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

$$\mathrm{E}\left(\varepsilon_{\mathrm{i}}\right)=0$$

→ 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.

$$V\left(\varepsilon_i\right) = E\left(\varepsilon_i{}^2\right) = \sigma_\varepsilon^2 = \mathrm{cons\,tan\,t}$$

→ 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.

$$\mathrm{Cov}\left(\varepsilon_i, \varepsilon_j\right) = E\left(\varepsilon_i \varepsilon_j\right) = 0,\ i \neq j$$

→ 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 xi is deterministic:

$$\begin{aligned}\mathrm{Cov}\left(X_i, \varepsilon_i\right) &= E\left(X_i \varepsilon_i\right) - E\left(X_i\right) * E\left(\varepsilon_i\right) \\ &= X_i E\left(\varepsilon_i\right) - X_i E\left(\varepsilon_i\right) \rightarrow \text{sin ce } X_i \text{ is det} \\ &= 0\end{aligned}$$

→ Problems: omitted variable bias, endogeneity and simultaneity

6. Other problems: measurement errors, multicolinearity

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.

Inference

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:

$$\sigma_{\beta}^{2} = \frac{\sigma_{\varepsilon}^{2}}{N * Var\left(X\right)} \quad , \quad s.e.(\beta) = \sqrt{\frac{SSR}{N * Var(X)}}$$

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

$$\frac{\hat{\beta}-r}{SE\left(\hat{\beta}\right)}=\frac{\hat{\beta}-r}{\sqrt{\frac{SSR}{N*Var\left(X\right)}}}\sim t_{(n-2)}$$

$$\frac{\hat{\beta}}{SE\left(\hat{\beta}\right)}=\frac{\hat{\beta}}{\sqrt{\frac{SSR}{N*Var\left(X\right)}}}\sim t_{(n-2)}$$

- 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.

Assume beta is 1 and the estimated standard error is 0.8
The critical value of the two-sided symmetric student t-distribution for n=∞ and alpha=5% is 1.96

Acceptance at the 5% level: $t = \dfrac{\hat{\beta} - 0}{SE\left(\hat{\beta}\right)}$

The Null (no significant relationship) will not be rejected if: $-1.96 \le t \le 1.96$

This condition can be expressed in terms of beta by substituting for t: $-1.96 \le \dfrac{\hat{\beta} - 0}{SE\left(\hat{\beta}\right)} \le 1.96$

Multiplying through by the SE of beta: $-1.96 * SE\left(\hat{\beta}\right) \le \hat{\beta} - 0 \le 1.96 * SE\left(\hat{\beta}\right)$

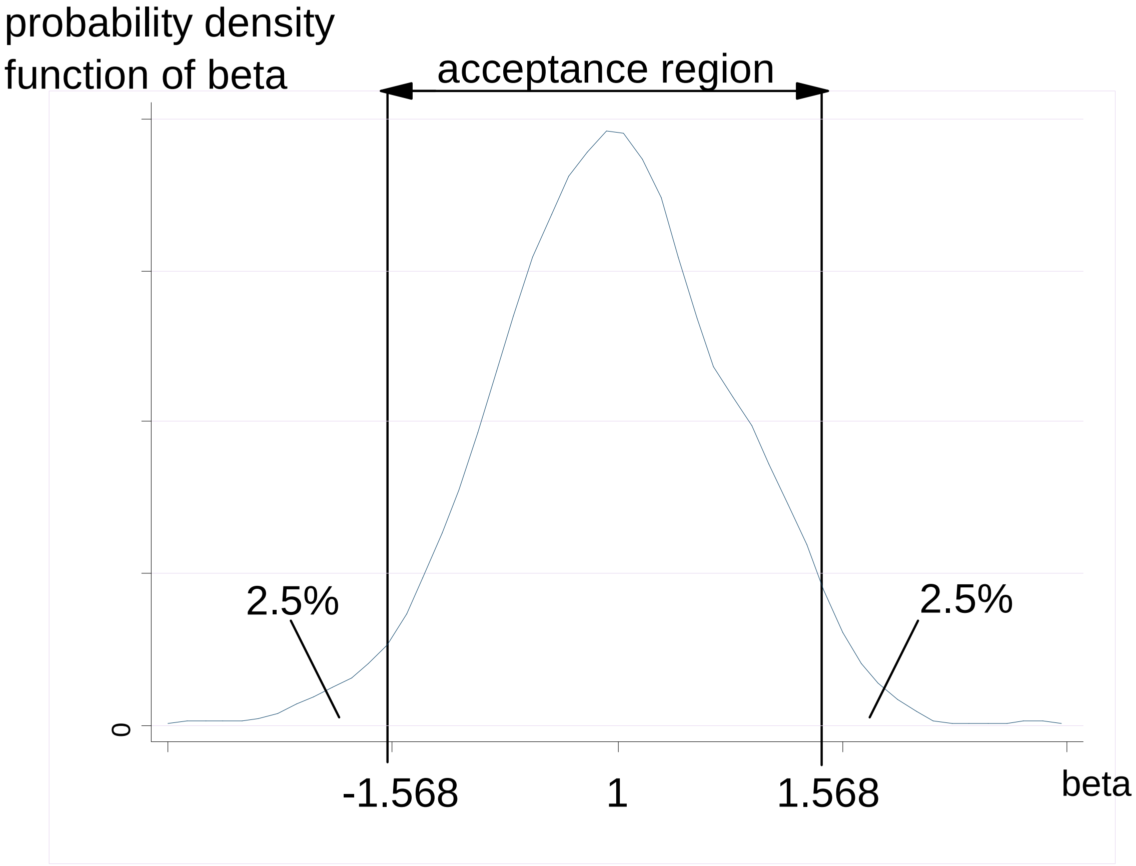
Then: $0 - 1.96 * SE\left(\hat{\beta}\right) \le \hat{\beta} \le 0 + 1.96 * SE\left(\hat{\beta}\right)$

Substituting beta=1 and se(beta)=0.8:

$$0 - 1.96 * 0.8 \le 1 \le 0 + 1.96 * 0.8$$

$$-1.568 \le 1 \le 1.568$$

Since this in-equality holds true, the null-hypothesis is not rejected. Thus, we accept that there is rather no relationship between x and y and beta equals with a high probability (95%) zero.

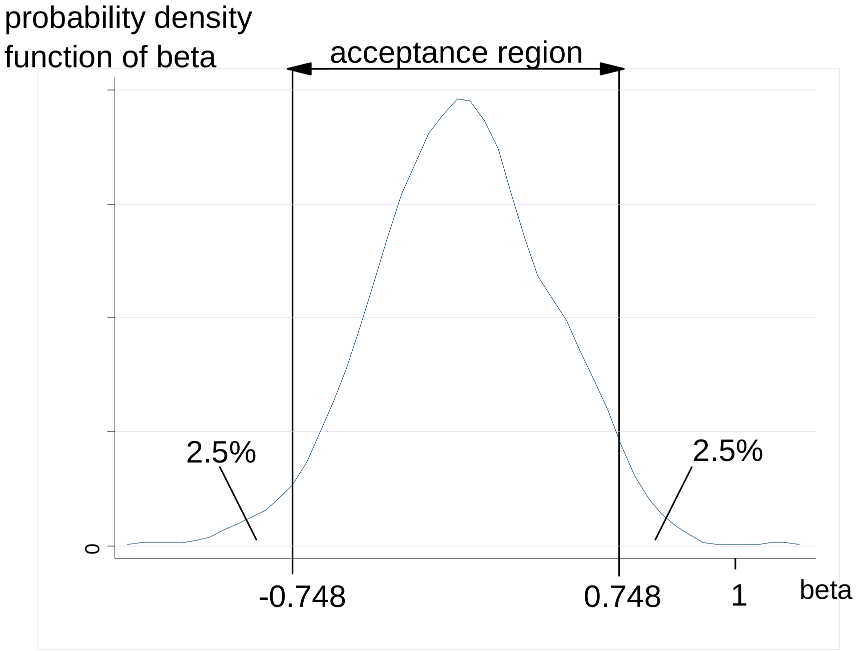

probability density
function of beta
acceptance region
2.5%
2.5%
0
-1.568
1
1.568
beta

Now assume that the standard error of beta is 0.4 instead of 0.8, we get:

$$0-1.96*0.4 \leq 1 \leq 0+1.96*0.4$$

$$-0.784 \leq 1 \leq 0.784$$

This in-equality is wrong, therefore we reject the null-hypothesis that beta equals zero and decide in favour of the alternative hypothesis that there is a significant positive relationship between x and y.



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[Type I Error]

  =Pr[rejecting the H0 | H0 is true]

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

  beta=Pr[Type II Error]

  =Pr[accepting the H0 | Ha is true]

Type I and Type II errors

Alpha and beta errors: an example:

A criminal trial: taking as the null hypothesis that the defendant is innocent, type I error occurs when the jury wrongly decides that the defendant is guilty. A type two error occurs when the jury wrongly acquits the defendant.

In significance test:

H0: beta is insignificant = 0:

Type I error: wrongly rejecting the null hypothesis

Type II error: wrongly accepting the null that the coefficient is zero.

Selection of significance levels increase or decrease the probability of type I and type II errors.

The smaller the significance level (5%, 1%) the lower the probability of type I and the higher the probability of type II errors.

Confidence Intervals

Significance tests assume that hypotheses come before the test: beta≠0. however, the significance test leaves us with some vacuum since we know that beta is different from zero but since we have a probabilistic theory we are not sure what the exact value should be.

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

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

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%).

Since the critical value of the t distribution is greater for the 1% level than for the 5% level, for any given number of degrees of freedom, it follows that the 99% interval is wider than the 95% interval and encompasses al the hypothetical values of beta in the 95% confidence interval plus some more on either side

Example: b=1, se(b)=0.4, 95% confidence interval, t_critical=1.96,

-1.96:

$$1-0.4*1.96 \leq \text{beta} \leq 1+0.4*1.96$$

95% confidence interval:

$$0.216 \leq \text{beta} \leq 1.784$$

Thus, all values between 0.216 and 1.784 are theoretically possible and would not be rejected. They are compatible to the estimated b = 1. 1 is the central value of the confidence interval.

Interpretation of regression results:

```
 reg y x

  Source |      SS           df       MS                    Number of obs =   100
-------------+------------------------------------------    F(  1,    98)      =   89.78
    Model |  1248.96129     1     1248.96129                Prob > F           =   0.0000
 Residual |  1363.2539     98     13.9107541                R-squared          =   0.4781
-------------+------------------------------------------    Adj R-squared  =   0.4728
      Total |  2612.21519   99     26.386012                Root MSE           =   3.7297

------------------------------------------------------------------------------
          y |     Coef.        Std. Err.      t       P>|t|      [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x |   1.941914     .2049419     9.48   0.000     1.535213    2.348614
  _cons |     .8609647   .4127188     2.09   0.040     .0419377    1.679992
------------------------------------------------------------------------------
```

Degrees of freedom: number of observations minus number of estimated parameters: in this case alpha and beta: 100-2=98. If we had 2 explanatory variables the number of degrees of freedom would decrease to 97, 3 – 96, etc.

The concept of DoF implies that you cannot have more explanatory variables than observations!

Definitions

Total Sum of Squares (SST):

$$SST = \sum_{i=1}^{n} \left(y_i - \overline{y} \right)^2$$

Explained (Estimation) Sum of Squares (SSE):

$$SSE = \sum_{i=1}^{n} \left(\hat{y}_i - \overline{y} \right)^2$$

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

$$SSR = \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \sum_{i=1}^{n} \left(y_i - \alpha - \beta x_i \right)^2$$

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 R² 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:

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

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_hat (which all lie on the regression line)
- R^2=(corr(y,yhat))^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.

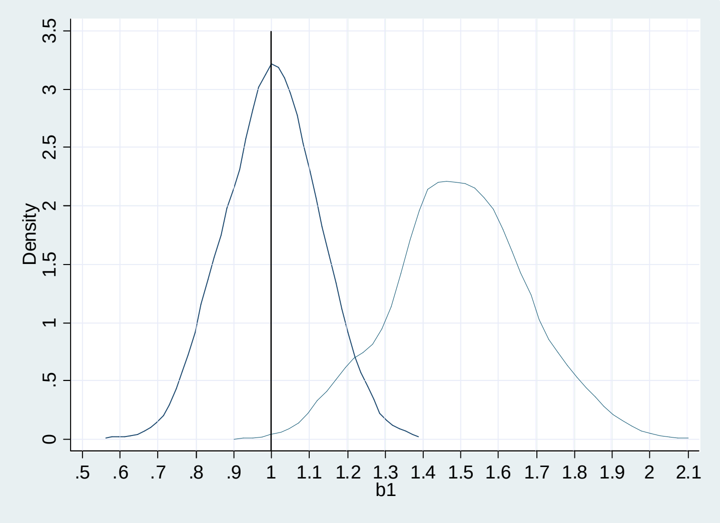
Unbiasedness

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.

$$E\left(\beta-\hat{\beta}\right)=0$$

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.

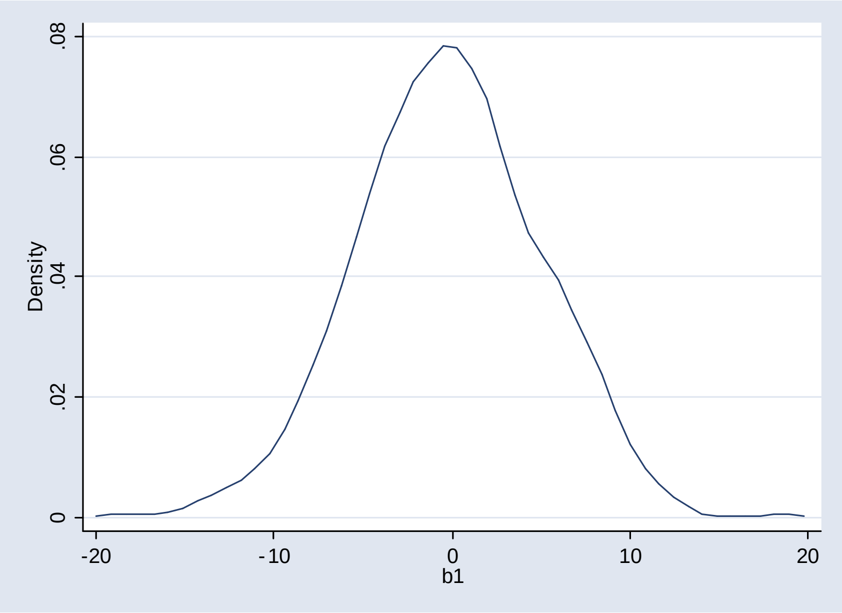


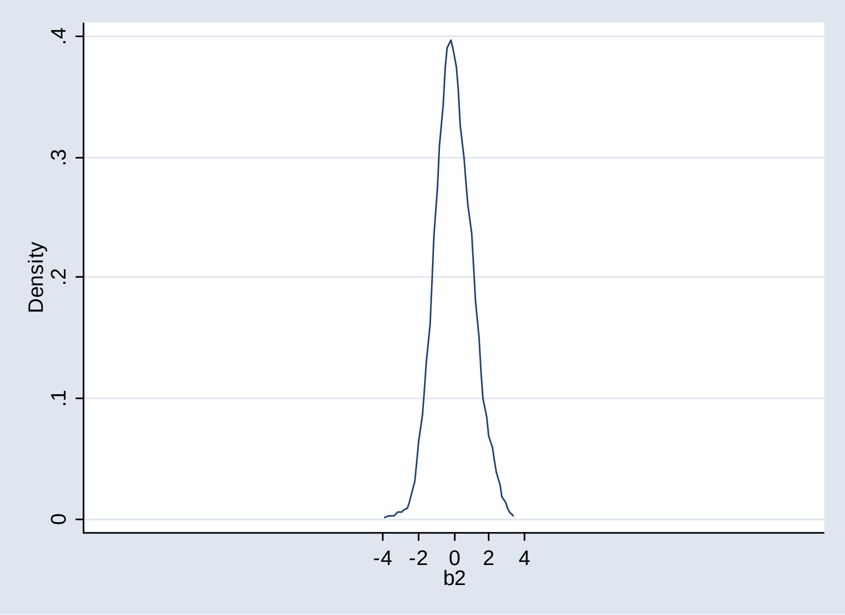
Sampling Variance of an estimator

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 β. With variances $V\left[\hat{\beta}\right]$ and $V\left[\tilde{\beta}\right]$. Then $\hat{\beta}$ is called to be relative more efficient than $\tilde{\beta}$ if $V\left[\hat{\beta}\right]$ is smaller than $V\left[\tilde{\beta}\right]$.

The property of relative efficiency only helps us to rank two unbiased estimators.

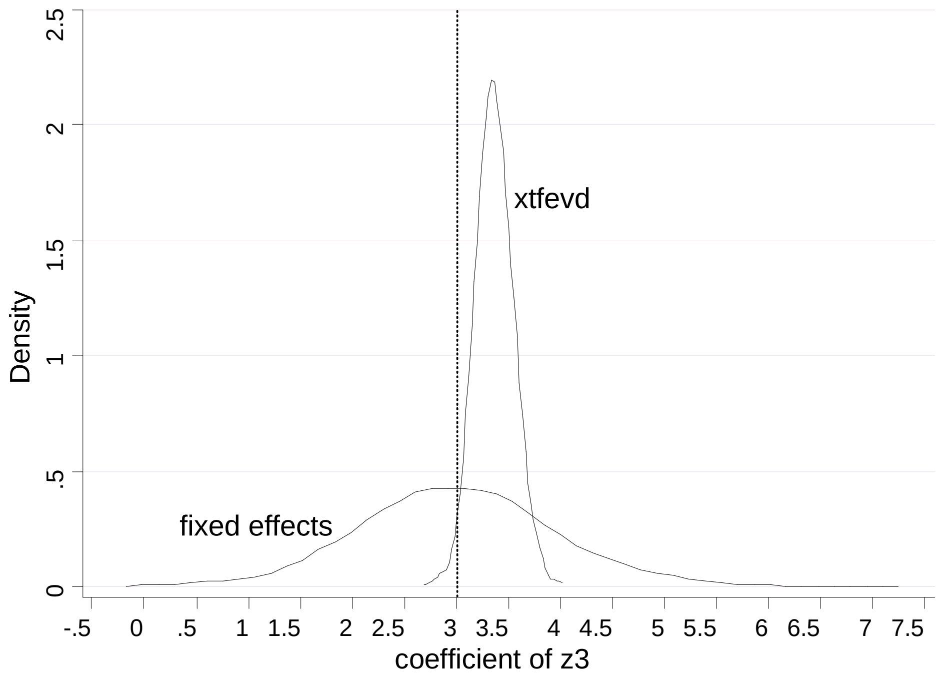




Trade-off between Bias and Efficiency

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:



$$\mathrm{RMSE}\left(\hat{\beta}\right)=\frac{1}{K}\sum_{k=1}^{K}\sqrt{\left(\hat{\beta}-\beta_{\mathrm{true}}\right)^2}$$

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

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

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→∞ the estimated beta equals the true beta:

$$\lim_{n\to\infty} \Pr\left[\left|\hat{\beta}_n - \beta\right| > \varepsilon\right] = 0, \quad \mathrm{p}\lim \hat{\beta}_n = \beta,$$

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

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:

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

- Control for omitted variable bias
- But: increases inefficiency of the regression since explanatory variables might be collinear.

Obtaining OLS Estimates in Multiple Regressions

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

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

$$\hat{\beta} = X\,'X^{-1}X\,'y$$

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

$$Var\left(\hat{\beta}_1\right)=\frac{\sigma^2}{SST_1\left(1-R_1^2\right)} \quad \Rightarrow \quad \hat{\sigma}^2=\frac{1}{n-\left(k+1\right)}\sum_{i=1}^{n}\hat{\varepsilon}_i^2$$

$$SST_1=\sum_{i=1}^{n}\left(x_{i1}-\overline{x}_1\right)^2$$

$$R_1^2 \;\; for\, the\, regression\, of \;\; x_{i1} \; on \;\; x_{i2} : \; R_1^2=\frac{SSE}{SST}=\frac{\sum_{i=1}^{n}\left(\hat{x}_{i1}-\overline{x}_1\right)^2}{\sum_{i=1}^{n}\left(x_{i1}-\overline{x}_1\right)^2}$$

$$SD\left(\hat{\beta}_1\right)=SE\left(\hat{\beta}_1\right)=\frac{\sigma}{\sqrt{SST_1\left(1-R_1^2\right)}}$$

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:
- The F-statistic is F distributed under the Null-Hypothesis.

$$F = \frac{\left(SSR_r - SSR_{ur}\right)/q}{SSR_{ur}/\left(n-k-1\right)}$$

$$F \sim F_{q,n-k-1}$$

- 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 R² of the Regression:

$$\frac{R^2/k}{\left(1-R^2\right)/\left(n-k-1\right)}$$

- 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 R² in the same way.

BUT: R² never decreases but tends to increase with the number of explanatory variables.

THUS, R² 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 R²: takes the number of explanatory variables into account since the R² increases with the number of regressors:

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

k is the number of explanatory variables and n the number of observations

Comparing Coefficients

The size of the slope parameters depends on the scaling of the variables (on which scale a variables 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:

$$\hat{b}_j = \frac{\hat{\sigma}_{x_j} \hat{\beta}_j}{\hat{\sigma}_y} \quad for \;\; j = 1, \ldots, k$$

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 b_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).

Problems in Multiple Regressions:
1. Multicolinearity

- Perfect multicolinearity 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. Multicolinearity 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 multicolinearity.

Testing for Multicolinearity

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 *Rj* 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 *Rj* equals zero (i.e., no correlation between *Xj* 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. Omitted Variable Bias

- The effect of omitted variables that ought to be included:
- Suppose the dependent variable y depends on two explanatory variables:

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

- But you are unaware of the importance of x2 and only include x

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

- If x2 is omitted from the regression equation, x1 will have a “double” effect on y (a direct effect and one mimicking x2)
- The mimicking effect depends on the ability of x1 to mimic x2 (the correlation) and how much x2 would explain y
- Beta1 in the second equation is biased upwards in case x1 and x2 are positively correlated and downward biased otherwise
- Beta1 is only unbiased if x1 and x2 are not related (corr(x1,x2)=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)

3. 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 ~ Chi² 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)

$$-2\ln(\lambda) = (NT)\ln\left(\sigma^2\right) - \Sigma\left(T\ln\left(\sigma_i^2\right)\right) \sim \chi^2\left(dF = n-1\right)$$

c. White test if form of Heteroskedasticity is unknown:

- H0: $V\left[\varepsilon_i \mid x_i\right] = \sigma^2$
- Ha: $V\left[\varepsilon_i \mid x_i\right] = \sigma_i^2$

1. Estimate the model under H0
2. Compute squared residuals: e_i^2
3. Use squared residuals as dependent variable of auxiliary regression: RHS: all regressors, their quadratic forms and interaction terms

$$e_i^2 = \delta_0 + \delta_1 x_{i2} + \ldots + \delta_{k-1} x_{ik} + \delta_{k-1} x_{i2}^2 + \delta_{k+1} x_{i2} x_{i3} + \ldots + \delta_q x_{ik}^2 + \xi_i$$

4. Compute White statistic from R² of auxiliary regression:

$$n * R^2 \xrightarrow{a} \chi^2_{(q)}$$

5. Use one-sided test and check if n*R² is larger than 95% quantile of Chi²-distribution

Robust White Heteroskedasticity Consistent Variance-Covariance Estimator:

Normal Variance of beta: $\sigma_\beta^2 = \frac{\sigma_\varepsilon^2}{N * Var\left(X\right)}$,

Robust White VC matrix: $\hat{V}\left[\hat{\beta}\right] = \frac{1}{n}\hat{\Omega}$

$$\hat{\Omega} = n\left(X\,'X\right)^{-1} X\,'\hat{D}X\left(X\,'X\right)^{-1}$$

$$\hat{D} = diag\left[e_i^2\right]$$

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:

$$\beta = \left(\sum_{i=1}^{N} X_i '\Omega^{-1} X_i \right)^{-1} \left(\sum_{i=1}^{N} X_i '\Omega^{-1} y_i \right) \Rightarrow$$

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

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…:

$$\Omega = \begin{bmatrix} \varepsilon_1^2 & \varepsilon_{21} & \varepsilon_{31} & \cdots & \varepsilon_{n1} \\ \varepsilon_{12} & \varepsilon_2^2 & \varepsilon_{32} & \cdots & \varepsilon_{n2} \\ \varepsilon_{13} & \varepsilon_{23} & \varepsilon_3^2 & \cdots & \varepsilon_{n3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{1n} & \varepsilon_{2n} & \varepsilon_{3n} & \cdots & \varepsilon_n^2 \end{bmatrix}$$

4. Autocorrelation

The observation of the residual in t1 is dependent on the observation in t0: not controlling for autocorrelation violates on 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 constitent VC matrix

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

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 on of the basic assumptions of OLS and may bias the estimation of the beta coefficients

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

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:

$$d=\frac{\sum_{t=2}^{T}\left(e_t-e_{t-1}\right)^2}{\sum_{t=1}^{T}e_t^2}$$

- 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)

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

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

$$\omega_l=1-\frac{1}{p+1}$$

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

$$\mathrm{y}_{\mathrm{it}} = \alpha + \beta_0 \mathrm{y}_{\mathrm{it}-1} + \beta_{\mathrm{k}} \mathrm{x}_{\mathrm{it}} + \varepsilon_{\mathrm{i\,t}}$$

- The interpretation of the LDV as measure of time-persistency is missleading
- 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→ ∞.
- Consistent

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

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

Prais-Winsten Transformation

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

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

The ξ_{it} are iid – with $N\left(0,\sigma^2\right)$

- 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,…,n

$$y_{it} - \rho y_{it-1} = \beta\left(x_{it} - \rho x_{it-1}\right) + \varsigma_{it}$$

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

$$\sqrt{1-\rho^2}\, y_1 = \beta\left(\sqrt{1-\rho^2}\, x_1\right) + \sqrt{1-\rho^2}\, \varsigma_1$$

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:

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

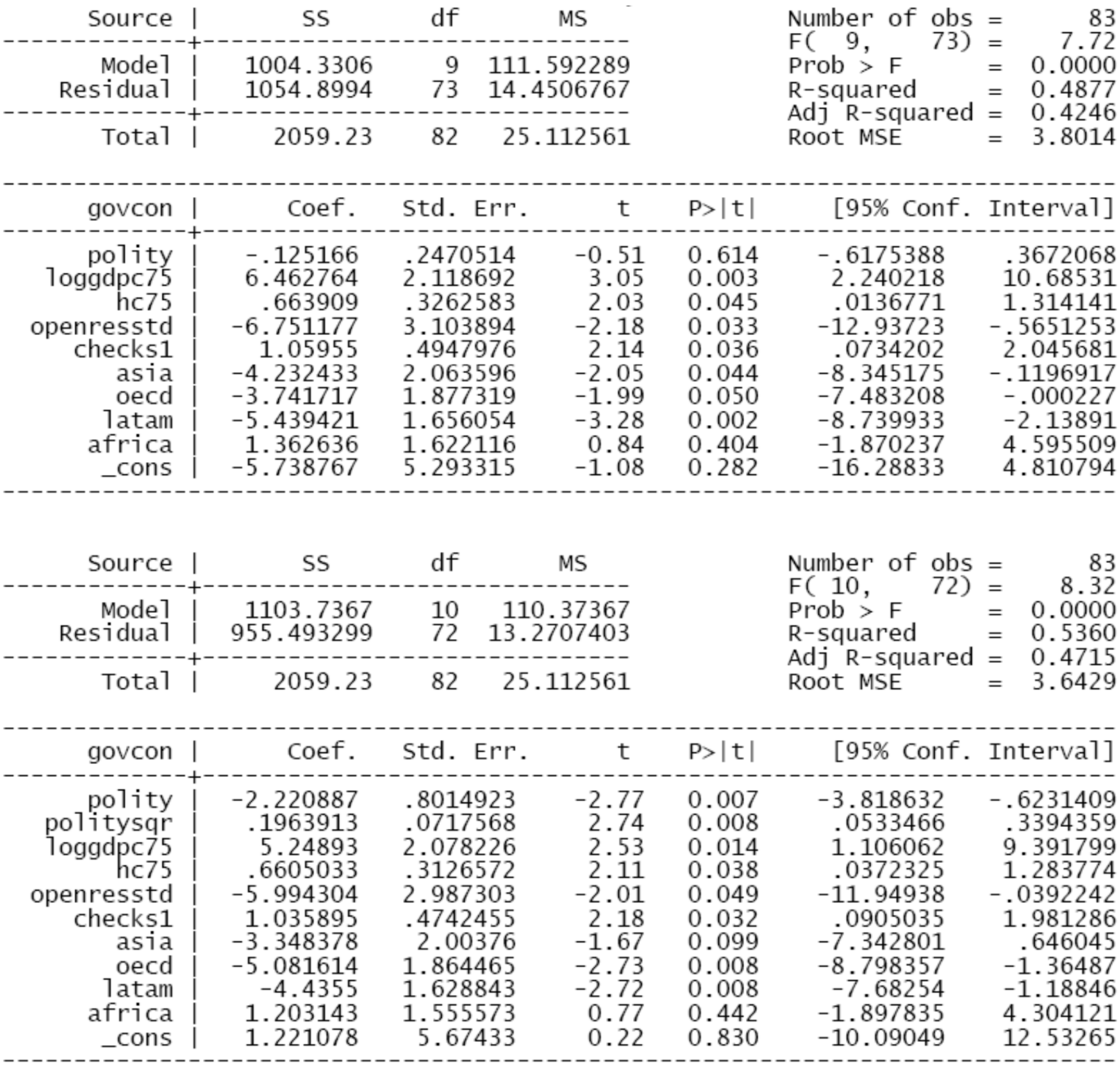
$$n = 1, \ldots, t-1$$

Specification Issues in Multiple Regressions: 1. Non-Linearity

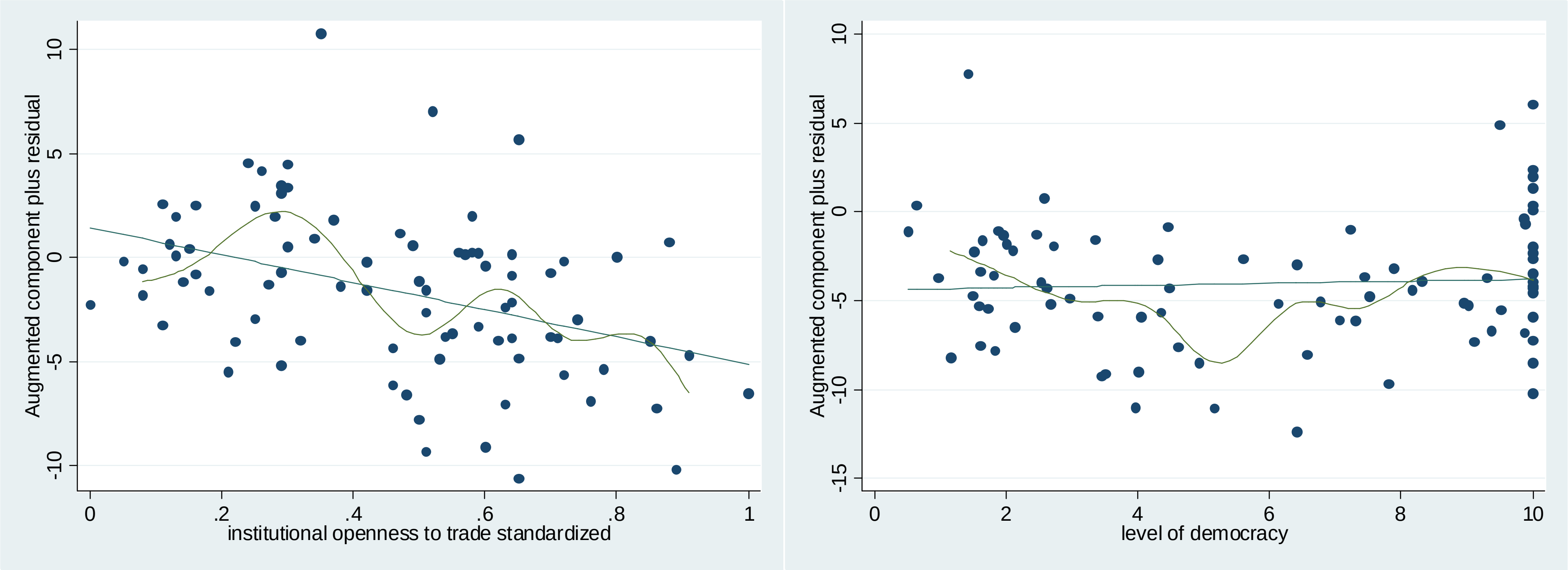
One or more explanatory variables have a non-linear effect on the dependent variable: estimating a linear model would lead to wrong or/and insignificant results. Thus, even though in the population there exist a relationship between an explanatory variable and the dependent variable, but this relationship 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.



```
      Source |       SS       df       MS              Number of obs =      83
-------------+------------------------------           F(  9,    73) =    7.72
       Model |  1004.3306     9  111.592289           Prob > F      =  0.0000
    Residual |  1054.8994    73  14.4506767           R-squared     =  0.4877
-------------+------------------------------           Adj R-squared =  0.4246
       Total |    2059.23    82   25.112561           Root MSE      =  3.8014

------------------------------------------------------------------------------
      govcon |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      polity |   -.125166   .2470514    -0.51   0.614    -.6175388    .3672068
    loggdpc75 |   6.462764   2.118692     3.05   0.003     2.240218    10.68531
        hc75 |    .663909   .3262583     2.03   0.045     .0136771    1.314141
   openresstd |  -6.751177   3.103894    -2.18   0.033    -12.93723   -.5651253
     checks1 |    1.05955   .4947976     2.14   0.036     .0734202    2.045681
        asia |  -4.232433   2.063596    -2.05   0.044    -8.345175   -.1196917
        oecd |  -3.741717   1.877319    -1.99   0.050    -7.483208    -.000227
       latam |  -5.439421   1.656054    -3.28   0.002    -8.739933    -2.13891
      africa |   1.362636   1.622116     0.84   0.404    -1.870237    4.595509
       _cons |  -5.738767   5.293315    -1.08   0.282    -16.28833    4.810794
------------------------------------------------------------------------------

      Source |       SS       df       MS              Number of obs =      83
-------------+------------------------------           F( 10,    72) =    8.32
       Model |  1103.7367    10   110.37367           Prob > F      =  0.0000
    Residual | 955.493299    72  13.2707403           R-squared     =  0.5360
-------------+------------------------------           Adj R-squared =  0.4715
       Total |    2059.23    82   25.112561           Root MSE      =  3.6429

------------------------------------------------------------------------------
      govcon |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      polity |  -2.220887   .8014923    -2.77   0.007    -3.818632   -.6231409
    politysqr |   .1963913   .0717568     2.74   0.008     .0533466    .3394359
    loggdpc75 |    5.24893   2.078226     2.53   0.014     1.106062    9.391799
        hc75 |   .6605033   .3126572     2.11   0.038     .0372325    1.283774
   openresstd |  -5.994304   2.987303    -2.01   0.049    -11.94938   -.0392242
     checks1 |   1.035895   .4742455     2.18   0.032     .0905035    1.981286
        asia |  -3.348378    2.00376    -1.67   0.099    -7.342801     .646045
        oecd |  -5.081614   1.864465    -2.73   0.008    -8.798357    -1.36487
       latam |    -4.4355   1.628843    -2.72   0.008     -7.68254    -1.18846
      africa |   1.203143   1.555573     0.77   0.442    -1.897835    4.304121
       _cons |   1.221078    5.67433     0.22   0.830    -10.09049    12.53265
------------------------------------------------------------------------------
```


Augmented component plus residual
10
5
0
-5
-10
0
.2
.4
.6
.8
1
institutional openness to trade standardized
Augmented component plus residual
10
5
0
-5
-10
-15
0
2
4
6
8
10
level of democracy

Solutions

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 $\log y_i = \log \alpha + \beta * \log x_i + \log \varepsilon_i$
  - 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.

$$y_i = \alpha + \beta_1 * x_i + \beta_2 * x_i^2 + \varepsilon_i$$

Hausken, Martin, Plümper 2004: Government Spending and Taxation in Democracies and Autocracies, *Constitutional Political Economy 15, 239-59.*

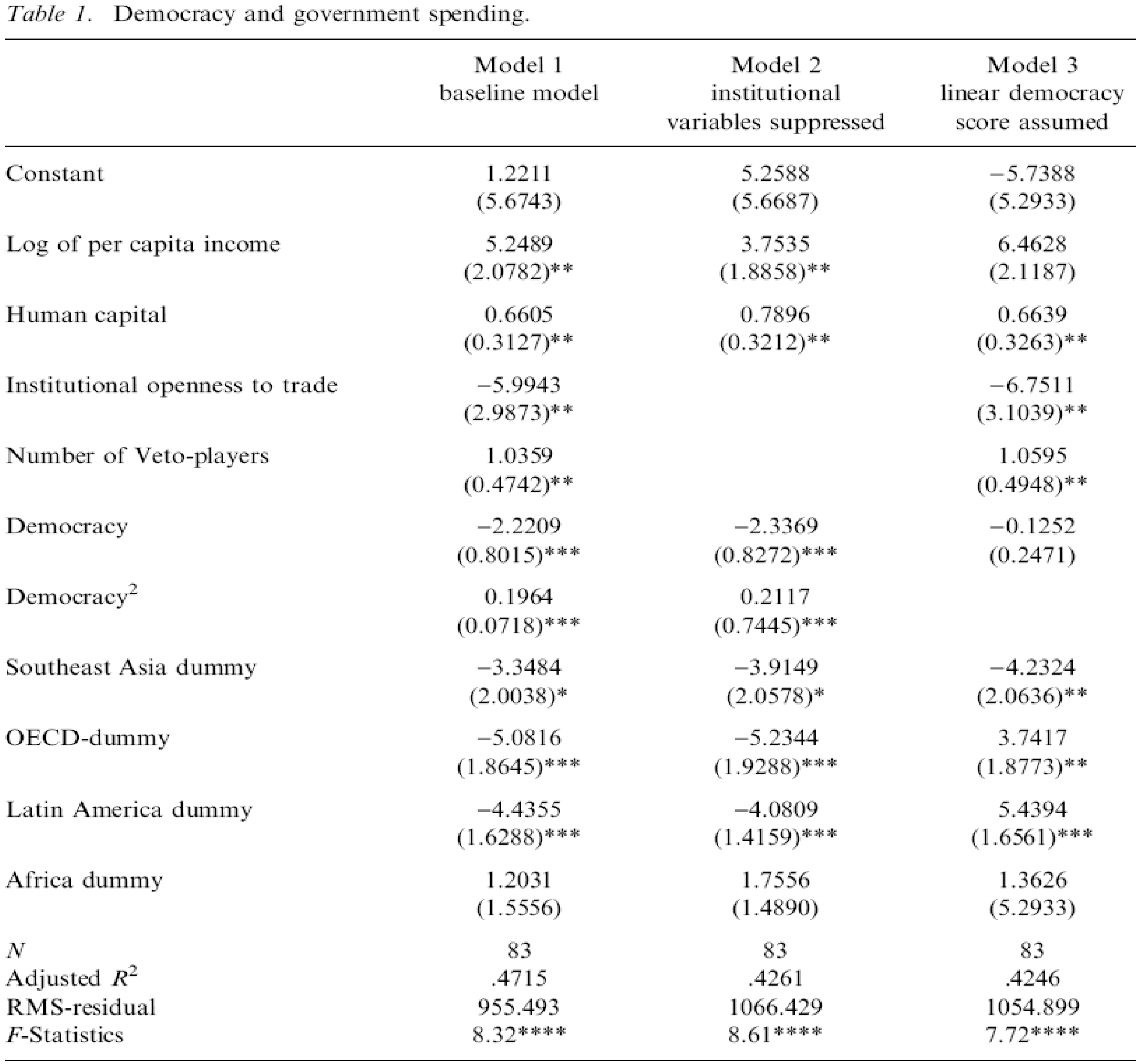
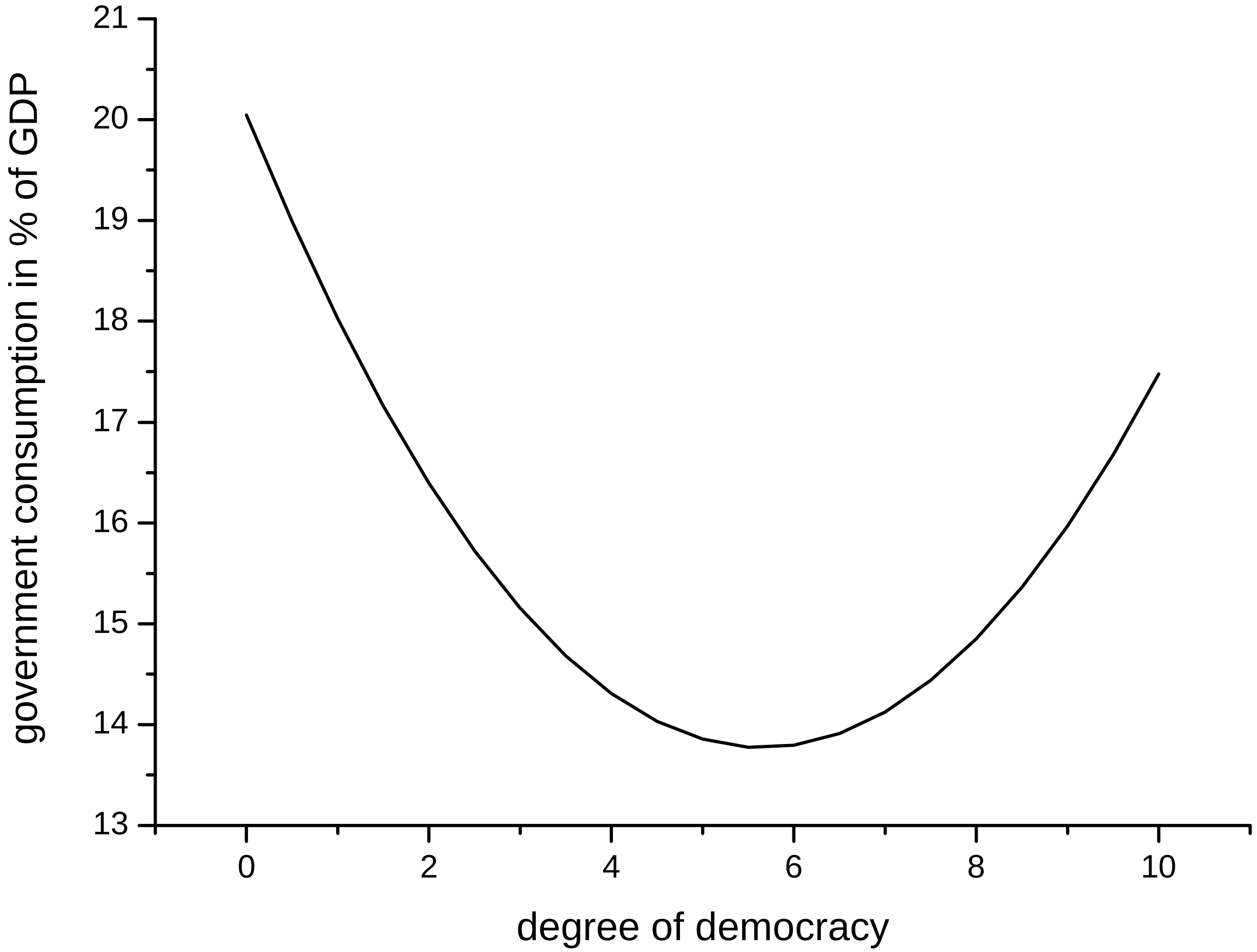


Table 1. Democracy and government spending.

	Model 1 baseline model	Model 2 institutional variables suppressed	Model 3 linear democracy score assumed
Constant	1.2211 (5.6743)	5.2588 (5.6687)	−5.7388 (5.2933)
Log of per capita income	5.2489 (2.0782)**	3.7535 (1.8858)**	6.4628 (2.1187)
Human capital	0.6605 (0.3127)**	0.7896 (0.3212)**	0.6639 (0.3263)**
Institutional openness to trade	−5.9943 (2.9873)**		−6.7511 (3.1039)**
Number of Veto-players	1.0359 (0.4742)**		1.0595 (0.4948)**
Democracy	−2.2209 (0.8015)***	−2.3369 (0.8272)***	−0.1252 (0.2471)
Democracy^2	0.1964 (0.0718)***	0.2117 (0.7445)***	
Southeast Asia dummy	−3.3484 (2.0038)*	−3.9149 (2.0578)*	−4.2324 (2.0636)**
OECD-dummy	−5.0816 (1.8645)***	−5.2344 (1.9288)***	3.7417 (1.8773)**
Latin America dummy	−4.4355 (1.6288)***	−4.0809 (1.4159)***	5.4394 (1.6561)***
Africa dummy	1.2031 (1.5556)	1.7556 (1.4890)	1.3626 (5.2933)
N	83	83	83
Adjusted R^2	.4715	.4261	.4246
RMS-residual	955.493	1066.429	1054.899
F-Statistics	8.32****	8.61****	7.72****

polity_sqr	polity	govcon
0	0	20.049292
0.25	0.5	18.987946
1	1	18.024796
2.25	1.5	17.159842
4	2	16.393084
6.25	2.5	15.724521
9	3	15.154153
12.25	3.5	14.681982
16	4	14.308005
20.25	4.5	14.032225
25	5	13.85464
30.25	5.5	13.775251
36	6	13.794057
42.25	6.5	13.911059
49	7	14.126257
56.25	7.5	14.43965
64	8	14.851239
72.25	8.5	15.361024
81	9	15.969004
90.25	9.5	16.67518
100	10	17.479551

The „u“ shaped relationship between democracy and government spending:



2. Interaction Effects

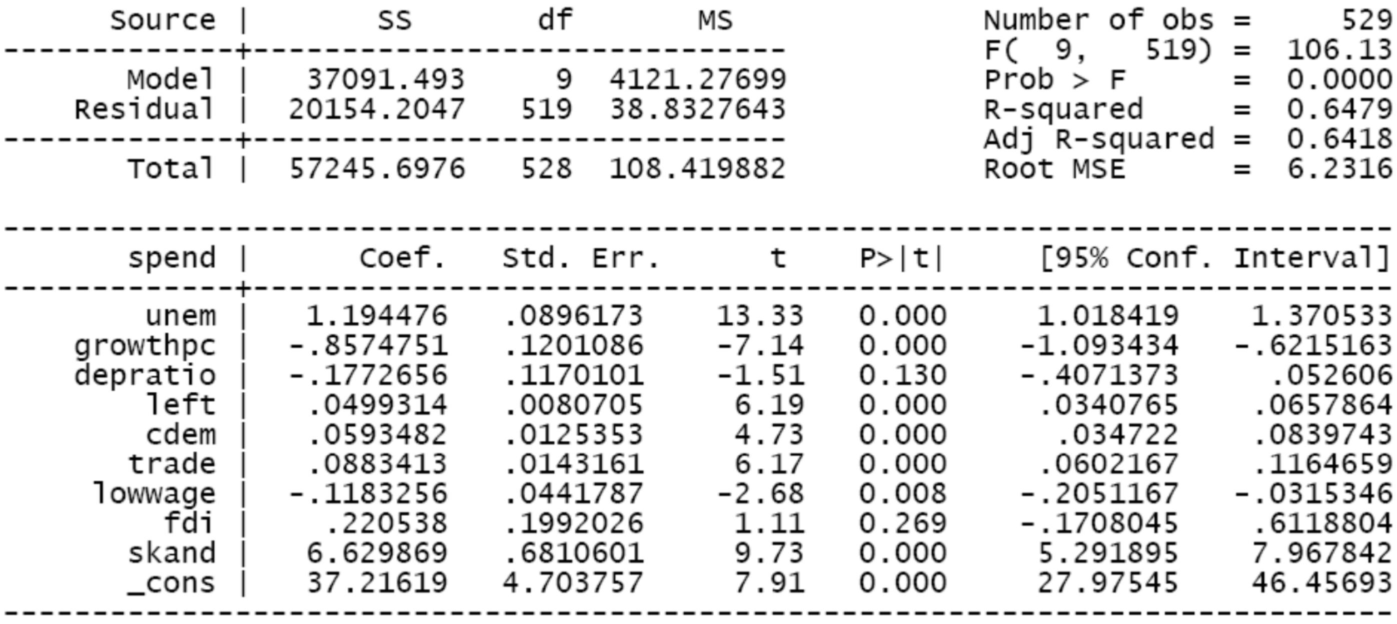
Two explanatory variables do not only have a direct effect on the dependent variable but also a combined effect

$$y_i = \alpha + \beta_1 * x_{1i} + \beta_2 * x_{2i} + \beta_3 * x_{1i} * x_{2i} + \varepsilon_i$$

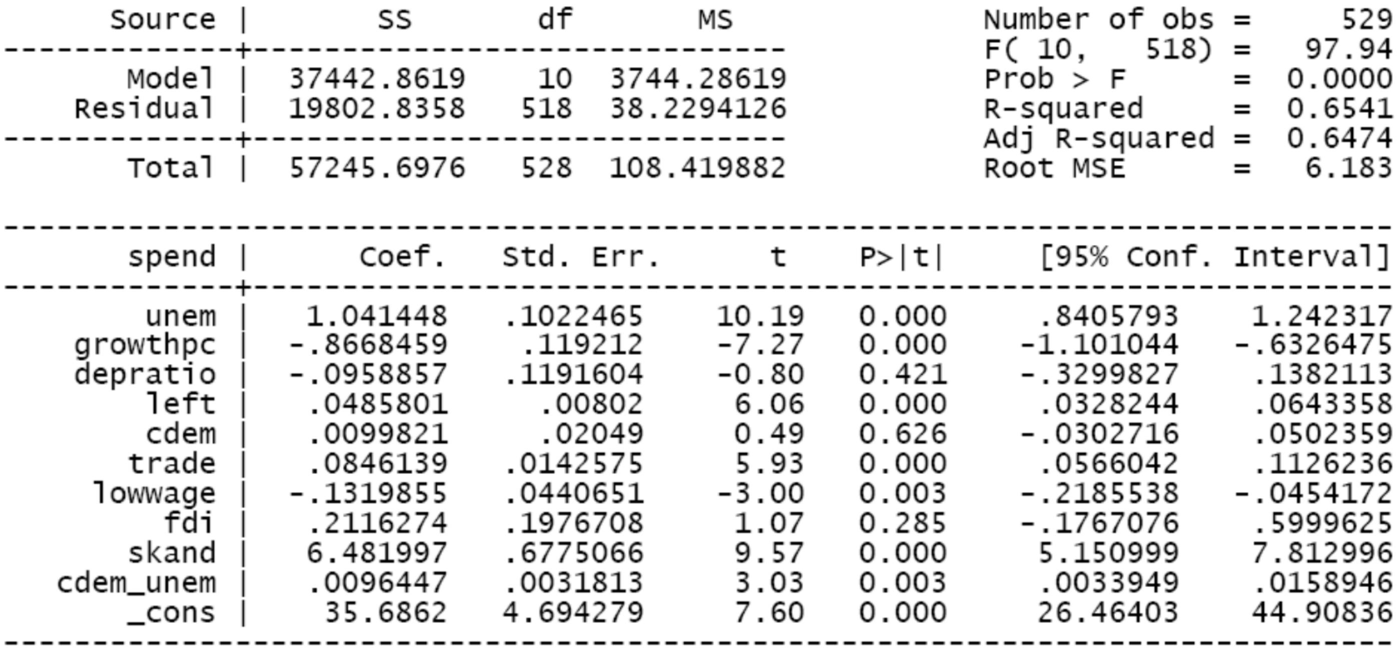
Interpretation: combined effect: b1*SD(x1)+b2*SD(x2)+b3*SD(x1*x2)

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

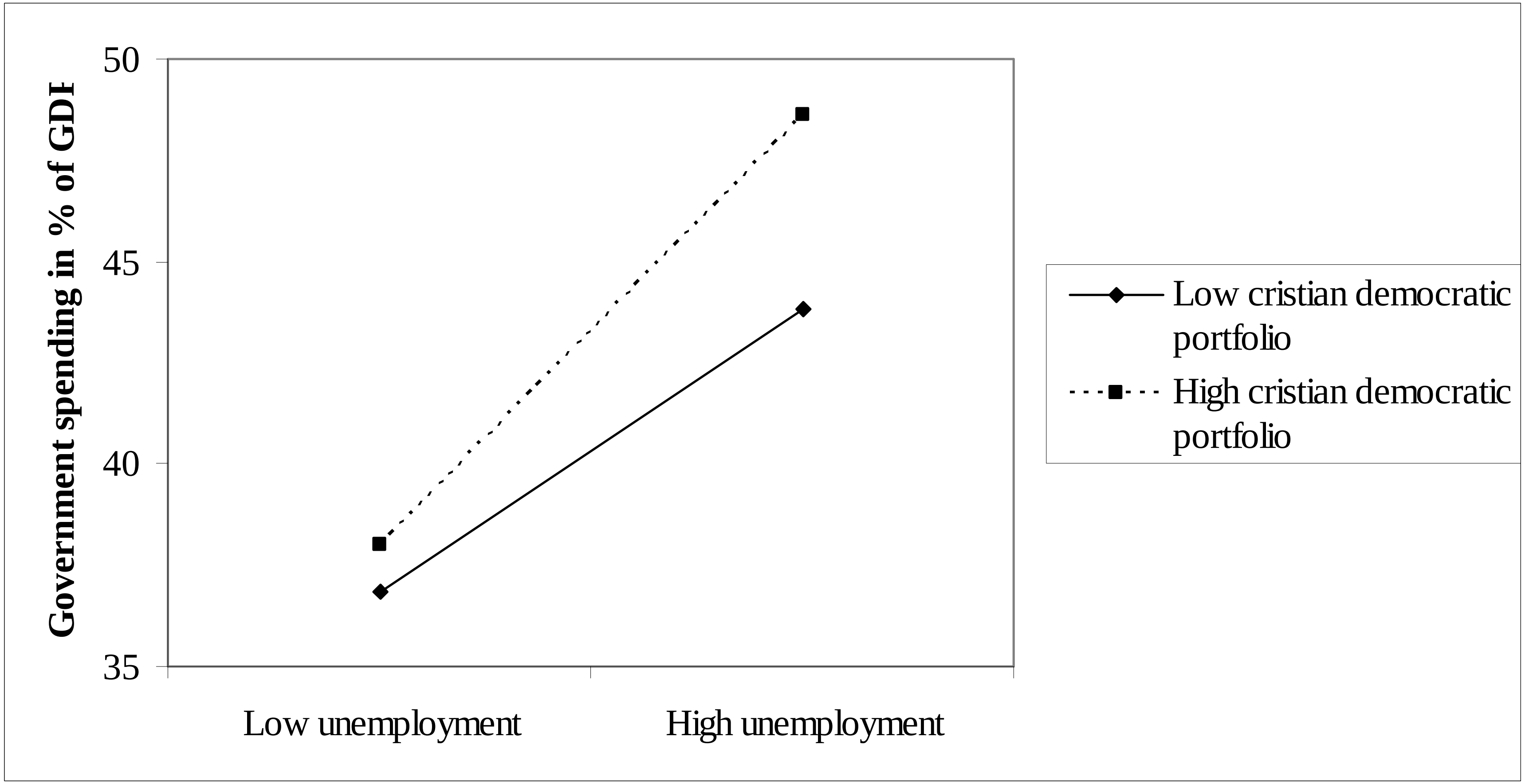
Example: government spending

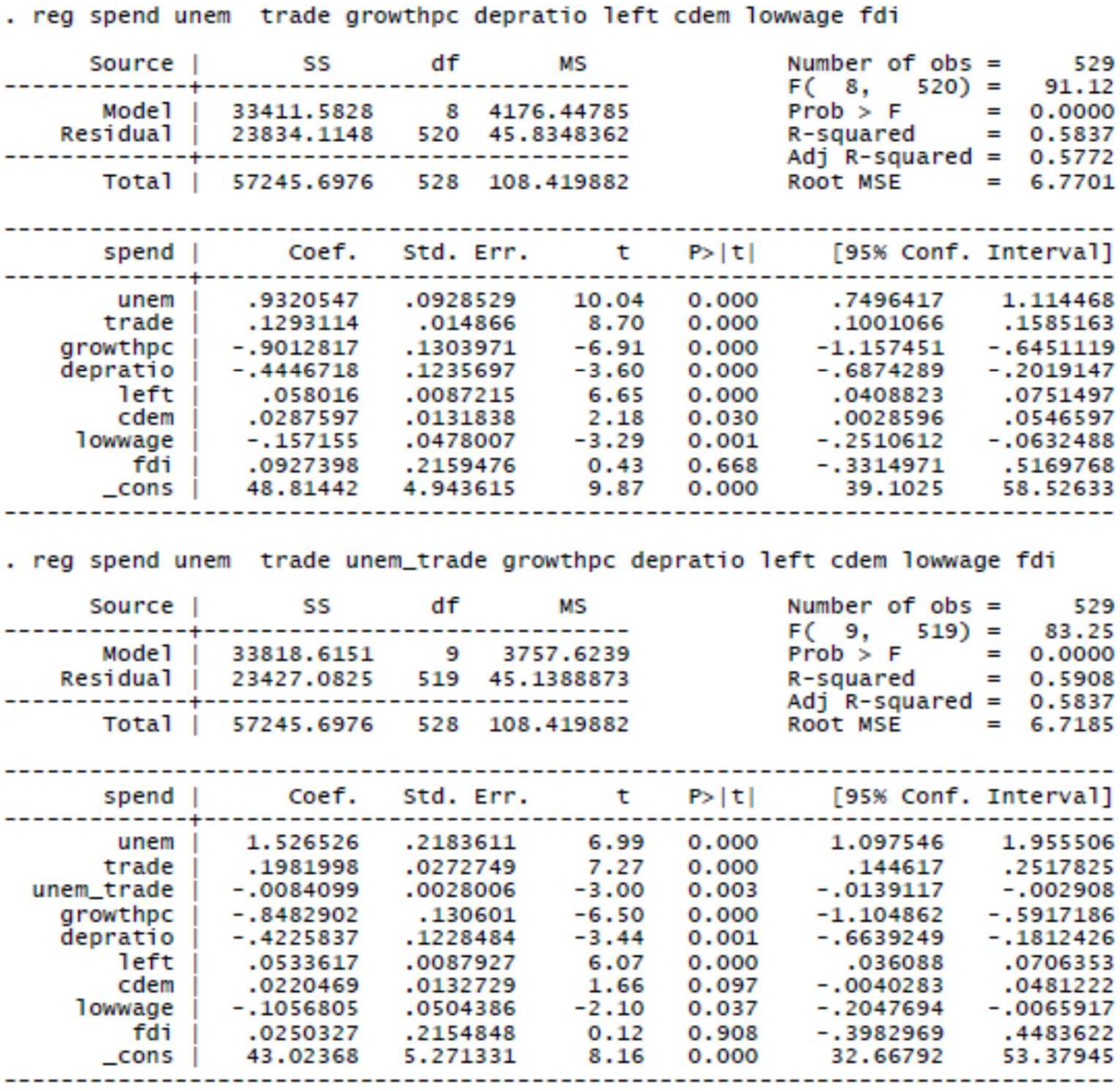


```
      Source |       SS       df       MS              Number of obs =     529
-------------+------------------------------           F(  9,   519) =  106.13
       Model |  37091.493     9  4121.27699           Prob > F      =  0.0000
    Residual | 20154.2047   519  38.8327643           R-squared     =  0.6479
-------------+------------------------------           Adj R-squared =  0.6418
       Total | 57245.6976   528  108.419882           Root MSE      =  6.2316

------------------------------------------------------------------------------
       spend |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        unem |   1.194476   .0896173    13.33   0.000     1.018419    1.370533
    growthpc |  -.8574751   .1201086    -7.14   0.000    -1.093434   -.6215163
    depratio |  -.1772656   .1170101    -1.51   0.130    -.4071373     .052606
        left |   .0499314   .0080705     6.19   0.000     .0340765    .0657864
        cdem |   .0593482   .0125353     4.73   0.000      .034722    .0839743
       trade |   .0883413   .0143161     6.17   0.000     .0602167    .1164659
     lowwage |  -.1183256   .0441787    -2.68   0.008    -.2051167   -.0315346
         fdi |    .220538   .1992026     1.11   0.269    -.1708045    .6118804
       skand |   6.629869   .6810601     9.73   0.000     5.291895    7.967842
       _cons |   37.21619   4.703757     7.91   0.000     27.97545    46.45693
------------------------------------------------------------------------------
```

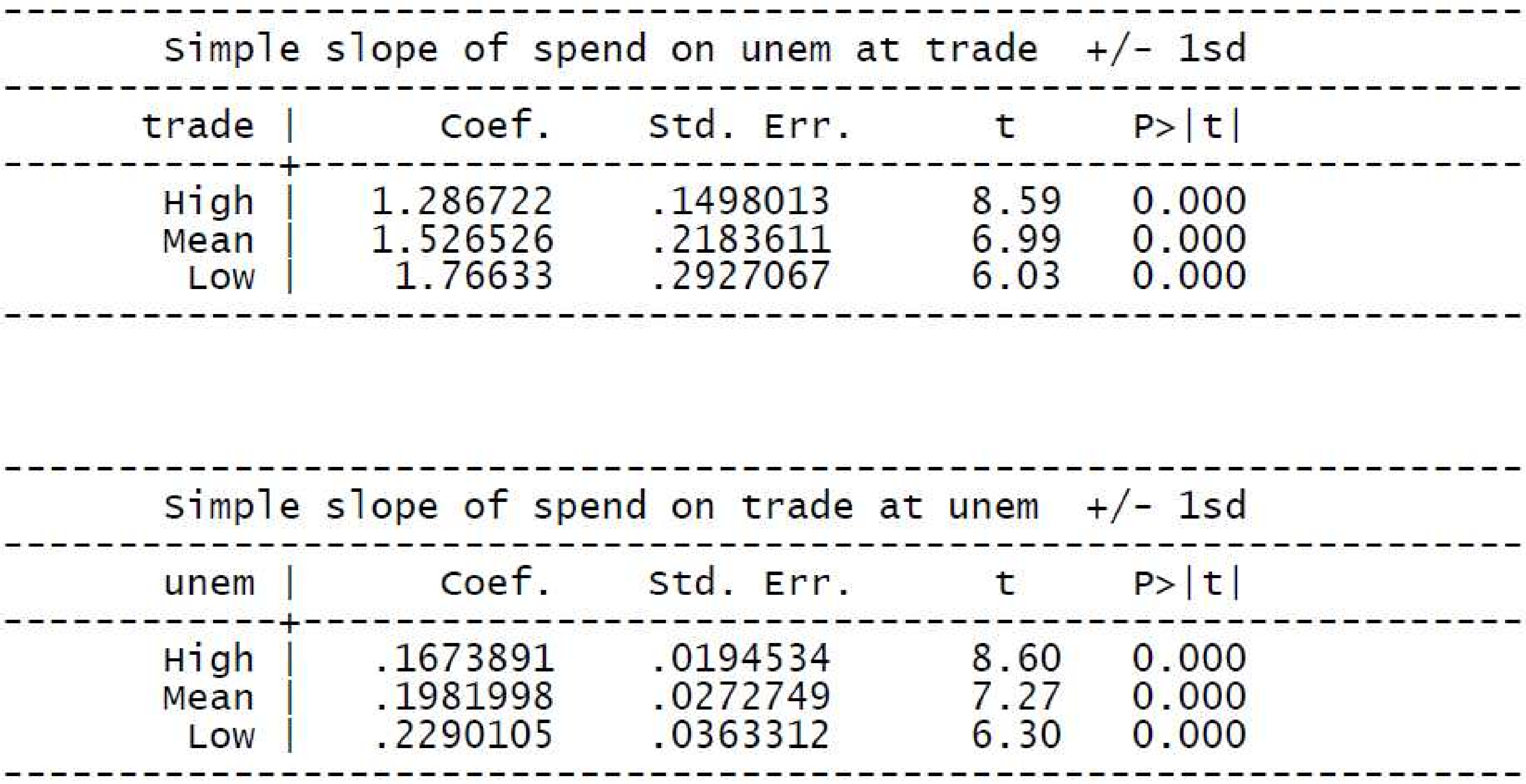


```
      Source |       SS       df       MS              Number of obs =     529
-------------+------------------------------           F( 10,   518) =   97.94
       Model | 37442.8619    10  3744.28619           Prob > F      =  0.0000
    Residual | 19802.8358   518  38.2294126           R-squared     =  0.6541
-------------+------------------------------           Adj R-squared =  0.6474
       Total | 57245.6976   528  108.419882           Root MSE      =   6.183

------------------------------------------------------------------------------
       spend |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        unem |   1.041448   .1022465    10.19   0.000     .8405793    1.242317
    growthpc |  -.8668459    .119212    -7.27   0.000    -1.101044   -.6326475
    depratio |  -.0958857   .1191604    -0.80   0.421    -.3299827    .1382113
        left |   .0485801     .00802     6.06   0.000     .0328244    .0643358
        cdem |   .0099821     .02049     0.49   0.626    -.0302716    .0502359
       trade |   .0846139   .0142575     5.93   0.000     .0566042    .1126236
     lowwage |  -.1319855   .0440651    -3.00   0.003    -.2185538   -.0454172
         fdi |   .2116274   .1976708     1.07   0.285    -.1767076    .5999625
       skand |   6.481997   .6775066     9.57   0.000     5.150999    7.812996
   cdem_unem |   .0096447   .0031813     3.03   0.003     .0033949    .0158946
       _cons |    35.6862   4.694279     7.60   0.000     26.46403    44.90836
------------------------------------------------------------------------------
```


Government spending in % of GDI
50
45
40
35
Low unemployment
High unemployment
Low cristian democratic portfolio
High cristian democratic portfolio

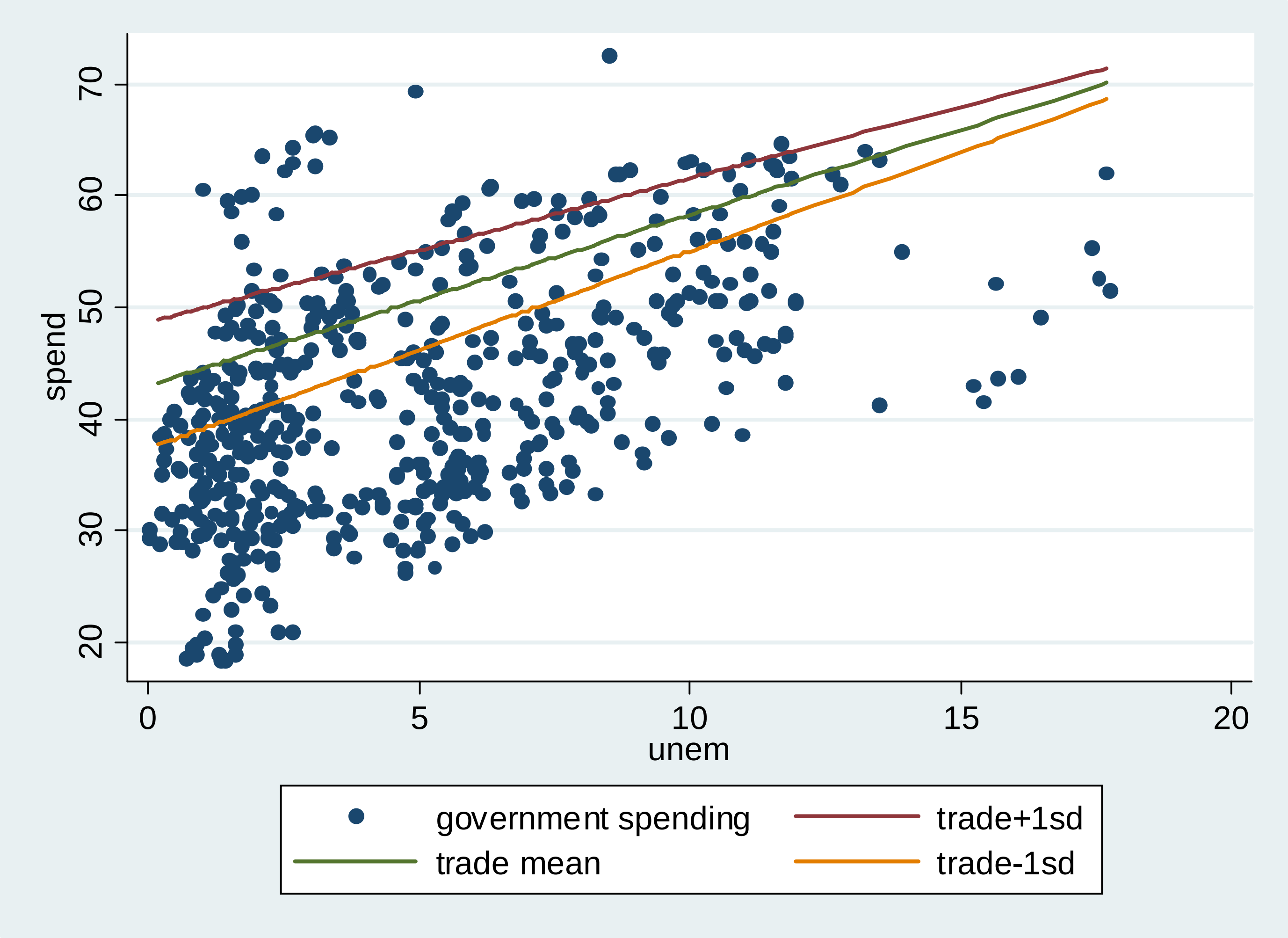


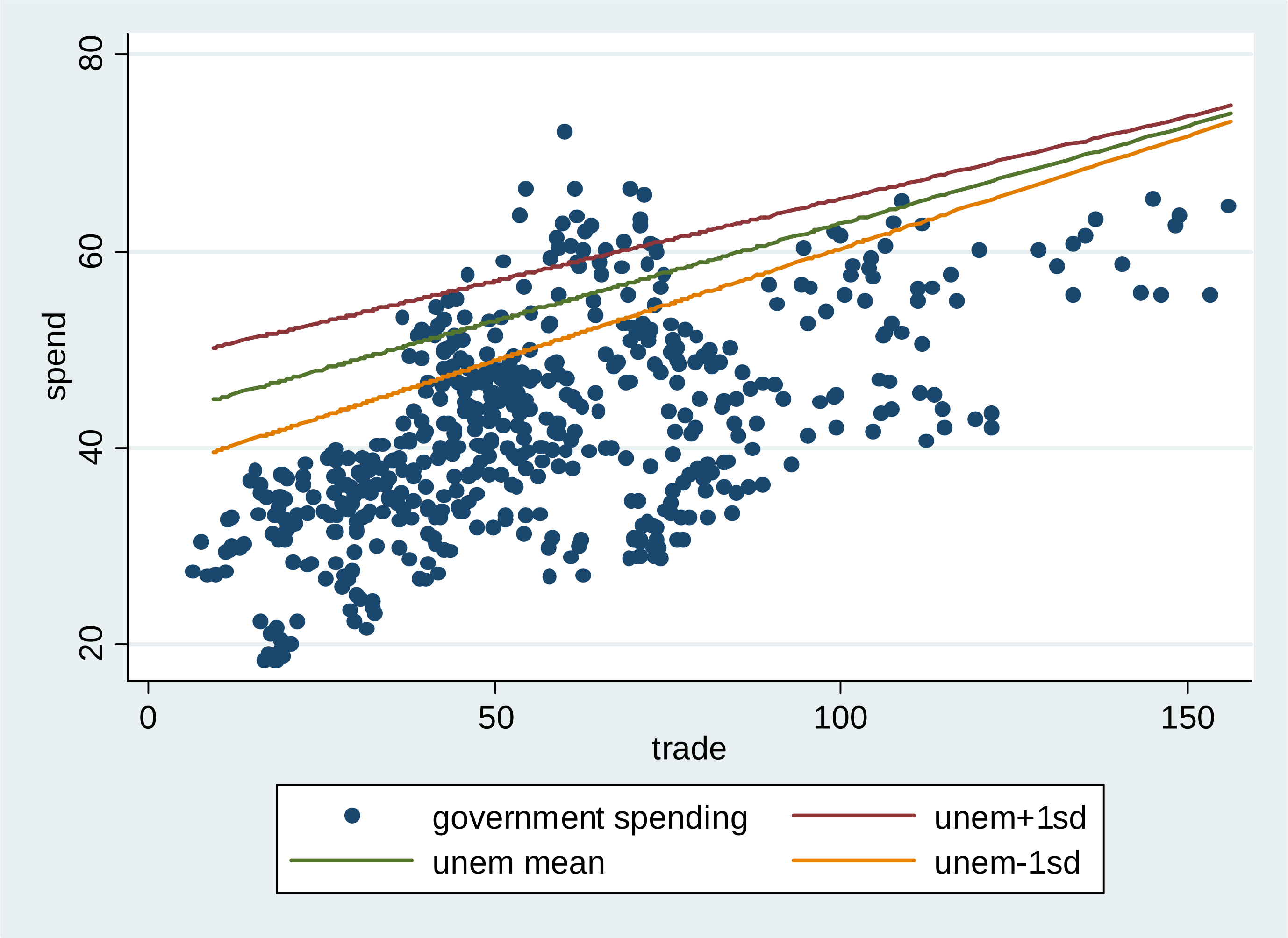
```
. reg spend unem  trade growthpc depratio left cdem lowwage fdi

      Source |       SS       df       MS              Number of obs =     529
-------------+------------------------------           F(  8,   520) =   91.12
       Model |  33411.5828     8  4176.44785           Prob > F      =  0.0000
    Residual |  23834.1148   520  45.8348362           R-squared     =  0.5837
-------------+------------------------------           Adj R-squared =  0.5772
       Total |  57245.6976   528  108.419882           Root MSE      =  6.7701

------------------------------------------------------------------------------
       spend |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        unem |   .9320547   .0928529    10.04   0.000     .7496417    1.114468
       trade |   .1293114    .014866     8.70   0.000     .1001066    .1585163
    growthpc |  -.9012817   .1303971    -6.91   0.000    -1.157451   -.6451119
    depratio |  -.4446718   .1235697    -3.60   0.000    -.6874289   -.2019147
        left |    .058016   .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
------------------------------------------------------------------------------

. reg spend unem  trade unem_trade growthpc depratio left cdem lowwage fdi

      Source |       SS       df       MS              Number of obs =     529
-------------+------------------------------           F(  9,   519) =   83.25
       Model |  33818.6151     9   3757.6239           Prob > F      =  0.0000
    Residual |  23427.0825   519  45.1388873           R-squared     =  0.5908
-------------+------------------------------           Adj R-squared =  0.5837
       Total |  57245.6976   528  108.419882           Root MSE      =  6.7185

------------------------------------------------------------------------------
       spend |      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    .0706353
        cdem |   .0220469   .0132729     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
------------------------------------------------------------------------------
```



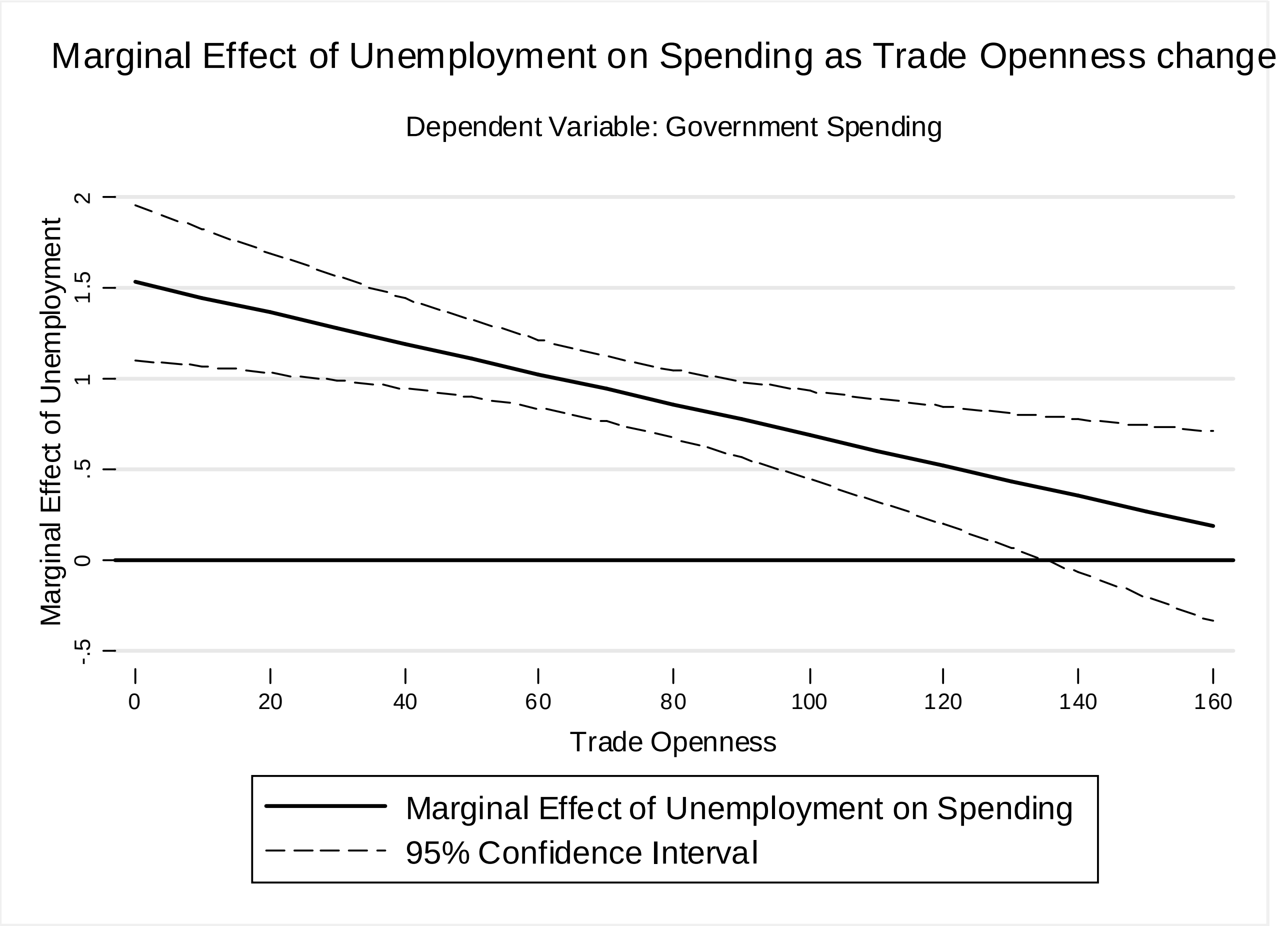
```
------------------------------------------------------------------
       Simple slope of spend on unem at trade  +/- 1sd
------------------------------------------------------------------
      trade |      Coef.    Std. Err.      t     P>|t|
------------+-----------------------------------------------------
       High |   1.286722    .1498013      8.59   0.000
       Mean |   1.526526    .2183611      6.99   0.000
        Low |    1.76633    .2927067      6.03   0.000
------------------------------------------------------------------
```

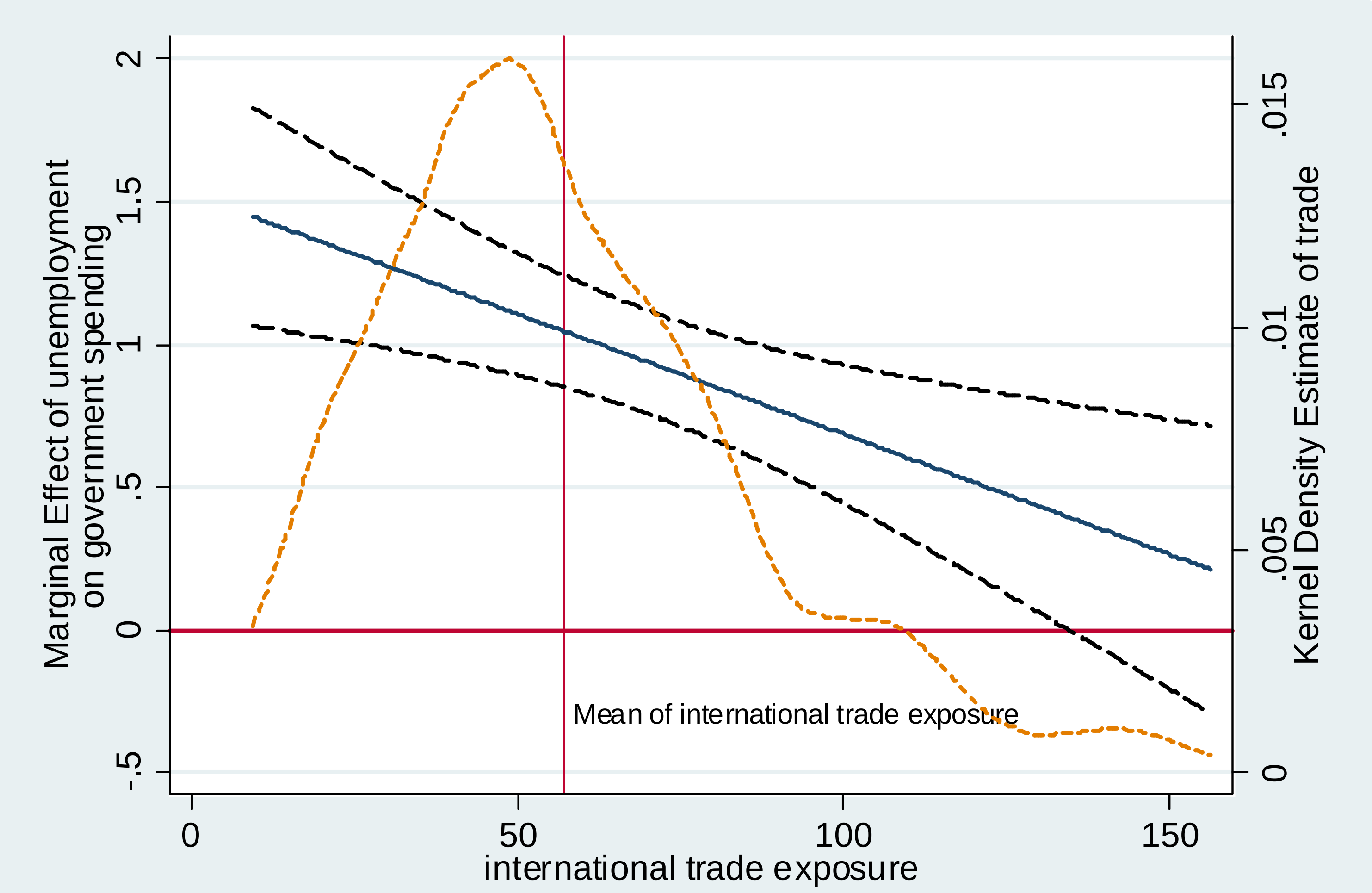
```
------------------------------------------------------------------
       Simple slope of spend on trade at unem  +/- 1sd
------------------------------------------------------------------
       unem |      Coef.    Std. Err.      t     P>|t|
------------+-----------------------------------------------------
       High |   .1673891    .0194534      8.60   0.000
       Mean |   .1981998    .0272749      7.27   0.000
        Low |   .2290105    .0363312      6.30   0.000
------------------------------------------------------------------
```


spend
70
60
50
40
30
20
0
5
10
15
20
unem
● government spending
trade+1sd
trade mean
trade-1sd


80
60
40
20
spend
0
50
100
150
trade
government spending
unem+1sd
unem mean
unem-1sd

Interaction Effects of Continuous Variables





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).

$$y_i = \alpha + \beta * D + \varepsilon_i$$

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

4. Outliers

Problem:

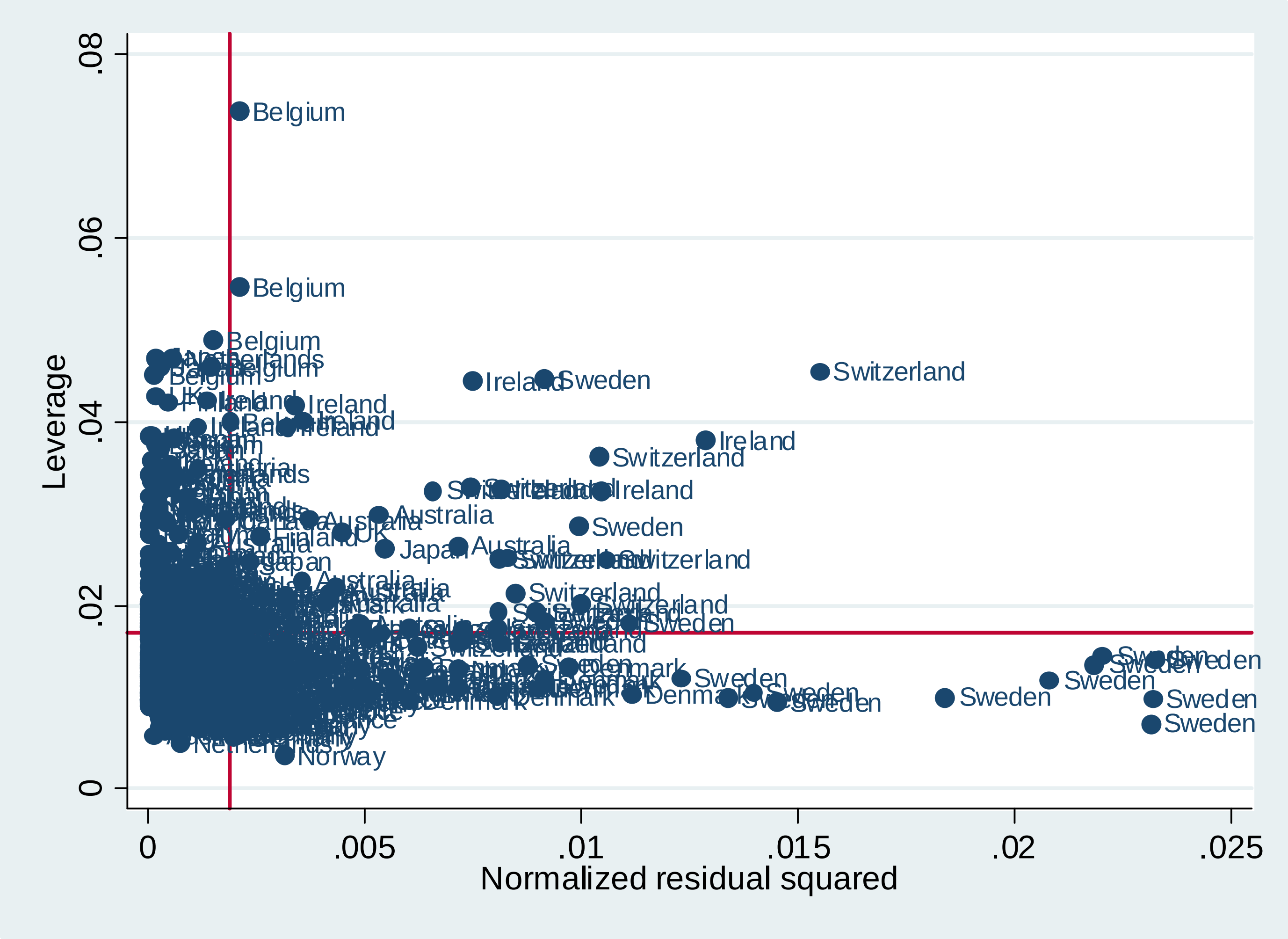
The OLS principle implies the minimization of squared residuals. From this follows that extreme cases can have a strong impact on the regression line. Inclusion/exclusion of extreme cases might change the results significantly.

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
- lvr2plot 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?

  The measure that measures how much impact each observation has on a particular coefficient is DFBETAs. 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.


Leverage
0
.02
.04
.06
.08
Normalized residual squared
0
.005
.01
.015
.02
.025
Belgium
Belgium
Belgium
Netherlands
Belgium
Ireland
Sweden
Switzerland
Ireland
Ireland
Ireland
Switzerland
Switzerland
Ireland
Australia
Sweden
Japan
Australia
Switzerland
Switzerland
Australia
Switzerland
Switzerland
Sweden
Denmark
Sweden
Denmark
Sweden
Sweden
Sweden
Sweden
Sweden
Sweden
Sweden
Sweden
Norway

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 “