Quantitative Social Science Methods, I, Lecture Notes: Multiple Equation Models

Gary King[1]
Institute for Quantitative Social Science
Harvard University

August 17, 2020

[1] GaryKing.org

Identification

Seemingly Unrelated Regression Models

Reciprocal Causation (Endogeneity)

Multinomial Choice Models

Identification

- Definition
  - Qualitative: we can learn the parameter when $n \longrightarrow \infty$
  - Mathematical: Each parameter value produces unique likelihood value
  - Graphical: A likelihood with a unique maximum point
- Partially identified models: the likelihood is informative but only about a range of parameter values
- Non-identified models
  - Likelihood with nonunique maximum
  - Models that often make little sense, even if hard to tell.
- Reading: *Unifying Political Methodology*, Chapter 8

Example 1: Flat Likelihoods

- A (dumb) model: $Y_i \sim f_p(y_i|\lambda_i), \quad \lambda_i = 1 + 0\beta$
- What do we know about β? (from the likelihood perspective)

$$L(\lambda|y) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!}$$

$$\ln L(\beta|y) = \sum_{i=1}^{n} \{-(0\beta + 1) - y_i \ln(0\beta + 1)\} = \sum_{i=1}^{n} -1 = -n$$

- Identified likelihood
  - Has a unique maximum
- Likelihood with a plateau
  - Not identified
  - No unique MLE
  - Can be informative

Example 2: Non-unique Reparameterization

The problem of collinearity derived from the likelihood theory of inference

- A model

$$
\begin{aligned}
Y_i &\sim f_N(y_i|\mu_i, \sigma^2) \\
\mu_i &= x_{1i}\beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3, \qquad \text{where } x_{2i} = x_{3i} \\
&= x_{1i}\beta_1 + x_{2i}(\beta_2 + \beta_3)
\end{aligned}
$$

- What is the (unique) MLE of β_2 and β_3? Different parameter values lead to the same values of μ and thus the same likelihood values:

$$
\begin{aligned}
\mu_i &= x_{1i}\beta_1 + x_{2i}(5 + 3) \\
\mu_i &= x_{1i}\beta_1 + x_{2i}(3 + 5) \\
\mu_i &= x_{1i}\beta_1 + x_{2i}(7 + 1)
\end{aligned}
$$

- Likelihood of $\{\beta_2 = 3, \beta_3 = 5\}$ is same as $\{\beta_2 = 5, \beta_3 = 3\}$

Identification

Seemingly Unrelated Regression Models

Reciprocal Causation (Endogeneity)

Multinomial Choice Models

A Multiple Equation Model

- Joint Stochastic Component

$$\underset{N\times 1}{Y_i} \sim f(\underset{N\times 1}{\theta_i}, \underset{N\times N}{\alpha}) \quad \text{for each } i \; (i = 1, \ldots, n)$$

- Systematic components

$$\begin{aligned} \theta_{1i} &= g_1(x_{1i}, \beta_1) \\ \theta_{2i} &= g_2(x_{2i}, \beta_2) \\ &\vdots \\ \theta_{Ni} &= g_N(x_{Ni}, \beta_N) \end{aligned}$$

- Independence Assumption
  - $Y_i \perp\!\!\!\perp Y_j \mid \forall i \neq j$
  - Not an assumption about relationships among $Y_{i1}, \ldots, Y_{iN}$

When is a M.E. Model Better than N Separate Models?

When elements of $Y_i|X$ are *stochastically* or *parametrically* dependent

- Full Probability, with $N = 2$: $f(y|\theta) = \prod_{i=1}^{n} f(y_{1i}, y_{2i}|\theta_{1i}, \theta_{2i})$
- Assume stochastic independence so we can factor f

$$
\begin{aligned}
P(y|\theta) &= \prod_{i=1}^{n} f(y_{1i}, y_{2i}|\theta_{1i}, \theta_{2i}) \\
&= \prod_{i=1}^{n} f(y_{1i}|\theta_{1i}) f(y_{2i}|\theta_{2i})
\end{aligned}
$$

with log-likelihood

$$
\ln L(\theta_1, \theta_2|y) = \sum_{i=1}^{n} \ln f(y_{1i}|\theta_{1i}) + \sum_{i=1}^{n} \ln f(y_{2i}|\theta_{2i})
$$

- Also assume parametric independence so θ_1 and θ_2 are distinct
  - $\rightsquigarrow$ Equivalent to maximizing equation-by-equation likelihoods

Seemingly Unrelated Regression Model (SURM)

- The Model
  1. $\underset{N\times 1}{Y_i} \sim N(\underset{N\times 1}{\mu_i}, \underset{N\times N}{\Sigma})$
  2. $\underset{N\times 1}{\mu_{ij}} = \underset{N\times k_j}{X_{ij}} \underset{k_j\times 1}{\beta_j}$ for equation j ($j = 1, \dots, N$)
  3. $Y_i \perp\!\!\!\perp Y_j \mid \forall i \neq j$

- Likelihood

$$L(\beta, \Sigma) = \prod_{i=1}^{n} N(y_i | \mu_i, \Sigma)$$
$$= \prod_{i=1}^{n} (2\pi)^{-1} |\Sigma|^{-1/2} \exp\left[-\frac{1}{2}(y_i - \mu_i)' \Sigma^{-1} (y_i - \mu_i)\right]$$

SURM Notes

- If $Y|X$ are *stochastically independent*, and β's are *parametrically independent*: SURM = equation-by-equation LS.
- In any PDF
  - stochastic independence $[\mathrm{P}(a, b) = \mathrm{P}(a)\mathrm{P}(b)] \implies$
  - mean independence $[E(ab) = E(a)E(b)] \implies$
  - uncorrelatedness (no linear relationship) $[\mathrm{Corr}(a, b) = 0]$.
- If Y is multivariate normal
  - Uncorrelatedness $\implies$ stochastic independence
  - Identical X's: SURM = equation-by-equation LS
- $\implies$ identification of extra parameters with identical X: solely from modeling assumptions (i.e., lots of model dependence)
- Programming is more complicated: Y is $n \times N$ instead of $n \times 1$
- Computational tricks: can make estimation a lot faster

Identification

Seemingly Unrelated Regression Models

Reciprocal Causation (Endogeneity)

Multinomial Choice Models

Model for Reciprocal Causation

- Stochastic component: $(Y_{1i}, Y_{2i}) \sim N(\mu_{1i}, \mu_{2i}, \sigma_1, \sigma_2, \sigma_{12})$
- Systematic components

$$\begin{aligned} \text{Vote:} \quad & \mu_{1i} = x_{1i}\beta_1 + x_{2i}\beta_2 + \mu_{2i}\beta_3 \\ \text{PID:} \quad & \mu_{2i} = x_{1i}\gamma_1 + x_{3i}\gamma_2 + \mu_{1i}\gamma_3 \end{aligned}$$

Where,

β_3 and γ_3: QOIs
x_1 demographics (in both equations)
x_2 candidate characteristics (affecting vote but not PID)
x_3 parents PID (affecting PID but not vote)

- Likelihood

$$\begin{aligned} f(y|\mu, \Sigma) &= \prod_{i=1}^{n} N(y_i|\mu_i, \Sigma) \\ &= \prod_{i-1}^{n} (2\pi)^{N/2} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(y_i - \mu_i)'\Sigma^{-1}(y_i - \mu_i)\right\} \end{aligned}$$

- What do we do with $\mu_i = (\mu_{i1}, \mu_{i2})$?

How to substitute in the systematic component?

- Goal of reparameterization: Reduce parameter dimensionality
- Standard models: reparameterized (substitute μ for $X\beta$)
- Time series models: recursively reparameterized
- M.E. Models: *multiple equation reparameterization*

$$\begin{aligned}
\mu_{1i} &= x_{1i}\beta_1 + x_{2i}\beta_2 + \mu_{2i}\beta_3 \\
&= x_{1i}\beta_1 + x_{2i}\beta_2 + (x_{1i}\gamma_1 + x_{3i}\gamma_2 + \mu_{1i}\gamma_3)\beta_3 \\
&= x_{1i}\beta_1 + x_{2i}\beta_2 + x_{1i}\gamma_1\beta_3 + x_{3i}\gamma_2\beta_3 + \mu_{1i}\gamma_3\beta_3 \\
&= \left(\frac{1}{1-\gamma_3\beta_3}\right)\left[x_{1i}\beta_1 + (x_{1i}\gamma_1 + x_{3i}\gamma_2)\beta_3 + x_{2i}\beta_2\right]
\end{aligned}$$

$$\begin{aligned}
\mu_{2i} &= x_{1i}\gamma_1 + x_{3i}\gamma_2 + \mu_{1i}\gamma_3 \\
&= x_{1i}\gamma_1 + x_{3i}\gamma_2 + (x_{1i}\beta_1 + x_{2i}\beta_2 + \mu_{2i}\beta_3)\gamma_3 \\
&= x_{1i}\gamma_1 + x_{3i}\gamma_2 + (x_{1i}\beta_1 + x_{2i}\beta_2)\gamma_3 + \mu_{2i}\beta_3\gamma_3 \\
&= \left(\frac{1}{1-\beta_3\gamma_3}\right)\left[x_{1i}\gamma_1 + x_{3i}\gamma_2 + (x_{1i}\beta_1 + x_{2i}\beta_2)\gamma_3\right]
\end{aligned}$$

What Happens With no Variable Exclusions?

- Suppose we drop X_2 and X_3 from systematic components

$$\begin{aligned}\mu_{1i} &= \left(\frac{1}{1-\gamma_3\beta_3}\right)\left[x_{1i} + x_{1i}\gamma_1\beta_3\right]\\ &= x_{1i}\left(\frac{\beta_1+\gamma_1\beta_3}{1-\gamma_3\beta_3}\right)\end{aligned}$$

- Likelihood not identified
    - $\{\beta_1 = 1, \gamma_1 = 30, \beta_3 = 555, \gamma_3 = -30\} \implies \mu_{1i} = -x_i$.
    - $\{\beta_1 = 1, \gamma_1 = 1, \beta_3 = -999, \gamma_3 = -1\} \implies \mu_{1i} = -x_i$.
    - As long as $\beta_1 = 1$ and $\gamma_1 = -\gamma_3$, for any value of β_3, $\mu_{1i} = -x_i$.
- $\rightsquigarrow$ High levels of model dependence wrt choices of X_2 and X_3
- What about other approaches?
    - Lots of choices (IV, 2SLS, 3SLS, etc.)
    - Should we believe them? (The girl in the mirror)
    - My advice: wherever possible, choose a different project

Identification

Seemingly Unrelated Regression Models

Reciprocal Causation (Endogeneity)

Multinomial Choice Models

Multinomial Choice Models

1. $k \geq 2$ nominal choices, from which one is chosen.
2. Example: choice among candidates by voters; dishes at a restaurant by customers; war/peace/limited war by countries, etc
3. Generalizations of (binary) logit and probit models

Multinomial Probit

$$U_i^* \sim N(u_i^* | \mu_i, \Sigma)$$
$$\mu_{ij} = x_{ij}\beta_j$$

with observation mechanism:

$$Y_{ij} = \begin{cases} 1 & \text{if } U_{ij}^* > U_{ij'}^* \ \forall\ j \neq j' \\ 0 & \text{otherwise} \end{cases}$$

Stochastic component:

$$\Pr(Y_{ij} = 1) = \pi_{ij}, \quad \text{s.t.} \quad \sum_{j=1}^{J} \pi_{ij} = 1 \quad \text{for} \quad i = 1, \ldots, n$$

Systematic component: Let $Y^*_{ij} = U^*_{ij} - U^*_{ij'}$, so the observation mechanism is

$$Y_{ij} = \begin{cases} 1 & \text{if } Y^*_{ij} > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned}
\pi_{ij} &= \Pr(y_{ij} = 1) \\
&= \Pr(Y^*_{i1} \le 0, \ldots, Y^*_{ij} > 0, \ldots, Y^*_{iJ} \le 0) \\
&= \int_{-\infty}^{0} \cdots \int_{0}^{\infty} \cdots \int_{-\infty}^{0} N(y|\mu_i, \Sigma)\, dy_{i1} \cdots dy_{ij} \cdots dy_{iJ}
\end{aligned}$$

Computational and Estimation issues

- No analytical solution is known to the integral
- Doing it numerically with > 4 or 5 choices would take forever
- The problem has been solved to at least 9-12 choices by simulation (simple version: draw from normal and count fraction in regions)
- All elements of Σ are not identified.
- The entire model is only weakly identified
- Model is a straightforward generalization of SURM, or of univariate probit

Multinomial Logit

- The Binary logit Model:
  - $$\Pr(Y_i = 1|X) = \pi_i \quad \Pr(Y_i = 0|X) = 1 - \pi_i$$
  - $$\pi_i = \Pr(Y_i = 1|X) = \frac{1}{1 + e^{-X_i\beta}} = \frac{e^{X_i\beta}}{1 + e^{X_i\beta}}$$
- Multinomial Logit Model:
  - $$\Pr(Y_{ij} = 1) = \pi_{ij}, \quad \text{s.t.} \quad \sum_{j=1}^{J} \pi_{ij} = 1, \quad \text{for} \quad i = 1, \ldots, n$$
  - $$\pi_{ij} = \frac{e^{x_{ij}\beta_j}}{\sum_{k=1}^{J} e^{x_{ik}\beta_k}} \quad \text{for } j = 1, \ldots, J$$
  - Identical to binary logit when $J = 2$, but it generalizes
- The Likelihood: $L(\beta|y) = \prod_{i=1}^{n} \left[\prod_{j=1}^{J} \pi_{ij}^{y_{ij}}\right]$

Independence of Irrelevant Alternatives (IIA)

- Under MNL:

$$\frac{\pi_{i1}}{\pi_{i2}} = \frac{\frac{e^{x_{i1}\beta_1}}{\sum_{k=1}^{J} e^{x_{ik}\beta_k}}}{\frac{e^{x_{i2}\beta_2}}{\sum_{k=1}^{J} e^{x_{ik}\beta_k}}} = \frac{e^{x_{i1}\beta_1}}{e^{x_{i2}\beta_2}}$$

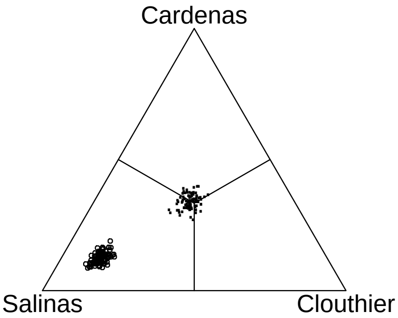
  which is not a function of choices 3,4,5, etc.
- Under MNP, probability ratios are always a function of all alternatives
- The red-bus-blue-bus problem: buses and candidates.
- Does it matter empirically? It can, but less often given estimation uncertainty

Replicating Domínguez and McCann (1996) from King, Tomz, and Wittenberg (2000)

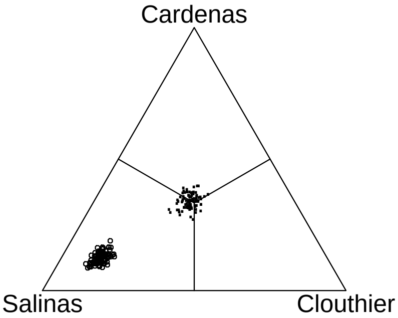
- Salinas (of the ruling PRI) won the 1988 Mexican presidential election.
- If voters had thought the PRI was weakening, who would have won?
- Model: MNL with 3 choices (Salinas from the PRI, Clouthier (from the PAN, a right-wing party), and Cuauhtémoc Cárdenas (head of a leftist coalition) and 31 control vars:

$$Y_i \sim \text{Multinomial}(\pi_i)$$

$$\pi_i = \frac{e^{X_i\beta_j}}{\sum_{k=1}^{3} e^{X_i\beta_k}} \qquad \text{where } j = 1, 2, 3 \text{ candidates.}$$



- Each point: an election outcome where all voters believe Salinas' PRI is strengthing (for the "o"'s in the bottom left) or weakening (for the "."'s in the middle), with other variables held constant at their means.



- If voters believe PRI is strengthing, PRI wins easily.
- If voters believe PRI is weakening, its a tossup.
- $\rightsquigarrow$ Despite voter fraud, Salinas probably did defeat a divided opposition in 1988. The PRI lost the next election (finally, after 72 years in power)