Roy Models: Construction, Calibration & Applications

Martin Shu

Boston University

Discussion Session
STEG Virtual Course on Macro Development
March 29, 2021

Roy Models

- Also known as models of sorting, models of (self-)selection
- Unobserved factors could determine economic decisions
  - Ability, preference, intangible costs, productivity, etc.
  - Decisions over occupation, education, location, program participation, firm entry, foreign market entry, etc.
- A bit of history
  - Roy (1951): Unobserved occupation specific skills and the correlation between skill distributions determine observed distribution of earnings
  - Heckman (1979): Two-step estimation method to correct for sample selection bias
  - Heckman & Sedlacek (1985): Empirical model of selection in U.S. labor market
  - Borjas (1987): Immigrants self-select into moving to another country

Content of the Session

- Basic setup of a 2-sector Roy model
  - Also providing alternative recipes
- Calibration/Structural Estimation of a Roy model
  - Focusing on parameters of the skill distribution
- Examples of applications

General Setup

- Two sectors: agriculture and non-agriculture
- Unit measure of workers
- Sector-specific skills of individual i, $\{z_a(i), z_n(i)\}$ drawn from joint distribution $F(z_a, z_n)$
- Wage rates for efficiency labor units: w_a, w_n
- Worker's problem:

$$\max\{w_a z_a(i), w_n z_n(i)\}$$

- Work in agriculture if

$$z_a(i) > \frac{w_n}{w_a} z_n(i)$$

Labor Market Outcomes

- Assume that joint distribution of skills is independent Fréchet

$$F(z_a, z_n) = e^{-z_a^{-\theta} - z_n^{-\theta}}$$

- Shape parameter θ: inverse measure of dispersion
- Scale normalized to 1
- Employment share of agriculture

$$\pi_a\left(\frac{w_a}{w_n}\right) = \int_0^\infty \int_{\frac{w_n z_n}{w_a}}^\infty dF(z_a)dF(z_n) = \frac{1}{\left(\frac{w_a}{w_n}\right)^{-\theta} + 1}$$

- Ratio of employment shares

$$\frac{\pi_a}{\pi_n} = \left(\frac{w_a}{w_n}\right)^\theta$$

Properties of Independent Fréchet Distribution

- Pros
  - Closed form density functions and allocation
  - Distribution of earnings outcome is still Fréchet
  - Could be thought of as maximum of several subsectors
  - Fat tail consistent with observed data
  - Log transformation is Gumbel – another distribution with closed form density functions
- Caution
  - n^{th} moment exists only if $\theta > n$
  - Average earnings in equilibrium is equal across sectors
- Alternative choices
  - Sector-specific shape parameters, log-normal distributions
  - Closed form properties no longer holds for these alternatives

Properties about Productivity

- $E[z_a|z_a/z_n > x]$ and $E[z_n|z_n/z_a > x]$ are strictly increasing in x

  $$E[z_a|z_a/z_n > x] = E[z_n|z_n/z_a > x] = (x^\theta + 1)^{\frac{1}{\theta}}\Gamma\left(1 - \frac{1}{\theta}\right)$$

  where $\Gamma(\cdot)$ is the gamma function
- Average skill level decreases in expanding sector, increases in contracting sector
- Restrictive property; could be a strong assumption depending on the economic question
- Solution: correlated skill distributions

Modeling Correlation

- Lagakos & Waugh (2013):

$$F(z_a, z_n) = C[F_a(z_a), F_n(z_n)]$$

where

$$F_a(z_a) = e^{-z_a^{-\theta_a}}, F_n(z_n) = e^{-z_n^{-\theta_n}}$$

and

$$C[u, v] = -\frac{1}{\rho} \ln \left[1 + \frac{(e^{-\rho u} - 1)(e^{-\rho v} - 1)}{e^{-\rho} - 1} \right]$$

- $C[\cdot, \cdot]$ is a Frank copula linking marginal distributions, $\rho \in (-\infty, \infty) \setminus \{0\}$ governs degree of correlation
- Symmetric tail dependence
- Asymmetric alternatives: Gumbel/Clayton copula
- Parameters to be calibrated: θ_a, θ_n, ρ

Modeling Correlation Ctnd.

- Alternative methods of specifying correlation between skills:
- Pulido & Święcki (2019): Bivariate lognormal distribution

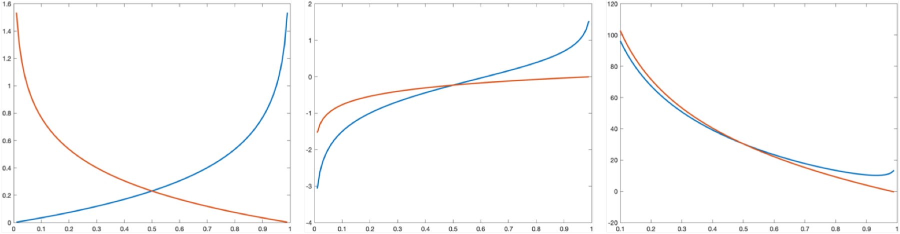
$$\{\ln z_a, \ln z_n\} \sim \mathcal{N}(0, \Sigma)$$

Parameters to be calibrated: 3 elements in the Σ matrix

- Question: does correlation between skills precisely depict the patterns of selection?

Generalized Schedules of Comp & Abs Advantages

- Define $\kappa(p) = \kappa \ln \frac{p}{1-p}$ and $\alpha(p) = \bar{\alpha} + \alpha \ln p$ where $\kappa(p)$ is comparative advantage $\ln(z_n/z_a)$ of worker at p^{th} percentile of comparative advantage, $\alpha(p)$ is $\ln z_a$ of the worker



Case 1: Fréchet $\alpha + \kappa = 0, \kappa \in (0,1)$

Case 2: $\alpha > 0$

Case 3: $\alpha + \kappa \ll 0$

Figure: Sectoral skills and percentile of comparative advantage

Modeling Correlation of Skills

- Correlation between skills is sufficient when reallocation involves only workers at the lower end of the income distribution in both sectors
- Need to verify the correlation between comparative and absolute advantages when workers among the best in either sector are reallocated
- The schedules for comparative and absolute advantages offer a starting point to model negative selection
- See Alvarez-Cuadrado, Amodio & Poschke (2020) and Adão (2016)
- See Appendix D of Adão (2016) for detailed derivations of expressions in last slide

Calibration/Structural Estimation

Data Structure and Identifiability

- Heckman and Honoré (1990):
- In general, sectoral wage data can be rationalized in a single cross section by a 2-sector Roy model with correlation greater than that of the actual data generating process
- Given multi-market data with sufficient variations in relative prices, a Roy model with non-normal distribution of skills can be identified, even when returns from one sector is unobservable (e.g. home production)
- Given panel data of individual earnings, a Roy model with non-normal distribution of skills can be identified over the range $\frac{w_n z_n}{w_a z_a} \leq 1 \leq \frac{w'_n z_n}{w'_a z_a}$
- Rule of thumb: use panel data whenever possible!

Moments to Match

- General practice: simulate data and search for parameter values that
  - match data moments and simulated moments, or
  - minimize distance between data moments and simulated moments
- LW: variance of permanent component of income inform θ_a and θ_n; choose ρ to match ratio of sectoral income
- Intuition:
  θ governs dispersion of skill distribution
  High ρ means greater gaps between mean sectoral income
  Low ρ means smaller gaps between mean sectoral income

Moments to Match Ctnd.

- Pulido & Święcki (2019)
  1. Regress log income on observables, controlling for interaction between sector and year
  2. Construct residual income net of observables
  3. Run auxiliary regressions of residual income on variables of sectoral choices with individual/time/individual-time/no fixed effects, and extract estimated coefficients
  4. Search for parameter values that minimize the weighted sum of squared distances between coefficients obtained in Step 3 and those from running the same regressions with simulated data
  5. Obtain standard errors with bootstrapping

Moments to Match Ctnd.

- Hsieh, Hurst, Jones & Klenow (2020)
  - Assumed independence; observed wages follow Fréchet, in particular, the ratio between variance and the square of mean has closed form expression as function of θ
  - Cross-checked with extensive margin elasticity of labor supply since home production is modeled
- Adão (2016)
  - Multi-region data with different exposure to world price shocks
  - Model is identified with cross-section data of income and labor allocation from different markets
  - Weaker assumption than in Heckman and Honoré (1990)

Summary

- No unified practice in general
- Fréchet shape parameters (standard deviations for lognormal distributions) informed by measures of dispersions in data
- Correlation parameters (covariance for lognormal distributions) informed by observations that switch sectors
- Look for variables in the model most directly related to the parameters and match their observable counterparts in data

Examples of Applications

Agricultural Productivity Gap

- Lagakos & Waugh (2013): With non-homothetic preference, selection is able to enlarge agricultural productivity gap when technological progress is sector-neutral
- Experiment: vary level of technology in the calibrated general equilibrium model and record sectoral productivity gaps

Table 2—90–10 Productivity Differences, Data and Benchmark Model

	Agriculture	Aggregate	Non-agriculture	Ag/non-agriculture ratio
Data	45	22	4	10.7
Model	29	22	13	2.2
Without selection	19	19	19	1.0

Notes: The aggregate productivity difference is the ratio of GDP per worker between the ninetieth and tenth percentile countries. Sector productivity differences are the ratio of sector output per worker in the ninetieth and tenth percentile countries. The ag/non-agriculture ratios are the agriculture productivity differences divided by the non-agriculture productivity differences.

Source: Authors' calculations and Caselli (2005).

Gains from Reduced Labor Market Discriminations

- Hsieh, Hurst, Jones & Klenow (2020): falling occupational barriers explain roughly 40% of aggregate growth in market GDP per person
- Selection based on ability fits data better than selection based on preferences

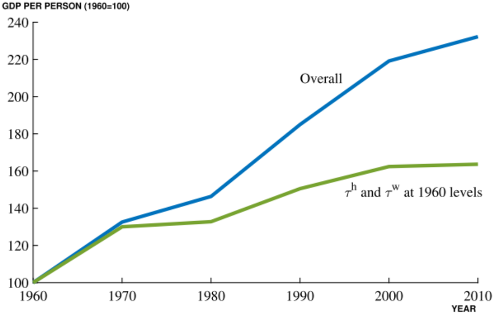


FIGURE 7.—GDP per person, data and model counterfactual. Note: The graph shows the cumulative growth in GDP per person (market), in the data (overall), and in the model with no changes in τ's as in Table V.

Intergenerational Evolution of Sector-Specific Skills

- Hobijn, Schoellman & Vindas (2018): when workers can costly train/retrain sector-specific skills, structural transformation is more rapid as people anticipate future growth in relative wages and train themselves in advance
- Adão, Beraja & Pandalai-Nayar (2019): consequences of cognitive-biased innovations can play out slowly when technology-skill specificity and skill investment cost jointly determine within-generation reallocation of workers and between-generation evolution of the skill distribution

Reference

1. Adão, Rodrigo. 2016 "Worker Heterogeneity, Wage Inequality, and International Trade: Theory and Evidence from Brazil." Working Paper.
2. Adão, Rodrigo, Martin Beraja, and Nitya Pandalai-Nayar. 2019 "Technological Transitions with Skill Heterogeneity Across Generations." Working Paper.
3. Alvarez-Cuadrado, Francisco, Francesco Amodio, and Markus Poschke. 2020. "Selection and Absolute Advantage in Farming and Entrepreneurship." Working Paper.
4. Borjas, George J. 1987. "Self-Selection and the Earnings of Immigrants." *American Economic Review*, 77(4): 531-553.
5. Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica*, 47(1): 153-161.
6. Heckman, James J., and Bo E. Honoré. 1990. "The Empirical Content of the Roy Model." *Econometrica*, 58(5): 1121-1149.
7. Heckman, James J., and Guilherme Sedlacek. 1985. "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market." *Journal of Political Economy*, 93(6): 1077-1125.
8. Hobijn, Bart, Todd Schoellman, and Alberto Vindas Q. 2018. "Structural Transformation by Cohort." Working Paper.
9. Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow. 2019. "The Allocation of Talent and U.S. Economic Growth." *Econometrica*, 87(5): 1439-1474.
10. Lagakos, David, and Michael Waugh. 2013. "Selection, Agriculture and Cross-Country Productivity Differences." *American Economic Review*, 103(2): 948-980.
11. Pulido, José, and Thomas Święcki. 2019. "Barriers to Mobility or Sorting? Sources and Aggregate Implications of Income Gaps across Sectors in Indonesia." Working Paper.
12. Roy, Andrew D. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers*, New Series, 3(2): 135-146.