MIT 14.662 Spring 2026: Lecture Slides 3 – Comparative Advantage, Self-Selection, and the Roy Model

Part 1: The normal selection model and a first application

David Autor, MIT and NBER

February 9, 2026 (rev 2026/02/10)

Agenda

Models of comparative advantage

- In models we've studied so far, workers are bound to skill type (e.g., H or L)
- Can differ by quantity of H, L within type, perfect substitutes within types
- In reality, workers may have multiple skills, but they may not be able to use them simultaneously
  - You might be a great ballroom dancer *and* a great economist
  - If you choose a career in ballroom dancing, your skills in economics will not be financially rewarded, and vice versa
- We want a model of people choosing which skills to sell to the market
- This naturally leads to the notion of *self-selection according to comparative advantage*
  - Roy '51 was the first paper to formalize this notion—in terms of labor supply rather than international trade

Models of comparative advantage

- The starting point of formal treatment of this topic in economics is Roy's (1951) "Thoughts on the Distribution of Earnings"
- Roy discusses the optimizing choices of 'workers' selecting between *fishing* and *hunting*
- Roy's question: Who chooses which activity, and what do those choices imply for the distribution of earnings?
- Roy's answer: Three core factors determine these outcomes
  1. The fundamental distribution of skills (i.e., in hunting and fishing)
  2. The technologies for applying these skills (translating ability into wages)
  3. The correlations among these skills in the population

Some natural 'Roy' questions

1. How much of the 'return' to education is due to positive (or negative) self-selection? (Willis and Rosen '79)
2. What is the 'return' to field of study, e.g., medicine, engineering, law, humanities? (Kirkeboen, Leuven, and Mogstad '16)
3. Why does the average U.S. immigrant from Asia earn more than the average U.S. native (Many papers by Borjas; Abramitzky and Boustan)
4. How much does race and gender discrimination reduce productivity through misallocation? (Hsieh, Hurst, Jones, and Klenow '19)
5. Would raising taxes on top earners distort talent allocation? (Scheuer and Werning '17)
6. Are medical patients (e.g., heart attack patients) given the 'right' treatments as a function of their underlying health and the hospital where they are treated? (Chandra and Staiger '07, '20)
7. When doctors interpret clinical test data, is it only preferences that determine systematic differences? (Gentzkow and Yu '22)

The ‘normal’ Roy model

- Willis and Rosen ’79 (“Education and Self-Selection”) is perhaps the first paper to write a formal Roy model (though there may be earlier examples)
- Borjas 1987 developed the ‘normal’ version (based on the bivariate normal distribution), which is intuitive and tractable
  - ‘Borjas selection model’ is a window into labor economics in the 1980s and 1990s
  - See James Heckman’s famous ’79 *Ecma* paper ($> 30K$ cites), “Sample Selection Bias as a Specification Error.”
- See Adäo ’16 for a fully modern non-parametric version
- Roy model has also caught on big-time in trade models
  - These models use Fréchet distribution of skills (the “Roy-Fréchet” model)
  - This has technical advantages and substantive disadvantages

Models of comparative advantage

In thinking about these models, three concepts to keep in mind

1. Comparative advantage, which determines self-selection
2. Absolute advantage, which determines wage distribution conditional on self-selection
3. General equilibrium, which is determined by both (1) and (2). (We will ignore general equilibrium for now, but we'll see it elsewhere soon.)

Agenda

Warm up: The normal selection model

- Before formalizing the Roy model, let's review the normal selection model
- Assume that the full earnings distribution is given by

$$w \sim N\left(\mu_0, \sigma_0^2\right)$$

- We observe wages only for those who work
- Simplest case: everyone has a reservation wage of κ
- Thus, the wage is observed iff $w > \kappa$, equivalently $\varepsilon_0 > \kappa - \mu_0$

The normal selection model

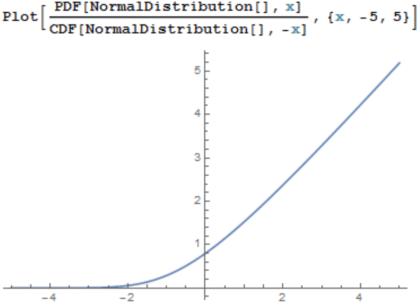
What is the expectation of observed wages?

$$\begin{aligned} E\left(w|w>\kappa\right) &= \mu_0 + E\left(\varepsilon_0|\varepsilon_0>\kappa-\mu_0\right) \\ &= \mu_0 + \sigma_0 E\left(\frac{\varepsilon_0}{\sigma_0}\middle|\frac{\varepsilon_0}{\sigma_0}>\frac{\kappa-\mu_0}{\sigma_0}\right) \\ &= \mu_0 + \sigma_0 \frac{\phi\left(z\right)}{1-\Phi\left(z\right)} \\ &= \mu_0 + \sigma_0 \frac{\phi\left(z\right)}{\Phi\left(-z\right)}, \end{aligned}$$

- Here $z=\left(\kappa-\mu_0\right)/\sigma_0$, and $\phi\left(\cdot\right)$ and $\Phi\left(\cdot\right)$ are the PDF and CDF of the standard normal distribution respectively
- We can flip the sign on $\Phi\left(-z\right)$ due to the symmetry of the normal dist'n. Also:
  - Conditional distribution of a normal random variable is *also* a normal random variable
  - Conditional expectations of normal vars are linear in parameters

The inverse Mills ratio (after Mills '26): $\frac{\phi(z)}{\Phi(-z)}$

- IMR provides the **mean of the truncated Normal distribution**
- IMR is a *hazard function*: event rate at time t conditional on survival to time t



- $\lambda(z) \geq 0, \lambda'(z) > 0, \lim_{z \to -\infty} \lambda_t(z) \to 0, \lim_{z \to \infty} \lambda'_t(z) \to 1$
- At low point of truncation, IMR is close to zero
- As truncation point rises, IMR approaches point of truncation b/c normal is thin-tailed
- Truncating right tail makes IMR negative. Intuition is familiar

Agenda

1. Introduction
2. Basics: The normal selection model
3. **The normal Roy model: selection with two sectors**
   - Two skills, each applicable in only one sector
   - Positive and negative selection
   - Absolute versus comparative advantage
   - Linking comparative to absolute advantage
4. A first application: Heart attack treatment
5. Labor-Roy application: Self-selection and returns to college major
6. Conclusions

Agenda

The normal Roy model

Now, we want to expand to two skills

- Instead of choosing work/no-work, workers i choose sector $j \in \{0, 1\}$
- Each worker endowed with two skills $s_i = \{\varepsilon_{0i}, \varepsilon_{1i}\}$ with

$$s_i \sim N\left(\begin{array}{c} 0 \\ 0 \end{array}, \begin{array}{cc} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{array}\right) = N\left(\begin{array}{c} 0 \\ 0 \end{array}, \begin{array}{cc} \sigma_0^2 & \rho_{01}\sigma_0\sigma_1 \\ \rho_{01}\sigma_0\sigma_1 & \sigma_1^2 \end{array}\right)$$

  where ρ is the bivariate correlation coefficient $\rho_{01} = \sigma_{01}/\sigma_0\sigma_1$

- Each worker faces two potential wages, expressed in logs:

$$\ln \text{wage}_{ij} \equiv w_{ij} = \mu_j + \varepsilon_{ij}.$$

- We will assume income maximization, though that's for simplicity (all we need is self-selection):

$$w_i = \begin{cases} w_{i0} & \text{if} \quad w_{i0} \geq w_{i1} \\ w_{i1} & \text{if} \quad w_{i0} < w_{i1} \end{cases}$$

The normal Roy model

- Income maximization implies

$$w_i = \begin{cases} \mu_1 + \varepsilon_{i1} & \text{if} \quad \varepsilon_{i1} - \varepsilon_{i0} \geq \mu_0 - \mu_1 \\ \mu_0 + \varepsilon_{i0} & \text{if} \quad \varepsilon_{i1} - \varepsilon_{i0} < \mu_0 - \mu_1 \end{cases}$$

- Recall that these expressions are in terms of *log* wages. It may be more intuitive to write

$$\mathrm{wage}_i = \begin{cases} \exp\left[\mu_1 + \varepsilon_{i1}\right] & \text{if} \quad \frac{\exp(\varepsilon_{i1})}{\exp(\varepsilon_{i0})} \geq \frac{\exp(\mu_0)}{\exp(\mu_1)} \\ \exp\left[\mu_0 + \varepsilon_{i0}\right] & \text{if} \quad \frac{\exp(\varepsilon_{i1})}{\exp(\varepsilon_{i0})} < \frac{\exp(\mu_0)}{\exp(\mu_1)} \end{cases}$$

- This equation captures comparative advantage

Agenda

1. **Introduction**
2. **Basics: The normal selection model**
3. **The normal Roy model: selection with two sectors**
   - Two skills, each applicable in only one sector
   - Positive and negative selection
   - Absolute versus comparative advantage
   - Linking comparative to absolute advantage
4. **A first application: Heart attack treatment**
5. **Labor-Roy application: Self-selection and returns to college major**
6. **Conclusions**

Comparative versus absolute advantage

- Worker compares

$$\frac{\exp\left(\varepsilon_{i1}\right)}{\exp\left(\varepsilon_{i0}\right)} \lesseqgtr \frac{\exp\left(\mu_0\right)}{\exp\left(\mu_1\right)}$$

  The choice of sector is entirely determined by *comparative* advantage $\left(\frac{\exp(\varepsilon_{i1})}{\exp(\varepsilon_{i0})}\right)$ not *absolute* advantage $(\max\{\varepsilon_{i0}, \varepsilon_{i1}\})$

- Worker with highest absolute advantage in sector 1 $(\varepsilon_{i'1} = \max_i\{\varepsilon_{i1}\})$ will choose sector 0 if $\frac{\exp(\varepsilon_{i'1})}{\exp(\varepsilon_{i'0})} < \frac{\exp(\mu_0)}{\exp(\mu_1)}$

- *Positive selection* means workers in a sector are "above average"

$$E\left[\varepsilon_0 \,|\text{sector } 0\right] > 0$$
$$E\left[\varepsilon_1 \,|\text{sector } 1\right] > 0$$

- Selection according to comparative advantage does *not* imply strictly positive self-selection, as we shall see

Agenda

1. **Introduction**
2. **Basics: The normal selection model**
3. **The normal Roy model: selection with two sectors**
   - Two skills, each applicable in only one sector
   - Positive and negative selection
   - Absolute versus comparative advantage
   - Linking comparative to absolute advantage
4. **A first application: Heart attack treatment**
5. **Labor-Roy application: Self-selection and returns to college major**
6. **Conclusions**

Absolute advantage and positive vs. negative self-selection

Are the 'best' workers employed in each sector?

1. Case 1: Workers in each sector are more productive than average
   - Absolute advantage mirrors comparative advantage
   - $E[\varepsilon_0|\text{sector } 0] > 0$ and $E[\varepsilon_1|\text{sector } 1] > 0$
2. Case 2: Workers in one sector more productive than average in that sector, workers in other sector less productive than average in that sector
   - Absolute advantage mirrors comparative advantage in *one* sector not *both*
   - $E[\varepsilon_0|\text{sector } 0] > 0$ and $E[\varepsilon_1|\text{sector } 1] < 0$ **or**
   - $E[\varepsilon_0|\text{sector } 0] < 0$ and $E[\varepsilon_1|\text{sector } 1] > 0$
3. Case 3: Can workers have *both* comparative adv. *and* absolute disadvantage in *each* sector?
   - $E[\varepsilon_0|\text{sector } 0] < 0$ **and** $E[\varepsilon_1|\text{sector } 1] < 0$

- *The normal selection model makes these questions tractable*

Jump to Roy figures

Calculating wages in each sector

Start by calculating observed wages by sector

- What is expected wage in sector 1 of those choosing sector 1?
- Let $\nu = \varepsilon_1 - \varepsilon_0$ and $z = \mu_0 - \mu_1$
- Choosing sector 1 means $\nu > z$
- Choosing sector 0 means $\nu < z$

What is the expected wage of workers who choose sector 1?

- **What is expected wage in sector 1 of those choosing sector 1?**

$$E(w_1|\text{sector } 1) = \mu_1 + E\left(\varepsilon_1 \left| \nu > \mu_0 - \mu_1 \right.\right) = \mu_1 + \sigma_1 E\left(\frac{\varepsilon_1}{\sigma_1} \left| \frac{\nu - z}{\sigma_\nu} > 0 \right.\right)$$

- **This equation depends on three things:**
  1. Mean earnings μ_1 and μ_0 in sectors 0 and 1 (through z)
  2. Both error terms ε_0 and ε_1 through ν
  3. It also implicitly depends on the correlation between ε_0 and ε_1 – since this affects the choice to work in sector 1 rather than 0

What is the expected wage of workers who choose sector 1?

- Before calculating $E(w_1|\text{sector } 1)$ let's consider a simpler problem, calculating $E(w_1|\nu)$
- We know that $E[w_1|\varepsilon_1] = \mu_1 + \varepsilon_1$. But we are given ν not ε
- Given the normality of $\varepsilon_0, \varepsilon_1$, $E(\varepsilon_1|\nu)$ is simply equal to the regression coefficient

$$E(\varepsilon_1|\nu) = \frac{\text{cov}(\varepsilon_1, \nu)}{\text{var}(\nu)} \times \nu = \frac{\sigma_{1\nu}}{\sigma_\nu^2}\nu$$

$$E\left(\frac{\varepsilon_1}{\sigma_1}\middle|\frac{\nu}{\sigma_\nu}\right) = \frac{\sigma_{1\nu}}{\sigma_1\sigma_\nu^2}\nu = \rho_{1\nu}\frac{\nu}{\sigma_\nu}$$

$$\text{where } \rho_{1\nu} \equiv \frac{\sigma_{1\nu}}{\sigma_1\sigma_\nu}$$

- Of course, we don't know exact value of ν, we know only its conditional expectation given $\nu - z > 0$
- That's where the IMR comes in

Expected wage of workers who choose sector 1

Since we don't know ν, we use the IMR to calculate $E\left[\nu|\nu > z\right]$

$$
\begin{aligned}
E\left[w_1 \left|\nu > z\right.\right] &= \mu_1 + E\left(\varepsilon_1 \left|\nu > z\right.\right) \\
&= \mu_1 + \sigma_1 E\left(\frac{\varepsilon_1}{\sigma_1}\left|\frac{\nu}{\sigma_\nu} > \frac{z}{\sigma_\nu}\right.\right) \\
&= \mu_1 + \frac{\sigma_{1\nu}}{\sigma_\nu\sigma_1}\sigma_1 E\left(\frac{\nu}{\sigma_\nu}\left|\frac{\nu}{\sigma_\nu} > \frac{z}{\sigma_\nu}\right.\right) \\
&= \mu_1 + \rho_{1\nu}\sigma_1\left(\frac{\phi\left(z/\sigma_\nu\right)}{\Phi\left(-z/\sigma_\nu\right)}\right)
\end{aligned}
$$

- Here we've expressed the conditional expectation of $\frac{\varepsilon_1}{\sigma_1}$ by substituting $\frac{\nu}{\sigma_\nu}$ and multiplying by the correlation coefficient, where $\rho_{1\nu} = \frac{\sigma_{1\nu}}{\sigma_\nu\sigma_1}$

But $\rho_{1\nu}$ is an opaque term. Let's get back to primitives

Continuing with $E\left[w_1 \,|\nu > z\right]$ to simplify further

$$
\begin{aligned}
E\left[w_1 \,|\nu > z\right] &= \mu_1 + \underbrace{\frac{\sigma_{1\nu}}{\sigma_\nu \sigma_1}}_{=\rho_{1\nu}} \sigma_1 E\left(\frac{\nu}{\sigma_\nu} \middle| \frac{\nu}{\sigma_\nu} > \frac{z}{\sigma_\nu}\right) \\
&= \mu_1 + \frac{\operatorname{cov}\left[\varepsilon_1, \left(\varepsilon_1 - \varepsilon_0\right)\right]\sigma_1}{\sigma_\nu \sigma_1}\left(\frac{\phi\left(z/\sigma_\nu\right)}{\Phi\left(-z/\sigma_\nu\right)}\right) \\
&= \mu_1 + \frac{\sigma_1^2 - \sigma_{01}}{\sigma_\nu}\left(\frac{\phi\left(z/\sigma_\nu\right)}{\Phi\left(-z/\sigma_\nu\right)}\right) \\
&= \mu_1 + \frac{\sigma_0\sigma_1}{\sigma_\nu}\left(\frac{\sigma_1}{\sigma_0} - \frac{\sigma_{01}}{\sigma_0\sigma_1}\right)\left(\frac{\phi\left(z/\sigma_\nu\right)}{\Phi\left(-z/\sigma_\nu\right)}\right) \\
&= \mu_1 + \frac{\sigma_0\sigma_1}{\sigma_\nu}\left(\frac{\sigma_1}{\sigma_0} - \rho_{01}\right)\left(\frac{\phi\left(z/\sigma_\nu\right)}{\Phi\left(-z/\sigma_\nu\right)}\right)
\end{aligned}
$$

where $\rho_{01} = {}^{\sigma_{01}}/_{\sigma_0\sigma_1}$

Calculate expected wages of sector 0 workers identically

Average wages of sector 0 workers

$$
\begin{aligned}
E\left[w_0 \left| \nu < z\right.\right] &= \mu_0 + E\left(\varepsilon_0 \left| \nu < z\right.\right) \\
&= \mu_0 + \sigma_0 E\left(\frac{\varepsilon_0}{\sigma_0} \left| \frac{\nu}{\sigma_\nu} < \frac{z}{\sigma_\nu}\right.\right) \\
&= \mu_0 + \frac{\sigma_{0\nu}}{\sigma_\nu \sigma_0} \sigma_0 E\left(\frac{\nu}{\sigma_\nu} \left| \frac{\nu}{\sigma_\nu} < \frac{z}{\sigma_\nu}\right.\right) \\
&= \mu_0 + \rho_{0\nu}\sigma_0 \left(\frac{-\phi\left(z/\sigma_\nu\right)}{\Phi\left(z/\sigma_\nu\right)}\right) \\
&= \mu_0 + \frac{\sigma_0\sigma_1}{\sigma_\nu}\left(\rho_{01} - \frac{\sigma_0}{\sigma_1}\right)\left(\frac{-\phi\left(z/\sigma_\nu\right)}{\Phi\left(z/\sigma_\nu\right)}\right)
\end{aligned}
$$

- IMR is *negative* in this expression due to right truncation
- We use z not $-z$ in denominator accordingly

Self-selection and wage comparisons between sectors

Observed wages in each sector

$$E\left[w_1 \,|\nu > z\right] = \mu_1 + \frac{\sigma_0\sigma_1}{\sigma_\nu}\left(\frac{\sigma_1}{\sigma_0} - \rho_{01}\right)\left(\frac{\phi\left(z/\sigma_\nu\right)}{\Phi\left(-z/\sigma_\nu\right)}\right)$$

$$E\left[w_0 \,|\nu < z\right] = \mu_0 + \frac{\sigma_0\sigma_1}{\sigma_\nu}\left(\rho_{01} - \frac{\sigma_0}{\sigma_1}\right)\left(\frac{-\phi\left(z/\sigma_\nu\right)}{\Phi\left(z/\sigma_\nu\right)}\right)$$

- Observed wages in each sector are *not* generally equal to latent mean wages, μ_0 or μ_1, due to self-selection
- Comparing *observed* wages between sectors tells you nothing about selection unless you happen to know latent values of μ_0 and μ_1 and ρ_{01}
- Intuition suggests: If workers are positively selected into sector 1, they are negatively selected into sector 0
- *This intuition is misleading*

Agenda

1. **Introduction**
2. **Basics: The normal selection model**
3. **The normal Roy model: selection with two sectors**
   - Two skills, each applicable in only one sector
   - Positive and negative selection
   - Absolute versus comparative advantage
   - Linking comparative to absolute advantage
4. **A first application: Heart attack treatment**
5. **Labor-Roy application: Self-selection and returns to college major**
6. **Conclusions**

Linking comparative to absolute advantage

$$E\left[w_1 \left| \nu > z\right.\right] = \mu_1 + \frac{\sigma_0\sigma_1}{\sigma_\nu}\left(\frac{\sigma_1}{\sigma_0} - \rho_{01}\right)\left(\frac{\phi\left(z/\sigma_\nu\right)}{\Phi\left(-z/\sigma_\nu\right)}\right)$$

$$E\left[w_0 \left| \nu < z\right.\right] = \mu_0 + \frac{\sigma_0\sigma_1}{\sigma_\nu}\left(\rho_{01} - \frac{\sigma_0}{\sigma_1}\right)\left(\frac{-\phi\left(z/\sigma_\nu\right)}{\Phi\left(z/\sigma_\nu\right)}\right)$$

Define positive **self-selection** in sector s as $Q_s = 1$

$$Q_0 = \mathbb{1}E\left[\varepsilon_0 \middle| \text{sector } 0\right] > 0$$

$$Q_1 = \mathbb{1}E\left[\varepsilon_1 \middle| \text{sector } 1\right] > 0$$

The relevant cases

Selection equations

$$E\left[w_1 \,|\nu > z\right] = \mu_1 + \frac{\sigma_0\sigma_1}{\sigma_\nu}\left(\frac{\sigma_1}{\sigma_0} - \rho_{01}\right)\left(\frac{\phi\left(z/\sigma_\nu\right)}{\Phi\left(-z/\sigma_\nu\right)}\right)$$

$$E\left[w_0 \,|\nu < z\right] = \mu_0 + \frac{\sigma_0\sigma_1}{\sigma_\nu}\left(\rho_{01} - \frac{\sigma_0}{\sigma_1}\right)\left(\frac{-\phi\left(z/\sigma_\nu\right)}{\Phi\left(z/\sigma_\nu\right)}\right)$$

1. **Comparative advantage and absolute advantage align** $Q_0 = Q_1 = 1$

$$\rho_{01} < \min\left\{\frac{\sigma_1}{\sigma_0}, \frac{\sigma_0}{\sigma_1}\right\}$$

2. $(+)$ **selection in sector** 1**,** $(-)$ **in sector** 0**:** $Q_0 = 0$, $Q_1 = 1$

$$\frac{\sigma_1}{\sigma_0} > 1 \text{ and } \rho_{01} > \frac{\sigma_0}{\sigma_1}$$

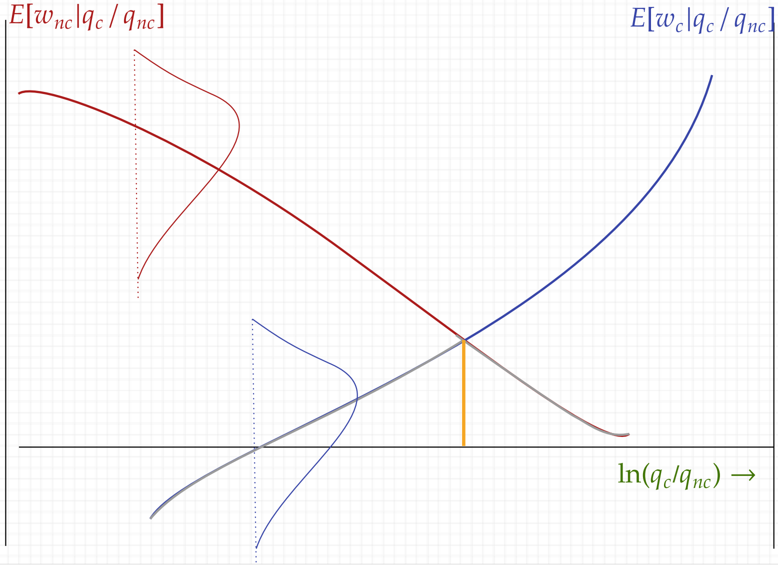
3. $(-)$ **selection in sector** 1**,** $(+)$ **in sector** 0**:** $Q_0 = 1$, $Q_1 = 0$

$$\frac{\sigma_0}{\sigma_1} > 1 \text{ and } \rho_{01} > \frac{\sigma_1}{\sigma_0}$$

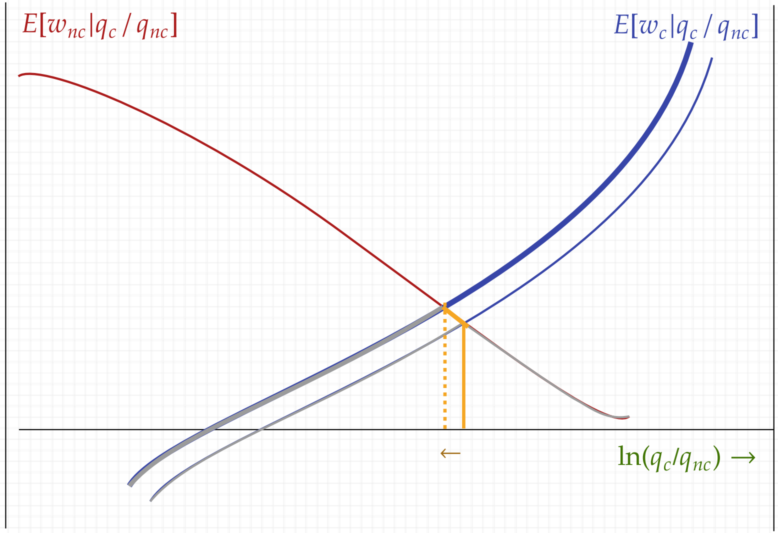
Can we have negative self-selection on absolute advantage in both sectors?

- **A fourth case? Can we have** $Q_0 = Q_1 = 0$?
  - **Nope**. This would require $\left(\rho_{01} > \frac{\sigma_1}{\sigma_0} \text{ and } \rho_{01} > \frac{\sigma_0}{\sigma_1}\right) \Rightarrow \rho_{01} > 1$. That's not feasible: $\rho_{01} \in [-1, 1]$
  - Inconsistent with income maximization: workers with absolute advantage in each sector choose the opposite sector
  - In at least one sector, comparative and absolute advantage must align

Positive Selection into College and Non-College



Positive Selection into College and Non-College



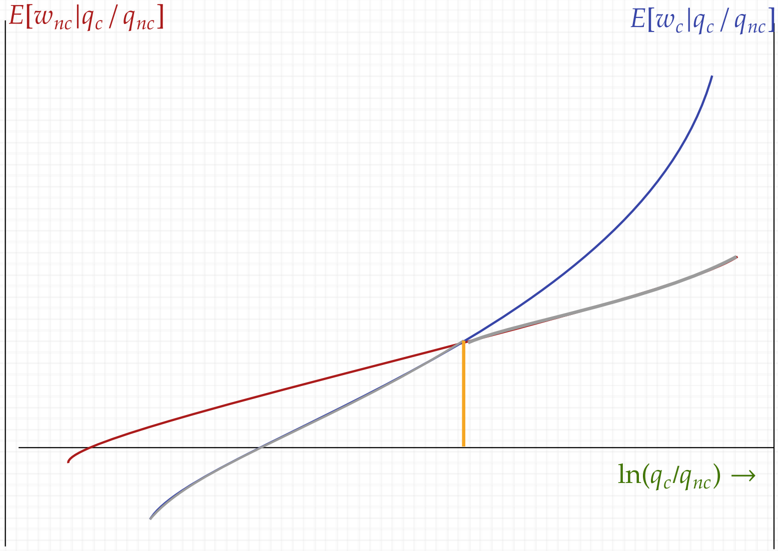
Effect of rise in the return to 'college' on selection

Example: Two States, College and Non-College

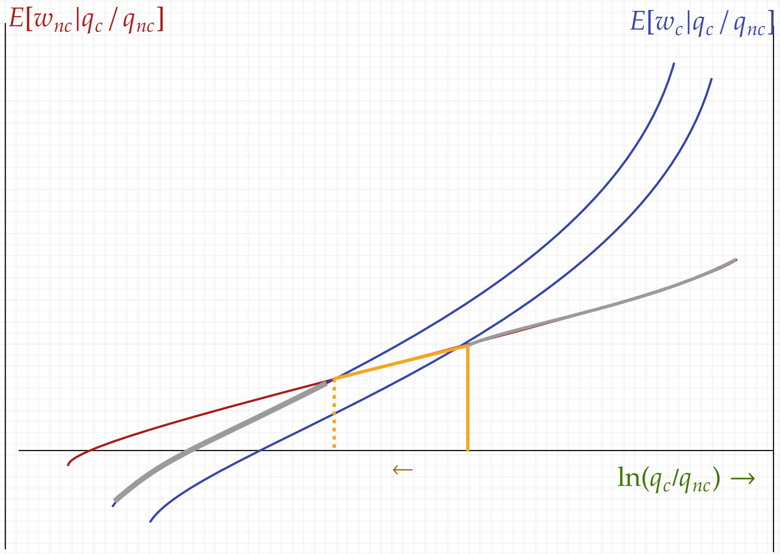
Effect of a Rise in Return to "College Skill"

	$\Delta E(Q_c \mid \text{College})$	$\Delta E(Q_{nc} \mid \text{Non-college})$	$E(W_c)$	$E(W_{nc})$
+ select college + select non-coll	< 0	> 0	≥ 0	> 0
+ select college – select non-coll				
– select college + select non-coll				

Positive Selection into College, Neg Selection into Non-College



Positive Selection into College, Neg Selection into Non-College



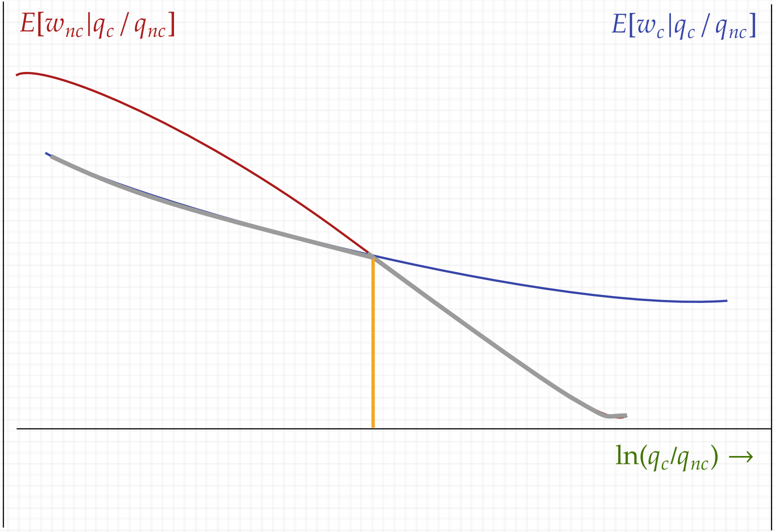
Effect of rise in the return to 'college' on selection

Example: Two States, College and Non-College

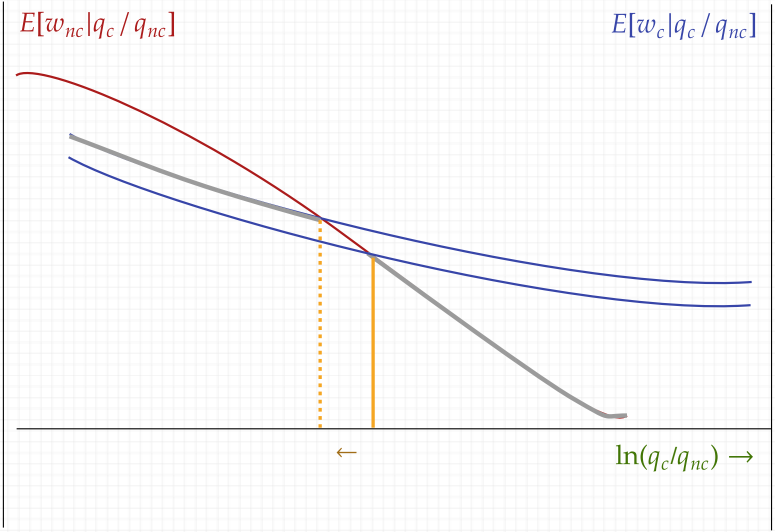
Effect of a Rise in Return to "College Skill"

	$\Delta E(Q_c \mid \text{College})$	$\Delta E(Q_{nc} \mid \text{Non-college})$	$E(W_c)$	$E(W_{nc})$
$+$ select college $+$ select non-coll	< 0	> 0	≥ 0	> 0
- select college $-$ select non-coll	< 0	< 0	$<> 0$	< 0
$-$ select college $+$ select non-coll				

Negative Selection into College, Pos Selection into Non-College



Negative Selection into College, Pos Selection into Non-College



Effect of rise in the return to 'college' on selection

Example: Two States, College and Non-College

Effect of a Rise in Return to "College Skill"

	$\Delta E(Q_c \mid \text{College})$	$\Delta E(Q_{nc} \mid \text{Non-college})$	$E(W_c)$	$E(W_{nc})$
+ select college + select non-coll	< 0	> 0	≥ 0	> 0
+ select college − select non-coll	< 0	< 0	$<> 0$	< 0
− select college + select non-coll	> 0	> 0	> 0	> 0

Jump back to equations

Agenda

1. **Introduction**
2. **Basics: The normal selection model**
3. **The normal Roy model: selection with two sectors**
   - Two skills, each applicable in only one sector
   - Positive and negative selection
   - Absolute versus comparative advantage
   - Linking comparative to absolute advantage
4. **A first application: Heart attack treatment**
5. **Labor-Roy application: Self-selection and returns to college major**
6. **Conclusions**

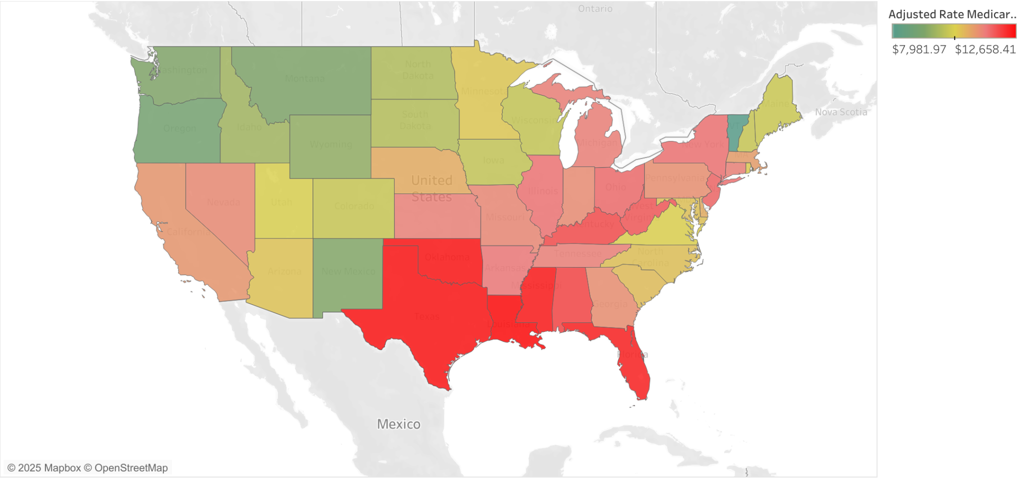
Startling facts about US healthcare: Chandra-Staiger ’07

1. Huge geographic variation in the use of intensive treatments (surgical or technology-intensive regimens) for specific diseases
2. Use of more intensive procedures *is* associated with significantly higher costs
3. Use of more intensive procedures *not* associated with improved satisfaction, outcomes, or survival

Demographic & price adjusted Medicare expenditure per enrollee 2019: States

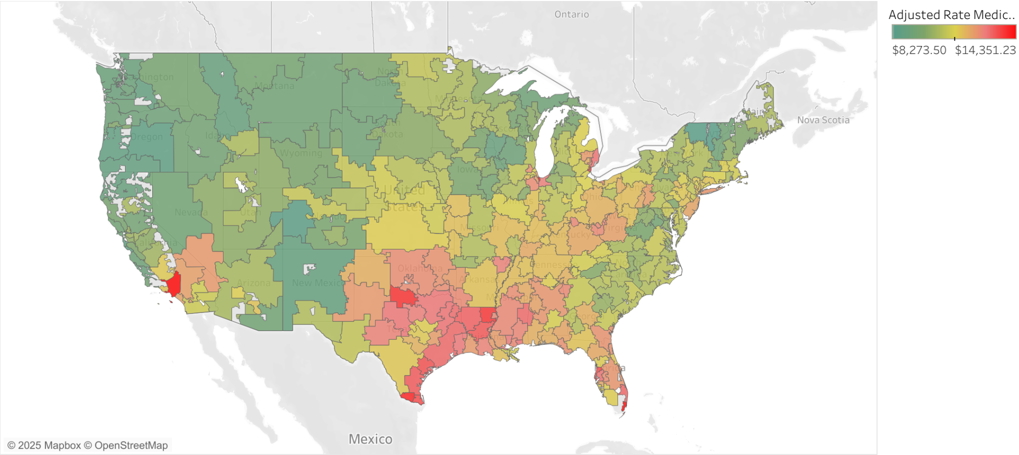
Map: Price-Adjusted Total Medicare Reimbursements per Enrollee (Parts A and B), by State (2019)

(Price, Age, Sex, and Race adjusted)



Demographic & price adjusted Medicare expenditure per enrollee 2019: HRRs

Map: Price-Adjusted Total Medicare Reimbursements per Enrollee (Parts A and B), by HRR (2019)
(Price, Age, Sex, and Race adjusted)



Startling facts about US healthcare: Chandra-Staiger '07

1. Huge geographic variation in the use of intensive treatments (surgical or technology-intensive regimens) for specific diseases

2. Use of more intensive procedures *is* associated with significantly higher costs

3. Use of more intensive procedures *not* associated with improved satisfaction, outcomes, or survival

4. Yet, much evidence finds that the marginal benefit of more intensive treatment *increases* where it is used more frequently – suggesting learning by doing

How do we reconcile these facts?

What's going on?

1. **Flat of the curve (FoC) medicine**
   - FoC predicts that marginal benefit from more intensive patient treatment should be lower in areas that are more aggressive
   - But areas that do more of a treatment are almost always better at it
   - Still puzzling that *average* outcome is no better
   - More intensive treatment should still have positive effects *on average*

What's going on?

1. Flat of the curve medicine
2. **Treatment intensity reflects comparative advantage**
   - Hospitals that are good at high intensity treatments *do more of them*
   - For *inframarginal* patients—those most suitable for intensive treatment—benefits are very high
   - But for *marginal* patients—those least suited for intensive treatment—benefits may be negative
   - Least suited patients could be better off at hospital specializing in non-intensive treatment
   - Even so, hospitals maximizing patient outcomes *given* their comparative advantage
   - Key mechanism: 'productivity spillovers' or 'increasing returns' or 'learning by doing' through specialization

Empirical focus: Treatment of acute myocardial infarction

Treatment of acute myocardial infarction (AMI = heart attack)

1. Cardiovascular disease leading cause of death in U.S.
   - One year survival rate is 70 percent
2. Markets for heart attack treatment are geographically distinct
   - You go to the nearest hospital
3. Two potential interventions
   - Intensive (surgical) management: angioplasty or bypass
   - Non-intensive (medial) management: fibrinolytics therapy (thrombolytics), which is a drug treatment rather than surgery
4. No region can fully specialize
   - Intensive intervention *risky* for frail and elderly
   - Thus, both treatments observed in all markets
5. **Key idea: Learning-by-doing spillovers**. The more that regions do intensive (or non-intensive) management, the better they get at it

A Roy model of heart attack treatment: Setup

Heart attack patient's survival probability and treatment cost if given treatment $i \in \{0, 1\}$

$$\text{Survival}_i = \beta_i^s Z + \alpha_i^s P_i + \epsilon_i^s \text{ for } i = 1, 2$$

$$\text{Cost}_i = \beta_i^c Z + \alpha_i^c P_i + \epsilon_i^c \text{ for } i = 1, 2$$

$$U_i = \text{Survival}_i - \lambda \text{Cost}_i = \beta_i Z + \alpha_i P_i + \epsilon_i \text{ for } i = 1, 2$$

- i refers to the two treatment regimes (not patients): medical and surgical management
- $\beta_i \equiv \beta_i^s - \lambda \beta_i^c$, and similarly for α_i and ϵ_i
- $\beta_i Z$ is a measure of appropriateness of treatment regime i for a patient with characteristics Z
- P_i is the fraction of patients in a region receiving treatment i
- α_i is a measure of **productivity spillovers**. If $\alpha_i > 0$, the more a region does procedure i the better it is at this procedure (in equilibrium, P_i will depend on α_i)

A Roy model of heart attack treatment: Setup

Treatment decision – Hospital makes best choice given its capabilities

- Let $i = 2$ denote intensive treatment, and $P_2 = 1 - P_1$

$$\begin{aligned}
\Pr\{\text{intensive treatment}\} &= \Pr\{i = 2\} \\
&= \Pr\{U_2 - U_1 > 0\} \\
&= \Pr\{(\alpha_1 + \alpha_2) P_2 - \alpha_1 + (\beta_2 - \beta_1) Z > (\epsilon_1 - \epsilon_2)\} \\
&= \Pr\{\alpha P_2 - \alpha_1 + \beta Z > \epsilon\}
\end{aligned}$$

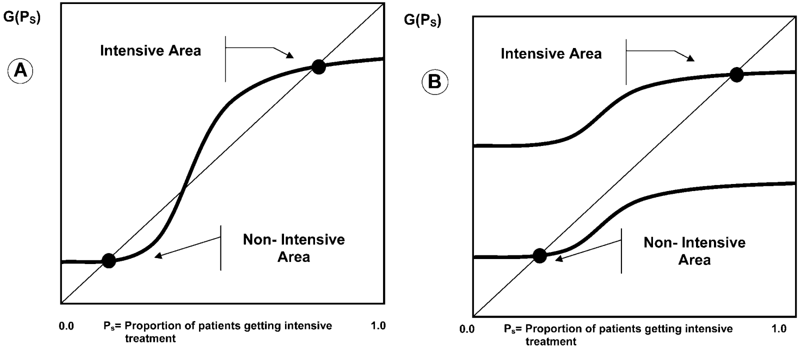
- Among patients who get the intensive treatment, expected utility gain is

$$E[U_2 - U_1 | U_2 - U_1 > 0] = \beta Z + \alpha P_2 - \alpha_1 + E[\epsilon | U_2 - U_1 > 0]$$

- Eqm is **fixed point** where probability of choosing P_2 equals fraction of patients receiving P_2

$$\begin{aligned}
P_2 &= \int_Z \Pr(\alpha P_2 - \alpha_1 + \beta Z > \epsilon) f(Z)\, dZ \\
&\equiv G(P_2)
\end{aligned}$$

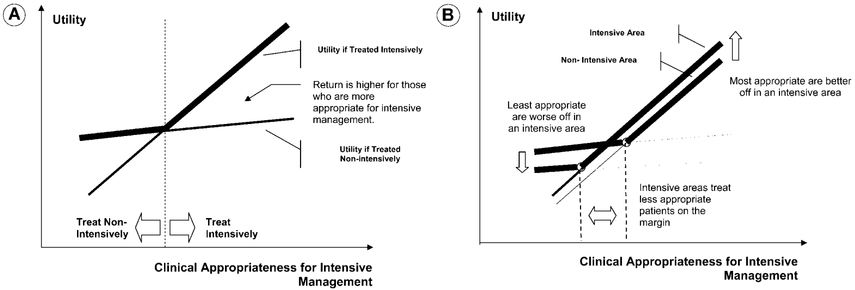
Two scenarios for patterns of specialization by region



Market demand curves for intensive treatment

1. Left: Multiple equilibria due to productivity spillovers
2. Right: Underlying differences in patient population suitability for intensive treatment leads to eq'm specialization that magnifies these diffs

Informal Roy framework for regional specialization in intensive treatment



- Left: Comparative advantage without increasing returns
- Right: Comparative advantage with increasing returns to specialization
- Notice that threshold patient treated appropriately in both regions, but threshold patient may be better off in less-intensive region

Implications of a Roy framework with spillovers

Implications

1. Identical patients will be more likely to be treated intensively in regions that do more of that treatment
2. The benefit to receiving intensive treatment in intensive regions is larger than the benefit in non-intensive regions
3. Outcomes for non-intensive management are worse in regions that are intensive
4. Marginal patients patients receiving intensive treatment in intensive regions are *less appropriate* than marginal patients receiving intensive treatment in non-intensive regions

Implications of a Roy framework with spillovers

Empirical approach

- Unit of analysis is Hospital Referral Region (HRR)
- Identification of causal effect of HRR intensity exploits *differential distance* to HRR by residential zipcode

Balance test and reduced form comparisons

No difference in suitability among patients 'assigned' by DD to low versus high-cath regions (col 4 - 8), but major diffs in treatment and survival (cols 1 - 4): *Low DD → close to high-cath hospital*

TABLE 2

Relationship between Differential Distance (DD) and Probability of Catheterization and Survival, and Differential Distance and Observable Characteristics (%)

Sample	One-Year Predicted Survival		30-Day Predicted Cath Rate for Patients Getting Cath	
	DD Below Median (5)	DD Above Median (6)	DD Below Median (7)	DD Above Median (8)
All patients (N = 129,997)	67.5	67.2	63.3	63.2
By cath propensity:				
Above the median (N = 64,733)	83.4	83.5	72.6	72.6
Below the median (N = 65,244)	51.1	51.6	32.3	32.5
By age:				
65–80 (N = 90,016)	73.9	73.9	67.4	67.3
Over 80 (N = 39,961)	52.6	52.7	34.6	34.1

Note.—Cath propensity is an empirical measure of patient appropriateness for intensive treatments. We define this measure by using fitted values from a logit model of the receipt of cardiac catheterization on all the CCP risk adjusters. Differential distance is measured as the distance between the patient's zip code of residence and the nearest catheterization hospital minus the distance to the nearest hospital.

Balance test and reduced form comparisons

No difference in suitability among patients 'assigned' by DD to low versus high-cath regions (col 4 - 8), **But** major diffs in treatment and survival (cols 1 - 4): *Low DD → close to high-cath hospital*

TABLE 2

Relationship between Differential Distance (DD) and Probability of Catheterization and Survival, and Differential Distance and Observable Characteristics (%)

	30-Day Cath Rate		One-Year Survival		One-Year Predicted Survival		30-Day Predicted Cath Rate for Patients Getting Cath	
Sample	DD Below Median (1)	DD Above Median (2)	DD Below Median (3)	DD Above Median (4)	DD Below Median (5)	DD Above Median (6)	DD Below Median (7)	DD Above Median (8)
All patients (N = 129,997)	48.9	42.8	67.6	66.7	67.5	67.2	63.3	63.2
By cath propensity:								
Above the median (N = 64,733)	74.0	67.1	84.6	83.8	83.4	83.5	72.6	72.6
Below the median (N = 65,244)	22.9	19.5	50.1	50.4	51.1	51.6	32.3	32.5
By age:								
65–80 (N = 90,016)	61.1	54.9	74.3	73.5	73.9	73.9	67.4	67.3
Over 80 (N = 39,961)	20.3	16.5	52.1	52.1	52.6	52.7	34.6	34.1

Note.—Cath propensity is an empirical measure of patient appropriateness for intensive treatments. We define this measure by using fitted values from a logit model of the receipt of cardiac catheterization on all the CCP risk adjusters. Differential distance is measured as the distance between the patient's zip code of residence and the nearest catheterization hospital minus the distance to the nearest hospital.

Using 'differential-distance' IV to identify effect of intensive-treat × suitability

Among patients suited for intensive treatment,
intensive treatment more beneficial in more intensive regions

TABLE 1
INSTRUMENTAL VARIABLE ESTIMATES OF INTENSIVE MANAGEMENT AND SPENDING ON ONE-YEAR SURVIVAL BY CLINICAL APPROPRIATENESS OF PATIENT

	Instrumental Variable Estimates of		
	Impact of Cath		
Sample	On One-Year Survival (1)	On One-Year Cost ($1,000s) (2)	Impact of $1,000 on One-Year Survival (3)
A. All patients (N = 129,895)	.142	9.086	.016
	(.036)	(1.810)	(.005)
B. By cath propensity:			
Above the median (N = 64,799)	.184	4.793	.038
	(.034)	(1.997)	(.017)
Below the median (N = 65,096)	.035	17.183	.002
	(.083)	(3.204)	(.005)
Difference	.149	−12.39	.036
	(.090)	(3.775)	(.018)
C. By age:			
65–80 (N = 89,947)	.171	6.993	.024
	(.037)	(1.993)	(.009)
Over 80 (N = 39,948)	.016	16.026	.001
	(.108)	(2.967)	(.007)
Difference	.155	−9.033	.023
	(.114)	(3.574)	(.011)

Comparing marginal versus average patient in high vs. low referral regions

- Let A_{ij} equal average suitability of a patients who receive surgical treatment in HRR j
- Regress suitability of patients treated surgically on number (or rate) treated surgically

$$A_{ij} = \mu_0 + \mu_1 \ln \left(\text{Cath Rate}\right)_j + e_i$$

$$E\left[A_{ij}\right] = \frac{\sum_i A_{ij}}{N_j}$$

$$\frac{\partial E\left[A_{ij}\right]}{\partial \ln N} = \frac{\partial \left(A/N\right)}{\partial \ln N} = N\partial(A/N)/\partial N = N\left(\frac{N\frac{\partial A}{\partial N} - A}{N^2}\right)$$

$$\rightarrow \hat{\mu}_1 = \frac{\partial A}{\partial N} - \frac{A}{N}$$

- Thus, log derivative estimates difference between marginal and average
- if $\hat{\mu}_1 < 0$, implies that marginal patient is less appropriate than average

Comparing marginal versus average patient in high vs. low referral regions

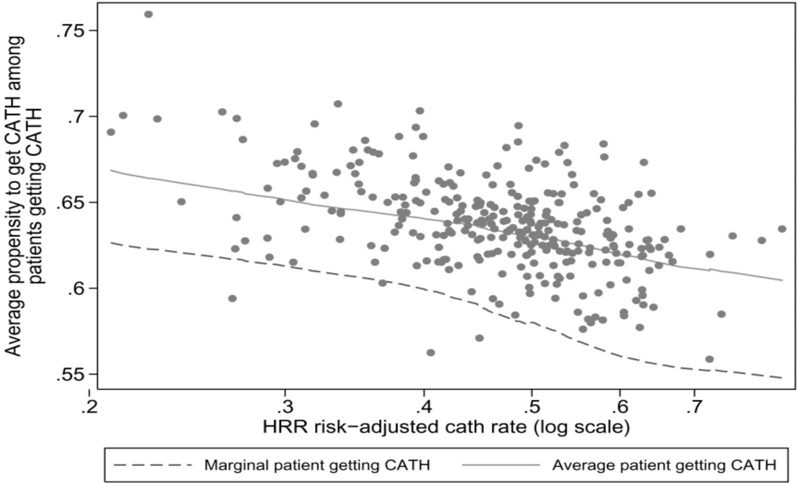
Log derivative identifies difference between marginal and average suitability of patients receiving treatment as f'n of number (or fraction) treated intensively

TABLE 3
RELATIONSHIP BETWEEN THE AVERAGE AND MARGINAL PATIENT RECEIVING CARDIAC CATHETERIZATION ($N = 303$)

Patient Characteristic	Characteristic of Average Patient Getting Cath across All Areas (1)	Difference between Marginal Patient and Average Patient Getting Cath in Higher-Cath HRRs (2)
Cath propensity	.633	−.045
	(.002)	(.008)
Over age 80	.125	.063
	(.002)	(.012)
Not eligible for cath using ACC/AHA guidelines	.028	.010
	(.001)	(.003)

NOTE.—Cath propensity is an empirical measure of patient appropriateness for intensive treatments. We define this measure by using fitted values from a logit model of the receipt of cardiac catheterization on all the CCP risk adjusters. The sample is restricted to patients receiving cardiac catheterization within 30 days of an AMI. ACC/AHA guidelines reflect a binary variable assigned to each patient in the CCP that measures whether the patient is ideal, appropriate, or not eligible for catheterization on the basis of a review of the patient's chart.

Comparing marginal versus average patient cath-suitability as a function of hospitals' risk-adjusted cath rates



Patients most suited for cath benefit most from treatment in high-cath regions. Patients least suited for cath who receive cath in low-cath regions fare worst

TABLE 6
INSTRUMENTAL VARIABLE ESTIMATES OF INTENSIVE MANAGEMENT AND SPENDING ON SURVIVAL, BY SURGICAL INTENSITY OF HOSPITAL REFERRAL REGION

	Instrumental Variable Estimates of		
	Impact of Cath		
Sample	On One-Year Survival (1)	On One-Year Cost ($1,000s) (2)	Impact of $1,000 on One-Year Survival (3)
A. All patients:			
HRR risk-adjusted cath rate:			
Above the median ($N =$ 63,771)	.256	6.691	.038
	(.061)	(3.510)	(.021)
Below the median ($N =$ 66,124)	.09	9.835	.009
	(.059)	(3.155)	(.007)
Difference	.166	−3.144	.029
	(.085)	(4.720)	(.022)
B. Patients above the median cath propensity:			
HRR risk-adjusted cath rate:			
Above the median ($N =$ 32,388)	.271	.347	.78
	(.064)	(4.370)	(9.820)
Below the median ($N =$ 32,411)	.168	4.962	.034
	(.046)	(2.890)	(.021)
C. Patients below the median cath propensity:			
HRR risk-adjusted cath rate:			
Above the median ($N =$ 31,383)	.206	16.21	.013
	(.129)	(5.130)	(.009)
Below the median ($N =$ 33,713)	−.139	22.064	−.006
	(.165)	(6.870)	(.007)

Survival rates and costs of intensive treatment differ systematically with HRR risk-adjusted intensity rates

Area intensity predicts higher costs but not improved outcomes among patients as a whole.
But (1) patients appropriate for intensive treatment *benefit* from treatment in high intensity areas; (2) patients least appropriate for intensive treatment are *harmed*

TABLE 7
Relationship between HRR Catheterization Rate, Survival, and Costs, by Clinical Appropriateness for Intensive Management

	OLS Estimates of the Relationship between HRR Risk-Adjusted Cath Rate and			
Sample	One-Year Survival (1)	One-Year Cost ($1,000s) (2)	Beta-Blocker in Hospital (3)	Catheterization within 30 Days (4)
A. All patients (N = 138,873)	.007	8.093	−.28	.702
	(.019)	(1.410)	(.073)	(.004)
B. By cath propensity:				
Top tercile (N = 46,287)	.052	10.012	−.366	.802
	(.019)	(1.439)	(.073)	(.032)
Middle tercile (N = 46,295)	.03	11.154	−.271	.906
	(.030)	(1.784)	(.082)	(.021)
Bottom tercile (N = 46,291)	−.075	2.763	−.209	.369
	(.028)	(1.612)	(.073)	(.021)
Difference (top − bottom)	.127	7.249	−.157	.433
	(.034)	(2.161)	(.103)	(.038)

Summary: Implications of a Roy Framework with spillovers

1. Identical patients will be more likely to be treated intensively in regions that do more of that treatment
2. The benefit to receiving intensive treatment in intensive regions is larger than the benefit in non-intensive regions
3. Outcomes for non-intensive management are worse in regions that are intensive
4. Marginal patients patients receiving intensive treatment in intensive regions are *less appropriate* than marginal patients receiving intensive treatment in non-intensive regions

Agenda

What are the ‘returns’ to field of study?

Norway’s centralized admissions cover almost all post-secondary schools

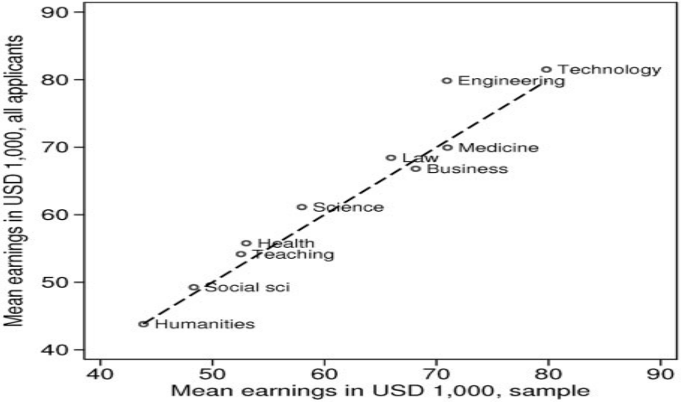
- Prospective students apply to a field and institution simultaneously (e.g., teaching, at the University of Oslo)
- Applicants can rank up to 15 choices
- They are scored by central organization based on high school grade point average (GPA)
- Applicants ranked by application score:
  - Highest-ranked applicant gets her preferred choice
  - Next-highest ranked applicant gets the highest available choice remaining, and so on...

Do 'returns' to field positively covary with preference? They should if prefs reflect comparative advantage in earnings

What is special about this setting

- We know the choice each applicant receives *and* their next best alternative
- For example
  - A prefers medicine to law and is assigned to law (2^{nd} choice)
  - B prefers law to medicine and is *also* assigned to law (1^{st} choice)
- Is the payoff to law versus medicine comparable for A and B?
  - Roy model says *probably not*
  - Expectation: payoff to law vs. medicine is *higher* for B (1^{st} choice law) than for A (1^{st} choice medicine)

Very large earnings diffs by field of study (as expected)



(a) By field

FIGURE II

Mean Earnings by Completed Field and Institution: Sample and All Applicants

First stage: Ranking, threshold, field of study, degree completion

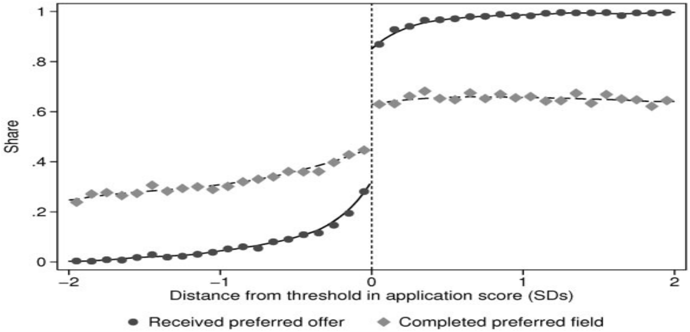


FIGURE III

Admission Thresholds and Preferred Field Offer and Completion

No evidence of manipulation of running variable

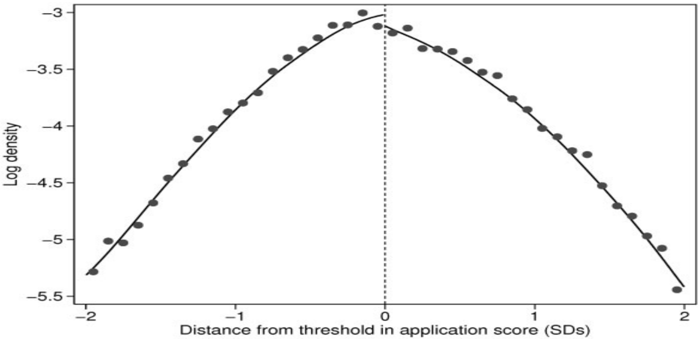


FIGURE IV

Bunching Check around the Admission Cutoffs

Appear to be large field effects conditional on preference. But preference not synonymous with income maximization

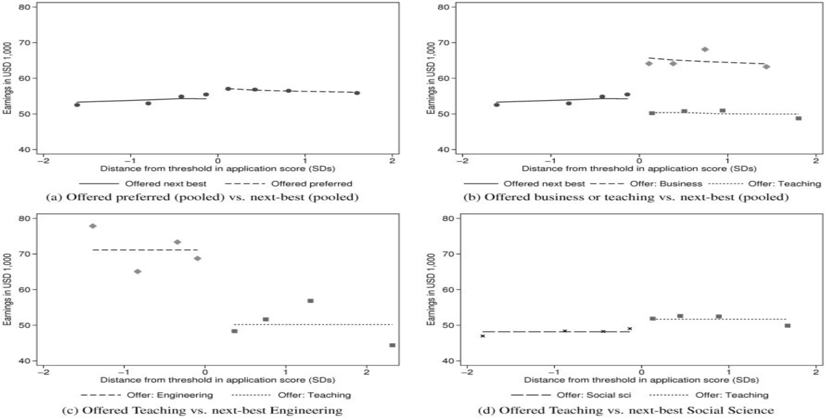


FIGURE VI

Average Earnings around Admission Cutoffs

Counterfactual fields of study for compliers (shares)

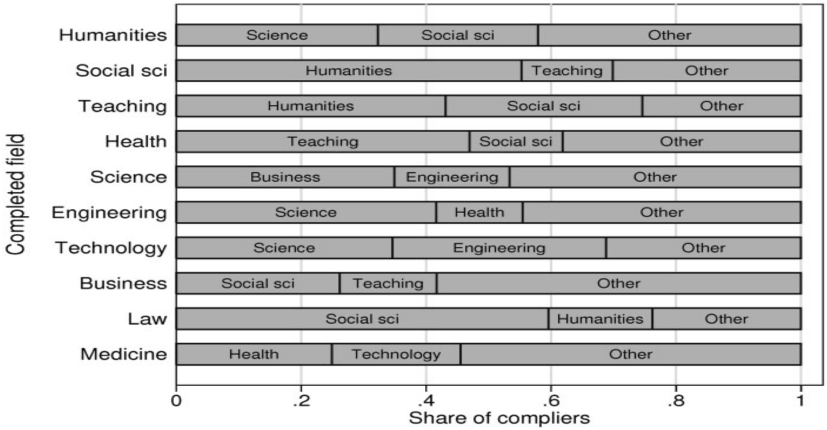


FIGURE VII

Complier Weights of Alternative Fields by Completed Fields

Payoff to field of study–relative to alternatives

TABLE IV

2SLS ESTIMATES OF THE PAYOFFS TO FIELD OF STUDY (USD 1,000)

	Next Best Alternative (k)								
	Humanities	Soc Science	Teaching	Health	Science	Engineering	Technology	Business	Law
Completed field (j)									
Humanities		21.4*	−4.7	−22.9*	5.0	−38.5**	6.9	−42.2**	−156.3
		(11.0)	(9.8)	(12.1)	(11.9)	(14.7)	(48.3)	(10.6)	(437.3)
Social Science	18.7**		9.8	−10.8	55.5**	−55.4**	−110.4	−28.4**	−76.1
	(6.7)		(11.6)	(13.0)	(21.5)	(20.6)	(103.0)	(10.7)	(86.4)
Teaching	22.3**	31.4**		1.8	23.5**	−33.9**	−35.3	−21.1**	22.8
	(5.0)	(7.9)		(6.6)	(9.5)	(12.5)	(37.1)	(7.1)	(127.9)
Health	18.8**	30.7**	7.7**		28.9**	−27.9**	−43.4**	−17.4**	−55.2
	(6.3)	(7.6)	(2.8)		(7.6)	(10.4)	(20.8)	(4.0)	(97.7)
Science	53.7**	69.6**	38.6**	29.6**		−2.2	16.8	−4.9	148.3
	(18.4)	(22.4)	(14.2)	(11.5)		(14.6)	(18.1)	(10.5)	(276.2)
Engineering	59.8	−5.5	75.2**	0.2	52.4**		−46.0	−13.0	−57.7
	(50.6)	(58.2)	(37.5)	(16.4)	(21.0)		(43.9)	(23.7)	(166.6)
Technology	41.9**	58.7**	22.1*	32.5**	68.1**	−5.6		7.0	−53.1
	(10.8)	(10.1)	(12.4)	(10.1)	(9.6)	(12.0)		(9.5)	(147.5)

Payoff to field of study–relative to alternatives (part II

TABLE IV

(CONTINUED)

	Next Best Alternative (k)								
	Humanities	Soc Science	Teaching	Health	Science	Engineering	Technology	Business	Law
Business	48.1**	61.9**	31.0**	30.2**	58.0**	−3.4	28.5*		3.5
	(11.3)	(12.0)	(8.8)	(10.9)	(10.5)	(12.6)	(15.6)		(83.0)
Law	46.3**	55.6**	36.6**	21.5*	40.1**	−27.5	−15.6	−1.4	
	(7.2)	(8.3)	(11.6)	(11.5)	(9.7)	(18.3)	(18.0)	(8.7)	
Medicine	83.3**	79.4**	62.6**	45.6**	81.3**	21.1	40.1**	23.3**	14.8
	(9.8)	(10.7)	(9.0)	(7.0)	(9.7)	(20.7)	(11.7)	(8.8)	(83.6)
Female	−7.0**	−6.3**	−10.3*	−5.6**	−5.3**	−5.1**	−4.1**	−7.0**	−10.6
	(1.1)	(1.6)	(1.3)	(0.9)	(1.3)	(1.0)	(1.6)	(3.5)	(6.9)
Application score	−0.6	4.3**	4.0**	1.6**	−0.7	1.1*	−0.1	0.1	13.8
	(0.8)	(1.6)	(0.9)	(0.6)	(0.7)	(0.6)	(1.3)	(2.8)	(14.6)
Average y^k	30.0	23.4	46.2	51.8	27.3	87.9	78.4	75.6	105.8
Observations	8,391	11,030	10,987	3,269	6,422	3,085	1,245	4,403	1,251

Notes. From 2SLS estimation of equations (14)–(15), we obtain a matrix of the payoffs to field j as compared to k for those who prefer j and have k as next-best field. Each cell is a 2SLS estimate (with standard errors in parenthesis) of the payoff to a given pair of preferred field and next-best field. The rows represent completed fields and the columns represent next-best fields. The row labeled average y^k reports the weighted average of the levels of potential earnings for compliers in the given next-best field. The final row reports the number of observations for every next-best field. Stars indicate statistical significance, at the *10% level and **5% level.

Average estimated payoff by completed field relative to next best alternative

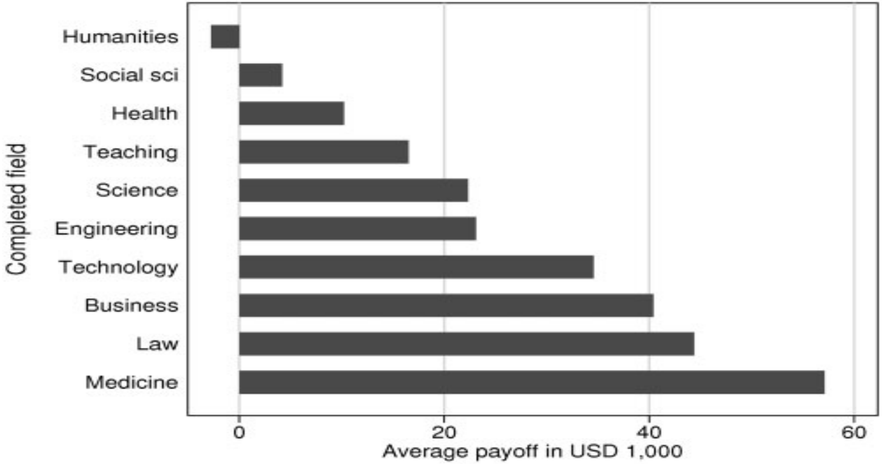


FIGURE IX

Average Estimated Payoffs ($1,000) by Completed Field

Implications of comparative advantage

Strong implication of selection on comparative advantage

- Person 1 has comparative advantage in field j relative to k
- Person 2 has comparative advantage in field k relative to j

$$\frac{Y_1^j}{Y_1^k} > \frac{Y_2^j}{Y_2^k} \Longleftrightarrow \ln Y_1^j - \ln Y_1^k > \ln Y_2^j - \ln Y_2^k$$

- Empirical analogue

$$E\left[\ln Y_i^j - \ln Y_i^k | j \succ k\right] > E\left[\ln Y_i^j - \ln Y_i^k | k \succ j\right]$$

Payoff to field j vs. k higher among applicants with $j \succ k$, lower among applicants with $k \succ j$

Most compliers earn more in their preferred field than they would have earned in next-best alternative

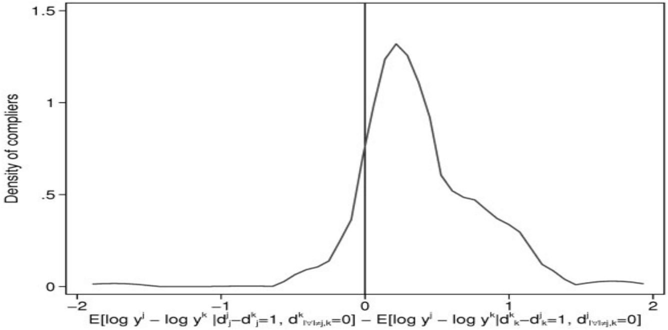


FIGURE XII

Testable Implication of Sorting Based on Comparative Advantage

This figure graphs the distribution of the differences in relative payoffs to field j versus k between individuals whose preferred field is j and next-best alternative is k, and those with the reverse ranking. To construct this graph, we use the complier-weighted distribution of estimates in Online Appendix Table B.VI.

Agenda

Comparative advantage a crucial concept in labor economics — and economics more generally

- In international economics, comparative advantage is why **gains from trade** arise
- In labor economics, there's a different implication: comparative advantage implies that the **mapping between earnings and 'ability' is not one-to-one**
- Comparative advantage + self-selection (or 'assignment') **often magnifies the dispersion of wages relative to the dispersion of underlying ability**
- We will talk about this vis-a-vis Sattinger '75 when we get to 'superstars' shortly