Toulouse School of Economics

M1 INTERMEDIATE ECONOMETRICS

Linear regression: Mechanics

Koen Jochmans François Poinas

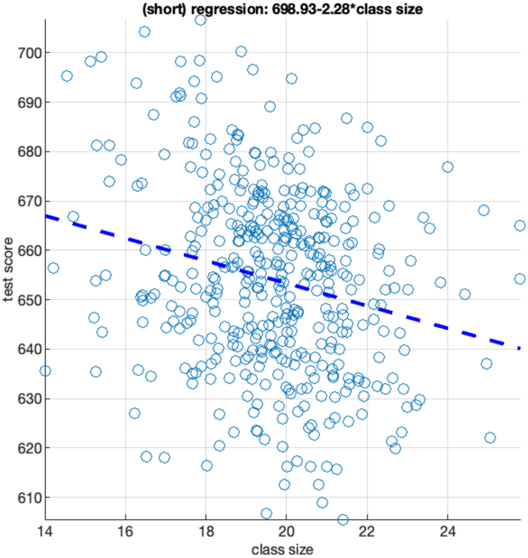
2025 — 2026

Californian data on 420 school districts.

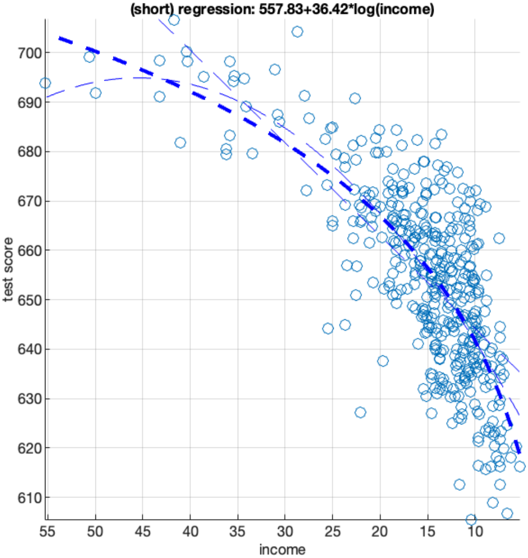
We observe, for each district, the average of

- test score,
- class size,
- income (in 1000 USD).

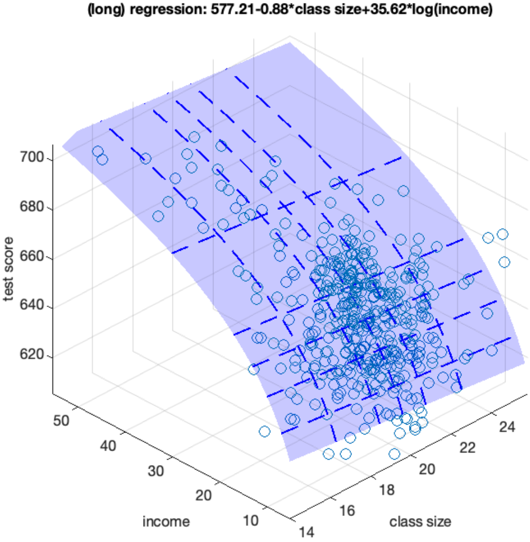
We consider regressions of test scores on class size, income, and both.



We predict a decrease in test score of 2.28 points for an increase of class size by one student.



We predict an increase of 0.36 points in test scores for a 1% increase in income.



We predict a decrease in test score of 0.88 points for an increase of class size by one student, holding fixed income.

The ordinary least squares estimator

We look to predict an outcome based on a vector of k regressors.

The data consist of n observations on the dependent variable, $Y_1, Y_2, \ldots, Y_n$ and on a vector of independent variables, $X_1, X_2, \ldots, X_n$.

A linear predictor at x is $x'\hat{\beta}$, i.e., a linear combination of the regressors.

Ordinary least-squares uses

$$\hat{\beta} = \arg\min_b \sum_{i=1}^{n} (Y_i - X_i' b)^2.$$

It minimizes the sum of squared errors. Many other possibilities; one such alternative would be

$$\arg\min_b \sum_{i=1}^{n} |Y_i - X_i' b|,$$

which minimizes least absolute deviations.

The algebra of ordinary least squares

Quadratic problem is easy to optimize.

The first-order condition for a minimum is

$$\sum_{i=1}^{n} X_i(Y_i - X_i'b) = 0.$$

Provided that the matrix $\sum_{i=1}^{n} X_i X_i'$ is invertible the unique solution is

$$\hat{\beta} = \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1} \left(\sum_{i=1}^{n} X_i Y_i\right).$$

If not there are many linear combinations that yield the same predictor.

Under such colinearity, can drop regressors until the rank condition is satisfied.

Again, now with matrix notation

Collects the n observations in

$\boldsymbol{Y} = (Y_1 \ , \ldots, Y_n \ )'$, an n-by-1 vector, and

$\boldsymbol{X} = (X_1', \ldots, X_n')'$, an n-by-k matrix .

Then

$$\hat{\beta} = \arg\min_b (\boldsymbol{Y} - \boldsymbol{X}b)'(\boldsymbol{Y} - \boldsymbol{X}b) = \arg\min_b \|\boldsymbol{Y} - \boldsymbol{X}b\|^2,$$

for $\|\cdot\|$ the Euclidean norm.

The normal equations for a minimum become

$$\boldsymbol{X}'\boldsymbol{Y} = \boldsymbol{X}'\boldsymbol{X}\, b,$$

with unique solution

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{Y})$$

when the columns of $\boldsymbol{X}$ are linearly independent.

Fitted values and residuals

Can define the residual vector $\hat{\boldsymbol{e}} = \boldsymbol{Y} - \boldsymbol{X}\hat{\beta}$. Then

$$\hat{\boldsymbol{e}}'\hat{\boldsymbol{e}} = \|\hat{\boldsymbol{e}}\|^2 = \min_b \|\boldsymbol{Y} - \boldsymbol{X}b\|^2.$$

By construction,

$$\boldsymbol{X}'\hat{\boldsymbol{e}} = \boldsymbol{X}'(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}) = (\boldsymbol{X}'\boldsymbol{Y}) - (\boldsymbol{X}'\boldsymbol{X})\,(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{Y}) = 0.$$

The residual vector $\hat{\boldsymbol{e}}$ is orthogonal to the regressor matrix $\boldsymbol{X}$.

The fitted values are $\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\beta}$. These are all points on the fitted hyperplane.

Because

$$\hat{\boldsymbol{Y}}'\hat{\boldsymbol{e}} = \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\hat{\boldsymbol{e}} = 0$$

the fitted values and residuals from an ordinary least-squares regression are orthogonal to each other.

Projection matrices

Define the n-by-n matrices

$$\boldsymbol{P_X} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}', \qquad \boldsymbol{M_X} = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'.$$

These are projection matrices; they are

symmetric ($\boldsymbol{P}'_{\boldsymbol{X}} = \boldsymbol{P_X}$), and

idempotent ($\boldsymbol{P}^2_{\boldsymbol{X}} = \boldsymbol{P_X}\boldsymbol{P_X} = \boldsymbol{P_X}$).

We have

$$\hat{\boldsymbol{Y}} = \boldsymbol{P_X}\boldsymbol{Y}, \qquad \hat{\boldsymbol{e}} = \boldsymbol{M_X}\boldsymbol{Y},$$

and $\boldsymbol{P_X} + \boldsymbol{M_X} = \boldsymbol{I}_n$ and $\boldsymbol{P_X}\boldsymbol{M_X} = 0$.

So

$$\boldsymbol{Y} = \boldsymbol{P_X}\boldsymbol{Y} + \boldsymbol{M_X}\boldsymbol{Y}$$

with both parts orthogonal to one another.

Short and long regressions

Partition $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$.

The long regression is the regression of $\boldsymbol{Y}$ on both $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, that is

$$\boldsymbol{Y} = \boldsymbol{X}_1\hat{\beta}_1 + \boldsymbol{X}_2\hat{\beta}_2 + \hat{\boldsymbol{e}},$$

with

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_1'\boldsymbol{X}_1 & \boldsymbol{X}_1'\boldsymbol{X}_2 \\ \boldsymbol{X}_2'\boldsymbol{X}_1 & \boldsymbol{X}_2'\boldsymbol{X}_2 \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{X}_1'\boldsymbol{Y} \\ \boldsymbol{X}_2'\boldsymbol{Y} \end{pmatrix}.$$

The short regression is the regression of $\boldsymbol{Y}$ on $\boldsymbol{X}_1$ alone. Its slope equals

$$(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}(\boldsymbol{X}_1'\boldsymbol{Y}) = \hat{\beta}_1 + (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}(\boldsymbol{X}_1'\boldsymbol{X}_2)\hat{\beta}_2.$$

This is not equal to $\hat{\beta}_1$ unless the two sets of regressors are orthogonal, $\boldsymbol{X}_1'\boldsymbol{X}_2 = 0$, in general.

Partitioned regression

The long regression was

$$\boldsymbol{Y} = \boldsymbol{X}_1\hat{\beta}_1 + \boldsymbol{X}_2\hat{\beta}_2 + \hat{\boldsymbol{e}}.$$

We can get an expression for $\hat{\beta}_1$ by working out the inverse on the previous slide.

Alternatively: Premultiply with $\boldsymbol{M}_{\boldsymbol{X}_2}$ to partial-out $\boldsymbol{X}_2$ to get

$$\boldsymbol{M}_{\boldsymbol{X}_2}\boldsymbol{Y} = \boldsymbol{M}_{\boldsymbol{X}_2}\boldsymbol{X}_1\hat{\beta}_1 + \boldsymbol{M}_{\boldsymbol{X}_2}\boldsymbol{X}_2\hat{\beta}_2 + \boldsymbol{M}_{\boldsymbol{X}_2}\hat{\boldsymbol{e}} = \boldsymbol{M}_{\boldsymbol{X}_2}\boldsymbol{X}_1\hat{\beta}_1 + \hat{\boldsymbol{e}},$$

using that $\boldsymbol{M}_{\boldsymbol{X}_2}\boldsymbol{X}_2 = 0$ and $\boldsymbol{M}_{\boldsymbol{X}_2}\hat{\boldsymbol{e}} = \hat{\boldsymbol{e}} - \boldsymbol{P}_{\boldsymbol{X}_2}\hat{\boldsymbol{e}} = \hat{\boldsymbol{e}}$.

Then premultiplying with $\boldsymbol{X}_1'$ and re-arranging we obtain

$$\hat{\beta}_1 = (\boldsymbol{X}_1'\boldsymbol{M}_{\boldsymbol{X}_2}\boldsymbol{X}_1)^{-1}(\boldsymbol{X}_1'\boldsymbol{M}_{\boldsymbol{X}_2}\boldsymbol{Y}),$$

because $\boldsymbol{X}_1'\hat{\boldsymbol{e}} = 0$.

Note that

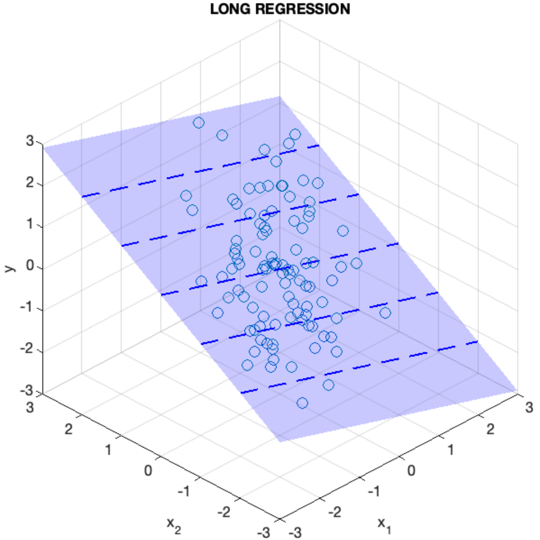
$$\boldsymbol{M}_{\boldsymbol{X}_2}\boldsymbol{Y} = \boldsymbol{M}_{\boldsymbol{X}_2}\boldsymbol{X}_1\hat{\beta}_1 + \hat{\boldsymbol{e}}$$

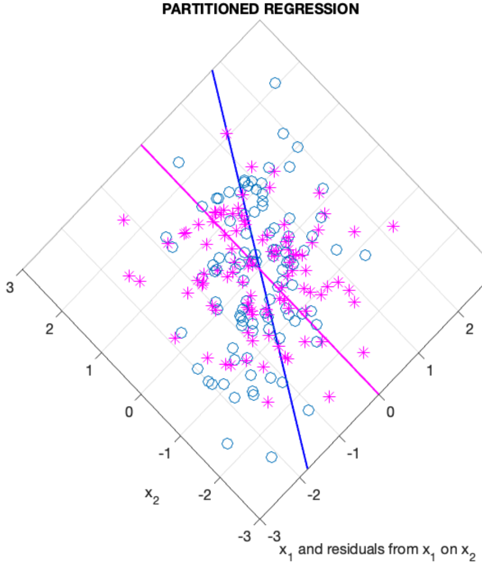
is a short regression, in that it does not regress on $\boldsymbol{X}_2$.

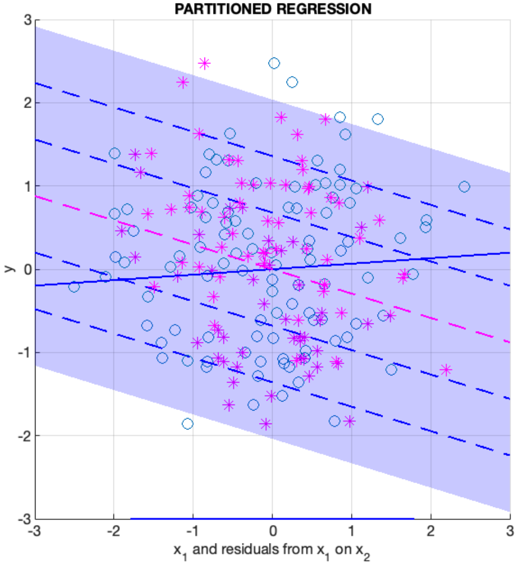
Including $\boldsymbol{X}_2$ is unnecessary here. By construction, it does not correlate with either the dependent or the independent variable in this regression.

The linear dependence of both $\boldsymbol{Y}$ and $\boldsymbol{X}_1$ on $\boldsymbol{X}_2$ has been removed (or 'partialled-out') by working with the residuals $\boldsymbol{M}_{\boldsymbol{X}_2}\boldsymbol{Y}$ and $\boldsymbol{M}_{\boldsymbol{X}_2}\boldsymbol{X}_1$ instead.

From the expression for $\hat{\beta}_1$ (because $\boldsymbol{M}_{\boldsymbol{X}_2}$ is idempotent) it suffices to partial-out $\boldsymbol{X}_2$ only from $\boldsymbol{X}_1$. So we can directly regress $\boldsymbol{Y}$ on $\boldsymbol{M}_{\boldsymbol{X}_2}\boldsymbol{X}_1$.


LONG REGRESSION
y
3
2
1
0
-1
-2
-3
3
2
1
0
-1
-2
-3
x2
-3
-2
-1
0
1
2
3
x1


PARTITIONED REGRESSION
3
2
1
0
-1
-2
-3
x2
-3
-2
-1
0
1
2
3
x1 and residuals from x1 on x2


PARTITIONED REGRESSION
3
2
1
0
-1
-2
-3
y
-3
-2
-1
0
1
2
3
x1 and residuals from x1 on x2