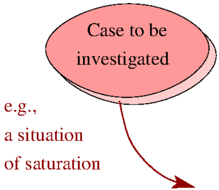
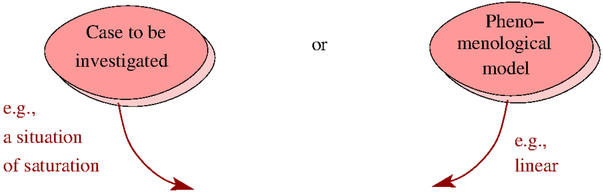
Lecture 02: Linear (Regression) Models

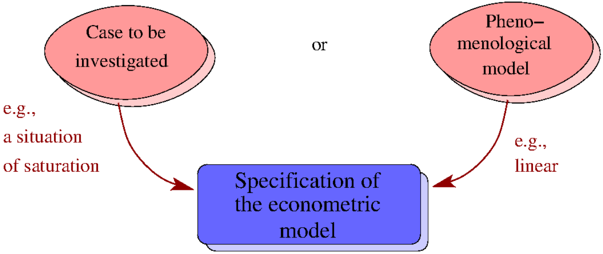
2.1 Flow Chart of the Econometric Method



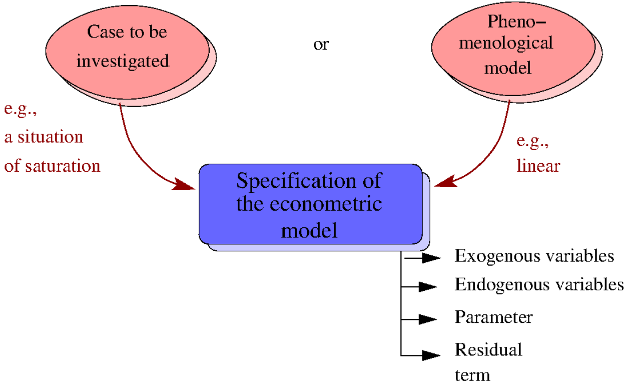
2.1 Flow Chart of the Econometric Method



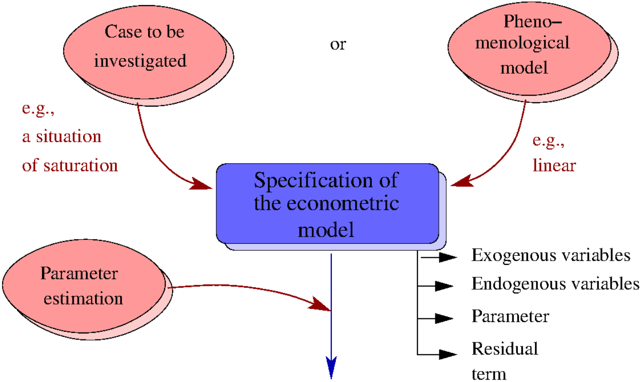
2.1 Flow Chart of the Econometric Method



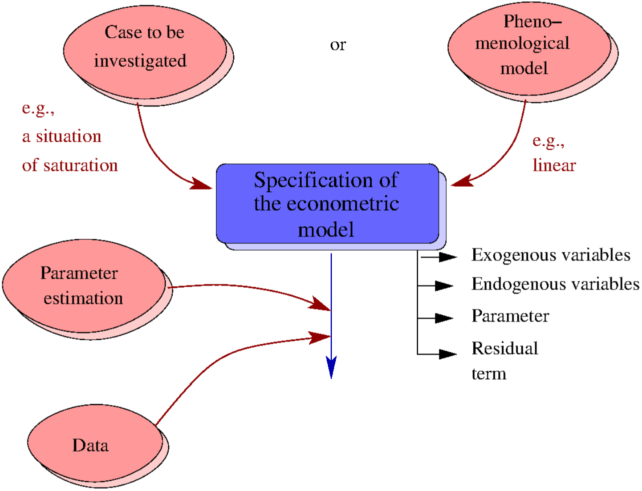
2.1 Flow Chart of the Econometric Method



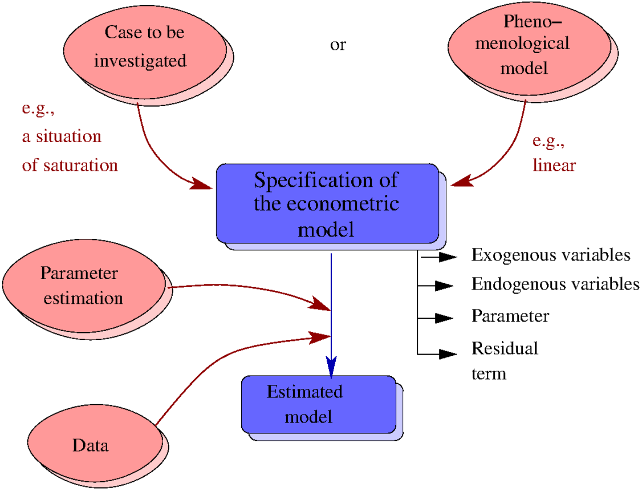
2.1 Flow Chart of the Econometric Method



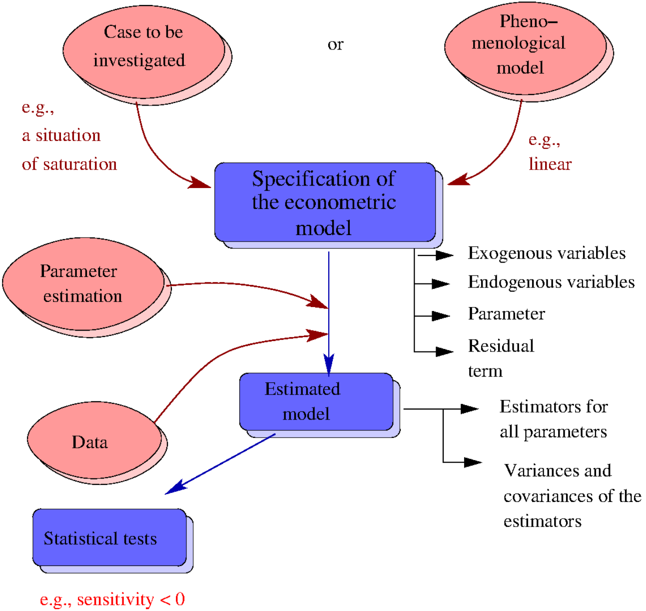
2.1 Flow Chart of the Econometric Method



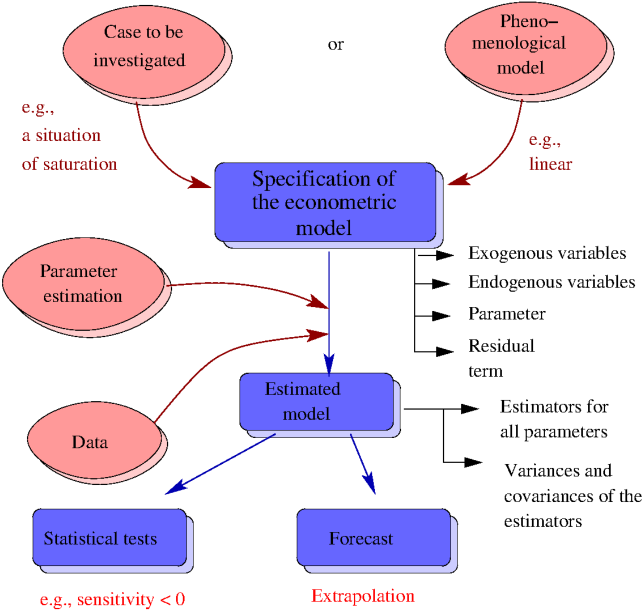
2.1 Flow Chart of the Econometric Method



2.1 Flow Chart of the Econometric Method



2.1 Flow Chart of the Econometric Method



2.2. Model Specification

As **model specification**, we denote the *complete structural definition* of the model and its consistency with the available data. There are three aspects:

- **Functional specification**: The model's exogenous and endogenous variables and the functional form in which they appear, particularly how the original exogenous variables $\tilde{x}$ are expressed in terms of linear **factors** $x_j = g_j(\tilde{x})$ by fixed, generally nonlinear functions $g_j(.)$
- **Statistical specification**: If the model contains stochastic elements, e.g., residual "error" terms we want to know how they are distributed and correlated with each other
- The **data specification** should ensure that the available data can be used to analyze the data, for example, sufficient number of data sets, check if each set contains all the exogenous and endogenous variables

2.2. Model Specification

As **model specification**, we denote the *complete structural definition* of the model and its consistency with the available data. There are three aspects:

- **Functional specification**: The model's exogenous and endogenous variables and the functional form in which they appear, particularly how the original exogenous variables $\tilde{\boldsymbol{x}}$ are expressed in terms of linear **factors** $x_j = g_j(\tilde{\boldsymbol{x}})$ by fixed, generally nonlinear functions $g_j(.)$
- **Statistical specification**: If the model contains stochastic elements, e.g., residual "error" terms we want to know how they are distributed and correlated with each other
- The **data specification** should ensure that the available data can be used to analyze the data, for example, sufficient number of data sets, check if each set contains all the exogenous and endogenous variables

2.2. Model Specification

As **model specification**, we denote the *complete structural definition* of the model and its consistency with the available data. There are three aspects:

- **Functional specification**: The model's exogenous and endogenous variables and the functional form in which they appear, particularly how the original exogenous variables $\tilde{\boldsymbol{x}}$ are expressed in terms of linear **factors** $x_j = g_j(\tilde{\boldsymbol{x}})$ by fixed, generally nonlinear functions $g_j(.)$
- **Statistical specification**: If the model contains stochastic elements, e.g., residual "error" terms we want to know how they are distributed and correlated with each other
- The **data specification** should ensure that the available data can be used to analyze the data, for example, sufficient number of data sets, check if each set contains all the exogenous and endogenous variables

2.2. Model Specification

As **model specification**, we denote the *complete structural definition* of the model and its consistency with the available data. There are three aspects:

- **Functional specification**: The model's exogenous and endogenous variables and the functional form in which they appear, particularly how the original exogenous variables $\tilde{\boldsymbol{x}}$ are expressed in terms of linear **factors** $x_j = g_j(\tilde{\boldsymbol{x}})$ by fixed, generally nonlinear functions $g_j(.)$
- **Statistical specification**: If the model contains stochastic elements, e.g., residual "error" terms we want to know how they are distributed and correlated with each other
- The **data specification** should ensure that the available data can be used to analyze the data, for example, sufficient number of data sets, check if each set contains all the exogenous and endogenous variables

WARNING

If the econometric model is not specified correctly, all sorts of problems occur, from mild to nasty:

- irrelevant: some mis-specification are detected automatically during model estimation producing "zero/zero" errors and the like
- mild: a mis-specification is not detected automatically but there is no bias and the estimation method is even efficient. However, inferential conclusions may be incorrect
- medium: the results are still unbiased but the inferential analysis is not efficient and generally gives erroneous conclusions (higher significance than in reality)
- nasty: the results are biased in an unpredictable way

Junk in, junk out!

There are lies, damned lies, and **statistics!**

WARNING

If the econometric model is not specified correctly, all sorts of problems occur, from mild to nasty:

- irrelevant: some mis-specification are detected automatically during model estimation producing "zero/zero" errors and the like
- mild: a mis-specification is not detected automatically but there is no bias and the estimation method is even efficient. However, inferential conclusions may be incorrect
- medium: the results are still unbiased but the inferential analysis is not efficient and generally gives erroneous conclusions (higher significance than in reality)
- nasty: the results are biased in an unpredictable way

Junk in, junk out!

There are lies, damned lies, and **statistics!**

WARNING

If the econometric model is not specified correctly, all sorts of problems occur, from mild to nasty:

- irrelevant: some mis-specification are detected automatically during model estimation producing "zero/zero" errors and the like
- mild: a mis-specification is not detected automatically but there is no bias and the estimation method is even efficient. However, inferential conclusions may be incorrect
- medium: the results are still unbiased but the inferential analysis is not efficient and generally gives erroneous conclusions (higher significance than in reality)
- nasty: the results are biased in an unpredictable way

Junk in, junk out!

There are lies, damned lies, and **statistics!**

WARNING

If the econometric model is not specified correctly, all sorts of problems occur, from mild to nasty:

- irrelevant: some mis-specification are detected automatically during model estimation producing "zero/zero" errors and the like
- mild: a mis-specification is not detected automatically but there is no bias and the estimation method is even efficient. However, inferential conclusions may be incorrect
- medium: the results are still unbiased but the inferential analysis is not efficient and generally gives erroneous conclusions (higher significance than in reality)
- nasty: the results are biased in an unpredictable way

Junk in, junk out!

There are lies, damned lies, and **statistics!**

WARNING

If the econometric model is not specified correctly, all sorts of problems occur, from mild to nasty:

- irrelevant: some mis-specification are detected automatically during model estimation producing "zero/zero" errors and the like
- mild: a mis-specification is not detected automatically but there is no bias and the estimation method is even efficient. However, inferential conclusions may be incorrect
- medium: the results are still unbiased but the inferential analysis is not efficient and generally gives erroneous conclusions (higher significance than in reality)
- nasty: the results are biased in an unpredictable way

Junk in, junk out!

There are lies, damned lies, and **statistics!**

WARNING

If the econometric model is not specified correctly, all sorts of problems occur, from mild to nasty:

- irrelevant: some mis-specification are detected automatically during model estimation producing "zero/zero" errors and the like
- mild: a mis-specification is not detected automatically but there is no bias and the estimation method is even efficient. However, inferential conclusions may be incorrect
- medium: the results are still unbiased but the inferential analysis is not efficient and generally gives erroneous conclusions (higher significance than in reality)
- nasty: the results are biased in an unpredictable way

Junk in, junk out!

There are lies, damned lies, and **statistics!**

WARNING

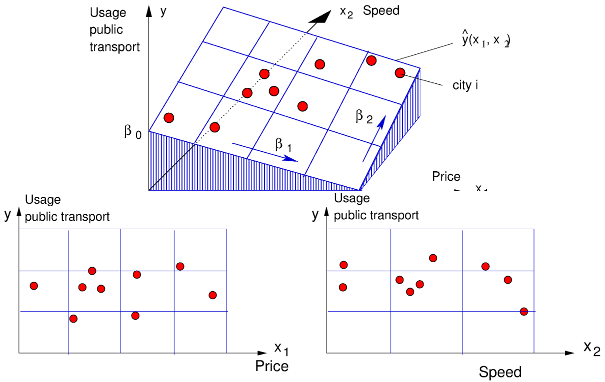
If the econometric model is not specified correctly, all sorts of problems occur, from mild to nasty:

- irrelevant: some mis-specification are detected automatically during model estimation producing "zero/zero" errors and the like
- mild: a mis-specification is not detected automatically but there is no bias and the estimation method is even efficient. However, inferential conclusions may be incorrect
- medium: the results are still unbiased but the inferential analysis is not efficient and generally gives erroneous conclusions (higher significance than in reality)
- nasty: the results are biased in an unpredictable way

Junk in, junk out!

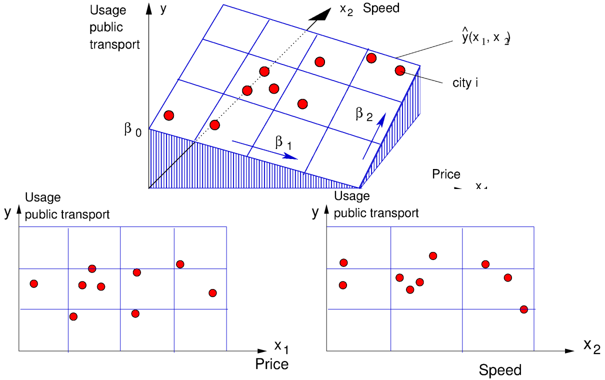
There are lies, damned lies, and **statistics!**

2.2.1 Functional specification 1: relevant factors



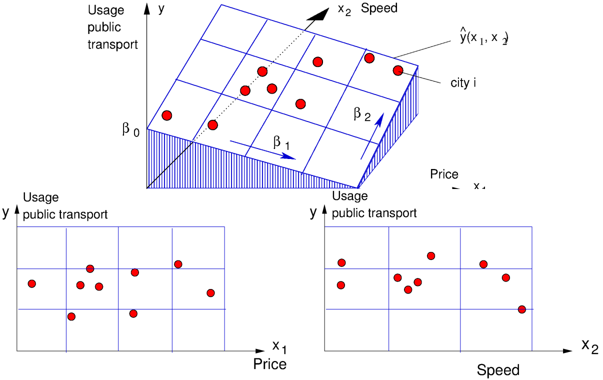
- All relevant influencing factors should be taken into account (top), no one missed (bottom).
- **Consequences of missing factors:** a bias, i.e., "junk in, junk out"
- **Consequences of superfluous factors:** no bias, higher estimation errors
- **Solution:** check for superfluous factors: *F-test*; finding missing factors: *your expertise!*

2.2.1 Functional specification 1: relevant factors



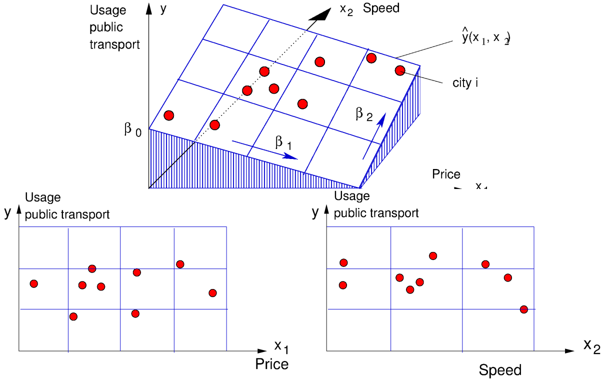
- All relevant influencing factors should be taken into account (top), no one missed (bottom).
- **Consequences of missing factors:** a **bias**, i.e., **"junk in, junk out"**
- **Consequences of superfluous factors:** **no bias, higher estimation errors**
- **Solution:** check for superfluous factors: *F-test*; finding missing factors: *your expertise!*

2.2.1 Functional specification 1: relevant factors



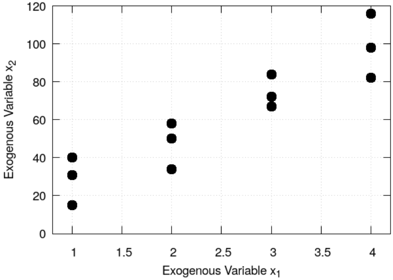
- All relevant influencing factors should be taken into account (top), no one missed (bottom).
- **Consequences of missing factors:** a **bias**, i.e., **"junk in, junk out"**
- **Consequences of superfluous factors:** **no bias, higher estimation errors**
- **Solution:** check for superfluous factors: *F-test*; finding missing factors: *your expertise!*

2.2.1 Functional specification 1: relevant factors

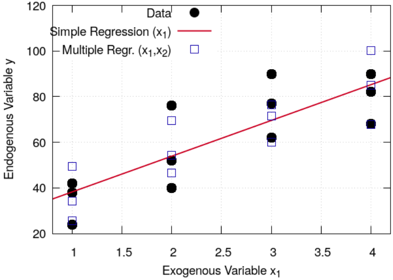


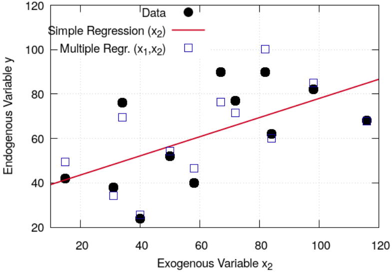
- All relevant influencing factors should be taken into account (top), no one missed (bottom).
- **Consequences of missing factors:** a **bias**, i.e., **"junk in, junk out"**
- **Consequences of superfluous factors:** **no bias, higher estimation errors**
- **Solution:** check for superfluous factors: *F-test*; finding missing factors: *your expertise!*

Example: modeling the demand for hotel rooms

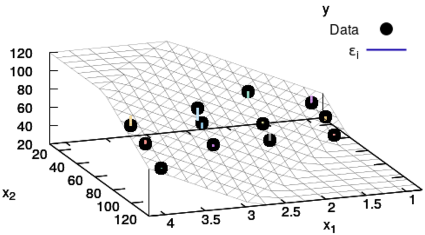


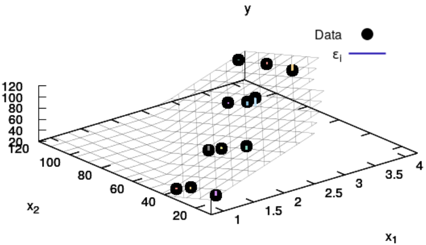
- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$ with the factors: $x_0 = 1$, x_1: proxy for quality [# stars]; x_2: price [€/night].
- The exogenous variables/factors are non-perfectly correlated: ✓
- Endogenous variable: booking rate [%]
- The demand is positively correlated with both the quality and the price (!)





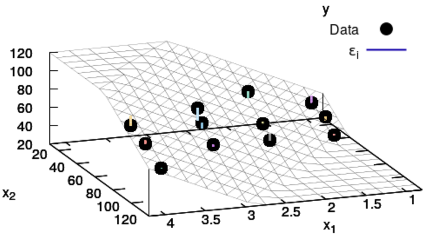
Visualization of the fit quality

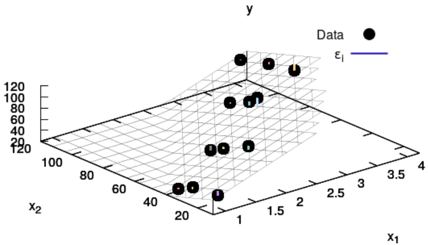




- Surface: model $\hat{y}(\boldsymbol{x}) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$
- Black bullets: data (right graphics: twice mirrored)
- OLS estimate: $\hat{\beta}_0 = 25.5$, $\hat{\beta}_1 = 38.2$, $\hat{\beta}_2 = -0.953$.
- Blue or pink bars: residuals ϵ_i (≤ 0 if below the model plane)

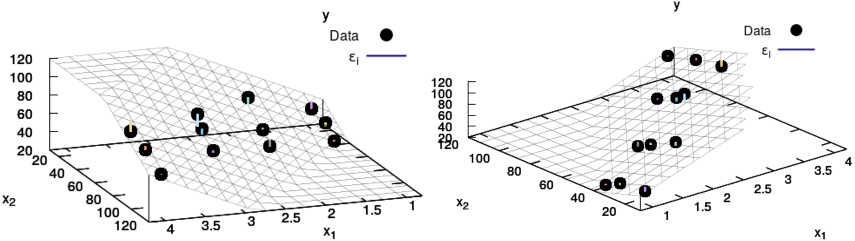
Visualization of the fit quality





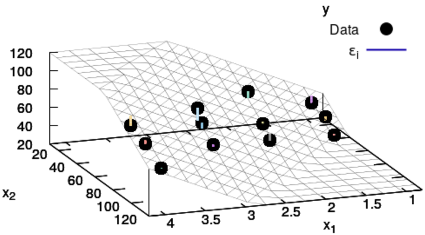
- Surface: model $\hat{y}(\boldsymbol{x}) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$
- Black bullets: data (right graphics: twice mirrored)
- OLS estimate: $\hat{\beta}_0 = 25.5$, $\hat{\beta}_1 = 38.2$, $\hat{\beta}_2 = -0.953$.
- Blue or pink bars: residuals ϵ_i (≤ 0 if below the model plane)

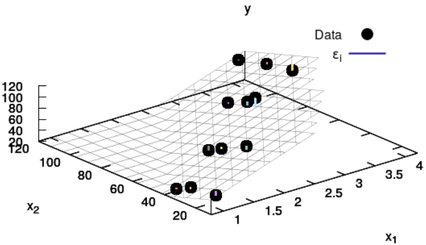
Visualization of the fit quality



- Surface: model $\hat{y}(\boldsymbol{x}) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$
- Black bullets: data (right graphics: twice mirrored)
- OLS estimate: $\hat{\beta}_0 = 25.5$, $\hat{\beta}_1 = 38.2$, $\hat{\beta}_2 = -0.953$.
- Blue or pink bars: residuals ϵ_i (≤ 0 if below the model plane)

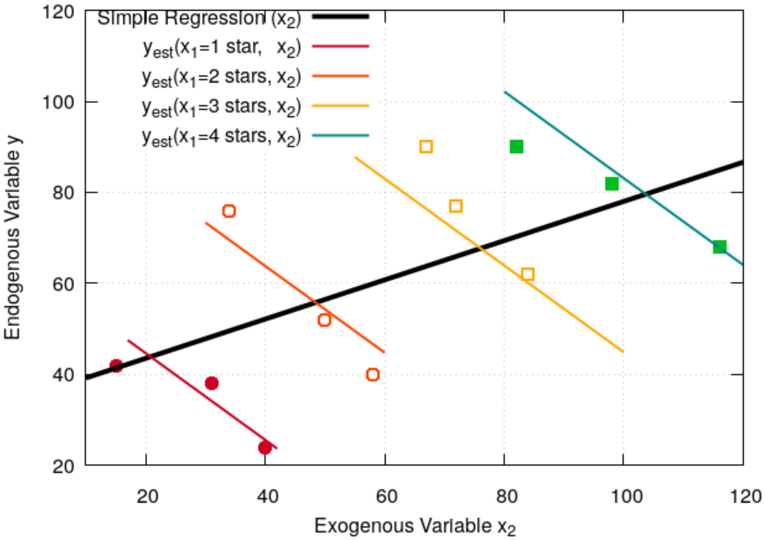
Visualization of the fit quality



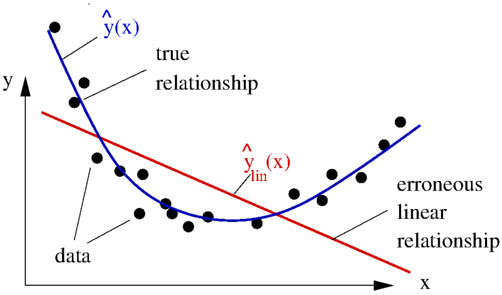


- Surface: model $\hat{y}(\boldsymbol{x}) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$
- Black bullets: data (right graphics: twice mirrored)
- OLS estimate: $\hat{\beta}_0 = 25.5$, $\hat{\beta}_1 = 38.2$, $\hat{\beta}_2 = -0.953$.
- Blue or pink bars: residuals ϵ_i (≤ 0 if below the model plane)

Effect of the correlations between the exogenous variables

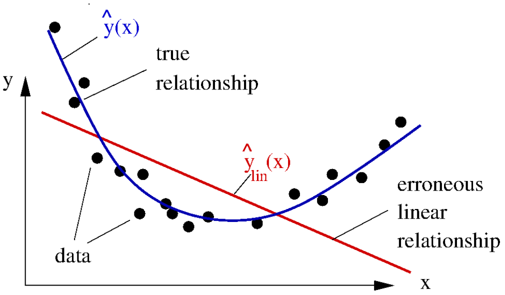


Functional specification 2: linearity



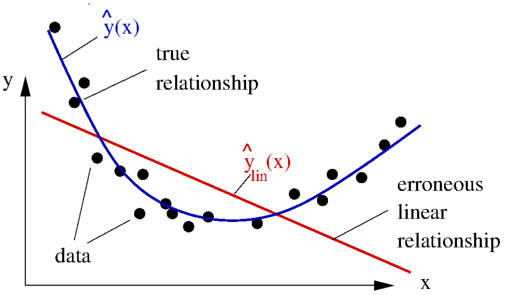
- The model should be linear which is not fulfilled here.
- **Consequences of violation:** **"junk in, junk out"**
- **Solution:** A change of the independent variable into several **factors** would be a solution here, e.g. $x'_0 = 1, x'_1 = 1/x, x'_2 = x^2$ or $x'_0 = 1, x'_1 = x, x'_2 = x^2$.

Functional specification 2: linearity



- The model should be linear which is not fulfilled here.
- **Consequences of violation: "junk in, junk out"**
- **Solution:** A change of the independent variable into several **factors** would be a solution here, e.g. $x'_0 = 1, x'_1 = 1/x, x'_2 = x^2$ or $x'_0 = 1, x'_1 = x, x'_2 = x^2$.

Functional specification 2: linearity



- The model should be linear which is not fulfilled here.
- **Consequences of violation: "junk in, junk out"**
- **Solution:** A change of the independent variable into several **factors** would be a solution here, e.g. $x'_0 = 1, x'_1 = 1/x, x'_2 = x^2$ or $x'_0 = 1, x'_1 = x, x'_2 = x^2$.

Example: fuel consumption

Assuming a constant efficiency chemical energy $\rightarrow$ mechanical energy, the required fuel per 100 km, y, is proportional to the driving resistance with the contributions

- Friction tire-road: contributions independent of the speed $\tilde{x}_1$ and proportional to the mass $\tilde{x}_2$.
- Air drag: proportional to speed squared, $\tilde{x}_1^2$, and independent from mass
- Gradient: proportional to mass times gradient $\tilde{x}_3$

In addition, there is a base consumption rate (about 0.6 liters/h) when the car is idling/driving very slowly $\Rightarrow$ contribution proportional to 1/speed [liters/km=liters/h * h/km] $\Rightarrow$ model

$$y(x) = \sum_{j=1}^{4} \beta_j x_j + \epsilon, \quad x_1 = \tilde{x}_2, \quad x_2 = \tilde{x}_1^2, \quad x_3 = \tilde{x}_2 \tilde{x}_3, \quad x_4 = \frac{1}{\tilde{x}_1}$$

Example: fuel consumption

Assuming a constant efficiency chemical energy $\rightarrow$ mechanical energy, the required fuel per 100 km, y, is proportional to the driving resistance with the contributions

- Friction tire-road: contributions independent of the speed $\tilde{x}_1$ and proportional to the mass $\tilde{x}_2$.
- Air drag: proportional to speed squared, $\tilde{x}_1^2$, and independent from mass
- Gradient: proportional to mass times gradient $\tilde{x}_3$

In addition, there is a base consumption rate (about 0.6 liters/h) when the car is idling/driving very slowly $\Rightarrow$ contribution proportional to 1/speed [liters/km=liters/h * h/km] $\Rightarrow$ model

$$y(x) = \sum_{j=1}^{4} \beta_j x_j + \epsilon, \quad x_1 = \tilde{x}_2, \quad x_2 = \tilde{x}_1^2, \quad x_3 = \tilde{x}_2 \tilde{x}_3, \quad x_4 = \frac{1}{\tilde{x}_1}$$

Example: fuel consumption

Assuming a constant efficiency chemical energy $\to$ mechanical energy, the required fuel per 100 km, y, is proportional to the driving resistance with the contributions

- Friction tire-road: contributions independent of the speed $\tilde{x}_1$ and proportional to the mass $\tilde{x}_2$.
- Air drag: proportional to speed squared, $\tilde{x}_1^2$, and independent from mass
- Gradient: proportional to mass times gradient $\tilde{x}_3$

In addition, there is a base consumption rate (about 0.6 liters/h) when the car is idling/driving very slowly $\Rightarrow$ contribution proportional to 1/speed [liters/km=liters/h * h/km] $\Rightarrow$ model

$$y(\boldsymbol{x}) = \sum_{j=1}^{4} \beta_j x_j + \epsilon, \quad x_1 = \tilde{x}_2, \quad x_2 = \tilde{x}_1^2, \quad x_3 = \tilde{x}_2 \tilde{x}_3, \quad x_4 = \frac{1}{\tilde{x}_1}$$

Example: fuel consumption

Assuming a constant efficiency chemical energy $\to$ mechanical energy, the required fuel per 100 km, y, is proportional to the driving resistance with the contributions

- Friction tire-road: contributions independent of the speed $\tilde{x}_1$ and proportional to the mass $\tilde{x}_2$.
- Air drag: proportional to speed squared, $\tilde{x}_1^2$, and independent from mass
- Gradient: proportional to mass times gradient $\tilde{x}_3$

In addition, there is a base consumption rate (about 0.6 liters/h) when the car is idling/driving very slowly $\Rightarrow$ contribution proportional to 1/speed [liters/km=liters/h * h/km] $\Rightarrow$ model

$$y(\boldsymbol{x}) = \sum_{j=1}^{4} \beta_j x_j + \epsilon, \quad x_1 = \tilde{x}_2, \quad x_2 = \tilde{x}_1^2, \quad x_3 = \tilde{x}_2 \tilde{x}_3, \quad x_4 = \frac{1}{\tilde{x}_1}$$

Example: fuel consumption

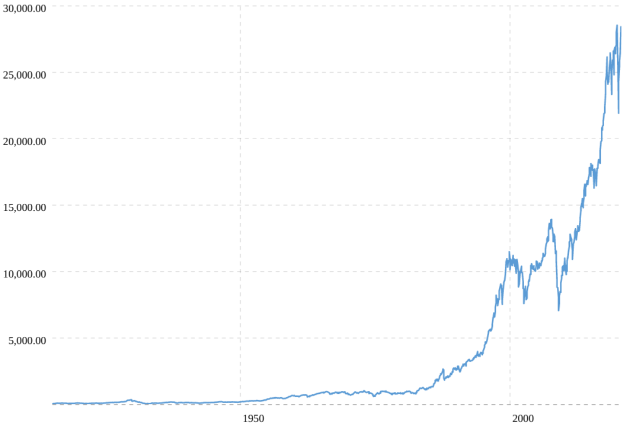
Assuming a constant efficiency chemical energy $\rightarrow$ mechanical energy, the required fuel per 100 km, y, is proportional to the driving resistance with the contributions

- Friction tire-road: contributions independent of the speed $\tilde{x}_1$ and proportional to the mass $\tilde{x}_2$.
- Air drag: proportional to speed squared, $\tilde{x}_1^2$, and independent from mass
- Gradient: proportional to mass times gradient $\tilde{x}_3$

In addition, there is a base consumption rate (about 0.6 liters/h) when the car is idling/driving very slowly $\Rightarrow$ contribution proportional to 1/speed [liters/km=liters/h * h/km] $\Rightarrow$ model

$$y(\boldsymbol{x}) = \sum_{j=1}^{4} \beta_j x_j + \epsilon, \quad x_1 = \tilde{x}_2, \quad x_2 = \tilde{x}_1^2, \quad x_3 = \tilde{x}_2 \tilde{x}_3, \quad x_4 = \frac{1}{\tilde{x}_1}$$

Transformation of the endogenous variable I

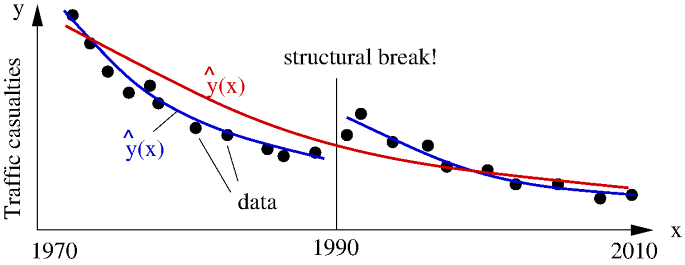


Transformation of time $\tilde{x}$ to a factor $x = \exp(\tilde{x})$ would linearize the model but the fluctuations are not i.i.d (see statistical specification below)

Transformation of the endogenous variable II

Transformation of the endogenous variable $y \to u = \ln(y)$ and $x = \tilde{x}$ gives a properly specified linear model $u(x) = \beta_0 + \beta_1 x + \epsilon$, $\epsilon \sim$ i.i.d.

Functional specification 3: homogeneity

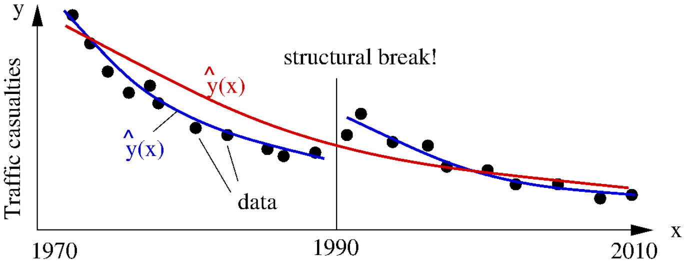


- **Consequences:** an untreated discontinuity ("structural break") in the space of the exogenous variables leads to a **bias**, i.e., **junk in, junk out**
- **Solution:** a *dummy variable* with values 0 before, 1 after the break.

? What could possibly cause a structural break?

! 1. new data basis (GDR+West Germany → Germany); 2. Redefinition of a variable (e.g., seriously injured from visit to hospital to overnight visit)

Functional specification 3: homogeneity

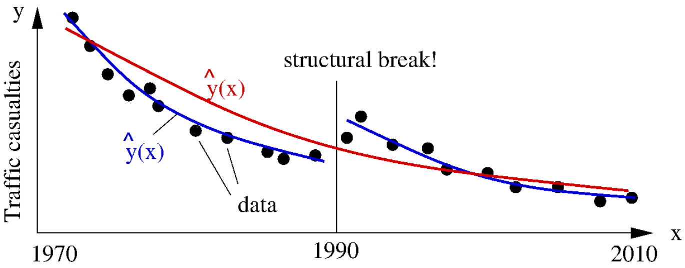


- **Consequences:** an untreated discontinuity ("structural break") in the space of the exogenous variables leads to a **bias**, i.e., **junk in, junk out**
- **Solution:** a *dummy variable* with values 0 before, 1 after the break.

? What could possibly cause a structural break?

! 1. new data basis (GDR+West Germany → Germany); 2. Redefinition of a variable (e.g., seriously injured from visit to hospital to overnight visit)

Functional specification 3: homogeneity

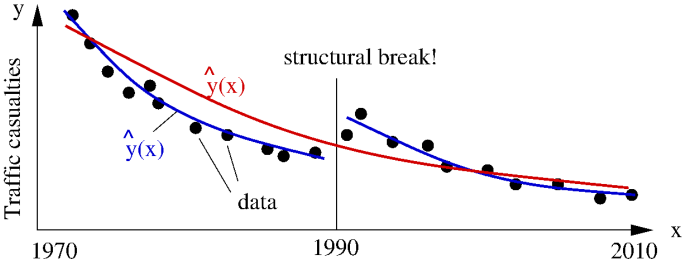


- **Consequences:** an untreated discontinuity ("structural break") in the space of the exogenous variables leads to a **bias**, i.e., **junk in, junk out**
- **Solution:** a *dummy variable* with values 0 before, 1 after the break.

? What could possibly cause a structural break?

! 1. new data basis (GDR+West Germany → Germany); 2. Redefinition of a variable (e.g., seriously injured from visit to hospital to overnight visit)

Functional specification 3: homogeneity

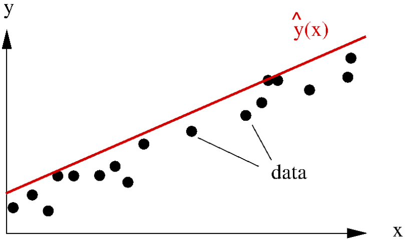


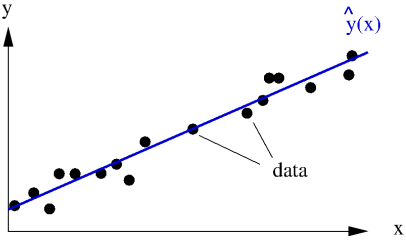
- **Consequences:** an untreated discontinuity ("structural break") in the space of the exogenous variables leads to a **bias**, i.e., **junk in, junk out**
- **Solution:** a *dummy variable* with values 0 before, 1 after the break.

? What could possibly cause a structural break?

! 1. new data basis (GDR+West Germany → Germany); 2. Redefinition of a variable (e.g., seriously injured from visit to hospital to overnight visit)

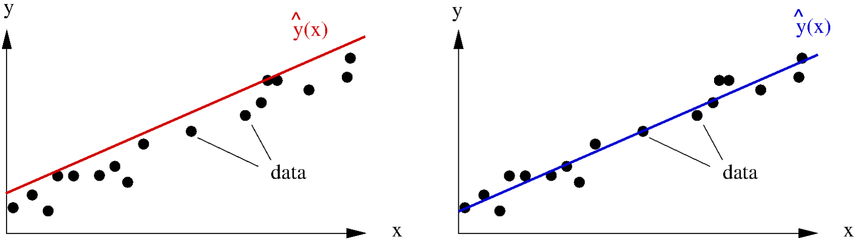
2.2.2 Statistical Specification
1. the residual ϵ has zero expectation





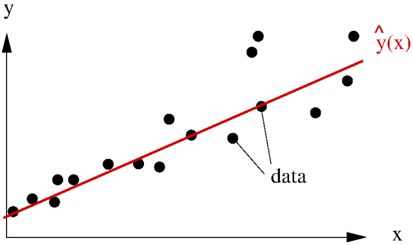
- The expectation value of the residual deviation should be $E(\epsilon) = 0$.
- **Consequences: None**: The Ordinary Least Squares (OLS) method takes care for you. If only differences matter (discrete-choice theory), this is even not relevant at all.

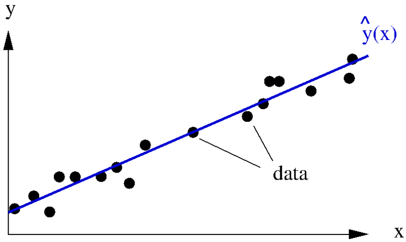
2.2.2 Statistical Specification
1. the residual ϵ has zero expectation



- The expectation value of the residual deviation should be $E(\epsilon) = 0$.
- **Consequences: None**: The Ordinary Least Squares (OLS) method takes care for you. If only differences matter (discrete-choice theory), this is even not relevant at all.

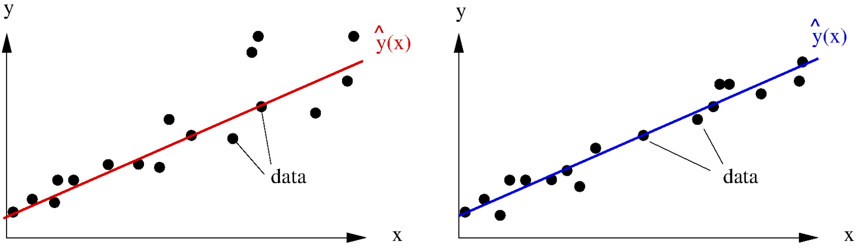
Statistical specification 2: homoskedasticity





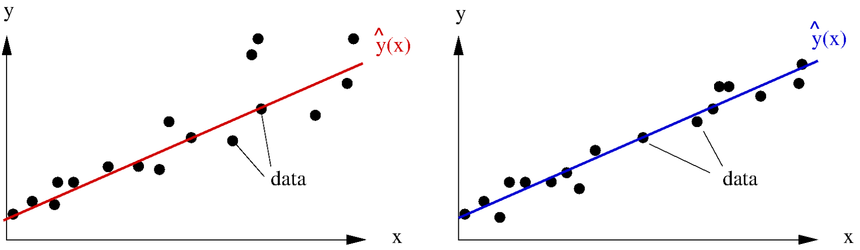
- The residual ϵ should be homoscedastic (on the right), not heteroscedastic (left).
- **Consequences:** if violated, OLS estimation remains **unbiased but is no longer efficient** (a relatively "mild" error).
- **Solution:** Advanced methods, e.g. weighted OLS; sometimes automatically resolved when transforming y as in the Dow-Jones example

Statistical specification 2: homoskedasticity



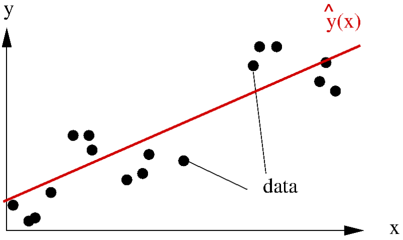
- The residual ϵ should be homoscedastic (on the right), not heteroscedastic (left).
- **Consequences:** if violated, OLS estimation remains **unbiased but is no longer efficient** (a relatively "mild" error).
- **Solution:** Advanced methods, e.g. weighted OLS; sometimes automatically resolved when transforming y as in the Dow-Jones example

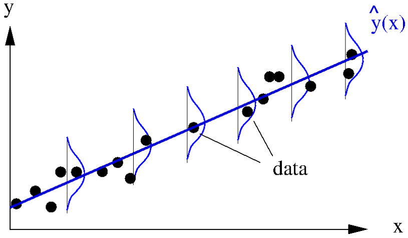
Statistical specification 2: homoskedasticity



- The residual ϵ should be homoscedastic (on the right), not heteroscedastic (left).
- **Consequences:** if violated, OLS estimation remains **unbiased but is no longer efficient** (a relatively "mild" error).
- **Solution:** Advanced methods, e.g. weighted OLS; sometimes automatically resolved when transforming y as in the Dow-Jones example

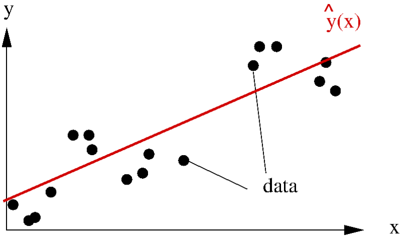
Statistical specification 3: no correlations

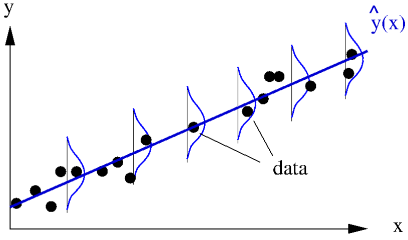




- There should be no correlation of ϵ relative to x_i or y (on the right). The model on the left is mis-specified.
- **Consequences:** relatively **mild:** (OLS estimator not efficient; underestimation of estimation errors; possibly a small bias).
- **Solution:** try identify a missing systematic factor such as a periodicity.

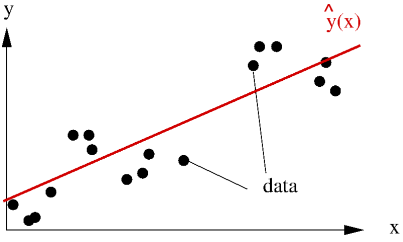
Statistical specification 3: no correlations

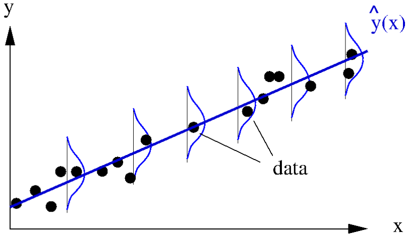




- There should be no correlation of ϵ relative to x_i or y (on the right). The model on the left is mis-specified.
- **Consequences:** relatively **mild:** (OLS estimator not efficient; underestimation of estimation errors; possibly a small bias).
- **Solution:** try identify a missing systematic factor such as a periodicity.

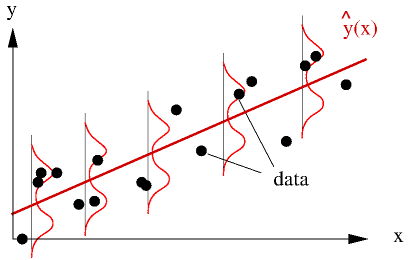
Statistical specification 3: no correlations

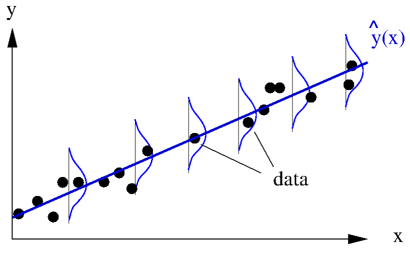




- There should be no correlation of ϵ relative to x_i or y (on the right). The model on the left is mis-specified.
- **Consequences:** relatively **mild:** (OLS estimator not efficient; underestimation of estimation errors; possibly a small bias).
- **Solution:** try identify a missing systematic factor such as a periodicity.

Statistical specification 4: Gaussian distribution

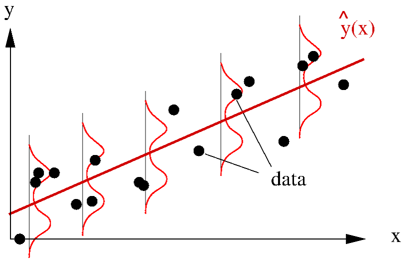


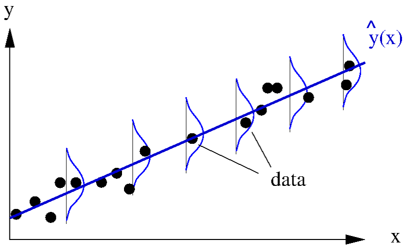


- The residual ϵ should be Gaussian distributed (right), not, e.g., bimodally distributed (left).
- **Consequences:** a violation has **very mild** consequences: OLS remains unbiased *and* efficient but the error estimates are wrong).
- All four statistical specifications can be summarized by requiring

$\epsilon \sim$ i.i.d.$N(0, \sigma^2)$ i.i.d.: **i**dentical **i**ndependent **d**istributions

Statistical specification 4: Gaussian distribution

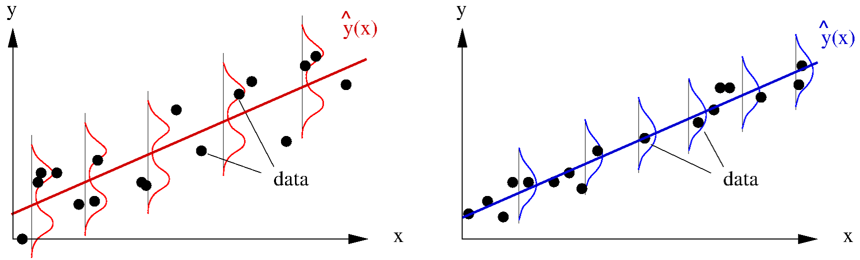




- The residual ϵ should be Gaussian distributed (right), not, e.g., bimodally distributed (left).
- **Consequences:** a violation has **very mild** consequences: OLS remains unbiased *and* efficient but the error estimates are wrong).
- All four statistical specifications can be summarized by requiring

$\epsilon \sim \text{i.i.d.} N(0, \sigma^2)$ i.i.d.: **i**dentical **i**ndependent **d**istributions

Statistical specification 4: Gaussian distribution



- The residual ϵ should be Gaussian distributed (right), not, e.g., bimodally distributed (left).
- **Consequences:** a violation has **very mild** consequences: OLS remains unbiased *and* efficient but the error estimates are wrong).
- All four statistical specifications can be summarized by requiring

$\epsilon \sim$ i.i.d.$N(0, \sigma^2)$ i.i.d.: **i**dentical **i**ndependent **d**istributions

Data specification 1: enough data

- There must be more data sets (containing all exogenous variables and the endogenous variale, each) than model parameters: $n > J + 1$
- This means, the data should overdetermine the model which is the basis for fitting.
- **Consequence of a violation:** If $n = J + 1$, the data determine the model exactly, i.e., it can be calibrated to zero residuals $\epsilon_i = 0$: *overfitting*. This is still **harmless** since OLS will detect it for you (zero residuals) and the inferential analysis will retur a "0/0 error"
- **Consequence of satisfying the requirement borderline:** If there are only a few more data points than parameters, i.e., only a few **degrees of freedom**, the data specification is OK, the estimation unbiased and efficient but the **estimation errors are big**
- **Solution:** Get more data ...

Data specification 1: enough data

- There must be more data sets (containing all exogenous variables and the endogenous variale, each) than model parameters: $n > J + 1$
- This means, the data should overdetermine the model which is the basis for fitting.
- **Consequence of a violation:** If $n = J + 1$, the data determine the model exactly, i.e., it can be calibrated to zero residuals $\epsilon_i = 0$: *overfitting*. This is still **harmless** since OLS will detect it for you (zero residuals) and the inferential analysis will retur a "0/0 error"
- **Consequence of satisfying the requirement borderline:** If there are only a few more data points than parameters, i.e., only a few **degrees of freedom**, the data specification is OK, the estimation unbiased and efficient but the **estimation errors are big**
- **Solution:** Get more data ...

Data specification 1: enough data

- There must be more data sets (containing all exogenous variables and the endogenous variale, each) than model parameters: $n > J + 1$
- This means, the data should overdetermine the model which is the basis for fitting.
- **Consequence of a violation:** If $n = J + 1$, the data determine the model exactly, i.e., it can be calibrated to zero residuals $\epsilon_i = 0$: *overfitting*. This is still **harmless** since OLS will detect it for you (zero residuals) and the inferential analysis will retur a "0/0 error"
- **Consequence of satisfying the requirement borderline:** If there are only a few more data points than parameters, i.e., only a few **degrees of freedom**, the data specification is OK, the estimation unbiased and efficient but the **estimation errors are big**
- **Solution:** Get more data ...

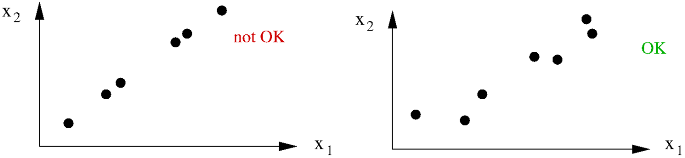
Data specification 1: enough data

- There must be more data sets (containing all exogenous variables and the endogenous variale, each) than model parameters: $n > J + 1$
- This means, the data should overdetermine the model which is the basis for fitting.
- **Consequence of a violation:** If $n = J + 1$, the data determine the model exactly, i.e., it can be calibrated to zero residuals $\epsilon_i = 0$: *overfitting*. This is still **harmless** since OLS will detect it for you (zero residuals) and the inferential analysis will retur a "0/0 error"
- **Consequence of satisfying the requirement borderline:** If there are only a few more data points than parameters, i.e., only a few **degrees of freedom**, the data specification is OK, the estimation unbiased and efficient but the **estimation errors are big**
- **Solution:** Get more data ...

Data specification 1: enough data

- There must be more data sets (containing all exogenous variables and the endogenous variale, each) than model parameters: $n > J + 1$
- This means, the data should overdetermine the model which is the basis for fitting.
- **Consequence of a violation:** If $n = J + 1$, the data determine the model exactly, i.e., it can be calibrated to zero residuals $\epsilon_i = 0$: *overfitting*. This is still **harmless** since OLS will detect it for you (zero residuals) and the inferential analysis will retur a "0/0 error"
- **Consequence of satisfying the requirement borderline:** If there are only a few more data points than parameters, i.e., only a few **degrees of freedom**, the data specification is OK, the estimation unbiased and efficient but the **estimation errors are big**
- **Solution:** Get more data ...

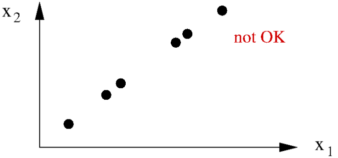
Data specification 2: no multicollinearity

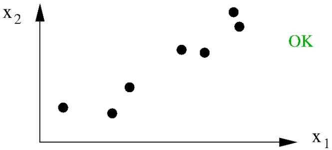


- A given exogenous variable must not be represented as a linear combination of other exogenous variables. Otherwise, the data matrix is *singular*
- However, nonperfect correlations $\neq \pm 1$ are allowed.
- Nonperfect correlations appear regularly, e.g., price vs quality
- **Consequences:** OLS **detects a perfect multicollinearity for you** by a "division by zero" error. A nearly perfect multicollinearity will lead to **large estimation errors**

If all items of all three specification categories are fulfilled, the econometric problem satisfies the **Gauß-Markov assumptions**

Data specification 2: no multicollinearity

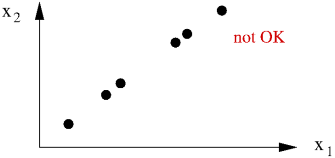


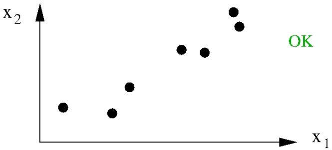


- A given exogenous variable must not be represented as a linear combination of other exogenous variables. Otherwise, the data matrix is *singular*
- However, nonperfect correlations $\neq \pm 1$ are allowed.
- Nonperfect correlations appear regularly, e.g., price vs quality
- **Consequences:** OLS **detects a perfect multicollinearity for you** by a "division by zero" error. A nearly perfect multicollinearity will lead to **large estimation errors**

If all items of all three specification categories are fulfilled, the econometric problem satisfies the **Gauß-Markov assumptions**

Data specification 2: no multicollinearity

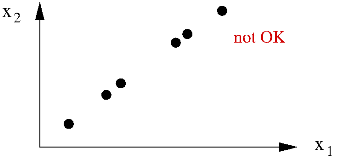


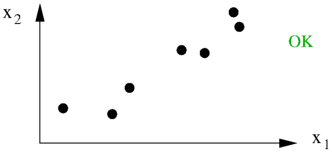


- A given exogenous variable must not be represented as a linear combination of other exogenous variables. Otherwise, the data matrix is *singular*
- However, nonperfect correlations $\neq \pm 1$ are allowed.
- Nonperfect correlations appear regularly, e.g., price vs quality
- **Consequences:** OLS **detects a perfect multicollinearity for you** by a "division by zero" error. A nearly perfect multicollinearity will lead to **large estimation errors**

If all items of all three specification categories are fulfilled, the econometric problem satisfies the **Gauß-Markov assumptions**

Data specification 2: no multicollinearity

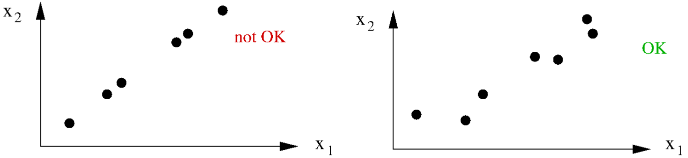




- A given exogenous variable must not be represented as a linear combination of other exogenous variables. Otherwise, the data matrix is *singular*
- However, nonperfect correlations $\neq \pm 1$ are allowed.
- Nonperfect correlations appear regularly, e.g., price vs quality
- **Consequences:** OLS **detects a perfect multicollinearity for you** by a "division by zero" error. A nearly perfect multicollinearity will lead to **large estimation errors**

If all items of all three specification categories are fulfilled, the econometric problem satisfies the **Gauß-Markov assumptions**

Data specification 2: no multicollinearity



- A given exogenous variable must not be represented as a linear combination of other exogenous variables. Otherwise, the data matrix is *singular*
- However, nonperfect correlations $\neq \pm 1$ are allowed.
- Nonperfect correlations appear regularly, e.g., price vs quality
- **Consequences:** OLS **detects a perfect multicollinearity for you** by a "division by zero" error. A nearly perfect multicollinearity will lead to **large estimation errors**

If all items of all three specification categories are fulfilled, the econometric problem satisfies the **Gauß-Markov assumptions**

How to detect multicollinearity

- Assume n data sets $\{x_{i0}, ..., x_{ij}, ..., x_{iJ}\}$, $i = 1, ..., n$ (the data sets also contain the endogenous variable but it is not relevant here)
- x_{ij} is the j^{th} exogenous factor in the i^{th} data set
- Multicollinearity exists if there is one exogenous factor x_k that can be expressed as a linear combination of all other factors $j \neq k$ in *all* data sets:

$$x_{ik} = \sum_{j \neq k} c_j x_{ij} \quad \forall i = 1, ..., n, \quad \text{constant } c_j$$

- A linear relation $x_2 = c_0 x_1$ is easy to detect but this is not the case for more complex relationships
- Solution: Check whether the **descriptive variance-covariance matrix**

$$S_{jk} = \frac{1}{n} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$$

has the full rank $J + 1$, i.e., $\det \mathbf{S} \neq 0$

- For $n < J + 1$, this is not satisfied trivially

How to detect multicollinearity

- Assume n data sets $\{x_{i0}, ..., x_{ij}, ..., x_{iJ}\}$, $i = 1, ..., n$ (the data sets also contain the endogenous variable but it is not relevant here)
- x_{ij} is the j^{th} exogenous factor in the i^{th} data set
- Multicollinearity exists if there is one exogenous factor x_k that can be expressed as a linear combination of all other factors $j \neq k$ in *all* data sets:
$$x_{ik} = \sum_{j \neq k} c_j x_{ij} \quad \forall i = 1, ..., n, \quad \text{constant } c_j$$
- A linear relation $x_2 = c_0 x_1$ is easy to detect but this is not the case for more complex relationships
- Solution: Check whether the **descriptive variance-covariance matrix**
$$S_{jk} = \frac{1}{n} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$$
has the full rank $J + 1$, i.e., $\det \mathbf{S} \neq 0$
- For $n < J + 1$, this is not satisfied trivially

How to detect multicollinearity

- Assume n data sets $\{x_{i0}, ..., x_{ij}, ..., x_{iJ}\}$, $i = 1, ..., n$ (the data sets also contain the endogenous variable but it is not relevant here)
- x_{ij} is the j^{th} exogenous factor in the i^{th} data set
- Multicollinearity exists if there is one exogenous factor x_k that can be expressed as a linear combination of all other factors $j \neq k$ in *all* data sets:
$$x_{ik} = \sum_{j \neq k} c_j x_{ij} \quad \forall i = 1, ..., n, \quad \text{constant } c_j$$
- A linear relation $x_2 = c_0 x_1$ is easy to detect but this is not the case for more complex relationships
- Solution: Check whether the **descriptive variance-covariance matrix**
$$S_{jk} = \frac{1}{n} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$$
has the full rank $J + 1$, i.e., det $\mathbf{S} \neq 0$
- For $n < J + 1$, this is not satisfied trivially

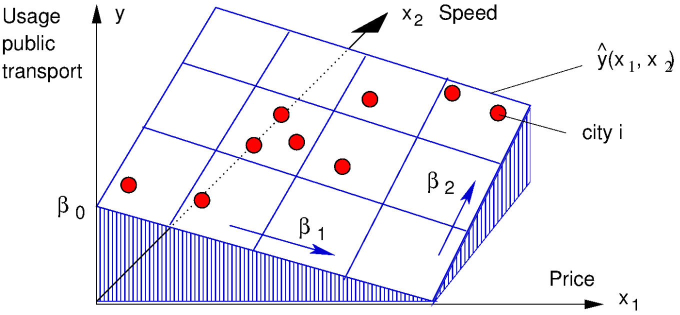
How to detect multicollinearity

- Assume n data sets $\{x_{i0}, ..., x_{ij}, ..., x_{iJ}\}$, $i = 1, ..., n$ (the data sets also contain the endogenous variable but it is not relevant here)
- x_{ij} is the j^{th} exogenous factor in the i^{th} data set
- Multicollinearity exists if there is one exogenous factor x_k that can be expressed as a linear combination of all other factors $j \neq k$ in *all* data sets:
$$x_{ik} = \sum_{j \neq k} c_j x_{ij} \quad \forall i = 1, ..., n, \quad \text{constant } c_j$$
- A linear relation $x_2 = c_0 x_1$ is easy to detect but this is not the case for more complex relationships
- Solution: Check whether the **descriptive variance-covariance matrix**
$$S_{jk} = \frac{1}{n} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$$
has the full rank $J + 1$, i.e., det $\mathbf{S} \neq 0$
- For $n < J + 1$, this is not satisfied trivially

How to detect multicollinearity

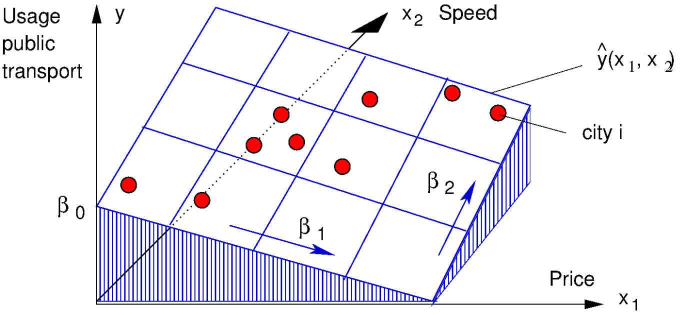
- Assume n data sets $\{x_{i0}, ..., x_{ij}, ..., x_{iJ}\}$, $i = 1, ..., n$ (the data sets also contain the endogenous variable but it is not relevant here)
- x_{ij} is the j^{th} exogenous factor in the i^{th} data set
- Multicollinearity exists if there is one exogenous factor x_k that can be expressed as a linear combination of all other factors $j \neq k$ in *all* data sets:
$$x_{ik} = \sum_{j \neq k} c_j x_{ij} \quad \forall i = 1, ..., n, \quad \text{constant } c_j$$
- A linear relation $x_2 = c_0 x_1$ is easy to detect but this is not the case for more complex relationships
- Solution: Check whether the **descriptive variance-covariance matrix**
$$S_{jk} = \frac{1}{n} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$$
has the full rank $J + 1$, i.e., $\det \mathbf{S} \neq 0$
- For $n < J + 1$, this is not satisfied trivially

Data specification 2: example



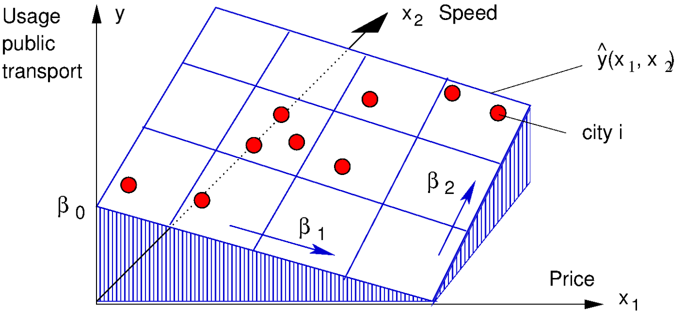
- The normalized demand y_i for public transport in city i depends on the price x_{i1} and the quality x_{i2} (proxy: speed) of the service.
- Parameters: intercept β_0, price sensitivity β_1, appraisal for quality β_2.
- Price and quality are correlated but not perfectly so.
- This model structure is quite generic for products and services.

Data specification 2: example



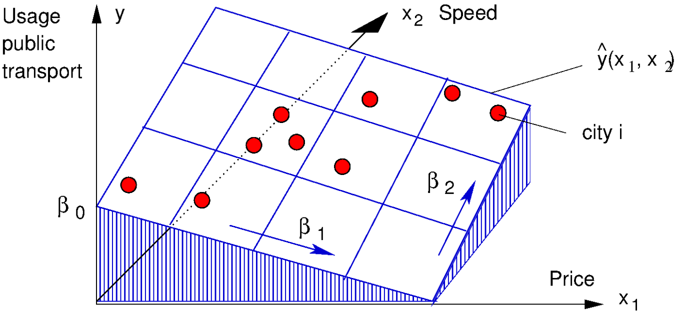
- The normalized demand y_i for public transport in city i depends on the price x_{i1} and the quality x_{i2} (proxy: speed) of the service.
- Parameters: intercept β_0, price sensitivity β_1, appraisal for quality β_2.
- Price and quality are correlated but not perfectly so.
- This model structure is quite generic for products and services.

Data specification 2: example



- The normalized demand y_i for public transport in city i depends on the price x_{i1} and the quality x_{i2} (proxy: speed) of the service.
- Parameters: intercept β_0, price sensitivity β_1, appraisal for quality β_2.
- Price and quality are correlated but not perfectly so.
- This model structure is quite generic for products and services.

Data specification 2: example



- The normalized demand y_i for public transport in city i depends on the price x_{i1} and the quality x_{i2} (proxy: speed) of the service.
- Parameters: intercept β_0, price sensitivity β_1, appraisal for quality β_2.
- Price and quality are correlated but not perfectly so.
- This model structure is quite generic for products and services.

2.3. Ordinary Least Squares (OLS) Estimation

- Given is a linear model of the form

$$y(\boldsymbol{x}) = \boldsymbol{\beta}'\boldsymbol{x} + \epsilon = \hat{y}(\boldsymbol{x}) + \epsilon, \quad \epsilon \sim i.i.d.\, N(0, \sigma^2)$$

  satisfying the Gauß-Markow specifications (the Gaussian distribution of the ϵ_i is not required, here)
- Given is also data in the form of n multidimensional data points containing all observations and satisfying the Gauß-Markow specifications as well:

$$\{\boldsymbol{p}_i = (x_{i0}, ..., x_{iJ}, y_i)', \quad i = 1, ..., n\}$$

- Searched for is a parameter estimator $\hat{\beta}$ minimizing the sum of squared errors between data and model prediction with respect to the parameters:

$$\hat{\beta} = \text{arg min}_{\beta}\, S(\beta)$$

  where

$$S(\beta) = \epsilon'\epsilon = (\boldsymbol{y} - \mathbf{X}\beta)'(\boldsymbol{y} - \mathbf{X}\beta).$$

2.3. Ordinary Least Squares (OLS) Estimation

- Given is a linear model of the form

$$y(\boldsymbol{x}) = \boldsymbol{\beta}'\boldsymbol{x} + \epsilon = \hat{y}(\boldsymbol{x}) + \epsilon, \quad \epsilon \sim i.i.d.\, N(0, \sigma^2)$$

  satisfying the Gauß-Markow specifications (the Gaussian distribution of the ϵ_i is not required, here)

- Given is also data in the form of n multidimensional data points containing all observations and satisfying the Gauß-Markow specifications as well:

$$\{\boldsymbol{p}_i = (x_{i0}, ..., x_{iJ}, y_i)', \quad i = 1, ..., n\}$$

- Searched for is a parameter estimator $\hat{\beta}$ minimizing the sum of squared errors between data and model prediction with respect to the parameters:

$$\hat{\beta} = \text{arg min}_{\beta}\, S(\beta)$$

  where

$$S(\beta) = \boldsymbol{\epsilon}'\boldsymbol{\epsilon} = (\boldsymbol{y} - \mathbf{X}\,\beta)'(\boldsymbol{y} - \mathbf{X}\,\beta).$$

2.3. Ordinary Least Squares (OLS) Estimation

- Given is a linear model of the form

$$y(\boldsymbol{x}) = \boldsymbol{\beta}'\boldsymbol{x} + \epsilon = \hat{y}(\boldsymbol{x}) + \epsilon, \quad \epsilon \sim i.i.d.\, N(0, \sigma^2)$$

  satisfying the Gauß-Markow specifications (the Gaussian distribution of the ϵ_i is not required, here)

- Given is also data in the form of n multidimensional data points containing all observations and satisfying the Gauß-Markow specifications as well:

$$\{\boldsymbol{p}_i = (x_{i0}, ..., x_{iJ}, y_i)', \quad i = 1, ..., n\}$$

- Searched for is a parameter estimator $\hat{\beta}$ minimizing the sum of squared errors between data and model prediction with respect to the parameters:

$$\hat{\beta} = \text{arg min}_{\beta} S(\beta)$$

  where

$$S(\beta) = \epsilon'\epsilon = (\boldsymbol{y} - \mathbf{X}\,\beta)'(\boldsymbol{y} - \mathbf{X}\,\beta).$$

2.3. Ordinary Least Squares (OLS) Estimation

- Given is a linear model of the form

$$y(\boldsymbol{x}) = \boldsymbol{\beta}'\boldsymbol{x} + \epsilon = \hat{y}(\boldsymbol{x}) + \epsilon, \quad \epsilon \sim i.i.d.\, N(0, \sigma^2)$$

  satisfying the Gauß-Markow specifications (the Gaussian distribution of the ϵ_i is not required, here)

- Given is also data in the form of n multidimensional data points containing all observations and satisfying the Gauß-Markow specifications as well:

$$\{\boldsymbol{p}_i = (x_{i0}, ..., x_{iJ}, y_i)', \quad i = 1, ..., n\}$$

- Searched for is a parameter estimator $\hat{\boldsymbol{\beta}}$ minimizing the sum of squared errors between data and model prediction with respect to the parameters:

$$\boxed{\hat{\boldsymbol{\beta}} = \text{arg min}_{\boldsymbol{\beta}}\, S(\boldsymbol{\beta})}$$

  where

$$S(\boldsymbol{\beta}) = \boldsymbol{\epsilon}'\boldsymbol{\epsilon} = (\boldsymbol{y} - \mathbf{X}\,\boldsymbol{\beta})'(\boldsymbol{y} - \mathbf{X}\,\boldsymbol{\beta}).$$

Determining the OLS estimator

$$S = (\boldsymbol{y} - \mathbf{X}\,\boldsymbol{\beta})'\,(\boldsymbol{y} - \mathbf{X}\,\boldsymbol{\beta})$$

Determining the OLS estimator

$$
\begin{aligned}
S &= (\boldsymbol{y} - \mathbf{X}\,\boldsymbol{\beta})'\,(\boldsymbol{y} - \mathbf{X}\,\boldsymbol{\beta}) \\
\text{[distributivity} \rightarrow\text{]} \quad &= \boldsymbol{y}'\boldsymbol{y} - (\mathbf{X}\,\boldsymbol{\beta})'\boldsymbol{y} - \boldsymbol{y}'\mathbf{X}\,\boldsymbol{\beta} + (\mathbf{X}\,\boldsymbol{\beta})'\mathbf{X}\,\boldsymbol{\beta}
\end{aligned}
$$

Determining the OLS estimator

$$\begin{aligned} S &= (\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})'(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}) \\ \text{[distributivity} \rightarrow\text{]} &= \boldsymbol{y}'\boldsymbol{y} - (\mathbf{X}\boldsymbol{\beta})'\boldsymbol{y} - \boldsymbol{y}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}\boldsymbol{\beta})'\mathbf{X}\boldsymbol{\beta} \\ \text{[transpose rule} \rightarrow\text{]} &= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{\beta}'\mathbf{X}'\boldsymbol{y} - \boldsymbol{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \end{aligned}$$

Determining the OLS estimator

$$
\begin{aligned}
S &= (\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})'(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}) \\
\text{[distributivity} \rightarrow\text{]} &= \boldsymbol{y}'\boldsymbol{y} - (\mathbf{X}\boldsymbol{\beta})'\boldsymbol{y} - \boldsymbol{y}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}\boldsymbol{\beta})'\mathbf{X}\boldsymbol{\beta} \\
\text{[transpose rule} \rightarrow\text{]} &= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{\beta}'\mathbf{X}'\boldsymbol{y} - \boldsymbol{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\
\text{[transpose rule} \rightarrow\text{]} &= \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'(\mathbf{X}'\boldsymbol{y}) + \boldsymbol{\beta}'\left(\mathbf{X}'\mathbf{X}\right)\boldsymbol{\beta}
\end{aligned}
$$

Determining the OLS estimator

$$
\begin{aligned}
S &= (\boldsymbol{y} - \mathbf{X}\,\boldsymbol{\beta})'\,(\boldsymbol{y} - \mathbf{X}\,\boldsymbol{\beta}) \\
\text{[distributivity} \rightarrow] &= \boldsymbol{y}'\boldsymbol{y} - (\mathbf{X}\,\boldsymbol{\beta})'\boldsymbol{y} - \boldsymbol{y}'\mathbf{X}\,\boldsymbol{\beta} + (\mathbf{X}\,\boldsymbol{\beta})'\mathbf{X}\,\boldsymbol{\beta} \\
\text{[transpose rule} \rightarrow] &= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{\beta}'\mathbf{X}'\boldsymbol{y} - \boldsymbol{y}'\mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\,\boldsymbol{\beta} \\
\text{[transpose rule} \rightarrow] &= \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'(\mathbf{X}'\boldsymbol{y}) + \boldsymbol{\beta}'\left(\mathbf{X}'\mathbf{X}\right)\boldsymbol{\beta}
\end{aligned}
$$

Taking the derivative $\frac{\partial}{\partial \boldsymbol{\beta}}$ respecting $\frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}'\boldsymbol{a}) = \boldsymbol{a}$ and $\frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}'\mathbf{A}\,\boldsymbol{\beta}) = (\mathbf{A} + \mathbf{A}')\boldsymbol{\beta}$ with $\mathbf{A} = \mathbf{X}'\mathbf{X}$ symmetric:

Determining the OLS estimator

$$
\begin{aligned}
S &= (\boldsymbol{y} - \mathbf{X}\,\boldsymbol{\beta})'\,(\boldsymbol{y} - \mathbf{X}\,\boldsymbol{\beta}) \\
\text{[distributivity} \rightarrow\text{]} &= \boldsymbol{y}'\boldsymbol{y} - (\mathbf{X}\,\boldsymbol{\beta})'\boldsymbol{y} - \boldsymbol{y}'\mathbf{X}\,\boldsymbol{\beta} + (\mathbf{X}\,\boldsymbol{\beta})'\mathbf{X}\,\boldsymbol{\beta} \\
\text{[transpose rule} \rightarrow\text{]} &= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{\beta}'\mathbf{X}'\boldsymbol{y} - \boldsymbol{y}'\mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\,\boldsymbol{\beta} \\
\text{[transpose rule} \rightarrow\text{]} &= \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'(\mathbf{X}'\boldsymbol{y}) + \boldsymbol{\beta}'\left(\mathbf{X}'\mathbf{X}\right)\boldsymbol{\beta}
\end{aligned}
$$

Taking the derivative $\frac{\partial}{\partial \boldsymbol{\beta}}$ respecting $\frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}'\boldsymbol{a}) = \boldsymbol{a}$ and $\frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}'\mathbf{A}\,\boldsymbol{\beta}) = (\mathbf{A} + \mathbf{A}')\boldsymbol{\beta}$ with $\mathbf{A} = \mathbf{X}'\mathbf{X}$ symmetric:

$$
\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0 - 2\mathbf{X}'\boldsymbol{y} + 2\mathbf{X}'\mathbf{X}\,\boldsymbol{\beta} \stackrel{!}{=} 0
$$

Determining the OLS estimator

$$
\begin{aligned}
S &= (\boldsymbol{y} - \mathbf{X}\,\boldsymbol{\beta})'\,(\boldsymbol{y} - \mathbf{X}\,\boldsymbol{\beta}) \\
\text{[distributivity} \rightarrow\text{]} &= \boldsymbol{y}'\boldsymbol{y} - (\mathbf{X}\,\boldsymbol{\beta})'\boldsymbol{y} - \boldsymbol{y}'\mathbf{X}\,\boldsymbol{\beta} + (\mathbf{X}\,\boldsymbol{\beta})'\mathbf{X}\,\boldsymbol{\beta} \\
\text{[transpose rule} \rightarrow\text{]} &= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{\beta}'\mathbf{X}'\boldsymbol{y} - \boldsymbol{y}'\mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\,\boldsymbol{\beta} \\
\text{[transpose rule} \rightarrow\text{]} &= \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'(\mathbf{X}'\boldsymbol{y}) + \boldsymbol{\beta}'\left(\mathbf{X}'\mathbf{X}\right)\boldsymbol{\beta}
\end{aligned}
$$

Taking the derivative $\frac{\partial}{\partial \boldsymbol{\beta}}$ respecting $\frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}'\boldsymbol{a}) = \boldsymbol{a}$ and $\frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}'\mathbf{A}\,\boldsymbol{\beta}) = (\mathbf{A} + \mathbf{A}')\boldsymbol{\beta}$ with $\mathbf{A} = \mathbf{X}'\mathbf{X}$ symmetric:

$$
\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0 - 2\mathbf{X}'\boldsymbol{y} + 2\mathbf{X}'\mathbf{X}\,\boldsymbol{\beta} \stackrel{!}{=} 0
$$

$$
\mathbf{X}'\mathbf{X}\,\boldsymbol{\beta} = \mathbf{X}'\boldsymbol{y}
$$

Determining the OLS estimator

$$\begin{array}{rcl} S & = & (\boldsymbol{y} - \mathbf{X}\,\boldsymbol{\beta})'\,(\boldsymbol{y} - \mathbf{X}\,\boldsymbol{\beta}) \\ \text{[distributivity} \rightarrow\text{]} & = & \boldsymbol{y}'\boldsymbol{y} - (\mathbf{X}\,\boldsymbol{\beta})'\boldsymbol{y} - \boldsymbol{y}'\mathbf{X}\,\boldsymbol{\beta} + (\mathbf{X}\,\boldsymbol{\beta})'\mathbf{X}\,\boldsymbol{\beta} \\ \text{[transpose rule} \rightarrow\text{]} & = & \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{\beta}'\mathbf{X}'\boldsymbol{y} - \boldsymbol{y}'\mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\,\boldsymbol{\beta} \\ \text{[transpose rule} \rightarrow\text{]} & = & \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'(\mathbf{X}'\boldsymbol{y}) + \boldsymbol{\beta}'\left(\mathbf{X}'\mathbf{X}\right)\boldsymbol{\beta} \end{array}$$

Taking the derivative $\frac{\partial}{\partial\boldsymbol{\beta}}$ respecting $\frac{\partial}{\partial\boldsymbol{\beta}}(\boldsymbol{\beta}'\boldsymbol{a}) = \boldsymbol{a}$ and $\frac{\partial}{\partial\boldsymbol{\beta}}(\boldsymbol{\beta}'\mathbf{A}\,\boldsymbol{\beta}) = (\mathbf{A} + \mathbf{A}')\boldsymbol{\beta}$ with $\mathbf{A} = \mathbf{X}'\mathbf{X}$ symmetric:

$$\frac{\partial S(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = 0 - 2\mathbf{X}'\boldsymbol{y} + 2\mathbf{X}'\mathbf{X}\,\boldsymbol{\beta} \stackrel{!}{=} 0$$

$$\mathbf{X}'\mathbf{X}\,\boldsymbol{\beta} \;=\; \mathbf{X}'\boldsymbol{y} \quad \mid \cdot\, (\mathbf{X}'\mathbf{X})^{-1}$$

Determining the OLS estimator

$$
\begin{aligned}
S &= (\boldsymbol{y} - \mathbf{X}\,\boldsymbol{\beta})'\,(\boldsymbol{y} - \mathbf{X}\,\boldsymbol{\beta}) \\
\text{[distributivity} \rightarrow\text{]} &= \boldsymbol{y}'\boldsymbol{y} - (\mathbf{X}\,\boldsymbol{\beta})'\boldsymbol{y} - \boldsymbol{y}'\mathbf{X}\,\boldsymbol{\beta} + (\mathbf{X}\,\boldsymbol{\beta})'\mathbf{X}\,\boldsymbol{\beta} \\
\text{[transpose rule} \rightarrow\text{]} &= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{\beta}'\mathbf{X}'\boldsymbol{y} - \boldsymbol{y}'\mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\,\boldsymbol{\beta} \\
\text{[transpose rule} \rightarrow\text{]} &= \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'(\mathbf{X}'\boldsymbol{y}) + \boldsymbol{\beta}'\left(\mathbf{X}'\mathbf{X}\right)\boldsymbol{\beta}
\end{aligned}
$$

Taking the derivative $\frac{\partial}{\partial \boldsymbol{\beta}}$ respecting $\frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}'\boldsymbol{a}) = \boldsymbol{a}$ and $\frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}'\mathbf{A}\,\boldsymbol{\beta}) = (\mathbf{A} + \mathbf{A}')\boldsymbol{\beta}$ with $\mathbf{A} = \mathbf{X}'\mathbf{X}$ symmetric:

$$
\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0 - 2\mathbf{X}'\boldsymbol{y} + 2\mathbf{X}'\mathbf{X}\,\boldsymbol{\beta} \stackrel{!}{=} 0
$$

$$
\mathbf{X}'\mathbf{X}\,\boldsymbol{\beta} = \mathbf{X}'\boldsymbol{y} \quad |\cdot\left(\mathbf{X}'\mathbf{X}\right)^{-1}
$$

$$
\boxed{\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{y}}
$$