Multiple Regression

Apply and extend the simple regression concepts to multiple regression. Evaluation of multivariate model.

Tomáš Oleš
Department of Economic Policy
Faculty of Economics and Finance

February 1, 2025

Agenda

- Understand the reasoning behind multiple regression.
- Apply and extend simple regression concepts to multiple regression.
- Learn to evaluate the quality of a multiple regression model.
- Understand and interpret multiple linear regression analysis.
- Develop a multiple regression model and estimate it using STATA.

Suppose you're making a dish and someone asks, "What makes it taste good?

Suppose you're making a dish and someone asks, "What makes it taste good?

You might say "Salt!"—but salt alone isn't enough. You need spices, oil, and other ingredients to get the full flavor.
Can adding too many ingredients (predictors) sometimes make the dish (model) worse?

From Simple to Multiple Regression

Extending the Regression Model

- In **simple regression**, we used one independent variable to predict Y:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$$

- In **multiple regression**, we expand this to include multiple predictors:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i$$

- This allows us to better model complex relationships by considering multiple explanatory variables at once.

Key Question: Why might a single variable be insufficient to explain Y?

Mathematical Formulation

Expected Value of Y

- In expectation, we assume the systematic part of the model explains Y:

$$E[Y_i] = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}$$

- The coefficients $\beta_1, \beta_2, \ldots, \beta_k$ represent the effect of each predictor on Y, holding other variables constant.

Special Case:

- If $X_1 = X_2 = \cdots = X_k = 0$, then $E[Y_i] = \beta_0$ (the intercept).
- This is the predicted value of Y when all independent variables are zero.

Role of the Error Term

Why Do We Need an Error Term?

- The model does not explain all variation in Y, so we introduce an error term ε_i:

$$Y_i = E[Y_i] + \varepsilon_i$$

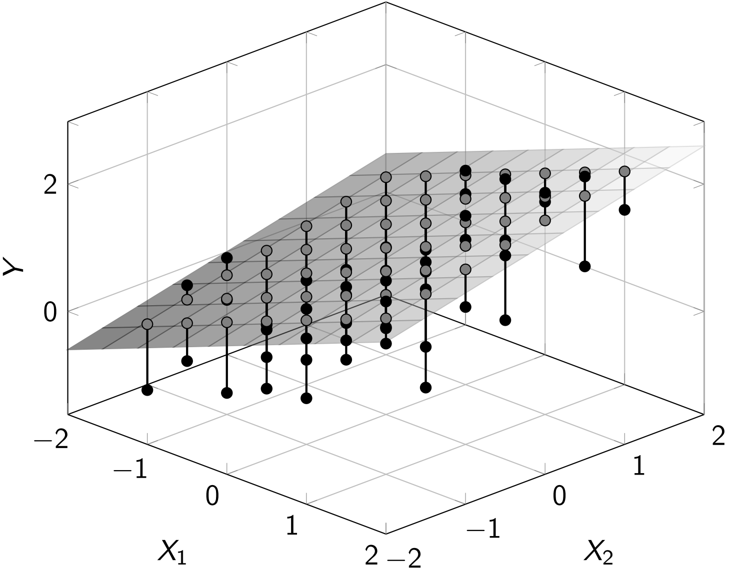
- Substituting for $E[Y_i]$:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i$$

- The error term accounts for:
  - Omitted variables
  - Measurement errors
  - Randomness in the data

The error term should be randomly distributed and uncorrelated with the explanatory variables for valid inference.

The Least-Squares Principle in Multiple Regression



$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}$$

$\hat{Y}_i$ (predicted)

$\hat{\epsilon}_i$ (residual)

Y_i (observed)

Estimation in Multiple Regression

The Least-Squares Principle

- The Least Squares Method minimizes the sum of squared residuals to find the best-fitting regression plane.
- In a model with two independent variables:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}$$

- The regression coefficients $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ are estimated by minimizing:

$$\min \sum \hat{\varepsilon}_i^2, \quad \text{where } \hat{\varepsilon}_i = Y_i - \hat{Y}_i$$

Key Insight:

- In simple regression, the least-squares method finds the best-fitting line.
- In multiple regression, it finds the best-fitting regression plane (for two predictors) or hyperplane (for more than two predictors).

Why does minimizing residuals lead to the best estimates of the coefficients?

Interpreting Regression Coefficients

How to Interpret β_k in Multiple Regression?

- Each coefficient β_k represents the effect of the corresponding variable X_k on Y, while keeping all other variables constant.
- Example Model:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$

- Interpretation:
  - β_1: Change in Y for a one-unit increase in X_1, holding X_2 and X_3 constant.
  - β_2: Change in Y for a one-unit increase in X_2, holding X_1 and X_3 constant.
  - β_3: Change in Y for a one-unit increase in X_3, holding X_1 and X_2 constant.

Why is it important to interpret coefficients in the context of holding other variables constant?

Goodness of Fit and the F-Test

Assessing Model Fit in Multiple Regression

- Residual Standard Deviation ($\hat{\sigma}_\varepsilon$)
  - Measures how much the observed Y_i values deviate from predicted $\hat{Y}_i$.
  - Lower values indicate a better model fit.
- Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{\sum \hat{\varepsilon}_i^2}{\sum (Y_i - \bar{Y})^2}$$

  - Measures the proportion of total variation in Y explained by the regression model.
  - Higher values indicate a better fit, but adding more variables can artificially inflate R^2.

The F-Test in Multiple Regression

Testing Overall Model Significance

- The F-test checks whether at least one independent variable significantly explains variation in Y.
- Null Hypothesis (H_0): $\beta_1 = \beta_2 = \cdots = \beta_k = 0$ (None of the independent variables explain Y)
- Alternative Hypothesis (H_1): At least one $\beta_k \neq 0$ (At least one predictor explains Y)

F-statistic formula:

$$F = \frac{R^2/(K-1)}{(1-R^2)/(n-K)}$$

where:

- $K - 1 =$ degrees of freedom for the model
- $n - K =$ degrees of freedom for the residuals

Decision Rule:

- If the p-value from the F-test is less than 0.05, reject H_0.

Adjusted R^2: A More Reliable Measure of Fit

Why Adjust R^2?

- Issue with R^2: Adding more independent variables **always** increases R^2, even if the new variables are irrelevant.
- This can create an illusion of a better model when the added variables do not truly improve explanatory power.
- **Solution:** Adjusted R^2 penalizes unnecessary variables.

Formula for Adjusted R^2:

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

where:

- K = number of predictors (independent variables)
- n = number of observations

Key Properties:

Partial Slope Coefficients in Multiple Regression

Understanding Statistical Control

- In experimental research, we ensure comparability by keeping conditions the same (e.g., random assignment).
- In non-experimental settings, multiple regression helps achieve **statistical control**.
- This means estimating the effect of an independent variable while holding other variables constant.

Mathematical Representation:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}$$

- $\hat{\beta}_1$ is the **partial slope coefficient** for X_1—it represents the change in $\hat{Y}$ for a one-unit increase in X_1, controlling for X_2.
- $\hat{\beta}_2$ is the partial effect of X_2 on Y, controlling for X_1.

Example: Calculating Predicted Values

$$\hat{Y}_i = -90.3 + 16.8X_1 + 2.23X_2$$

- If X_1 increases from 14 to 15, while holding X_2 constant, the predicted $\hat{Y}_i$ increases by 16.8 (which is $\hat{\beta}_1$).
- This confirms that the effect of X_1 remains the same at different values of X_2.

Key Insight:

- Multiple regression removes confounding effects, making estimates less biased than simple regression.
- Same significance testing procedures apply as in simple regression, but with adjusted degrees of freedom.

Why is it important to control for other variables when estimating the effect of X_1 on Y?

Prediction in Multiple Regression

Extending Prediction from Simple to Multiple Regression

- In simple regression, we predict mean-Y for a single independent variable X_1.
- In multiple regression, we predict mean-Y based on multiple independent variables $X_1, X_2, \ldots, X_k$.

Example: Predicting Mean-Y

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}$$

Given the estimated regression equation:

$$\hat{Y}_i = 2,674,274 + 151,855 X_1 + 100,000 X_2$$

If $X_1 = 3$ and $X_2 = 5$, we compute:

$$\hat{Y} = 2,674,274 + (151,855 \times 3) + (100,000 \times 5)$$

$$= 3,629,839$$

Key Takeaways:

- We can predict $\hat{Y}$ for any combination of $X_1, X_2, \ldots, X_k$ within the observed data range.
- The accuracy of predictions depends on how well the regression model fits the data.
- As in simple regression, we can construct confidence intervals for these predictions.

Why Standardize Regression Coefficients?

- In multiple regression, independent variables often have different measurement units.
- Directly comparing unstandardized coefficients can be misleading.
- Standardized coefficients (Beta Coefficients) allow direct comparisons by expressing effects in standard deviation units.

Standardization is useful when:

- Variables are measured on different scales (e.g., age in years vs. income in thousands).
- We want to compare the relative importance of predictors in explaining Y.

Why can't we directly compare raw regression coefficients when variables have different units?

Computing Standardized Coefficients

Standardizing Variables:

$$Z_{X_i} = \frac{X_i - \bar{X}}{\hat{\sigma}_X}, \quad Z_Y = \frac{Y - \bar{Y}}{\hat{\sigma}_Y}$$

- This converts raw values into z-scores (standard deviations from the mean).

Formula for Standardized Coefficients (Beta Coefficients):

$$\hat{b}_k = \hat{\beta}_k \times \left(\frac{\hat{\sigma}_{X_k}}{\hat{\sigma}_Y} \right)$$

where:

- $\hat{\beta}_k$ = unstandardized regression coefficient
- $\hat{\sigma}_{X_k}$, $\hat{\sigma}_Y$ = standard deviations of predictor and dependent variable

Key Interpretation: - $\hat{b}_k$ represents the number of standard deviations Y changes for a one-standard-deviation increase in X_k.

Interpreting Standardized Coefficients

How Do We Interpret $\hat{b}_k$?

- Standardized coefficients range between -1 and 1.
- Larger absolute values indicate stronger relationships.
- Interpretation: - $\hat{b}_k = 0.5 \rightarrow$ A one standard deviation increase in X_k leads to a 0.5 standard deviation increase in Y.

Effect Size Guidelines:

- $\hat{b}_k < 0.09 \rightarrow$ Small effect
- $0.1 \leq \hat{b}_k \leq 0.2 \rightarrow$ Moderate effect
- $\hat{b}_k \geq 0.2 \rightarrow$ Large effect

Key Question: How does standardization help us determine which predictor has the greatest impact on Y?

Excercise: Presents for a Partner

Making use of ' datasets/present.dta' data estimate the following model:

$$E[Present\_Value_i] = \beta_0 + \beta_1 Attractiveness_i + \beta_2 Kindness_i + \beta_3 Age_i$$

How would you interpret the β_0 coefficient? What about the significance? Is this making any sense?
And what about β_1, β_2, and β_3? Are they statistically significant?
What is the explained variation in present value give the attractiveness and kindness of the partner?

```
. regress Present_Value Attractiveness Kindness Age

      Source |       SS           df       MS      Number of obs   =        20
-------------+----------------------------------   F(3, 16)        =     14.39
       Model |  698034.608         3  232678.203   Prob > F        =    0.0001
    Residual |  258623.592        16  16163.9745   R-squared       =    0.7297
-------------+----------------------------------   Adj R-squared   =    0.6790
       Total |    956658.2        19  50350.4316   Root MSE        =    127.14

--------------------------------------------------------------------------------
 Present_Value |      Coef.   Std. Err.      t    P>|t|     [95\% Conf. Interval]
---------------+----------------------------------------------------------------
Attractiveness |   49.48174   19.41294     2.55   0.021     8.328152    90.63532
      Kindness |   33.93441   14.94049     2.27   0.037      2.26198    65.60683
           Age |   5.595464    2.58597     2.16   0.046     .1134516    11.07748
         _cons |  -66.95579   84.62477    -0.79   0.440    -246.3523    112.4407
--------------------------------------------------------------------------------
```

```
. regress Present_Value Attractiveness Kindness Age, beta *to estimate standardized
   coefficient add option ‘beta‘

      Source |       SS           df       MS      Number of obs   =        20
-------------+----------------------------------   F(3, 16)        =     14.39
       Model |  698034.608         3  232678.203   Prob > F        =    0.0001
    Residual |  258623.592        16  16163.9745   R-squared       =    0.7297
-------------+----------------------------------   Adj R-squared   =    0.6790
       Total |    956658.2        19  50350.4316   Root MSE        =    127.14

--------------------------------------------------------------------------------
 Present_Value |      Coef.   Std. Err.      t    P>|t|                     Beta
---------------+----------------------------------------------------------------
Attractiveness |   49.48174   19.41294     2.55   0.021                  .3949602
      Kindness |   33.93441   14.94049     2.27   0.037                  .3534765
           Age |   5.595464    2.58597     2.16   0.046                  .3236054
         _cons |  -66.95579   84.62477    -0.79   0.440                         .
--------------------------------------------------------------------------------
```

How much variation is explained by our independent variables?

```
. pcorr Present_Value Attractiveness Kindness Age
(obs=20)

Partial and semipartial correlations of Present_Value with

               Partial   Semipartial      Partial   Semipartial   Significance
   Variable |    Corr.         Corr.      Corr.^2       Corr.^2          Value
------------+-----------------------------------------------------------------
Attractiv~s |   0.5374        0.3313       0.2888        0.1098         0.0215
   Kindness |   0.4938        0.2952       0.2438        0.0872         0.0373
        Age |   0.4758        0.2813       0.2264        0.0791         0.0460
```

Why these squared semi-partial correlations do not sum up to the R^2 value?

Prediction of the mean-Y values as specific X-values

```
. margins, at(Attractiveness = 7 Kindness = 7 Age = 30)

Adjusted predictions                           Number of obs     =         20
Model VCE    : OLS

Expression   : Linear prediction, predict()
at           : Attractive~s    =           7
               Kindness        =           7
               Age             =          30

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   684.8211   74.73415     9.16   0.000     526.3918    843.2504
------------------------------------------------------------------------------
```

Prediction of the mean-Y values as specific X_i-values

```
. margins, at(Attractiveness=(1(1)7))

Predictive margins                              Number of obs     =         20
Model VCE    : OLS

Expression   : Linear prediction, predict()

1._at        : Attractive~s    =           1

2._at        : Attractive~s    =           2

3._at        : Attractive~s    =           3

4._at        : Attractive~s    =           4

5._at        : Attractive~s    =           5

6._at        : Attractive~s    =           6

7._at        : Attractive~s    =           7

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   298.1216   57.08541     5.22   0.000     177.1059    419.1372
          2  |   347.6033   41.39577     8.40   0.000     259.8482    435.3584
          3  |    397.085   30.36774    13.08   0.000     332.7083    461.4618
          4  |   446.5668   29.74077    15.02   0.000     383.5192    509.6144
          5  |   496.0485   40.00689    12.40   0.000     411.2377    580.8593
          6  |   545.5303   55.41041     9.85   0.000     428.0654    662.9951
          7  |    595.012    72.7585     8.18   0.000     440.7709    749.2531
------------------------------------------------------------------------------
```

esttab to Export Your Results

```
reg Present_Value Attractiveness Kindness Age
estimates store my_regression
estadd beta
esttab my_regression, title (Regression Model) nonumber ///
    mlabel(Results) ///
    cells(b(star fmt(2)) ci(par) beta(par)) ///
    stats(N p r2 r2_a rmse, ///
    labels( Number of observations ///
      Model significance R -square ///
      Adjusted R-square Residual standard deviation )) ///
    varwidth(30) legend
```

Table: Regression Model

	Results
	$\beta / ci95 / \beta^{standardized}$
Attractiveness	49.48*
	[8.33,90.64]
	(0.39)
Kindness	33.93*
	[2.26,65.61]
	(0.35)
Age	5.60*
	[0.11,11.08]
	(0.32)
Constant	-66.96
	[-246.35,112.44]
Number of observations	20.00
Model significance	0.00
R-square	0.73
Adjusted R-square	0.68
Residual standard deviation	127.14

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

References I

Laffers, L. (2021). *Draft poznámok k predmetu Moderná Aplikovaná regresia 1*. UMB Banská Bystrica.

Mehmetoglu, M. and Jakobsen, T. G. (2022). *Applied Statistics using Stata: a Guide for the Social Sciences*. Sage.