varepsilon_{n3} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\varepsilon_{1T} & \varepsilon_{2T} & \varepsilon_{3T} & \cdots & \varepsilon_{NT}
\end{bmatrix}
\]

\[
\Omega = \sum \otimes V_{ij} = 
\begin{bmatrix}
\sigma_{11} V_{11} & \sigma_{12} V_{12} & \sigma_{13} V_{13} & \cdots & \sigma_{1n} V_{1n} \\
\sigma_{21} V_{21} & \sigma_{22} V_{22} & \sigma_{23} V_{23} & \cdots & \sigma_{2n} V_{2n} \\
\sigma_{31} V_{31} & \sigma_{32} V_{32} & \sigma_{33} V_{33} & \cdots & \sigma_{3n} V_{3n} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\sigma_{n1} V_{n1} & \sigma_{n2} V_{n2} & \sigma_{n3} V_{n3} & \cdots & \sigma_{nn} V_{nn}
\end{bmatrix}
\]

\[
V_{ij} = 
\begin{bmatrix}
1 & \rho_j & \rho_j & \cdots & \rho_j^{T-1} \\
\rho_i & 1 & \rho_j & \cdots & \rho_j^{T-2} \\
\rho_i^2 & \rho_i & 1 & \cdots & \rho_j^{T-3} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\rho_i^{T-1} & \rho_i^{T-2} & \rho_i^{T-3} & \cdots & 1
\end{bmatrix}
\]
The Parks method combines two sequential FGLS transformations, first eliminating serial correlation of the errors then eliminating contemporaneous correlation of the errors.

This is done by initially estimating an OLS equation. The residuals from this estimation are used to estimate the unit-specific serial correlation of the errors, which are then used to transform the model into one with serially independent errors.

Residuals from this estimation are then used to estimate the contemporaneous correlation of the errors, and the data is once again transformed to allow for OLS estimation with now spherical errors.

Assumptions $E\left( \varepsilon_{it}^2 \right) = \sigma_i^2$, cross section specific variation

$E\left( \varepsilon_{it} \varepsilon_{jt} \right) = \sigma_{ij}$, for $i \neq j$, cross-sectional correlation

$\varepsilon_{it} = \rho_i \varepsilon_{it-1} + \xi_{it}$, first-order serial correlation
PANEL CORRECTED STANDARD ERRORS

Beck and Katz 1995: APSR – stata command: xtpcse – as a critique of the parks method (FGLS) that only can be used if T is as least as large as N: MC experiments show that OLS PCSE is more efficient than FGLS and at the same time avoids over-confidence in the estimation results.

Coefficient estimates of OLS are consistent but inefficient in pooled data; degree of inefficiency depends on data and exact error process.

B&K suggest to use OLS and correct only for the panel structure in the standard errors:

\[
Var[\beta] = (X'X)^{-1}X'\Omega X(X'X)^{-1}
\]

where

\[
\Omega = (E'E/T) \otimes I_T
\]
For panel models with contemporaneously correlated and panel heteroscedastic errors, Omega is an NT x NT block diagonal matrix with an N x N matrix of contemporaneous covariances, Sigma, along the diagonal. To estimate Omega, we need an estimate of Sigma. Since the OLS estimates are consistent, one can use the OLS residuals to provide a consistent estimate of Sigma. Let $e_{it}$ be the OLS residual for unit $i$ at time $t$. A typical element of Sigma is estimated by:

$$\hat{\Sigma}_{i,j} = \frac{\sum_{t=1}^{T} e_{i,t}e_{j,t}}{T}$$

With the estimate Sigma_hat being comprised of all these elements. This is used to form the estimator Omega_hat by creating a block diagonal matrix with the Sigma matrices along the diagonal. As the number of time points increases, Sigma_hat becomes an increasingly better estimator of Sigma.
PCSE

Treats cross-sectional correlation as nuisance – however sometimes spatial effects are of theoretical interest

Manipulation of the standard errors of the OLS estimates: the size of the SE (beta) is corrected by accounting for the panel structure of the data

Serial correlation has to be treated beforehand

PCSE estimates are biased whenever OLS estimates are biased
5. AUTOCORRELATION

The observation of the residual in t1 is dependent on the observation in t0: not controlling for autocorrelation violates one of the basic assumptions of OLS and may bias the estimation of the beta coefficients.

Options:

- lagged dependent variable
- differencing the dependent variable
- differencing all variables
- Prais-Winston Transformation of the data
- HAC constituent VC matrix

Tests:

- Durbin-Watson, Durbin’s m, Breusch-Godfrey test
- Regress e on lag(e)
Autocorrelation

The error term in $t1$ is dependent on the error term in $t0$: not controlling for autocorrelation violates one of the basic assumptions of OLS and may bias the estimation of the beta coefficients.

$$\varepsilon_{it} = \rho \varepsilon_{i,t-1} + \xi_{it}$$

The residual of a regression model picks up the influences of those variables affecting the DV that have not been included in the regression equation. Thus, persistence in excluded variables is the most frequent cause of autocorrelation.

Autocorrelation does make no predictions about a trend, though a trend in the DV is often a sign for serial correlation.

Positive autocorrelation: rho is positive: it is more likely that a positive value of the error-term is followed by a one and a negative by a negative one.

Negative autocorrelation: rho is negative: it is more likely that a positive value of the error-term is followed by a negative one and vice versa.
DW test for first order AC:

\[ d = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} \]

Regression must have an intercept

Explanatory variables have to be deterministic

Inclusion of LDV biases statistic towards 2

Efficiency problem of serial correlation can be fixed by Newey-West

HAC consistent VC matrix for Heteroskedasticity of unknown form and

AC of order p: Problem: VC matrix consistent but coefficient can still be

biased! (HAC is possible with “ivreg2” in stata)

\[
\hat{V}_{NW} \left[ \hat{\beta} \right] = T \left( X'X \right)^{-1} S^* \left( X'X \right)^{-1}
\]

\[
S^* = \frac{1}{T} \sum_{t=1}^{T} e_t^2 x_t x_t' + \frac{1}{T} \sum_{t=1}^{p} \sum_{t=l+1}^{T} \omega_l e_t e_{t-l} \left( x_t x_{t-l} + x_{t-l} x_t \right)
\]

\[
\omega_l = 1 - \frac{1}{p+1}
\]
OR a simpler Test:

Estimate the model by OLS

compute the residuals

Regress the residuals on all independent variables (including the LDV if present) and the lagged residuals

If the coefficient on the lagged residual is significant (with the usual t-test), we can reject the null of independent errors.
LAGGED DEPENDENT VARIABLE

\[ y_{it} = \alpha + \beta_0 y_{it-1} + \beta_k x_{it} + \varepsilon_{it} \]

The interpretation of the LDV as measure of time-persistency is misleading

LDV captures average dynamic effect, this can be shown by Cochrane-Orcutt distributive lag models. Thus LDV assumes that all x-variables have an one period lagged effect on y

→ make sure interpretation is correct – calculating the real effect of x - variables

Is an insignificant coefficient really insignificant if coefficient of lagged y is highly significant?

\[ y_{it} = \alpha + \beta_0 y_{i,t-1} + \beta_1 x_{it} + \varepsilon_{it} \]

\[ \beta_1 = \frac{y_{it} - (\alpha + \beta_0 y_{i,t-1} + \varepsilon_{it})}{x_{it}} \]
FIRST DIFFERENCE MODELS

Differencing only the dependent variable – only if theory predicts effects of levels on changes

FD estimator assumes that the coefficient of the LDV is exactly 1 – this is often not true

Theory predicts effects of changes on changes

Suggested remedy if time series is non-stationary (has a single unit root), asymptotic analysis for $T \to \infty$.

Consistent

$$y_{it} - y_{it-1} = \beta_k \sum_{k=1}^{K} (x_{kit} - x_{kit-1}) + \varepsilon_{it} - \varepsilon_{it-1}$$

$$\equiv \Delta y_{it} = \beta_k \sum_{k=1}^{K} \Delta x_{kit} + \Delta \varepsilon_{it}$$
PRAIS-WINSTEN TRANSFORMATION

Models the serial correlation in the error term – regression results for X variables are more straightforwardly interpretable:

\[ y_{it} = x_{it} \beta + \varepsilon_{it} \quad \text{with} \quad \varepsilon_{it} = \rho \varepsilon_{it-1} + \xi_{it} \]

The \( \xi_{it} \) are iid – with \( N(0, \sigma^2) \)

The VC matrix of the error term is

\[
\Psi = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{bmatrix}
\]

The matrix is stacked for N units. Diagonals are 1.

Prais-Winston is estimated by GLS. It is derived from the AR(1) model for the error term. The first observation is preserved.
1. Estimation of a standard linear regression:

\[ y_{it} = x_{it} \beta + \varepsilon_{it} \]

2. An estimate of the correlation in the residuals is then obtained by the following auxiliary regression:

\[ \varepsilon_{it} = \rho \varepsilon_{it-1} + \xi_{it} \]

3. A Cochrane-Orcutt transformation is applied for observations \( t=2,\ldots,n \)

\[ y_{it} - \rho y_{it-1} = \beta \left( x_{it} - \rho x_{it-1} \right) + \zeta_{it} \]

4. And the transformation for \( t=1 \) is as follows:

\[ \sqrt{1-\rho^2} y_1 = \beta \left( \sqrt{1-\rho^2} x_1 \right) + \sqrt{1-\rho^2} \zeta_1 \]

5. With iterating to convergence, the whole process is repeated until the change in the estimate of rho is within a specified tolerance, the new estimates are used to produce fitted values for \( y \) and rho is re-estimated, by:

\[ y_{it} - \hat{y}_{it} = \rho \left( y_{it-1} - \hat{y}_{it-1} \right) + \varepsilon_{it} \]
DISTRIBUTED LAG MODELS

Simplest form is Cochrane-Orcutt – dynamic structure of all independent variables is captured by 1 parameter, either in the error term or as LDV.

If dynamics are that easy – LDV or Prais-Winston is fine – saves Degrees of Freedom.

Problem: if theory predicts different lags for different right hand side variables – than a miss-specified model leads necessarily to bias.

Test down – start with relatively large number of lags for potential candidates:

\[ y_{it} = x_{it} \beta_1 + x_{it-1} \beta_2 + x_{it-2} \beta_3 + \ldots + x_{it-n} \beta_{n+1} + \varepsilon_{it} \]

\[ n = 1, \ldots, t - 1 \]
SPECIFICATION ISSUES IN MULTIPLE REgressions: 1. NON-LINEARITY

One or more explanatory variables have a non-linear effect on the dependent variable: estimating a linear model would lead to wrong or/and insignificant results. Thus, even though in the population there exist a relationship between an explanatory variable and the dependent variable, but this relationship cannot be detected due to the strict linearity assumption of OLS.

Test:
Ramsay RESET F-test gives a first indication for the whole model.
In general, we can use acprplot to verify the linearity assumption against an explanatory variable – though this is just “eye-balling”.
Theoretical expectations should guide the inclusion of squared terms.
### ANOVA Table 1:

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 83</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1004.3306</td>
<td>9</td>
<td>111.5923</td>
<td>F(9, 73) = 7.72</td>
</tr>
<tr>
<td>Residual</td>
<td>1054.8994</td>
<td>73</td>
<td>14.4507</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>2059.23</td>
<td>82</td>
<td>25.1126</td>
<td>R-squared = 0.4877</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adj R-squared = 0.4246</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Root MSE = 3.8014</td>
</tr>
</tbody>
</table>

### ANOVA Table 2:

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 83</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1103.7367</td>
<td>10</td>
<td>110.3737</td>
<td>F(10, 72) = 8.32</td>
</tr>
<tr>
<td>Residual</td>
<td>955.4933</td>
<td>72</td>
<td>13.2707</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>2059.23</td>
<td>82</td>
<td>25.1126</td>
<td>R-squared = 0.5360</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adj R-squared = 0.4715</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Root MSE = 3.6429</td>
</tr>
</tbody>
</table>

### Coefficient Table 1:

|            | Coef.   | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|------------|---------|-----------|-------|-----|---------------------|
| polity     | -0.125166| 0.2470514 | -0.51 | 0.614 | -0.6175388 to 0.3672068 |
| loggdpc75  | 0.6462764| 0.2118392 | 3.05  | 0.003 | 2.240218 to 10.68531 |
| h75        | 0.663909 | 0.3262583 | 2.03  | 0.045 | 0.0136771 to 1.314141 |
| openresstd | -0.751177| 0.3103894 | -2.18 | 0.033 | -12.93723 to -5.651253 |
| checks1    | 1.05955  | 0.4947976 | 2.14  | 0.036 | 0.0734202 to 2.045681 |
| asia       | -4.232433| 2.063596 | -2.05 | 0.044 | -8.345175 to -1.1196917 |
| oecd       | -3.741717| 1.877319 | -1.99 | 0.050 | -7.483208 to -0.00227 |
| latam      | -5.439421| 1.656054 | -3.28 | 0.002 | -8.739933 to -2.13891 |
| africa     | 1.362636 | 1.622116 | 0.84  | 0.404 | -1.870237 to 4.595509 |
| _cons      | -5.738767| 5.293315 | -1.08 | 0.282 | -16.28833 to 4.810794 |

### Coefficient Table 2:

|            | Coef.   | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|------------|---------|-----------|-------|-----|---------------------|
| polity     | -2.220887| 0.8014923 | -2.77 | 0.007 | -3.818632 to -0.6231409 |
| politysqr  | 0.1963913| 0.0717568 | 2.74  | 0.008 | 0.0533466 to 0.3394359 |
| loggdpc75  | 5.24893  | 2.078226 | 2.53  | 0.014 | 1.106062 to 9.31799 |
| h75        | 0.6605033| 0.3126572 | 2.11  | 0.038 | 0.0372325 to 1.283774 |
| openresstd | -5.994304| 2.987303 | -2.01 | 0.049 | -11.94938 to -0.0392242 |
| checks1    | 1.035895 | 0.4742455 | 2.18  | 0.032 | 0.0905035 to 1.981286 |
| asia       | -3.348378| 2.00376  | -1.67 | 0.099 | -7.342801 to 0.640645 |
| oecd       | -5.081614| 1.864465 | -2.73 | 0.008 | -8.798357 to -1.36487 |
| latam      | -4.4355  | 1.628843 | -2.72 | 0.008 | -7.68254 to -1.18846 |
| africa     | 1.203143 | 1.555573 | 0.77  | 0.442 | -1.897835 to 4.304121 |
| _cons      | 1.221078 | 5.67433  | 0.22  | 0.830 | -10.09049 to 12.53265 |
Handy solutions without leaving the linear regression framework:

Logarithmize the IV and DV: gives you the elasticity, higher values are weighted less (engel curve – income elasticity of demand). This model is called a log-log model or a log-linear model

\[
\log y_i = \log \alpha + \beta \log x_i + \log \epsilon_i
\]

- Different functional forms give parameter estimates that have different substantial interpretations. The parameters of the linear model have an interpretation as marginal effects. The elasticities will vary depending on the data. In contrast the parameters of the log-log model have an interpretation as elasticities. So the log-log model assumes a constant elasticity over all values of the data set. Therefore the coefficients of a log-linear model can be interpreted as percentage changes – if the explanatory variable changes by one percent the dependent variable changes by beta percent.

- The log transformation is only applicable when all the observations in the data set are positive. This can be guaranteed by using a transformation like \( \log(X+k) \) where \( k \) is a positive scalar chosen to ensure positive values. However, careful thought has to be given to the interpretation of the parameter estimates.

- For a given data set there may be no particular reason to assume that one functional form is better than the other. A model selection approach is to estimate competing models by OLS and choose the model with the highest R-squared.

include an additional squared term of the IV to test for U-shape and inverse U-shape relationships. Careful with the interpretation! The size of the two coefficients (linear and squared) determines whether there is indeed a u-shaped or inverse u-shaped relationship.
Table 1. Democracy and government spending.

<table>
<thead>
<tr>
<th></th>
<th>Model 1 baseline model</th>
<th>Model 2 institutional variables suppressed</th>
<th>Model 3 linear democracy score assumed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1.2211</td>
<td>5.2588</td>
<td>−5.7388</td>
</tr>
<tr>
<td></td>
<td>(5.6743)</td>
<td>(5.6687)</td>
<td>(5.2933)</td>
</tr>
<tr>
<td>Log of per capita income</td>
<td>5.2489</td>
<td>3.7535</td>
<td>6.4628</td>
</tr>
<tr>
<td></td>
<td>(2.0782)**</td>
<td>(1.8858)**</td>
<td>(2.1187)</td>
</tr>
<tr>
<td>Human capital</td>
<td>0.6605</td>
<td>0.7896</td>
<td>0.6639</td>
</tr>
<tr>
<td></td>
<td>(0.3127)**</td>
<td>(0.3212)**</td>
<td>(0.3263)**</td>
</tr>
<tr>
<td>Institutional openness to trade</td>
<td>−5.9943</td>
<td>6.7511</td>
<td>−6.7511</td>
</tr>
<tr>
<td></td>
<td>(2.9873)**</td>
<td>(3.1039)**</td>
<td></td>
</tr>
<tr>
<td>Number of Veto-players</td>
<td>1.0359</td>
<td>1.0595</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.4742)**</td>
<td>(0.4948)**</td>
<td></td>
</tr>
<tr>
<td>Democracy</td>
<td>−2.2209</td>
<td>−2.3369</td>
<td>−0.1252</td>
</tr>
<tr>
<td></td>
<td>(0.8015)**</td>
<td>(0.8272)**</td>
<td>(0.2471)</td>
</tr>
<tr>
<td>Democracy²</td>
<td>0.1964</td>
<td>0.2117</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0718)**</td>
<td>(0.7445)**</td>
<td></td>
</tr>
<tr>
<td>Southeast Asia dummy</td>
<td>−3.3484</td>
<td>−3.9149</td>
<td>−4.2324</td>
</tr>
<tr>
<td></td>
<td>(2.0038)*</td>
<td>(2.0578)*</td>
<td>(2.0636)**</td>
</tr>
<tr>
<td>OECD-dummy</td>
<td>−5.0816</td>
<td>−5.2344</td>
<td>3.7417</td>
</tr>
<tr>
<td></td>
<td>(1.8645)**</td>
<td>(1.9288)**</td>
<td>(1.8773)**</td>
</tr>
<tr>
<td>Latin America dummy</td>
<td>−4.4355</td>
<td>−4.0809</td>
<td>5.4394</td>
</tr>
<tr>
<td></td>
<td>(1.6288)**</td>
<td>(1.4159)**</td>
<td>(1.6561)**</td>
</tr>
<tr>
<td>Africa dummy</td>
<td>1.2031</td>
<td>1.7556</td>
<td>1.3626</td>
</tr>
<tr>
<td></td>
<td>(1.5556)</td>
<td>(1.4890)</td>
<td>(5.2933)</td>
</tr>
<tr>
<td>N</td>
<td>83</td>
<td>83</td>
<td>83</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.4715</td>
<td>0.4261</td>
<td>0.4246</td>
</tr>
<tr>
<td>RMS-residual</td>
<td>955.493</td>
<td>1066.429</td>
<td>1054.899</td>
</tr>
<tr>
<td>F-Statistics</td>
<td>8.32****</td>
<td>8.61****</td>
<td>7.72****</td>
</tr>
<tr>
<td>polity_sqr</td>
<td>polity</td>
<td>govcon</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>--------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>20.049292</td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>0.5</td>
<td>18.987946</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>18.024796</td>
<td></td>
</tr>
<tr>
<td>2.25</td>
<td>1.5</td>
<td>17.159842</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>16.393084</td>
<td></td>
</tr>
<tr>
<td>6.25</td>
<td>2.5</td>
<td>15.724521</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>15.154153</td>
<td></td>
</tr>
<tr>
<td>12.25</td>
<td>3.5</td>
<td>14.681982</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>14.308005</td>
<td></td>
</tr>
<tr>
<td>20.25</td>
<td>4.5</td>
<td>14.032225</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>5</td>
<td>13.85464</td>
<td></td>
</tr>
<tr>
<td>30.25</td>
<td>5.5</td>
<td>13.775251</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>6</td>
<td>13.794057</td>
<td></td>
</tr>
<tr>
<td>42.25</td>
<td>6.5</td>
<td>13.911059</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>7</td>
<td>14.126257</td>
<td></td>
</tr>
<tr>
<td>56.25</td>
<td>7.5</td>
<td>14.43965</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>8</td>
<td>14.851239</td>
<td></td>
</tr>
<tr>
<td>72.25</td>
<td>8.5</td>
<td>15.361024</td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>9</td>
<td>15.969004</td>
<td></td>
</tr>
<tr>
<td>90.25</td>
<td>9.5</td>
<td>16.67518</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>17.479551</td>
<td></td>
</tr>
</tbody>
</table>
The „u“ shaped relationship between democracy and government spending:
2. INTERACTION EFFECTS

Two explanatory variables do not only have a direct effect on the dependent variable but also a combined effect:

\[ y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} \times x_{2i} + \varepsilon_i \]

Interpretation: combined effect:
\[ b_1 \text{SD}(x1) + b_2 \text{SD}(x2) + b_3 \text{SD}(x1 \times x2) \]

Example: monetary policy of currency union has a direct effect on monetary policy in outsider countries but this effect is increased by import shares.
### Example:

**government spending**

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 529</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>37091.493</td>
<td>9</td>
<td>4121.27699</td>
<td>F( 9, 519) = 106.13</td>
</tr>
<tr>
<td>Residual</td>
<td>20154.2047</td>
<td>519</td>
<td>38.8327643</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R-squared = 0.6479</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adj R-squared = 0.6418</td>
</tr>
<tr>
<td>Total</td>
<td>57245.6976</td>
<td>528</td>
<td>108.419882</td>
<td>Root MSE = 6.2316</td>
</tr>
</tbody>
</table>

| spend   | Coef.    | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|---------|----------|-----------|------|------|----------------------|
| unem    | 1.194476 | 0.0896173 | 13.33| 0.000| 1.018419 1.370533   |
| growthpc| -0.8574751| 0.1201086 | -7.14| 0.000| -1.093434  -0.621563|
| depreatio| -0.1772656| 0.1170101 | -1.51| 0.130| -0.407137 0.052606 |
| left    | 0.049314  | 0.0080705 | 6.19 | 0.000| 0.0340765 0.0645784 |
| cdem    | 0.0593482 | 0.0125353 | 4.73 | 0.000| 0.034722  0.0839743 |
| trade   | 0.0883413 | 0.0143161 | 6.17 | 0.000| 0.0602167 0.1164659 |
| lowwage | -0.1183256| 0.0441787 | -2.68| 0.008| -0.205167 -0.0315346|
| fdi     | 0.220538  | 0.192026  | 1.11 | 0.269| -0.1708045 0.6118804 |
| skand   | 6.629869  | 0.6810601 | 9.73 | 0.000| 5.291895 7.967842   |
| _cons   | 37.21619  | 4.703757 | 7.91 | 0.000| 27.97545 46.45693   |

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 529</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>37442.8619</td>
<td>10</td>
<td>3744.28619</td>
<td>F( 10, 518) = 97.94</td>
</tr>
<tr>
<td>Residual</td>
<td>19802.8358</td>
<td>518</td>
<td>38.2294126</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R-squared = 0.6541</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adj R-squared = 0.6474</td>
</tr>
<tr>
<td>Total</td>
<td>57245.6976</td>
<td>528</td>
<td>108.419882</td>
<td>Root MSE = 6.183</td>
</tr>
</tbody>
</table>

| spend   | Coef.    | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|---------|----------|-----------|------|------|----------------------|
| unem    | 1.041448 | 0.1022465 | 10.19| 0.000| 0.8405793 1.242317   |
| growthpc| -0.8668459| 0.119212 | -7.27| 0.000| -1.101044  -0.6326475|
| depreatio| -0.0958857| 0.1191604| -0.80| 0.421| -0.3299827 0.1382113|
| left    | 0.0485801| 0.00802 | 6.06 | 0.000| 0.0328244 0.0643358 |
| cdem    | 0.0099821| 0.02049 | 0.49 | 0.626| 0.00302716 0.0502359 |
| trade   | 0.0846139| 0.0142575| 5.93 | 0.000| 0.0566042 0.1126236 |
| lowwage | -0.1319855| 0.0440651| -3.00| 0.003| -0.2185538 0.054172 |
| fdi     | 0.2116274| 0.1976708 | 1.07 | 0.285| -0.1767076 0.5999625 |
| skand   | 6.481997| 0.6775066 | 9.57 | 0.000| 5.150999 7.812996   |
| cdem_unem| 0.0096447| 0.0031813| 3.03 | 0.003| 0.0033949 0.0158946 |
| _cons   | 35.6862| 4.694279 | 7.60 | 0.000| 26.46403 44.90836   |
Low unemployment High unemployment

Government spending in % of GDI

Low vocational portfolio
High cristian democratic portfolio
### Regressions

#### First Regression

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 529</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>33411.5828</td>
<td>8</td>
<td>4176.44785</td>
<td>F( 8, 520) = 91.12</td>
</tr>
<tr>
<td>Residual</td>
<td>23834.1148</td>
<td>520</td>
<td>45.8348362</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>57245.6976</td>
<td>528</td>
<td>108.419882</td>
<td>R-squared = 0.5837</td>
</tr>
</tbody>
</table>

| Variable | Coef. | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|----------|-------|-----------|-------|------|----------------------------|
| unem     | .9320547 | .0928529 | 10.04 | 0.000 | .7496417 - 1.144468 |
| trade    | .1293114 | .0148666 | 8.70  | 0.000 | .1001066 - .1585163 |
| growthpc | -.9012817 | .1309371 | -6.91 | 0.000 | -.1.157451 - .6451119 |
| depratio | .4446718 | .1296597 | -3.60 | 0.000 | -.6674289 - .2019147 |
| left     | .0580160 | .0087215 | 6.65  | 0.000 | .0408823 - .0751497 |
| cdem     | .0287597 | .0131838 | 2.18  | 0.030 | .0028596 - .0546597 |
| lowwage  | -.157155 | .0478007 | -3.29 | 0.001 | -.2510612 - .0632488 |
| fdi      | .0927398 | .2159476 | 0.43  | 0.668 | -.3314971 - .5169768 |
| _cons    | 48.81442 | 4.943615 | 9.87  | 0.000 | 39.1025 - 58.52633 |

#### Second Regression

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 529</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>33818.6151</td>
<td>9</td>
<td>3757.6239</td>
<td>F( 9, 519) = 83.25</td>
</tr>
<tr>
<td>Residual</td>
<td>23427.0825</td>
<td>519</td>
<td>45.1388873</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>57245.6976</td>
<td>528</td>
<td>108.419882</td>
<td>R-squared = 0.5837</td>
</tr>
</tbody>
</table>

| Variable | Coef. | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|----------|-------|-----------|-------|------|----------------------------|
| unem     | 1.526526 | .2183611 | 6.99  | 0.000 | 1.097546 - 1.955506 |
| trade    | .1981998 | .0272749 | 7.27  | 0.000 | .144617 - .2517825 |
| unem_trade | -.0084099 | .0028006 | -3.00 | 0.003 | -.0139117 - -.002908 |
| growthpc | -.8482902 | .130601 | -6.50 | 0.000 | -.1.104862 - .5917186 |
| depratio | -.4225837 | .1228484 | -3.44 | 0.001 | -.6639249 - -.1812426 |
| left     | .0533617 | .0087927 | 6.07  | 0.000 | .036088 - .0706533 |
| cdem     | .0220469 | .0027279 | 1.66  | 0.097 | -.0040283 - .0481222 |
| lowwage  | -.1056805 | .0504386 | -2.10 | 0.037 | -.2047694 - .0065917 |
| fdi      | .0250327 | .2154848 | 0.12  | 0.908 | -.3982969 - .4483622 |
| _cons    | 43.02368 | 5.271331 | 8.16  | 0.000 | 32.66792 - 53.37945 |
### Simple slope of spend on unem at trade  +/- 1sd

| trade | Coef.   | Std. Err. | t     | P>|t| |
|-------|---------|-----------|-------|-----|
| High  | 1.286722| .1498013  | 8.59  | 0.000 |
| Mean  | 1.526526| .2183611  | 6.99  | 0.000 |
| Low   | 1.76633 | .2927067  | 6.03  | 0.000 |

### Simple slope of spend on trade at unem  +/- 1sd

| unem | Coef.   | Std. Err. | t     | P>|t| |
|------|---------|-----------|-------|-----|
| High | .1673891| .0194534  | 8.60  | 0.000 |
| Mean | .1981998| .0272749  | 7.27  | 0.000 |
| Low  | .2290105| .0363312  | 6.30  | 0.000 |
The diagram shows a scatter plot with the following elements:

- The x-axis represents "trade".
- The y-axis represents "spend".
- The scatter plot includes data points labeled "government spending".
- There are trend lines for "unem+1sd", "unem mean", and "unem-1sd".
INTERACTION EFFECTS OF CONTINUOUS VARIABLES

Marginal Effect of Unemployment on Spending as Trade Openness changes

Dependent Variable: Government Spending

- Marginal Effect of Unemployment on Spending
- 95% Confidence Interval
Thick dashed lines give 95% confidence interval.
Thin dashed line is a kernel density estimate of trade.
3. DUMMY VARIABLES

An explanatory variable that takes on only the values 0 and 1

Example: DV: spending, IV: whether a country is a democracy (1) or not (0).

\[ y_i = \alpha + \beta \cdot D + \varepsilon_i \]

Alpha then is the effect for non-democracies and alpha+beta is the effect for democracies.