The Dynamic Labor Supply Model

Lecture Notes

The Dynamic Labor Supply Model

Starting Point: The Basic Model

- Without heterogeneity and with quasilinear preferences ($\sigma = 0$), labor supply is:

$$\log(H_{n,t}) = \psi \log(\alpha) + \psi \log(W_{n,t})$$

- This predicts a perfectly straight line between log hours and log wages
- Clearly, we need more assumptions to bring this model to data

---

Extending the Model for Estimation

Preference Heterogeneity

- Let preferences be heterogeneous and decomposed as:

$$\psi \log(\alpha_n) = \mu_\alpha + \varepsilon_n, \ \mathbb{E}[\varepsilon] = 0$$

Measurement Error in Hours

- Assume hours are observed with additive measurement error ($\xi_{n,t}$):

$$\log(H_{n,t}) = \mu_\alpha + \psi \log(W_{n,t}) + \varepsilon_n + \xi_{n,t}$$

Wage Equation

- Wages follow:

$$\log(W_{n,t}) = \gamma_0 + Z_{n,t}\gamma_1 + \zeta_n + \upsilon_{n,t}$$

- ζ_n: unobserved permanent component of n's productivity
- $\upsilon_{n,t}$: time-varying shock
- $Z_{n,t}$: variable that shifts labor demand in ways essentially random with respect to individual-level unobservables

---

Extending Models: Two Key Questions

When you extend a model to account for randomness in outcomes, keep two things in mind:

1. **What is your theory for why this residual exists?**
    - What is the structural error term in your model?
2. **What other components could explain part of this residual that are *not* in your model?**

Identification Tasks (in order of importance)

1. Craft an argument and approach consistent with the assumptions of your model
    - Your model may already pose important endogeneity problems to solve
2. Craft an argument and approach that is plausible and robust to potential mechanisms *not* in your model
    - If you don't have (1), there's no point in (2)
    - But addressing (2) goes a long way to convincing your audience

---

Simple Identification

The Naive Approach

- Assume unobservables are independent of each other:

$$(\varepsilon_n, \xi_{n,t}) \perp (\zeta_n, \upsilon_{n,t})$$

- This implies:

$$\mathbb{E}[\varepsilon_n + \xi_{n,t} | W_{n,t}] = 0$$

- Sufficient for OLS to consistently recover ψ
- Could estimate ψ with a single cross-section by regressing log hours on log wages

Assessment of This Approach

1. Easy to make an identification argument consistent *inside* the model
2. The modeling assumptions themselves are much harder to justify
    - Think in terms of modeled and unmodeled unobservables

STOP FOR DISCUSSION

- Think of all the reasons why wages vary across people
- Think of all the reasons why hours vary across individuals
- Recall that ψ is a *causal* parameter. Is there anything even remotely plausible about the assumption that the unobserved determinants of wages are uncorrelated with unobserved determinants of hours?

Identification with Instrumental Variables

The Key Difference

- In the naive OLS approach, the identification condition was:

$$\mathbb{E}[\varepsilon + \xi | W] = 0$$

- This implicitly assumed that *all* variation in W (Z, v, and ζ) is essentially random

The IV Approach

- Extract the "plausibly random" component of wages given by the instrument
- Requires instead:

$$\mathbb{E}[\varepsilon + \xi | Z] = 0$$

- Depending on the nature of Z, this can be a much easier assumption to believe and defend

Why IV is More Credible

- When people say this approach is more *credible*, they mean:
  - The required assumptions for identification are weaker
  - Easier to defend
  - Robust to the kinds of mechanisms that discredited the OLS approach

The 2SLS Estimand

- For one endogenous variable and one instrument:

$$\alpha_{2SLS} = \frac{\mathbb{C}(\log(H), Z)}{\mathbb{C}(\log(W), Z)}$$

- When $\mathbb{E}[\varepsilon + \xi | Z] = 0$ and $\gamma_1 \neq 0$:

$$\alpha_{2SLS} = \psi$$

Whether vs How

- Proving sufficient conditions for identification in either case is straightforward
  - This is the "whether"
  - Usually taken as given without further discussion
- The "how" is more interesting
  - Refers to the *nature* of the respective independence assumptions
  - The independence condition for IV is strictly weaker than for OLS
  - May (depending on Z) be much easier to defend *a priori*

Thinking Outside the Model

- Sometimes out of necessity, we write simple models that imply naive identification is valid *inside the model*
- Example: heterogeneous agent macro models often assume homogeneous preferences (no unobserved heterogeneity)
  - If the model generated the data, OLS would consistently recover elasticities
- **Key point:** Think outside the model
  - Ask whether the identification strategy is robust to mild extensions or mechanisms that were too complicated for your model

---

Identification of the Model with Income Effects

Adding Income Effects Back

- Assume $\sigma > 0$; labor supply becomes:

$$\log(H_{n,t}) = \mu_\alpha + \psi \log(W_{n,t}) - \sigma\psi \log(C_{n,t}) + \varepsilon_n + \xi_{n,t}$$

- Now have access to one cross-section: joint distribution $\mathbb{P}_{Z,W,H,A,C}$
  - A is assets, C is consumption

The Simple 2SLS Problem

- If we estimate by 2SLS with just Z as an instrument:

$$\log(H) = \beta_0 + \beta_1 \log(W) + \epsilon_0$$

$$\log(W) = \kappa_0 + \kappa_1 Z + \epsilon_1$$

- With binary $Z \in \{0, 1\}$:

$$\alpha_{2SLS} = \frac{\mathbb{E}[\log(H)|Z=1] - \mathbb{E}[\log(H)|Z=0]}{\mathbb{E}[\log(W)|Z=1] - \mathbb{E}[\log(W)|Z=0]}$$

- This identifies the effect of the policy on hours (a specific causal parameter), not directly ψ alone

Achieving Point Identification

- Rank conditions for IV suggest we need two instruments for two endogenous variables
- Define $\tilde{Z} = M \times Z$ where $M \in \{0, 1\}$ indicates whether assets are above/below median
- Key insight: the policy affects consumption differently for high vs low asset individuals
  - This interaction provides the second instrument
- Result: With the interaction $\tilde{Z}$, both ψ and σ are identified

---

Difference in Differences Example

Setup

- Two cross-sections from periods $t \in \{1, 2\}$
- Two demographic groups $G \in \{A, B\}$
- $Z \in \{0, 1\}$: presence of a proportional tax subsidy τ
- Only group B is eligible for the subsidy

Net Wages

$$\mathbb{E}[\log(W)|G, t] = \gamma_t + \log(1+\tau) Z_t \mathbf{1}\{G = B\} + \omega_B \mathbf{1}\{G = B\}$$

- ω_B: persistent differences in productivity between groups
- γ_t: aggregate trends

Hours

$$\mathbb{E}[\log(H)|G, t] = \mu + \kappa_B \mathbf{1}\{G = B\} + \psi \mathbb{E}[\log(W)|G, t] - \psi\sigma \mathbb{E}[\log(C)|G, t]$$

Parallel Trends and Identification

- Euler equation implies (under full information, no shocks):

$$\Delta \mathbb{E}[\log(C)|G] = \log(\beta(1+r))$$

- If the policy were never introduced:

$$\Delta \mathbb{E}[\log(H)|G] = \psi(\gamma_2 - \gamma_1)$$

- Thus **parallel trends holds** for both log hours and log consumption

The DD Estimand

- Suppose the policy is introduced **unexpectedly** in period 2
- Difference-in-differences estimand:

$$\alpha^H_{DD} = \Delta \mathbb{E}[\log(H)|B] - \Delta \mathbb{E}[\log(H)|A]$$

- Substituting terms:

$$\alpha^H_{DD} = \psi \log(1+\tau) - \sigma\psi\alpha^C_{DD}$$

  where α^C_{DD} is the effect on log consumption

Key Observations from the DD Example

1. α_{DD}^{H} identifies a very specific causal parameter:
   - The effect of an unannounced policy introduction on hours for group B
2. With data on consumption and hours, we could combine α_{DD}^{C} and α_{DD}^{H} to learn ψ and σ
3. Groups A and B may differ in preferences, wages, and assets
   - The policy likely has different effects on their consumption
   - Heterogeneous income effects mean these estimands don't identify the effect on group A
4. These estimands also don't identify the effect with different perceived policy persistence
5. If the policy was announced in period 1 and implemented in period 2:
   - **Parallel trends would be violated**
6. Identification here achieved without assuming Z is independent of observables
   - Instead, we exploit the existence of parallel trends

Identification with Panel Data

The Panel Data Approach

- Have **panel data** on hours and wages for each individual
- Observe the distribution $\mathbb{P}_{(H_t,W_t,C_t)_{t=1}^T}$ for T periods

First Differences

$$\Delta \log(H) = \psi \Delta \log(W) - \psi\sigma\Delta \log(C) + \Delta\xi_{n,t}$$

- Notice: unobserved permanent heterogeneity (ε_n) is differenced out!

Identification Condition

$$\mathbb{E}[\Delta\xi_{n,t}|\Delta \log(W), \Delta \log(C)] = 0$$

- Guarantees OLS on first differences recovers ψ and $\psi\sigma$

Panel vs Instrumental Approaches

A consistent theme for solving identification problems:

- **Unobserved heterogeneity** lies at the heart of causal inference problems
- Two main solutions:
  1. **IV approach:** Find variation that is plausibly random
  2. **Panel approach:** Use repeated observations to learn about and handle unobserved variation

Additional Comments on Panel Identification

1. Much like IV, this approach extracts a "more credible" source of variation
   - Uses year-to-year changes
   - Differences out permanent individual differences
2. *Inside* the model: since ξ is assumed iid measurement error, this is valid
3. *Outside* the model: Are there confounding mechanisms?
   - One view: assumptions weaker than cross-sectional OLS, but stronger than a good instrument
4. **Measurement error warning:**
   - If there is also measurement error in wages, the cure could be worse than the disease
   - Measurement error could drive most of the year-to-year wage variation

---

Whether vs How (Revisited)

- The panel data approach consistently recovers parameters whether or not unobserved heterogeneity is an issue
- One could use this approach even if planning to use parameters in a simpler model without unobserved heterogeneity
- This illustrates the value of robust identification strategies