Model Estimation when the Dependent Variable is Unobservable

B & R Research Brief
March 2025

What happens when the dependent variable in an economic model is unobservable? When the dependent variable is unobserved, it's like trying to solve a puzzle with missing pieces. But econometricians have developed clever techniques to estimate these models. These approaches include:

1. Latent Variable Models:

- **Idea:** Assume the unobserved dependent variable (latent variable) is related to observable variables through a specific function.
- **Example:** In a model of job satisfaction, the latent variable could be "true happiness at work," which is unobserved. But we can observe things like salary, commute time, and relationships with coworkers. The model links these observable factors to the unobserved "happiness."
- **Techniques:**
    - **Structural Equation Modeling (SEM):** Estimates relationships between multiple latent and observed variables, often with complex causal pathways.
    - **Item Response Theory (IRT):** Used when the unobserved variable is a trait or ability (e.g., intelligence), measured through multiple observed items (e.g., test questions).

2. Limited Dependent Variable Models:

- **Idea:** The dependent variable is observed, but its values are limited or censored in some way.
- **Examples:**
    - **Tobit Model:** Used when the dependent variable is continuous but has a lower or upper limit (e.g., hours worked, which cannot be negative).
    - **Heckman Selection Model:** Addresses sample selection bias, where the observed sample is not random (e.g., wages are only observed for those who are employed).
- **Techniques:** Maximum likelihood estimation is often used to estimate these models.

3. Proxy Variables:

- **Idea:** Find an observable variable that is highly correlated with the unobserved variable and use it as a substitute.
- **Example:** If you can't directly measure "air quality," you might use "particulate matter concentration" as a proxy.

- **Challenges:** Finding a good proxy can be difficult, and there's always some measurement error involved.

4. Instrumental Variables (IV):

- **Idea:** Find a variable (instrument) that is correlated with the unobserved dependent variable but not directly with the error term in the model.
- **Example:** In studying the impact of education on earnings, "distance to nearest college" could be an instrument for education level.
- **Use:** Helps address endogeneity, where the unobserved variable is correlated with other explanatory variables.

General Considerations:

- **Assumptions:** All these methods rely on assumptions about the relationship between the unobserved variable and the observed data. The validity of the estimates depends on how well these assumptions hold.
- **Data quality:** Having rich and reliable data on the observable variables is crucial for accurate estimation.
- **Model identification:** The model must be identified, meaning there's enough information in the data to estimate the parameters of interest.

Selective references

1. Latent Variable Models:

- **General Latent Variable Models:**
  - Bollen, K. A. (1989). *Structural Equations with Latent Variables*. Wiley. This is a classic text on structural equation modeling.
  - Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Chapman & Hall/CRC.[1] This book provides a comprehensive overview of various latent variable models.
- **Item Response Theory:**
  - Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Lawrence Erlbaum Associates. This is a widely used textbook on IRT.
  - Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Routledge. This book focuses on applications of IRT in educational testing.

2. Limited Dependent Variable Models:

- **Tobit Model:**
  - Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, 26(1), 24-36. This is the original paper that introduced the Tobit model.

    - Amemiya, T. (1984). Tobit models: A survey. *Journal of Econometrics*, 24(1-2), 3-61. This paper provides a comprehensive review of the Tobit model and its extensions.
- **Heckman Selection Model:**
    - Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1), 153-161. This is the seminal paper on the Heckman selection model.
    - Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press. This textbook includes a detailed discussion of the Heckman model and its applications.

3. Proxy Variables:

- **General Use of Proxies:**
    - Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press. This textbook discusses the use of proxy variables in various econometric contexts.
    - Angrist, J. D., & Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press. This book provides a practical guide to econometric methods, including the use of proxies.
- **Specific Applications:**
    - The use of proxy variables is often discussed within the context of specific research areas, such as environmental economics, health economics, and development economics. You can find relevant citations by searching for literature in these areas.

4. Instrumental Variables (IV):

- **General IV Methods:**
    - Angrist, J. D., & Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press. This book provides a thorough treatment of IV methods.
    - Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press. This textbook also covers IV estimation in detail.
- **Specific Applications:**
    - Miguel, E., Satyanath, S., & Sergenti, E. (2004). Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy*, 112(4),[2] 725-753. This paper uses IV to study the impact of economic shocks on civil conflict.
    - Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*,[3] 91(5), 1369-1401. This influential paper uses IV to examine the long-term effects of colonial institutions on economic development.

See also:

1. Well-being and Happiness:

- **Dependent Variable:** Overall well-being, happiness, or life satisfaction. These are subjective and hard to quantify directly.
- **Proxy Variables:**
  - Income or consumption levels
  - Health indicators (e.g., self-reported health, disability status)
  - Social indicators (e.g., marital status, social connections)
- **Example:** Frey, B. S., & Stutzer, A. (2002). *Happiness and Economics: How the Economy and Institutions Affect Well-Being*. Princeton University Press.[1] This book explores the relationship between economic factors and happiness, often using proxy variables like income and unemployment to measure well-being.

2. Environmental Quality:

- **Dependent Variable:** Overall environmental quality, which is a complex and multi-dimensional concept.
- **Proxy Variables:**
  - Air pollution levels (e.g., particulate matter concentration)[3]
  - Water quality indicators (e.g., levels of pollutants, biodiversity)
  - Forest cover or other land-use measures
- **Example:** Barbier, E. B. (2011). *Capitalizing on Nature: Ecosystems as Natural Assets*. Cambridge University Press. This book discusses the economic value of ecosystems and often uses proxy variables to measure environmental quality.

3. Human Capital:

- **Dependent Variable:** A person's stock of knowledge, skills, and abilities, which is difficult to measure directly.
- **Proxy Variables:**
  - Years of education
  - Test scores
  - Work experience
  - On-the-job training
- **Example:** Hanushek, E. A., & Woessmann, L. (2011). The economics of international differences in educational achievement. In *Handbook of the Economics of Education*[5] (Vol. 3, pp. 89-200). Elsevier. This chapter examines the relationship between human capital and economic growth, often using proxy variables like test scores to measure human capital.

4. Social Capital:

- **Dependent Variable:** The value of social networks and connections, which is hard to quantify.
- **Proxy Variables:**
  - Membership in civic organizations
  - Voter turnout
  - Levels of trust (measured through surveys)
  - Community participation
- **Example:** Putnam, R. D. (2000). *Bowling Alone: The Collapse and Revival of American Community*. Simon and Schuster. This book explores the decline of social capital in the United States, using various proxy variables to measure this concept.