Econometric Methods for Political Analysis

Volume II: Causal Estimation for Observational Studies

Kevin Lingfeng Li

Table of contents

Part I

Further Regression Models

Chapter 1

Transformations and Prediction

1.1 Polynomial Transformations

Sometimes, a linear (straight-line) best-fit line is a poor description of a relationship. We can model more flexible relationships that are not straight lines, by including a transformation of the variable X that we are interested in.

Quadratic Transformations

> 💡 Key Definition: Quadratic Transformation
>
> Quadratic transformations of X_j take the following form:
>
> $$\mathbb{E}[Y_i|\overline{X}] = \alpha + \beta_1 X_{ji} + \beta_2 X_{ji}^2 + \text{other explanatory variables}$$

If you recall from high-school algebra, an equation that takes the form of $y = ax^2 + bx + c$ creates a *parabola*. A true parabola has a domain of $(-\infty, \infty)$. However, our model often does not need to do this. The best-fit parabola is only used for the range of plausible X values, given the nature of our explanatory variable. For example, if X was age, a negative number would make no sense. Because the parabola's domain often exceeds our plausible range of X values, the vertex of the parabola (where it changes directions) may not be in our data.

We always include lower degree terms in our model. For example, in this quadratic (power 2) model, we also include the X term without the square. To fit a model like this, we simply do the same process of minimising the sum of squared errors. How do we interpret the coefficients β_1 and β_2?

> ℹ️ Interpretation of Quadratic Transformations
>
> β_1's value is no longer directly interpretable. This is because we cannot "hold all other coefficients constant", since β_2 also contains the same X variable.
>
> β_2's value also cannot be directly interpreted. If the coefficient of β_2 is statistically significant, we can conclude that there is a non-linear relationship between X and Y. If β_2 is negative, the best-fit parabola will open downwards, and if β_2 is positive, the best-fit parabola will open upwards.

If we want to interpret the magnitude of the model, we are best off using predicted values of Y (obtained using the model equation above).

There is one more thing we can interpret with the quadratic transformation: the **vertex** of the best-fit parabola. The vertex, if we remember our algebra, is either the maximum or minimum point of a parabola. Thus, if we remember from calculus and optimisation, we can find the maximum and minimums through setting the derivative equal to 0. For the quadratic model, this is as follows - we first find the derivative, then set the derivative equal to 0:

$$\hat{Y} = \hat{\alpha} + \hat{\beta}_1 X + \hat{\beta}_2 X^2$$
$$\frac{d\hat{Y}}{dX} = 0 + \hat{\beta}_1 + 2\hat{\beta}_2 X$$
$$0 = \hat{\beta}_1 + 2\hat{\beta}_2 X$$
$$-\hat{\beta}_1 = 2\hat{\beta}_2 X$$
$$X = \frac{-\hat{\beta}_1}{2\hat{\beta}_2}$$

This point is useful, as it is either the maximum or minimum of our best-fit parabola. This means that at the X value we calculate from this equation, we will either see the highest or lowest expected Y value.

General Polynomial Models

While quadratic models are the most common polynomial transformation, we do not have to stop there. We can continue to add further polynomials (although anything beyond cubic is exceedingly rare):

- Cubic: $E[Y] = \alpha + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$

- Quartic: $E[Y] = \alpha + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \beta_4 x^4$

Each higher order coefficient, if statistically significant, indicates that the relationship between X and Y, is not of the previous highest power. For example, if the cubic term β_3 is statistically significant, we can reject a quadratic relationship between X and Y

Remember to always include the lower power monomials within our polynomial model. For example, if you have a quartic transformation, you must also have the linear, quadratic, and cubic terms.

1.2   Logarithmic Transformations

Logarithmic transformations are another form of non-linear transformations. These are commonly used for heavily skewed variables, such as when the explanatory variable is income, wealth, and so on.

In situations with heavily skewed variables, we often replace X in our models with $\log(X)$. Note that in statistics, when we refer to logarithms, we are referring to natural logarithms, such that $\log(X) = \ln(X)$.

> 💡 Key Definition: Logarithmic Transformation
>
> The logarithmic transformation of explanatory variable X_j takes the following form:
>
> $$\mathbb{E}[Y_i | \overrightarrow{X}] = \alpha + \beta_j \log(X_j) + \text{other explanatory variables}$$

Interpretation of the β coefficient can be a little bit trickier for logarithmic transformations. We could interpret it in the same way we interpret linear regressions: given a one unit increase in the log of X, there is an expected β change in Y. However, this issue is that this does not really say much - I mean, who knows what a *one unit increase in the log of X* even means?

With some properties of logarithms, we can actually create a more useful interpretation. Based on logarithm rules, we know the following to be true:

$$\log(X) + A = \log(X) + \log(e^A)$$
$$= \log(e^A \times X)$$

Now, let us plug this into our original regression model:

$$E[Y|X] = \alpha + \beta \log(X)$$
$$E[Y|e^A \times X] = \alpha + \beta \log(e^A \times X)$$
$$= \alpha + \beta[\log(X) + A]$$
$$= \alpha + \beta A + \beta \log(X)$$

Now find the difference between $E[Y|e^A \times X]$ and $E[Y|X]$:

$$E[Y|e^A \times X] - E[Y|X] = [\alpha + \beta A + \beta \log(X)] - [\alpha + \beta \log(X)]$$
$$E[Y|e^A \times X] - E[Y|X] = \beta A$$

Thus, we can seen that when we multiply X by e^A, we get an expected βA change in Y. We can make this interpretation more useful by purposely choosing some value A that makes e^A make more sense. For example, if $A = 0.095$, then $e^A = 1.1$, and multiplying by 1.1 is a 10% increase.

> **i** Interpreation: Logarithmic Transformation
>
> When X_j increases by 10%, there is a expected $0.095\beta_j$ unit change in Y

Chapter 2

Binomial Logistic Regression

Chapter 3

Maximum Likelihood Estimation

Chapter 4

Multinomial and Ordinal Logistic Regression

Chapter 5

Regression for Counts

Part II

Multivariate Measurement

Chapter 6

Principle Components Analysis

Chapter 7

Cluster Analysis

Part III

Latent Variable Models

Chapter 8

Exploratory Factor Analysis

Chapter 9

Confirmatory Factor Analysis

Chapter 10

Latent Trait Models

Chapter 11

Latent Class Models

Chapter 12

Structural Equation Models