Math for the Social Sciences Module - Young Researchers Fellowship

Lecture 5 - Modelling in the Social Sciences

Daniel Sánchez Pazmiño

Laboratorio de Investigación para el Desarrollo del Ecuador

2024

What is a quantitative model?

- A model represents a simplified version of reality. - It is a tool to understand and predict the behavior of a system. - It is a way to test hypotheses and theories.
- We focus on simplified realities because sometimes we don't care about all the details. - Details are irrelevant to the question we are asking. - It is easier to first understand the simple situation, then extend.
- A quantitative model is a model that is a model that uses quantitative information (numbers, equations, etc) to represent the real-world system. - We can later use this model to make predictions and test hypotheses. - Data can be used to validate or invalidate the model. - Public policy decisions can be made based on the model.

Social science models

- Social scientists use models to understand how humans interact with their reality. - Demand and supply - Rational voting - Item response theory - Logistic growth - Epidemiological transition
- In this lecture we focus on the common language of social science models - "Parameters", "Variables", "Constants" - "Exogenous"/"Endogenous" - "Equilibrium", "Stability" - "Independent"/"Dependent" - "Explanatory"/"Explained" - "Functions", "Functional form", "Equations", "Assumptions" - Reduced form models, Structural models

The parts of a model: variables

- **Variables**: Represent the quantities that **change** in the model.
- Sometimes whether something in the model changes might not be obvious - Depends in the context of the model. - Depends on the question we are asking i.e. the focus of the model. - Depends on the **assumptions** we make.
- In a statistical validation of a model, we collect data on variables to see if the model is correct.
- Variables can be **exogenous** or **endogenous**.
  - **Exogenous**: Determined outside the model.
  - **Endogenous**: Determined inside the model.

Example: Supply and demand

- The supply and demand model has two endogenous variables: price and quantity.
  - Determined by the model: the intersection of the curves and the ultimate focus of the model.
- Exogenous variables include demand and supply shifters: income, preferences, technology, etc.

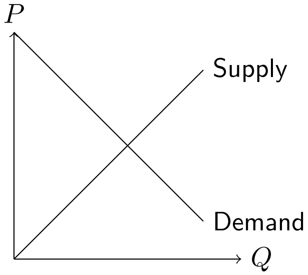


Figure 1: Supply and demand model

The parts of a model: assumptions

- **Assumptions**: Represent the **fixed** conditions in the model. - They are the conditions that must be met for the model to be valid. - They are the conditions that allow us to make predictions. - They are the conditions that allow us to test hypotheses.
- We often say that models "assume away" certain things. This is because we simplify the world so that our simpler logic applies.
- Assumptions create the model structure, and have a direct impact on the model's predictions.
    - Bad models often mean bad assumptions.
    - A lot of research work, even empirical, is about validating assumptions from theory.

Example: Supply and demand

- The very basic supply and demand model has several assumptions, among them that of perfect competition.
  - This assumption allows us to derive the supply and demand curves.
  - This assumption allows us to derive the equilibrium price and quantity.
- If we relax the assumption of perfect competition typically changes the conclusions of the model.
  - E.g. Monopoly: Higher prices, lower quantities, lower consumer surplus.

The parts of a model: parameters

- **Parameters**: Represent the **fixed** values in the model.
  - They are the values that determine the behavior of the model.
  - They are the values that we can change to see how the model behaves.
  - They are the values that we can estimate from data.
- Parameters are used to represent values of the model which do not vary.
  - We typically do not know the values of the parameters, but may have an idea of their range (i.e. positive or negative.)
  - Sometimes, we make simplifying assumptions about parameters (i.e. assume unitary elasticity).

Example: Demand curve estimation

- Assuming linear supply and demand curves, we can estimate the parameters of the model. For example, the demand curve can be written as:

$$Q = a + bP$$

- The parameters a and b are fixed values which we can't observe directly (so are not variables).
  - But we know that b is probably negative (law of demand).
  - We can estimate a and b from data, and use them to make predictions.

Application: The concept of elasticity

- **Elasticity**: A measure of how much one variable changes in response to a change in another variable.

$$\text{Elasticity of Y with respect to X} = \frac{\%\Delta Y}{\%\Delta X}$$

- Elasticity can be derived theoretically from models and estimated empirically from data.
  - It is a parameter of the model.
  - It is a measure of how sensitive the model is to changes in the exogenous variables.
- The price elasticity of demand is simply the b parameter times the price divided by the quantity.

Application: The concept of elasticity

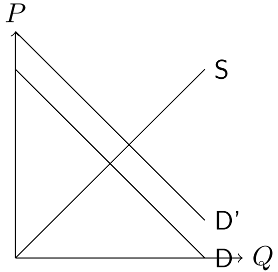
- Other important elasticities include income elasticity, wage income to education, education to grades, wage elasticity of labour supply, elasticity of intergenerational mobility, tax elasticity, fertility elasticity, housing elasticity, elasticity of crime with respect to police, etc.
- Public policy decisions can be made based on elasticities because of their predictive power.
  - For example, if the price elasticity of demand for cigarettes is high, then a tax increase will reduce consumption.
  - If the elasticity of intergenerational mobility is low, then policies to increase mobility are needed.
- Having estimated a value of elasticity (i.e. 20%), we multiply it times the change we believe something exert on the X variable to get the change in the Y variable we'd expect.
  - Not always the elasticity is linear, but it is an approximation.

Equilibrium

- **Equilibrium**: The point where the model is in balance.
  - The point where the model is stable.
  - The point where the model is not changing.
  - The point where the model is at rest.
- Equilibrium is a key concept in social science models.
  - Changes in the model are often about moving towards equilibrium.
  - It is hard to predict what will happen if the model is not in equilibrium.
  - Equilibrium is often the focus of the model.

Example: Shifts and movements along the supply and demand curves

- The equilibrium price and quantity are the points where the supply and demand curves intersect.
- An exogenous change (a change in consumer preferences, for example) will shift the demand curve. This moves the equilibrium point.



Independent and dependent variables

- In an estimation context, it is often heard that we categorize variables as independent and dependent.
- **Independent variables**: Variables that we have full control over, and are generally not affected by other variables in the model.
- **Dependent variables**: Variables that are affected by other variables in the model, generally the *independent* variables.
- In the social sciences, where we rely on observational rather than experimental studies, I argue that these are not the best terms to use.
  - Independent variables, while are often exogenous, are not always under our control or might be indirectly affected by other variables.
  - Observational studies need to apply sophisticated techniques to control for confounding variables.

Explanatory and explained variables

- Rather, one should use the term *explanatory* and *explained* variables.
- The explanatory variables are the ones that we use to explain the behavior of the explained variables.
- The explained variables are the ones that we are trying to understand (dependent variables).
- This is a more accurate description of the relationship between variables in a model.
- Other terms that are used are *predictor* and *response* variables.
  - Regressor and regressand.
  - Covariate and outcome.

Functions

- **Functions**: A relationship between variables.
- In an estimation context, we hear function a lot. For example, the demand function, the supply function, the production function, the utility function, etc.
  - These are relationships between variables that we can estimate from data.
  - These are relationships that we can use to make predictions.
- Within fields, a lot of functions are well-known and have been estimated many times, so it is important to know that researchers refer to them as functions.
- For example:
  - The demand function: $Q = a + bP$
  - The production function: $Q = f(K, L)$
  - The popularity function (presidential approval)

Identities

- **Identities**: A relationship between variables that is always true.
- An identity is a special type of function that is always true.
  - For example, the accounting identity in macroeconomics: $Y = C + I + G + NX$
  - Also, the accounting equation in a business context: $A = L + E$
- You can think of these as true by definition.
  - They are useful for understanding the relationships between variables, and for checking the validity of a model.
  - Often are restricting assumptions in a model.

Functional form

- **Functional form**: The way that a function is written, referring often to the polinoimial degree of the function or the type of function.
- The functional form of a function is important because it determines the shape of the relationship between variables.
  - For example, a linear function has a different shape than a quadratic function.
- The functional form is often determined by the theory behind the model, but researchers often try different functional forms to see which one fits the data best.

Reduced form vs. structural estimation of models

- In empirical work, we often hear about reduced form and structural models.
- Structural estimation means taking the model very seriously, and estimating the parameters of the model directly.
  - For example, estimating the parameters of the supply and demand model directly from data.
  - This is often difficult because the model is complex and the parameters are hard to estimate.
  - Requires a lot of strong math skills and often advanced statistical techniques.

Reduced form vs. structural estimation of models

- Reduced form estimation means estimating the parameters of the model indirectly, by estimating the relationship between the variables.
  - For example, estimating the price elasticity of demand from data.
  - This is often easier because the model is simpler and the parameters are easier to estimate.
  - Requires less math skills and often simpler statistical techniques.
- Both approaches have their advantages and disadvantages, and are often used together in empirical work.
- Modern work has seen a pronounced shift towards reduced form estimation, as it is often easier to estimate and more robust to assumptions.
  - Causal inference is often implemented in reduced form models.
  - In this program we will mostly learn about reduced form methods and read about reduced form papers.