Lecture 1: OLS – derivations and inference

Econometric Methods – Warsaw School of Economics

Andrzej Torój
Outline

1. Introduction
   - Course information
   - Econometrics: a reminder
   - Preliminary data exploration

2. OLS: theoretical reminder
   - Point estimation
   - Measuring precision
   - Model quality diagnostics under OLS

3. Multicollinearity
Outline

1. Introduction
2. OLS: theoretical reminder
3. Multicollinearity
Course information

- **lecturers:** Andrzej Torój & Marcin Owczarczuk
- **my website:** [http://web.sgh.waw.pl/~atoroj/](http://web.sgh.waw.pl/~atoroj/) (lecture slides, exercise files, literature, contact)
- **final grade:** homework (50%) + written exam (50%); details: website
Course information

- **Lecturers:** Andrzej Torój & Marcin Owczarczuk
- **My website:** [http://web.sgh.waw.pl/~atoroj/](http://web.sgh.waw.pl/~atoroj/) (lecture slides, exercise files, literature, contact)
- **Final grade:** homework (50%) + written exam (50%); details: website
Course information

- **Lecturers:** Andrzej Torój & Marcin Owczarczuk
- **My website:** [http://web.sgh.waw.pl/~atoroj/](http://web.sgh.waw.pl/~atoroj/) (lecture slides, exercise files, literature, contact)
- **Final grade:** homework (50%) + written exam (50%); details: website
Why econometrics?

- investigation of relationships
- finding parameter values in economic models (e.g. elasticities)
- confronting economic theories with data
- forecasting
- simulating policy scenarios
Why econometrics?

- investigation of relationships
- finding parameter values in economic models (e.g. elasticities)
- confronting economic theories with data
- forecasting
- simulating policy scenarios
Why econometrics?

- investigation of relationships
- finding parameter values in economic models (e.g. elasticities)
- confronting economic theories with data
- forecasting
- simulating policy scenarios
Why econometrics?

- investigation of relationships
- finding parameter values in economic models (e.g. elasticities)
- confronting economic theories with data
- forecasting
- simulating policy scenarios
Why econometrics?

- investigation of relationships
- finding parameter values in economic models (e.g. elasticities)
- confronting economic theories with data
- forecasting
- simulating policy scenarios
Data structure

1. **time series** (industrial production, 1995-2011, monthly index)
2. cross section (exit polls with 1000 respondents)
3. longitudinal data (quarterly GDP dynamics in EU states, 1995-2009)
Data structure

1. **time series** (industrial production, 1995-2011, monthly index)
2. **cross section** (exit polls with 1000 respondents)
3. **longitudinal data** (quarterly GDP dynamics in EU states, 1995-2009)
Data structure

1. **time series** (industrial production, 1995-2011, monthly index)
2. **cross section** (exit polls with 1000 respondents)
3. **longitudinal data** (quarterly GDP dynamics in EU states, 1995-2009)
Selection of explanatory variables

- theory, institutional setup, expert evaluation
- mechanical methods
  - data mining
  - correlation-matrix-based treatments
- *from general to specific* approach
- variable transformations
Selection of explanatory variables

- theory, institutional setup, expert evaluation
- mechanical methods
  - data mining
  - correlation-matrix-based treatments
- from general to specific approach
- variable transformations
Selection of explanatory variables

- theory, institutional setup, expert evaluation
- mechanical methods
  - data mining
  - correlation-matrix-based treatments
- *from general to specific* approach
- variable transformations
Selection of explanatory variables

- theory, institutional setup, expert evaluation
- mechanical methods
  - data mining
  - correlation-matrix-based treatments
- from general to specific approach
- variable transformations
Preliminary exploration (1)

- distribution – data errors, outliers?

![Histogram of residuals](chart.png)

- Series: Residuals
- Sample: 1997Q4 - 2009Q3
- Observations: 48
- Mean: -0.147728
- Median: -0.109705
- Maximum: 3.893341
- Minimum: -9.855776
- Std. Dev.: 2.006789
- Skewness: -2.165749
- Kurtosis: 12.70138
- Jarque-Bera: 225.7572
- Probability: 0.000000
Preliminary exploration (2)

- time series graph – trend? seasonality? volatility clustering?
Preliminary exploration (3)

- scatterplot – dependency, functional form, transformations?
Preliminary data exploration

Preliminary exploration (4)

- correlations?

<table>
<thead>
<tr>
<th></th>
<th>CAPITAL</th>
<th>LABOR</th>
<th>VALUEADD</th>
<th>TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAPITAL</td>
<td>1.000000</td>
<td>0.965568</td>
<td>0.976288</td>
<td>0.413928</td>
</tr>
<tr>
<td>LABOR</td>
<td>0.965568</td>
<td>1.000000</td>
<td>0.968783</td>
<td>0.355188</td>
</tr>
<tr>
<td>VALUEAD</td>
<td>0.976288</td>
<td>0.968783</td>
<td>1.000000</td>
<td>0.434433</td>
</tr>
<tr>
<td>TIME</td>
<td>0.413928</td>
<td>0.355188</td>
<td>0.434433</td>
<td>1.000000</td>
</tr>
</tbody>
</table>
Example (1/5)

**Student satisfaction survey**
Master students of Applied Econometrics at Warsaw School of Economics in Winter semester 2016/2017 were asked about their satisfaction from studying to be evaluated from 0 to 100. In addition, their average note from previous studies and their sex were registered.

1. What kind of data is this? Cross-section, time series, panel? Frequency? Micro- or macroeconomic?
2. How can we quickly visualise a hypothesised causality from average note to satisfaction from studying?
3. Does such a relationship seem to be there?
4. How can sex of the respondent potentially affect the satisfaction from studies or the relationship in question? How can we visualise this?
5. Bottom line, what is the right specification of the linear regression model?
Outline

1. Introduction

2. OLS: theoretical reminder

3. Multicollinearity
Introduction

OLS

Multicollinearity

Point estimation

Linear regression model

\[ y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \ldots + \beta_k x_{k,i} + \varepsilon_i = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \varepsilon_i = \mathbf{x}_i \beta + \varepsilon_i \]

Vector of parameters \( \begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & \ldots & \beta_k \end{bmatrix}^T \) is unknown. Minimize the dispersion of \( \varepsilon_i \) around zero, as measured e.g. by \( \sum_{t=1}^{n} \varepsilon_i^2 \).
Ordinary Least Squares (OLS)

\[ S = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1,i} - \beta_2 x_{2,i} - \ldots - \beta_k x_{k,i})^2 \rightarrow \min_{\beta_0, \beta_1, \ldots} \]

**FOC:** \( \frac{\partial S}{\partial \beta} = 0 \)

Denote:
\[
\begin{bmatrix}
y_1 \\
y_2 \\
\vdots \\
y_n
\end{bmatrix}, \quad
\begin{bmatrix}
1 & x_{1,1} & x_{2,1} & \ldots & x_{k,1} \\
1 & x_{1,2} & x_{2,2} & \ldots & x_{k,2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & x_{1,n} & x_{2,n} & \ldots & x_{k,n}
\end{bmatrix}, \quad
\begin{bmatrix}
\beta_0 \\
\beta_1 \\
\beta_2 \\
\vdots \\
\beta_k
\end{bmatrix}
\]

and obtain:
\[
\beta = (X^T X)^{-1} X^T y
\]
Proof

\[ S = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon^T \varepsilon = (y - X\beta)^T (y - X\beta) = \]
\[ = y^T y - \beta^T X^T y - y^T X\beta + \beta^T X^T X\beta = \]
\[ = y^T y - 2y^T X\beta + \beta^T X^T X\beta \]

(2. and 3. component were transposed scalars, so they were equal)

\[ \frac{\partial S}{\partial \beta} = 0 \iff \frac{\partial y^T y}{\partial \beta} - 2y^T X\beta + \beta^T X^T X\beta = 0 \]

According to the rules of matrix calculus:

\[ -2y^T X + \beta^T (2X^T X) = 0 \implies \beta^T (X^T X) = y^T X \implies \]
\[ (X^T X) \beta = X^T y \]
\[ \beta = (X^T X)^{-1} X^T y \]
Example (2/5)

Student satisfaction survey

1. Run the regression model with an automated command in R.
2. Write the equation and try to interpret the parameters. Be careful – it’s tricky! (Why?)
3. Manually replicate the parameter estimates.
Estimator as a random variable

- $\hat{\beta}$ is an estimator of the true parameter value $\beta$ (function of the random sample choice)
- samples, and hence the values of $\hat{\beta}$, can be different
- estimator as a (vector) random variable has its variance(-covariance matrix)

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}$$

$$\text{Var} (\hat{\beta}) = \begin{bmatrix} \text{var} (\hat{\beta}_0) & \text{cov} (\hat{\beta}_0, \hat{\beta}_1) & \text{cov} (\hat{\beta}_0, \hat{\beta}_2) & \cdots \\ \text{cov} (\hat{\beta}_0, \hat{\beta}_1) & \text{var} (\hat{\beta}_1) & \text{cov} (\hat{\beta}_1, \hat{\beta}_2) & \cdots \\ \text{cov} (\hat{\beta}_0, \hat{\beta}_2) & \text{cov} (\hat{\beta}_1, \hat{\beta}_2) & \text{var} (\hat{\beta}_2) & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ \text{cov} (\hat{\beta}_0, \hat{\beta}_k) & \text{cov} (\hat{\beta}_1, \hat{\beta}_k) & \text{cov} (\hat{\beta}_2, \hat{\beta}_k) & \cdots & \text{var} (\hat{\beta}_k) \end{bmatrix}$$
Estimator as a random variable

- $\hat{\beta}$ is an estimator of the true parameter value $\beta$ (function of the random sample choice)
- samples, and hence the values of $\hat{\beta}$, can be different
- estimator as a (vector) random variable has its variance(-covariance matrix)

\[
\hat{\beta} = \begin{bmatrix}
\hat{\beta}_0 \\
\hat{\beta}_1 \\
\hat{\beta}_2 \\
\vdots \\
\hat{\beta}_k
\end{bmatrix}
\]

\[
\text{Var}(\hat{\beta}) =
\begin{bmatrix}
\text{var}(\hat{\beta}_0) & \text{cov}(\hat{\beta}_0, \hat{\beta}_1) & \text{cov}(\hat{\beta}_0, \hat{\beta}_2) & \cdots \\
\text{cov}(\hat{\beta}_0, \hat{\beta}_1) & \text{var}(\hat{\beta}_1) & \text{cov}(\hat{\beta}_1, \hat{\beta}_2) & \cdots \\
\text{cov}(\hat{\beta}_0, \hat{\beta}_2) & \text{cov}(\hat{\beta}_1, \hat{\beta}_2) & \text{var}(\hat{\beta}_2) & \cdots \\
\vdots & \vdots & \vdots & \ddots \\
\end{bmatrix}
\]
Estimator as a random variable

- $\hat{\beta}$ is an estimator of the true parameter value $\beta$ (function of the random sample choice)
- samples, and hence the values of $\hat{\beta}$, can be different
- estimator as a (vector) random variable has its variance(-covariance matrix)

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}$$

$$\text{Var} (\hat{\beta}) = \begin{bmatrix} \text{var} (\hat{\beta}_0) & \text{cov} (\hat{\beta}_0, \hat{\beta}_1) & \text{cov} (\hat{\beta}_0, \hat{\beta}_2) & \cdots \\ \text{cov} (\hat{\beta}_0, \hat{\beta}_1) & \text{var} (\hat{\beta}_1) & \text{cov} (\hat{\beta}_1, \hat{\beta}_2) & \cdots \\ \text{cov} (\hat{\beta}_0, \hat{\beta}_2) & \text{cov} (\hat{\beta}_1, \hat{\beta}_2) & \text{var} (\hat{\beta}_2) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$
Measuring precision

Variance-covariance matrix of a random vector

Definition:
\[
\text{Var} (\beta) = E \left\{ [\beta - E(\beta)] [\beta - E(\beta)]^T \right\}
\]

For a centered variable, i.e. \( E(\varepsilon) = 0 \), this definition simplifies:
\[
\text{Var} (\varepsilon) = E (\varepsilon \varepsilon^T)
\]
OLS estimator: properties

\[ \hat{\beta} = (X^TX)^{-1}X^Ty \] is an estimator (function of the sample) of the “true”, unknown values \( \beta \) (population / data generating process). Under certain conditions (i.a. \( E(X^T \epsilon) = 0 \) \( E(\epsilon \epsilon^T) = \sigma^2 I \)), the OLS estimator is:

- **unbiased**: \( E(\hat{\beta}) = \beta \)
- **consistent**: \( \hat{\beta} \) converges to \( \beta \) with growing \( n \)
- **efficient**: least possible estimator variance (i.e. highest precision)
OLS estimator: properties

\[ \hat{\beta} = (X^T X)^{-1} X^T y \]

is an estimator (function of the sample) of the “true”, unknown values \( \beta \) (population / data generating process). Under certain conditions (i.a. \( E(X^T \varepsilon) = 0 \) \( E(\varepsilon \varepsilon^T) = \sigma^2 I \)), the OLS estimator is:

- **unbiased**: \( E(\hat{\beta}) = \beta \)
- **consistent**: \( \hat{\beta} \) converges to \( \beta \) with growing \( n \)
- **efficient**: least possible estimator variance (i.e. highest precision)
OLS estimator: properties

\[ \hat{\beta} = (X^T X)^{-1} X^T y \] is an estimator (function of the sample) of the “true”, unknown values \( \beta \) (population / data generating process). Under certain conditions (i.a. \( E(X^T \varepsilon) = 0 \) \( E(\varepsilon \varepsilon^T) = \sigma^2 I \)), the OLS estimator is:

- **unbiased**: \( E(\hat{\beta}) = \beta \)
- **consistent**: \( \hat{\beta} \) converges to \( \beta \) with growing \( n \)
- **efficient**: least possible estimator variance (i.e. highest precision)
Variance of the error term (1)

1. Variance of the error term (scalar): \( \hat{\sigma}^2 = \frac{1}{n-(k+1)} \sum_{i=1}^{n} \varepsilon_i^2 \)

Why such a formula if the general formula is

\[
Var(X) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2
\]

- First of all note that \( \bar{\varepsilon} = 0 \) (prove it on your own).
- Second, we need to know why 1 turned into \( (k + 1) \) in the denominator.
Variance of the error term (2)

By Your intuition, what is the standard deviation in the following dataset of 3 observation?

Without a correction in denominator:

$$\sqrt{\text{Var}} = \sqrt{\frac{1}{3} \left[ (3 - 2)^2 + (2 - 2)^2 + (1 - 2)^2 \right]} = \sqrt{\frac{2}{3}} \neq 1$$
Variance of the error term (2)

- By Your intuition, what is the standard deviation in the following dataset of 3 observation?

- Without a correction in denominator:
  \[
  \sqrt{\text{Var}} = \sqrt{\frac{1}{3} \left[ (3-2)^2 + (2-2)^2 + (1-2)^2 \right]} = \sqrt{\frac{2}{3}} \neq 1
  \]
Variance of the error term (3)

- The intuition behind the standard deviation of 1 is build upon an implicit, graphical calibration of mean based on the data sample.

- With an adequate correction for that in denominator:

$$\sqrt{\text{Var}} = \sqrt{\frac{1}{3-1} \left[ (3-2)^2 + (2-2)^2 + (1-2)^2 \right]} = \sqrt{\frac{2}{2}} = 1$$
Variance of the error term (3)

The intuition behind the standard deviation of 1 is build upon an implicit, graphical calibration of mean based on the data sample.

\[ \sqrt{\text{Var}} = \sqrt{\frac{1}{3-1} \left[ (3 - 2)^2 + (2 - 2)^2 + (1 - 2)^2 \right]} = \sqrt{\frac{2}{2}} = 1 \]
Variance of the error term (4)

- When $X$ is directly observed, the terms like $(x_i - \bar{x})$ consume one degree of freedom (there is one $\bar{x}$ estimated before).
- When $\varepsilon$ is not observed, the terms 
  \[ \varepsilon_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \ldots - \hat{\beta}_k x_{ki} \]
  consume $k + 1$ degrees of freedom (there are $k + 1$ elements in vector $\hat{\beta}$ estimated before).
Variance-covariance matrix of the estimator

\[ \text{Var}\left(\hat{\beta}\right) = E\left[\left(\hat{\beta} - \beta\right)\left(\hat{\beta} - \beta\right)^T\right] = \]
\[ = E\left\{\left[(X^TX)^{-1}X^T y - \beta\right] \cdot \left[(X^TX)^{-1}X^T y - \beta\right]^T\right\} = \]
\[ = E\left\{\left[(X^TX)^{-1}X^T (X\beta + \epsilon) - \beta\right] \cdot \left[(X^TX)^{-1}X^T (X\beta + \epsilon) - \beta\right]^T\right\} = \]
\[ = E\left\{(X^TX)^{-1}X^T \epsilon \cdot \epsilon^T X (X^TX)^{-1}\right\} = \]
\[ = (X^TX)^{-1}X^T E(\epsilon \epsilon^T) X (X^TX)^{-1} = \]
\[ = (X^TX)^{-1}X^T \sigma^2 I X (X^TX)^{-1} = \]
\[ = \sigma^2 (X^TX)^{-1}X^T X (X^TX)^{-1} = \]
\[ = \sigma^2 (X^TX)^{-1} \]

**Empirical counterpart:** \( \text{Var}\left(\hat{\beta}\right) = \hat{\sigma}^2 (X^TX)^{-1} \equiv [d_{i,j}]_{(k+1) \times (k+1)}\)
Standard errors of estimation

- **Standard errors of estimation** (vector – for each parameter):

  \[ s\left(\hat{\beta}_0\right) = \sqrt{d_{1,1}} \quad s\left(\hat{\beta}_1\right) = \sqrt{d_{2,2}} \quad s\left(\hat{\beta}_2\right) = \sqrt{d_{3,3}} \ldots \]

**Calculating S.E.**

1. estimate parameters, 2. compute the empirical error terms, 3. estimate their variance, 4. compute the variance-covariance matrix of the OLS estimator, 5. compute the SE as a square root of its diagonal elements.
Standard errors of estimation

- **Standard errors of estimation** (vector – for each parameter):
  \[
  s(\hat{\beta}_0) = \sqrt{d_{1,1}} \quad s(\hat{\beta}_1) = \sqrt{d_{2,2}} \quad s(\hat{\beta}_2) = \sqrt{d_{3,3}} \ldots
  \]

**Calculating S.E.**

1. estimate parameters, 2. compute the empirical error terms, 3. estimate their variance, 4. compute the variance-covariance matrix of the OLS estimator, 5. compute the SE as a square root of its diagonal elements.
**t-tests for variable significance**

**t-Student test**

- $H_0 : \beta_i = 0$, i.e. $i$-th explanatory variable does not significantly influence $y$
- $H_1 : \beta_i \neq 0$, i.e. $i$-th explanatory variable does not significantly influence $y$

Test statistic: $t = \frac{\hat{\beta}_i}{s(\hat{\beta}_1)}$ is distributed as $t(n - k - 1)$.

- $p$-value $< \alpha^*$ – reject $H_0$
- $p$-value $> \alpha^*$ – do not reject $H_0$
Example (3/5)

**Student satisfaction survey**

1. Compute the fitted values and the error terms.
2. Use this result to estimate the variance of the error term.
3. Estimate the variance-covariance matrix of the $\hat{\beta}$ estimates.
4. Derive the standard errors from this matrix.
5. Replicate and interpret the t statistics and the p-values.
R-squared (1)
R-squared (2)

\[ \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \]
R-squared (3)

\[ \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \]
R-squared (4)
R-squared (5)

\[ \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \]

\( y_i \)

\( \bar{y} \)

\( \hat{y}_i - \bar{y} \)

\( x_i \)
R-squared (6)
R-squared (7)

\[ R^2 \in [0; 1] \text{ is a share of } y_t \text{ volatility explained by the model in total } y_t \text{ volatility:} \]

\[
\sum_{t=1}^{T} (y_t - \bar{y})^2 = \sum_{t=1}^{T} (\hat{y}_t - \bar{y})^2 + \sum_{t=1}^{T} (y_t - \hat{y}_t)^2 \quad R^2 = \frac{\sum_{t=1}^{T} (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^{T} (y_t - \bar{y})^2}
\]

Standard goodness-of-fit measure in OLS regressions with a constant.
Wald test statistic

Wald test

\[ H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0, \text{ i.e. no explanatory variable influences } y \]
\[ H_1 : \exists i \quad \beta_i \neq 0, \text{ at least 1 explanatory variable influences } y \]

Test statistic: \[ F = \frac{R^2/k}{(1-R^2)/(T-k-1)} \] distributed as \( F(k, T-k-1) \).
Adjusted R-squared

\[ \bar{R}^2 = R^2_{\text{fit}} - \frac{k}{T - (k + 1)} (1 - R^2) \]

penalty for overparametrization
Example (4/5)

**Student satisfaction survey**

1. Interpret the F-test result.
2. Replicate the F statistic and its p-value manually.
3. Interpret the R-squared.
4. Replicate the R-squared and adjusted R-squared manually.
Outline

1. Introduction

2. OLS: theoretical reminder

3. Multicollinearity
Multicollinearity

Regressors are not independent:
- some are a linear combination of others (extreme case), then...
  - $X^T X$ is a singular matrix and we cannot compute
    \[ \beta = (X^T X)^{-1} X^T y \]
- some are highly correlated (usual case), then...
  - $X^T X$ may not be singular, but its diagonal elements still close
to zero...
  - ... then the diagonal elements of $(X^T X)^{-1}$ extremely high
    (and so the standard errors)
Multicollinearity – diagnostics

1. correlation matrix
   - only bilateral relationships
   - no cut-off value above which the problem can be considered serious

2. inflation variance factor (VIF) for regressor $j$
   - $VIF_j = \frac{1}{1-R_j^2}$
   - where $R_j^2$ is R-squared from the regression of variable $j$ on the rest of the explanatory variables
   - limit value: 10, above – multicollinearity

3. condition index
   - $\kappa = \sqrt{\frac{\lambda_{\text{max}}}{\lambda_{\text{min}}}}$
   - where $\lambda$ denotes eigenvalues of the matrix derived from $X^TX$ by division of every cell $(i,j)$ by the product of diagonal elements $(i,i)$ and $(j,j)$
   - limit value: 20, above – multicollinearity
Multicollinearity – solutions

- strengthen the precision of estimation by expanding sample size, removing a variable, imposing restrictions or calibrating the parameter
- "manually" increase the diagonal values in $X^TX$ (*ridge regression*)
- "squeeze" the common variance of the collinear variables into a lower number of new, independent ones (*principal components*)
Multicollinearity – solutions

- strengthen the precision of estimation by expanding sample size, removing a variable, imposing restrictions or calibrating the parameter
- ”manually” increase the diagonal values in $X^TX$ (ridge regression)
- ”squeeze” the common variance of the collinear variables into a lower number of new, independent ones (principal components)
Multicollinearity – solutions

- strengthen the precision of estimation by expanding sample size, removing a variable, imposing restrictions or calibrating the parameter
- "manually" increase the diagonal values in $X^TX$ \textit{(ridge regression)}
- "squeeze" the common variance of the collinear variables into a lower number of new, independent ones \textit{(principal components)}
Student satisfaction survey
Investigate the issue of multicollinearity in the proposed model.