

Econometrics



David Preinerstorfer

Institute for Statistics and Mathematics
Dept. of Finance, Accounting and Statistics

Spring 2026



Recap Lecture 1:

1. Linear models
2. OLS estimator $(X'X)^{-1}X'Y$, predicted values $\hat{Y} = X\hat{\beta}$, residuals $\hat{\varepsilon} = Y - \hat{Y}$.
3. Assumptions AL, AR, AX, AH (below); can you recall them?
4. Properties: sum of squares decomposition $TSS = RSS + ESS$, $R^2 = ESS/TSS$, consequences of the *normal equation* $0 = X'Y - X'X\hat{\beta} = X'\hat{\varepsilon}$.

Recap Lecture 2:

1. Statistical properties of the OLS estimator: (conditional) unbiasedness, variance covariance matrix.
2. Multicollinearity.
3. Estimating σ^2 .
4. Distribution of the OLS estimator under assumption AN and asymptotically.
5. Gauss-Markov-Theorem.
6. Testing.

To summarize the last part:

Under Assumptions AL, AR, AX, AH, AN the OLS estimator satisfies (conditionally on X) that:

$$\hat{\beta} - \beta = \text{Normal}(0, \sigma^2(X'X)^{-1}).$$

The transformation theorem (multiply by the $(j+1)$ -th unit vector) implies

$$\hat{\beta}_j - \beta_j = \text{Normal}(0, \sigma^2[(X'X)^{-1}]_{(j+1)(j+1)}).$$

Hence, upon knowing σ^2 , and for a given value β_j^* , say, one can use the test statistic

$$(\hat{\beta}_j - \beta_j^*) / \sqrt{\sigma^2[(X'X)^{-1}]_{(j+1)(j+1)}}, \quad (1)$$

to test the null hypothesis that $\beta_j = \beta_j^*$ against the alternative hypothesis that $\beta_j \neq \beta_j^*$ (reject if test statistic is large in absolute values), $\beta_j > \beta_j^*$ (reject if test statistic is large), or $\beta_j < \beta_j^*$ (reject if test statistic is small), respectively.

Note that under the null hypothesis $\beta_j = \beta_j^*$, the above test statistic is $\text{Normal}(0, 1)$, from which one obtains critical values.

If σ^2 is unknown, one uses $\hat{\sigma}^2$ instead of σ^2 in the above test statistic, i.e., uses the test statistic

$$(\hat{\beta}_j - \beta_j^*) / \sqrt{\hat{\sigma}^2 [(X'X)^{-1}]_{(j+1)(j+1)}}, \quad (2)$$

which, under Assumptions AL, AR, AX, AH, AN and under the null hypothesis $\beta_j = \beta_j^*$ is t -distributed with $N - K - 1$ degrees of freedom, from which one obtains critical values.

Example:

For testing $\beta_j = \beta_j^*$ against the alternative hypothesis that $\beta_j \neq \beta_j^*$ and for σ^2 unknown, one rejects if $|(\hat{\beta}_j - \beta_j^*) / \sqrt{\hat{\sigma}^2 [(X'X)^{-1}]_{(j+1)(j+1)}}| > c_\alpha$, for c_α the $1 - \alpha/2$ quantile of the t -distribution with $N - K - 1$ degrees of freedom.

Exercise: show that the probability to reject under the null hypothesis (i.e., the Type 1 error) then equals α (which is typically set to 5% or 1%)!

Under Assumptions AL, AR, AX, AH, AN the known σ^2 case, a confidence interval for β_j is given by

$$CI := \hat{\beta}_j \pm c_\alpha \times \sqrt{\sigma^2[(X'X)^{-1}]_{(j+1)(j+1)}}, \quad (3)$$

with c_α the $1 - \alpha/2$ quantile of the standard normal distribution.

The coverage probability $P(\beta_j \in CI) = 1 - \alpha$ (show this as an exercise).

If σ^2 is unknown, replace σ^2 above by $\hat{\sigma}^2$ and use the critical value from a t distribution.

Plan for today:

1. More testing.
2. Prediction.
3. Model specification: log-transformation, dummy variables, interactions, polynomial transformations.

Testing more than one coefficient I

- Testing the null hypothesis $\beta_j = 0$ based on t_j is only valid, if all other parameters remain in the model.
- Often, we want to test joint hypotheses about our parameters.
- E.g.; if the t_j -statistics is not significant for more than one parameter $j_1, \dots, j_q$, then one needs to test, if $\beta_{j_1} = 0, \dots, \beta_{j_q} = 0$ simultaneously.
- We cannot simply check each t_j -statistic separately. It is possible for jointly insignificant regressors to be individually significant (and vice versa).

Testing more than one coefficient II

Given the data, is it possible to reject the null hypothesis $\beta_{j_1} = 0, \dots, \beta_{j_q} = 0$?

Reject the null hypothesis, if the distance between the OLS estimator

$\tilde{\beta} = (\hat{\beta}_{j_1}, \dots, \hat{\beta}_{j_q})'$ and 0 is “large” (one-sided test).

The corresponding test statistic has to take into account that

- the standard deviations of the various OLS estimators are different.
- deviations of the OLS estimators from the true value are likely to be correlated.

Aggregate $t_{jl} = \hat{\beta}_{jl} / \text{SD}_{\hat{\beta}_{jl}}$ for $l = 1, \dots, q$, e.g., by taking the sum of squared t statistics? If the deviations of the OLS estimators $\hat{\beta}_{j1}, \dots, \hat{\beta}_{jq}$ from the true values are uncorrelated, then the aggregated test statistic

$$SSR_q = \sum_{l=1}^q \frac{\hat{\beta}_{jl}^2}{\text{SD}_{\hat{\beta}_{jl}}^2}$$

is the sum of q independent squared standard normal random variables, hence it follows a χ_q^2 distribution.

Obtaining uncorrelated coefficients

If $\beta \sim \text{Normal}_{K+1}(\mu, \Sigma)$, then the variable $\mathbf{H}\beta$, where $\mathbf{H}$ is a $q \times (K+1)$ -matrix follows also a normal distribution: $\mathbf{H}\beta \sim \text{Normal}_q(\mathbf{H}\mu, \mathbf{H}\Sigma\mathbf{H}')$. Obtaining uncorrelated coefficients:

- decompose $\Sigma = \mathbf{L}\mathbf{L}'$; then $\tilde{\beta} = \mathbf{L}^{-1}\beta \sim \text{Normal}_q(\mathbf{0}, \mathbf{I})$:

$$\begin{aligned}\tilde{\beta} &= \mathbf{L}^{-1}\beta \sim \text{Normal}_{K+1}(\mathbf{0}, \mathbf{L}^{-1}\Sigma(\mathbf{L}')^{-1}) \\ &\sim \text{Normal}_{K+1}(\mathbf{0}, \mathbf{L}^{-1}\mathbf{L}\mathbf{L}'(\mathbf{L}')^{-1}) \sim \text{Normal}_{K+1}(\mathbf{0}, \mathbf{I}).\end{aligned}$$

Therefore, $\tilde{\beta}_0, \dots, \tilde{\beta}_{K+1}$ are iid standard normal and $\tilde{\beta}'\tilde{\beta} = \beta'(\mathbf{L}')^{-1}\mathbf{L}^{-1}\beta = \beta'\Sigma^{-1}\beta$ follows a χ^2_{K+1} distribution.

Testing more than one coefficient - the Wald statistic

Under the nullhypothesis, the deviations of the OLS estimators $\hat{\beta}_{j_1}, \dots, \hat{\beta}_{j_q}$ from zero are usually correlated.

- The Wald-statistics transform the deviations to a coordinate system with independent standard normal random variables

$$W = \tilde{\beta}' \text{Cov}(\tilde{\beta})^{-1} \tilde{\beta} \sim \chi_q^2,$$

and follow a χ_q^2 -distribution with q degrees of freedom.

- Note: the χ_q^2 -distribution results only, if σ^2 is known.

Testing more than one coefficient - the F statistic I

The F-Test

- The F -statistic is obtained by substituting the unknown variance σ^2 by $\hat{\sigma}^2$ and dividing by q .
- The F -statistic is the ratio of two (independent) sum of squares, divided by the degrees of freedom, i.e., a χ_q^2/q and χ_{df}^2/df , where $df = N - K - 1$.
- If the null hypothesis $\beta_{j_1} = 0, \dots, \beta_{j_q} = 0$ is true, then the F -statistic follows a $F_{q,df}$ -distribution with parameters q (number of tested coefficients) and $df = N - K - 1$.
- Remark: for $q = 1$, $F = t_j^2$, where t_j is the t-statistic.

Testing more than one coefficient - the F statistic II

Reject the null-hypothesis, if

- the F -statistic is larger than the critical value from the corresponding $F_{q,df}$ -distribution (one-sided test).
- the corresponding p -value is smaller than the significance level. A p -value close to 0 shows that the value observed for the F -statistic is in conflict with the null hypothesis.

⇒ At least one of the coefficients $\beta_{j_1}, \dots, \beta_{j_q}$ is different from 0.

Testing more than one coefficient - the F statistic III

Do not reject the null-hypothesis, if

- the F -statistic is smaller than the critical value from the corresponding $F_{q,df}$ -distribution (one-sided test).
- The corresponding p -value is larger than the significance level. A p -value considerably larger than 0 shows that the observed value for the F -statistic is not in conflict with the null hypothesis.

⇒ There is no evidence in the data that we should reject the null hypothesis *that all coefficients $\beta_{j_1}, \dots, \beta_{j_q}$ are equal to 0.*

Testing more than one coefficient - the F statistic IV

An alternative form of the F-statistic: Equivalent forms of the F -statistic show that it measures the loss of fit from imposing the q restrictions on the model:

$$F = \frac{(\text{SSR}_r - \text{SSR})/q}{\text{SSR}/\text{df}},$$

$$F = \frac{(\text{R}^2 - \text{R}_r^2)/q}{(1 - \text{R}^2)/\text{df}},$$

where

- SSR is the minimum sum of squared residuals and R^2 is the coefficient of determination for the unrestricted regression model.
- SSR_r is the minimum sum of squared residuals and R_r^2 is the coefficient of determination for the restricted regression model.
- Note that $\text{SSR}_r > \text{SSR}$ and $\text{R}_r^2 < \text{R}^2$.

Suppose we want to test the hypothesis that two regression coefficients are equal, e.g., $\beta_1 = \beta_2$. This is equivalent to testing the following linear constraint (null hypothesis):

$$\beta_1 - \beta_2 = 0. \quad (44)$$

Test statistic based on the difference of the OLS estimators $\hat{\beta}_1 - \hat{\beta}_2$:

- If $\hat{\beta}_1 - \hat{\beta}_2$ is small, then the hypothesis (44) is not rejected.
- If $|\hat{\beta}_1 - \hat{\beta}_2|$ is large, then the hypothesis (44) is rejected.

What is the distribution of $\hat{\beta}_1 - \hat{\beta}_2$ under the null hypothesis? The distribution of

$\hat{\beta}_1 - \hat{\beta}_2$ under the null hypothesis (44) is equal to following univariate normal distribution:

$$\begin{aligned}\hat{\beta}_1 - \hat{\beta}_2 &\sim \text{Normal} \left(0, V \left(\hat{\beta}_1 - \hat{\beta}_2 \right) \right), \\ V \left(\hat{\beta}_1 - \hat{\beta}_2 \right) &= V \left(\hat{\beta}_1 \right) + V \left(\hat{\beta}_2 \right) - 2\text{Cov}(\hat{\beta}_1, \hat{\beta}_2).\end{aligned}\tag{45}$$

The test statistic

$$t = \frac{\hat{\beta}_1 - \hat{\beta}_2}{\text{se}(\hat{\beta}_1 - \hat{\beta}_2)} \sim t_{\text{df}},$$

where σ^2 is substituted by $\hat{\sigma}^2$, follows the t_{df} distribution with $\text{df} = (N - K - 1)$. Testing the linear constraint $\beta_1 - \beta_2 = 0$ for $\beta = (\beta_0, \beta_1, \dots, \beta_K)'$ is equivalent with

testing following null hypothesis:

$$R\beta = 0, \quad R = \begin{pmatrix} 0 & 1 & -1 & 0 & \cdots & 0 \end{pmatrix}.$$

Under the null hypothesis, the distribution of $R\hat{\beta}$ is derived using $\hat{\beta} - \beta \sim \text{Normal}_{K+1}(\mathbf{0}, \text{Cov}(\hat{\beta}))$:

$$R\hat{\beta} \sim \text{Normal}(\mathbf{0}, R\text{Cov}(\hat{\beta})R').$$

This yields (45).

Testing a set of linear constraints: Suppose we want to test simultaneously more than one linear constraint on the coefficients, i.e., $R\beta = d$, where R is $q \times (K + 1)$ -matrix (of full row rank). Under the null hypothesis, assumption **AN**,

$$R\hat{\beta} - d \sim \text{Normal}_q \left(\mathbf{0}, \text{RCov}(\hat{\beta})R' \right).$$

The corresponding Wald Statistics is defined as

$$W = (R\hat{\beta} - d)'(\text{RCov}(\hat{\beta})R')^{-1}(R\hat{\beta} - d),$$

follows a χ_q^2 statistics, where q is the number of linear constraints, if σ^2 is known. If σ^2 is estimated through $\hat{\sigma}^2$, then the F -statistic $F = W/q$ follows an $F_{q,df}$ -distribution.

Without Assumption AN (but grant the other assumptions), one has to resort to the asymptotic approximation

$$\sqrt{N}(\hat{\beta} - \beta) \rightarrow \text{Normal}(0, \sigma^2 E(X_i' X_i)^{-1}) \quad (4)$$

in the above arguments to conclude that the distributional results hold approximately.

This implies that the rejection probabilities and coverage probabilities are approximately the ones concluded above.

Regression models can be used for out-of-sample predictions, i.e., predict y_0 for a new observation of the regressor $\mathbf{x}_0$ which has not been included in the data. From the Gauss-Markov theorem it follows, that the prediction is the BLUE of $E(y_0)$.

The prediction error is given by:

$$e_0 = y_0 - \hat{y}_0 = \mathbf{x}_0' \boldsymbol{\beta} + \varepsilon_0 - \mathbf{x}_0' \hat{\boldsymbol{\beta}} = \varepsilon_0 + \mathbf{x}_0' (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}).$$

Under assumption **AH**, we obtain:

$$V(e_0) = \sigma^2 + \sigma^2 \mathbf{x}_0' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_0,$$

where the second term reflects the sampling error of $\hat{\boldsymbol{\beta}}$. In the simple regression model

we obtain:

$$V(e_0) = \sigma^2 \left[1 + \frac{1}{N} \left(1 + \frac{(x_0 - \bar{x})^2}{s_x^2} \right) \right]. \quad (46)$$

- The variance σ^2 of the disturbances is a lower bound for the variance of the prediction error.
- For fixed N , the variance of the prediction error increases with the distance of x_0 from the mean $\bar{x}$ of the observed regressors, in relation to s_x .
- Increasing N decreases this additional variance.

Under assumption **AN**, we obtain a $(1 - \alpha)$ prediction interval for y_0 from

$$\hat{y}_0 \pm t_{\text{df}, \alpha/2} \text{se}(e_0),$$

where

- the standard error $\text{se}(e_0)$ is the square root of the estimated variance $V(e_0)$, where σ^2 is estimated by $\hat{\sigma}^2$.
- $t_{\text{df}, \alpha/2}$ is the quantile from the t_{df} -distribution.

- Statistical Properties of OLS Estimation
- Specifications
- Regression diagnostics
- Time Series
- ARMA models

- The linearity assumption **AL** assumes that the regression model is linear in the parameters $\beta_1, \dots, \beta_K$ as well as the predictors $X_1, \dots, X_K$
- If we want to perform OLS estimation, linearity in the parameters ensures that the partial derivatives in (17) are linear in the unknown parameter β .
- The second assumption can be relaxed, i.e. for a given β , the relationship between the outcome Y and the predictors $X_1, \dots, X_K$ can be non-linear, e.g. a log-linear model.

The log-linear regression model I

Based on following nonlinear relation between Y and $X_1, \dots, X_K$:

$$Y = \tilde{\beta}_0 \cdot X_1^{\beta_1} \cdots X_K^{\beta_K} e^{\varepsilon}.$$

By taking the natural logarithm on both sides of this multiplicative model, we obtain the log-linear regression model for the transformed variables $\log Y$ and $\log X_1, \dots, \log X_K$, where $\beta_0 = \log \tilde{\beta}_0$:

$$\log Y = \beta_0 + \beta_1 \log X_1 + \dots + \beta_K \log X_K + \varepsilon. \quad (47)$$

The log-linear regression model II

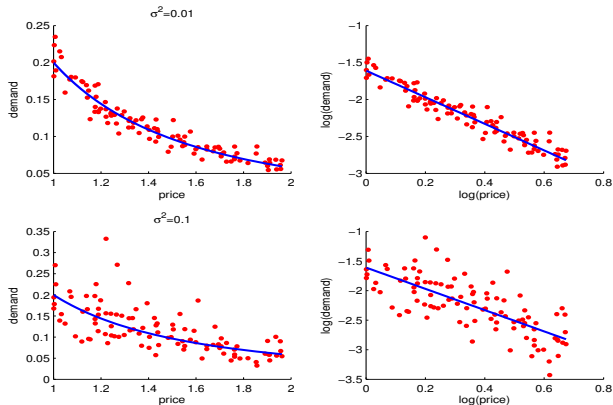
Consider, e.g. the simple log-linear regression model with $\tilde{\beta}_0 = 0.2$ and $\beta_1 = -1.8$:

$$Y = 0.2 \cdot X^{-1.8} e^{\varepsilon},$$

where $\varepsilon \sim \text{Normal}(0, \sigma^2)$. It is equivalent to:

$$\log Y = \log(0.2) - 1.8 \log X + \varepsilon.$$

The log-linear regression model III



The log-linear regression model IV

Understanding the parameters

Under **assumption AX**, the expectation of $E(\log Y)$ in the log-linear regression model is given by:

$$E(\log Y | X_1, \dots, X_K) = \beta_0 + \beta_1 \log X_1 + \dots + \beta_K \log X_K.$$

Elasticity measures how the relative change of one variable affects other variables. If $y = f(x)$, then the elasticity is defined as

$$\frac{\partial y}{y} / \frac{\partial x}{x} = \frac{\partial \log y}{\partial \log x}$$

The log-linear regression model

The expected elasticity with respect to changing X_j given by:

$$E \left(\frac{\partial \log Y}{\partial \log X_j} | X_1, \dots, X_K \right) = \beta_j.$$

Interpretation of β_j : a relative change of X_j by p percent, leads roughly to an expected relative change of Y by $\beta_j \cdot p$ percent.

Note that the expected absolute change of Y depends both on the levels of Y and X_j , whereas these levels are irrelevant in the linear model.

The log-linear regression model VI

OLS Estimation

OLS estimation is performed as before, using the transformed outcome $y'_i = \log(y_i)$ as dependent variable $\mathbf{y}$ and using the transformed predictors $x'_{j,i} = \log(x_{j,i})$ to build the matrix $\mathbf{X}$.

Note that the assumptions **AH** and **AN** concern the residuals in the transformed model (47), i.e.:

$$\log Y | X_1, \dots, X_K \sim \text{Normal}(\hat{y}^l, \sigma^2),$$

where $\hat{y}^l = \widehat{\log Y} = E(\log Y | X_1, \dots, X_K)$.

The log transformation often removes heteroskedasticity and non-normality, compared to fitting a linear regression directly to $Y, X_1, \dots, X_K$.

The log-linear regression model VII

Predictions

The log-linear models yields predictions on the log-scale. On the original scale, Y is non-Gaussian, even if assumption **AN** holds, and follows the **log normal distribution**. Therefore,

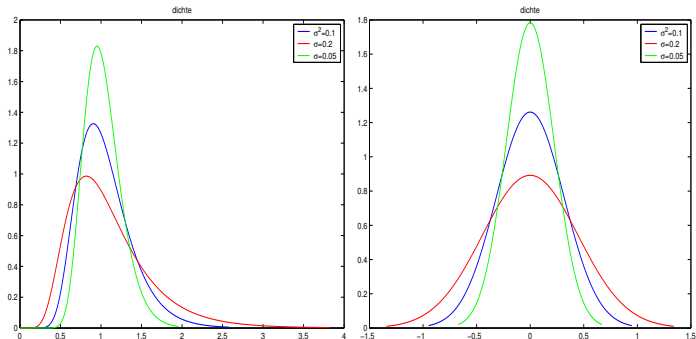
$$\hat{y} = E(Y|X_1, \dots, X_K) = \exp(\hat{y}^l) \exp(0.5 \cdot \hat{\sigma}^2).$$

$\hat{y}$ is bigger than the plug-in estimator $\exp(\hat{y}^l)$, due to Jensen's inequality:
 $\log E(Y) > E(\log Y)$.

$\hat{\sigma}^2$ is estimated from the OLS residuals in the log-linear model and is an approximate measure for the relative estimation error $(y - \hat{y})/\hat{y}$.

The log-normal distribution

Density of $\exp(\varepsilon')$ (left) and $\varepsilon' \sim \text{Normal}(0, \sigma^2)$ (right) for various values $\sigma^2 = 0.05$ (green), $\sigma^2 = 0.1$ (blue) and $\sigma^2 = 0.2$ (red)



In a semi-log models, only Y , but not the predictors:

$$\log Y = \beta_0 + \beta_1 X_1 + \dots + \beta_K X_K + \varepsilon.$$

Under the assumption **AX**,

$$E\left(\frac{\partial \log Y}{\partial X_j}\right) = \frac{\partial E(\log Y)}{\partial X_j} = \beta_j.$$

Interpretation of β_j : an absolute change of X_j by one unit leads roughly to an expected relative change of Y by $\beta_j \times 100$ percent.

Using the log-normal distribution, the expected relative change of Y can be computed exactly:

$$\frac{E(Y|X_j = x_j + 1, \dots) - E(Y|X_j = x_j, \dots)}{E(Y|X_j = x_j, \dots)} = (e^{\beta_j} - 1).$$

Another version is the following model:

$$Y = \beta_0 + \beta_1 \log X_1 + \dots + \beta_K \log X_K + \varepsilon.$$

Interpretation of β_j : a one percentage (relative) change in X_j yields an expected absolute change in Y of roughly $0.01\beta_j$ units.

- A ***dummy variable*** (binary variable) D is a variable that takes on the value 0 or 1.
- Examples: EU member ($D = 1$ if EU member, 0 otherwise), gender ($D = 1$ if male, 0 otherwise)
- Note that the labelling is not unique, a dummy variable could be labelled in two ways, i.e. for variable gender:
 - $D = 1$ if male, $D = 0$ if female;
 - $D = 1$ if female, $D = 0$ if male.

Consider a regression model with one continuous variable X and one dummy variable D :

$$Y = \beta_0 + \beta_1 D + \beta_2 X + \varepsilon.$$

If $D = 0$, then:

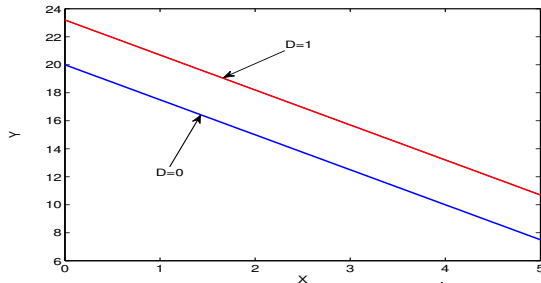
$$Y = \beta_0 + \beta_2 X + \varepsilon.$$

If $D = 1$, then:

$$Y = \beta_0 + \beta_1 + \beta_2 X + \varepsilon.$$

Regression Models with Dummy Variables III

Example: $Y = 20 + 3.2 \cdot D - 2.5 \cdot X$



The observed units are split into 2 groups according to D (e.g. into men and women). The group with $D = 0$ is called **baseline**.

Interpretation of the coefficients under assumption **AX**:

- For all values of X , the regression coefficient β_1 of D quantifies the expected difference in Y between the group corresponding to $D = 1$ and the baseline group.
- The null hypothesis $\beta_1 = 0$ corresponds to the assumption that there is no difference between the two groups.
- The effect of changing X by 1 unit is equal to β_2 and is the same for both groups.

Including more than one dummy variable

Estimate a model where D_1 is the gender (1: female, 0: male), D_2 is the brand (1: specific brand, 0: no-name), and P is the price:

$$Y = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \beta_3 P + \varepsilon,$$

- β_0 corresponds to the baseline (male, no-name product)
- β_1 : difference in the expected rating between male and female consumers (for all brands and prices).
- β_2 : difference in the expected rating between the specific brand and a no-name product (for all prices and both genders).

Categorical Variables

We can use dummy variables to control for characteristics with multiple categories ($m + 1$ categories).

Suppose one of the predictors stands for the industry. Such variables are often coded in the following way:

Coding	industry
1	Banking and Insurance
2	Consumer goods
3	IT

What is the effect of industry on a variable Y , e.g. hourly wages?

Including industry directly into a linear regression model would mean that the effect of "Consumer goods" compared to a "Banking and Insurance" is the same as the effect of "IT" compared to "Consumer goods".

To include industry as predictor in a regression model, define dummy variables D_1 , D_2 , and D_3 :

industry	D_1	D_2	D_3
Consumer goods	1	0	0
IT	0	1	0
Banking and Insurance	0	0	1

Include **two** of these dummy variables, e.g. D_1 and D_2 as predictors in a regression model:

$$Y = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \beta_3 X + \varepsilon.$$

The intercept β_0 corresponds to the baseline ($D_1 = 0, D_2 = 0$), in this case "Banking and Insurance".

Note that it is not possible to include all three dummy variables **and** an intercept, since $D_1 + D_2 + D_3 = 1$ and the corresponding matrix $\mathbf{X}$ would violate assumption **AR**.

Understanding the coefficients

- β_1 is the effect of "Consumer goods" compared to "Banking and Insurance" (baseline).
- β_2 is the effect of "IT" compared to "Banking and Insurance" (baseline).
- Are the wages the same for all industries? Test, if $\beta_1 = 0, \beta_2 = 0$.
- Is the wage gap for "Consumer goods" and "IT" (each compared to "Banking and Insurance") the same? Test, if $\beta_1 = \beta_2$, or equivalently, test the linear hypothesis $\beta_1 - \beta_2 = 0$.

Alternative specification

Including the dummy variable D_3 for "Banking and Insurance" is possible only, if no constant is included in the model, with a slightly different interpretation of the coefficients:

$$Y = \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 + \beta_4 X + \varepsilon.$$

- β_j is an industry specific intercept of the regression model for the industry corresponding to D_j . The difference in the expected outcome between two industries D_j and D_k is equal to $\beta_j - \beta_k$.
- Are the wages the same for all industries? Test, if $\beta_1 = \beta_2, \beta_2 = \beta_3$.

Example: Marketing

Baseline: unknown brand; 4 alternative brand; Y ... rating on a 20 point scale, X price (in Euro cent)

$$Y = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 + \beta_4 X + \varepsilon.$$

Estimated model from $N = 3195$ costumers:

$$Y = 20.5 + 0.88D_1 + 4.17D_2 + 0.29D_3 + 3.85D_4 - 0.3X + \varepsilon.$$

Hypothesis $\beta_2 = \beta_4, \beta_3 = 0$ cannot be rejected.

Dummy variables do not capture changes in the slope of a regression model (e.g. differences in price elasticities between men and women).

Consider interacting a dummy variable, D (e.g. gender), with a continuous variable, X (e.g. price):

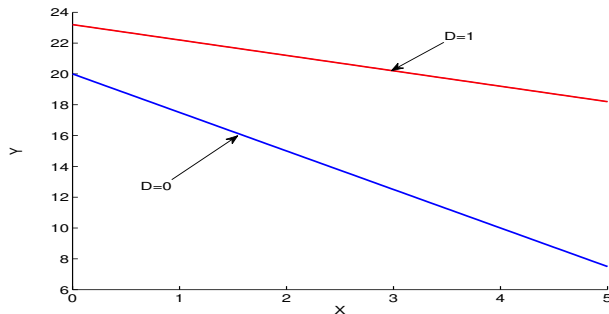
$$Y = \beta_0 + \beta_1 D + \beta_2 X + \beta_3 X \cdot D + \varepsilon. \quad (48)$$

The term $X \cdot D$ is an ***interaction term***.

The presence of the interaction term leads to non-linearity and changes the interpretation of the coefficients. The expected difference between the groups depends on the level of X .

An example of a switching regression:

Example: $Y = 20 + 3.2 \cdot D - 2.5 \cdot X + 1.5 \cdot D \cdot X$



The observed units are split into 2 groups according to D (e.g. into men and women).
If $D = 0$, then:

$$Y = \beta_0 + \beta_2 X + \varepsilon.$$

If $D = 1$, then:

$$Y = (\beta_0 + \beta_1) + (\beta_2 + \beta_3)X + \varepsilon.$$

Interpretation of the coefficients

The difference in the expected value of Y between the two groups for a given value of X is equal to:

$$E(Y|X, D = 1) - E(Y|X, D = 0) = \beta_1 + \beta_3 X.$$

An absolute change of X by ΔX leads to an expected change of Y equals to

- $\Delta E(Y) = \beta_2 \Delta X$, if $D = 0$;
- $\Delta E(Y) = (\beta_2 + \beta_3) \Delta X$, if $D = 1$.
- β_0 and β_2 are the regression coefficients of the baseline.
- The coefficient β_3 models the difference in the effect of changing X between the two groups.

Testing hypotheses

- The null hypothesis $\beta_3 = 0$ corresponds to the assumption that the effect of X is the same for both groups (interaction effect is not significant).
- The joint null hypothesis $\beta_2 = 0, \beta_2 + \beta_3 = 0$ corresponds to the assumption that the effect of X is zero for both groups.
- The joint null hypothesis $\beta_1 = 0, \beta_3 = 0$ corresponds to the assumption that the regression model is the same for both groups.

OLS estimation

OLS estimation of $\beta_0, \dots, \beta_3$ proceeds as discussed above, based on the predictors $X_1 = X$, $X_2 = D$, and $X_3 = X \cdot D$.

Note that the predictor X_3 takes the value 0 for $D = 0$. Hence,

- any variation of X_3 comes only from variation of X under $D = 1$. The SD of the corresponding coefficient β_3 will strongly depend on this variation.
- the rank condition may be even violated, if X is constant, whenever $D = 1$, as in this case the columns of $\mathbf{X}$ corresponding to $X_2 = D$ and $X_3 = XD$ are linearly dependent.

Example: Marketing

Estimate a model with price P and brand D interaction:

$$Y = \beta_0 + \beta_1 P + \beta_2 D + \beta_3 PD + \varepsilon.$$

- The price effect is negative ($\beta_1 = -0.296266$) for the baseline, and equal to $\beta_1 + \beta_3$ for brand D .
- If $\beta_2 = 0$, $\beta_3 = 0$, then brand D does not differ from the baseline.

Estimate a model with price P and several brand interactions (baseline: ordinary brand)

$$Y = \beta_0 + \beta_1 P + \beta_2 PD_1 + \beta_3 PD_2 + \beta_4 PD_3 + \varepsilon.$$

Difference-in-differences

As a special case, an interaction can be defined as the product of a time dummy T and another dummy D modelling group membership with respect to a treatment and a control group:

$$Y = \beta_0 + \beta_1 T + \beta_2 D + \beta_3 TD + \mathbf{X}\beta + \varepsilon,$$

where T is 0 before and 1 after an event and D is 1 for the treatment group and 0 for the control group. Interpretation of the coefficients (for a fixed $\mathbf{X}$):

- β_1 - average change for untreated before and after the event:

$$E(Y|T = 1, D = 0) - E(Y|T = 0, D = 0) = \beta_1.$$

- β_2 - average difference between treated and untreated before event:

$$E(Y|T = 0, D = 1) - E(Y|T = 0, D = 0) = \beta_2.$$

- β_3 - treatment effect - difference that is caused by the treatment:

$$E(Y|T = 1, D = 1) - E(Y|T = 0, D = 1) = \beta_1 + \beta_3.$$

Also in general regression models it makes sense to make the effect of a variable X_1 on Y dependent on another regressor X_2 . One way to capture such effects is including an interactions term between X_1 and X_2 :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 \cdot X_2 + \varepsilon. \quad (49)$$

Note that the relationship between X_1 and Y is non-linear as is the relationship between X_2 and Y .

OLS estimation of $\beta_0, \dots, \beta_3$ proceeds as discussed above, based on the predictors X_1 , X_2 , and $X_3 = X_1 \cdot X_2$.

Understanding the coefficients

In this framework, β_1 does not measure the expected change in Y with respect to X_1 , since $X_3 = X_1X_2$ cannot be held constant, if X_1 changes. The model implies a change in the slope of X_1 that depends on the level of X_2 :

$$\frac{\partial E(Y)}{\partial X_1} = \beta_1 + \beta_3 X_2.$$

Hence, if X_1 is changed by ΔX_1 , then the expected change of Y is approximately given by:

$$\Delta E(Y) \approx \Delta X_1 \frac{\partial E(Y)}{\partial X_1} = \Delta X_1 (\beta_1 + \beta_3 X_2).$$

Similarly, β_2 does not measure the expected change in Y with respect to X_2 , since $X_3 = X_1X_2$ cannot be held constant, if X_2 changes. The model implies a change in the slope of X_2 that depends on the level of X_1 :

$$\frac{\partial E(Y)}{\partial X_2} = \beta_2 + \beta_3 X_1.$$

Hence, if X_2 is changed by ΔX_2 , then the expected change of Y is approximately given by:

$$\Delta E(Y) \approx \Delta X_2 \frac{\partial E(Y)}{\partial X_2} = \Delta X_2 (\beta_2 + \beta_3 X_1).$$

To simplify the interpretation of the coefficients, it is useful to interpret the coefficients for typical values of the other variables such as the average values or typical deviations from the average.

The **average** effect δ_1 of X_1 on $E(Y)$ is evaluated at the sample mean $\overline{X_2}$ of X_2 :

$$\delta_1 = \beta_1 + \beta_3 \overline{X_2}.$$

Similarly, the **average** effect δ_2 of X_2 on $E(Y)$ is evaluated at the sample mean $\overline{X_1}$ of X_1 :

$$\delta_2 = \beta_2 + \beta_3 \overline{X_1}.$$

Centering the predictors

An alternative parametrization of the model is

$$Y = \delta_0 + \delta_1 X_1 + \delta_2 X_2 + \delta_3 (X_1 - \overline{X_1})(X_2 - \overline{X_2}) + \varepsilon,$$

where the interaction term involves the standardized predictors X_1 and X_2 :

- δ_1 is directly the average effect of X_1 on $E(Y)$ at the mean of X_2 .
- δ_2 is directly the average effect of X_2 on $E(Y)$ at the mean of X_1 .

Interest in modeling increasing or decreasing marginal effects of certain variables - capture such effects by including quadratic terms:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon. \quad (50)$$

Note that the relationship between X and Y is non-linear.

- OLS estimation of β_0 , β_1 , and β_2 proceeds as discussed above, based on the predictors $X_1 = X$ and $X_2 = X^2$.
- Although the relationship between X_1 and X_2 is deterministic (note that $X_2 = X_1^2$), the predictors X_1 and X_2 are typically not *linearly dependent* (exception exist, e.g. X is a dummy variable).

Testing for non-linearity

- Model (50) reduces to a model linear in X , if $\beta_2 = 0 \Rightarrow$ test the null hypothesis $H_0 : \beta_2 = 0$ to test for the presence of non-linear effects.
- If the null hypothesis $\beta_2 = 0$ is rejected, then non-linearity is present in model (50).
- Otherwise, there is no evidence that non-linearity is present.

Understanding the coefficients

The instantaneous change of $E(Y)$ is equal to the first derivative with respect to X :

$$\frac{\partial E(Y)}{\partial X} = \beta_1 + 2\beta_2 X.$$

- For $\beta_2 \neq 0$, the expected change of $E(Y)$ depends not only on β_1 , *but also on β_2 and the current value X .*
- The expected change of $E(Y)$ switches sign at the vertex of the quadratic function (50), i.e. $X_0 = -\beta_1/(2\beta_2)$.
- The model describes a monotonic behavior, if only values of X are considered, which lie on one side of the vertex.

Monotonic behaviour

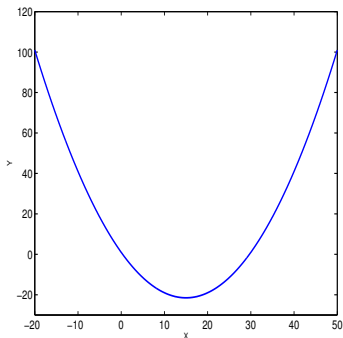
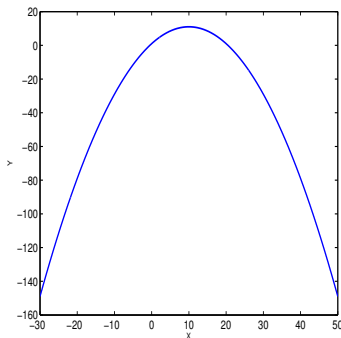
Does a monotonic behavior result, if only part of the parabola is used over the range of X , e.g. the smallest and the largest observed value x_i ?

The position of the vertex X_0 is important in this respect:

- Vertex *within* the relevant range of X - effect is not *monotone*.
- Vertex *outside* the relevant range of X - effect is *monotone*:
 - $\beta_2 > 0$ and $\min X > X_0$: increasing
 - $\beta_2 > 0$ and $\max X < X_0$: decreasing
 - $\beta_2 < 0$ and $\min X > X_0$: decreasing
 - $\beta_2 < 0$ and $\max X < X_0$: increasing

Quadratic terms **V**

Two examples, where X varies over all real numbers. Left hand side:
 $E(Y|X) = 1 + 2X - 0.1X^2$; right hand side: $E(Y|X) = 1 - 3X + 0.1X^2$



Quadratic terms VI

Same examples with X being limited to the range $[0,10]$. Left hand side:

$E(Y|X) = 1 + 2X - 0.1X^2$ ($X_0 = 10$); right hand side: $E(Y|X) = 1 - 3X + 0.1X^2$ ($X_0 = 15$)

