Self-selection: The Roy model

Heidi L. Williams

MIT 14.662

Spring 2015
1. Roy model application #1: Health care

2. Roy model application #2: Redistribution

3. Looking ahead
1. Roy model application #1: Health care

2. Roy model application #2: Redistribution

3. Looking ahead
Despite its origin in labor economics, the Roy model has been applied across a wide range of fields in economics.

Chandra and Staiger (2007): one of the most important papers in health economics in recent years, and one that has really changed how people think about a variety of issues.
Geographic variation in medical expenditures

- Earliest work I’m aware of: Glover (1938)
  - England and Wales: variation in small-area tonsillectomy rates
  - Looked for correlations with “any factor which might have some ætiological bearing on chronic tonsillitis and adenoidal growths - such factors for example as overcrowding...not the slightest suggestion of correlation has been obtained.”
  - Maybe not the regressions we would estimate, but the start of a puzzle!

- Skinner (2012) *Handbook* chapter provides overview of this literature
  - Adjusting for prices doesn’t really matter (Gottlieb *et al.* 2010)
  - Debate over relative importance of supply vs. demand
Geographic variation in medical expenditures (continued)

- “Fact” #1: geographic variation in health spending is not associated with improved satisfaction, outcomes, or survival
  - Consensus view from Dartmouth Atlas
  - Caveats: Cutler (2005), Joe Doyle’s line of research

- Surprising in light of Fact #2: many technologies shown to be associated with improved survival in randomized clinical trials

- Facts #1, 2 often reconciled by “flat of the curve” argument
  - RCTs run on patients most likely to benefit
  - Physicians may treat until marginal return is zero
Three problems with “flat of the curve” argument

1. No explanation of why we observe geographic variation
2. Still predicts positive relationship between spending, outcomes unless all areas in range of zero or negative marginal benefits
   ▶ Has never been documented in the literature
3. Predicts marginal benefit from more intensive treatment should be lower in areas that treat more aggressively
   ▶ Available US-Canada comparisons suggest the opposite: US treats heart attacks more intensively, yet marginal benefit from intensive heart attack treatments appears to be larger in US.
Chandra and Staiger (2007)

Chandra and Staiger present a Roy model with productivity spillovers that can reconcile these facts.

Their paper is an excellent illustration of how a set of facts can motivate a (relatively simple) theoretical framework producing testable implications that can then be taken back to the data.
What motivates a model with productivity spillovers?

- Patients receive either of two treatments:
  1. Nonintensive management (medical management; subscript 1)
  2. Intensive intervention (surgery; subscript 2)
- Physicians choose treatment to maximize utility over expected survival ($\text{Survival}_1$, $\text{Survival}_2$) and cost ($\text{Cost}_1$, $\text{Cost}_2$)
- Productivity spillovers: survival, cost positively related to share of patients receiving same treatment ($P_1$, $P_2 = 1 - P_1$)
  - As in Katz and Shapiro’s (1985) model of network externalities (telephones, hardware-software, foreign auto firms)
What motivates a model with productivity spillovers?

Why would this productivity spillovers assumption be plausible? Chandra and Staiger focus on three possible explanations:

1. **Knowledge spillovers.** Physicians may learn about new surgical techniques and procedures from direct contact with other physicians (“see one, do one, teach one”)

2. **Availability of support services.** Some places have cardiac catheterization labs whereas other don’t (choice variable)

3. **Selective migration.** Physicians more skilled at the intensive treatment may self-select into areas that treat more intensively
Model

To the basic framework outlined above, add heterogeneity across patients that affects expected survival and cost

- Some heterogeneity captured by observable characteristics (Z)
- Other factors ($\epsilon$) known to patient and physician at the time of choosing treatment, but not observed by econometrician

This is the Roy model component of the model: patients are sorted into the two treatments based on expected returns
Model (continued)

For treatments \( i \in \{\text{nonintensive, intensive}\} \), denote the survival rate and cost for each treatment as:

\[
\text{Survival}_i = \beta_i^s Z + \alpha_i^s P_i + \epsilon_i^s \quad \text{for } i = 1, 2
\]
\[
\text{Cost}_i = \beta_i^c Z + \alpha_i^c P_i + \epsilon_i^c \quad \text{for } i = 1, 2
\]

Denoting value of life (survival) by \( \lambda \), patient’s indirect utility \( U \) is:

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

where \( \beta_i = \beta_i^s - \lambda \beta_i^c \), \( \alpha_i = \alpha_i^s - \lambda \alpha_i^c \), and \( \epsilon_i = \epsilon_i^s - \lambda \epsilon_i^c \)

- \( \beta_i Z \): index of patient appropriateness for each treatment (e.g. age)
- \( \alpha_i P_i \): productivity spillover (\( \alpha \) could be zero)
- \( \epsilon_i \): unobservables that influence survival and cost
- Note \( \lambda \) could be 0 due to insurance
An individual is treated intensively \((i = 2)\) if \(U_2 > U_1\) (treatment maximizes patient \(U\), not accounting for externalities). Recall \(P_1 = 1 - P_2\):

\[
\Pr\{\text{intensive}\} = \Pr\{i = 2\} \\
= \Pr\{U_2 - U_1 > 0\} \\
= \Pr\{\beta_2 Z + \alpha_2 P_2 + \epsilon_2 - \beta_1 Z - \alpha_1 (1 - P_2) - \epsilon_1 > 0\} \\
= \Pr\{P_2 (\alpha_1 + \alpha_2) - \alpha_1 + (\beta_2 - \beta_1) Z > \epsilon_1 - \epsilon_2\} \\
= \Pr\{\alpha P_2 - \alpha_1 + \beta Z > \epsilon\}
\]

where \(\alpha = \alpha_1 + \alpha_2\), \(\beta = \beta_2 - \beta_1\), and \(\epsilon = \epsilon_1 - \epsilon_2\)
Among the patients who choose the intensive treatment, the expected utility gain is:

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

\[ \Rightarrow \] patients receiving treat_2 have higher expected utility gain if:

1. More appropriate (higher \( \beta Z \))
2. Live in a more intensive region (higher \( \alpha P_2 \))

Intuition: patients are given the best care conditional on where they live, but marginal patients would be better off in area with other specialization.
Equilibrium (fixed point) condition

Let $f(Z)$ denote distribution of $Z$. In equilibrium, fraction of patients choosing intensive treatment ($P_2$) must match demand equation for \( P \{ \text{intensive treatment} \} \).

That is, proportion of patients choosing intensive treatments must generate benefits (with productivity spillovers) consistent with proportion.

$$P_2 = \int Z \Pr \{ \alpha P_2 - \alpha_1 + \beta Z > \epsilon \} f(Z) dZ$$

$$= G(P_2)$$

Variation across areas in use of intensive treatment can arise for two reasons: multiple equilibria, or single equilibrium determined by small differences in patient characteristics.
Equilibrium

Variation across areas in treat$_2$ can arise for two reasons:

1(A): Multiple (here: two) stable equilibria: intensive (high returns to treat$_2$) and non-intensive (low returns to treat$_2$); no prediction on choice

1(B): Single equilibrium determined by small differences in patient characteristics: productivity spillovers can magnify small differences

© The University of Chicago Press. All rights reserved. This content is excluded from our Creative Commons license. For more information, see http://ocw.mit.edu/help/faq-fair-use/.
Ignore $\epsilon$, plot $U$ against $Z$

Think of $Z$ as propensity score of appropriateness for treat$_2$
(age, comorbidities)
Equilibrium: Key figure

Figure 2(a): within-area, gap between treat$_1$ and treat$_2$ larger for more appropriate patients (⇒ returns are higher for these patients)
Equilibrium: Key figure

Figure 2(b):

1. Less appropriate patients worse off in intensive areas
2. Marginal patient less appropriate in intensive areas
3. More appropriate patients better off in intensive areas

© The University of Chicago Press. All rights reserved. This content is excluded from our Creative Commons license. For more information, see http://ocw.mit.edu/help/faq-fair-use/.
Welfare

- Spillovers $\Rightarrow$ increase in $P_2$ has positive externality on some patients, negative externality on others
- Unsurprisingly, externalities $\Rightarrow$ equilibrium may not be optimal
- Single vs. multiple equilibrium cases matter for welfare
  - Multiple: “area approach” can determine optimal $P_2$
  - Single: *too little* area variation in treatment as long as marginal patient ignores externality she imposes
Data

- **Context:** heart attacks (‘acute myocardial infarctions’)
  - Common condition
  - Extensive data (Medicare claims + CCP chart data)
  - Relatively high mortality rate
  - Limited role for patients to select providers
- **Treatments:**
  - Non-intensive: beta blockers (note: should be prescribed to all)
  - Intensive: cardiac catheterization
- ‘Standard’ market definitions: 306 ‘hospital referral regions’
- Assign patients to HRR of residence, not treatment
Estimation

Partition patients into groups \((k)\) based on appropriateness for treat

For \(\text{Outcome}_{ijk} \in \{\text{Survival}_{ijk}, \text{Cost}_{ijk}\}\) for patient \(i\) in HRR \(j\) and group \(k\), key estimating equation is:

\[
\text{Outcome}_{ijk} = \beta_{0k} + \beta_{1k} \text{Intensive Treatment}_i + \mathbf{X}_i \Pi_k + u_{ijk}
\]

- What is the potential problem with this OLS regression?
- IV: ‘differential distance’ (McClellan et al. 1994)
  - Distance to nearest cath hospital minus distance to nearest noncath hospital (negative \(\Rightarrow\) nearest hospital is cath hospital)
- Appropriateness measure: \(\Pr(\text{Cardiac Cath}_{ij}) = \hat{G}(\theta_0 + \mathbf{X}_i \Phi)\)
Results

Two sets of results:

- **Testing implications of the Roy model**
  1. Returns to intensive treatment increase in appropriateness
  2. Marginal patient less appropriate in intensive areas

- **Testing implications of productivity spillovers**
  1. Quality of medical management worse in intensive areas
  2. Characteristics of other patients influence treatment
  3. Returns to intensive treatment higher in intensive areas
  4. Most appropriate patients better off in intensive areas
  5. Least appropriate patients worse off in intensive areas
Results: Table 1

- IV: outcome = f(cath), survival = f(spending)
- More appropriate patients benefit more from treat₂: 0.038 vs. 0.002 in Column 3; higher survival, lower costs
- Similar results with age
- Consistent with Roy model

### Table 1

**Instrumental Variable Estimates of Intensive Management and Spending on One-Year Survival by Clinical Appropriateness of Patient**

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

*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 (measured as the distance between the patient’s zip code of residence and the nearest catheterization hospital minus the distance to the nearest hospital) is the instrument. Each model includes all the CCP risk adjusters, and the standard errors are clustered at the level of each HRR.

© The University of Chicago Press. All rights reserved. This content is excluded from our Creative Commons license. For more information, see [http://ocw.mit.edu/help/faq-fair-use/](http://ocw.mit.edu/help/faq-fair-use/).
Results: Table 2

Split the sample by above/below median values of instrument

1. Predicts cath (first stage, 48.9 - 42.8 = 6.1pp)
2. Predicts survival (reduced form, 67.6 - 66.7 = 0.9pp)
3. *Doesn’t* predict ‘predicted survival’ (67.5 - 67.2 = 0.3pp)
4. Columns (7) and (8) argue marginal patients similar to average

### TABLE 2

<table>
<thead>
<tr>
<th>Sample</th>
<th>30-Day Cath Rate</th>
<th>One-Year Survival</th>
<th>One-Year Predicted Survival</th>
<th>30-Day Predicted Cath Rate for Patients Getting Cath</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DD Below Median</td>
<td>DD Above Median</td>
<td>DD Below Median</td>
<td>DD Above Median</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>All patients (N = 129,997)</td>
<td>48.9</td>
<td>42.8</td>
<td>67.6</td>
<td>66.7</td>
</tr>
<tr>
<td>By cath propensity:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Above the median (N = 64,733)</td>
<td>74.0</td>
<td>67.1</td>
<td>84.6</td>
<td>83.8</td>
</tr>
<tr>
<td>Below the median (N = 65,244)</td>
<td>22.9</td>
<td>19.5</td>
<td>50.1</td>
<td>50.4</td>
</tr>
<tr>
<td>By age:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65-80 (N = 90,016)</td>
<td>61.1</td>
<td>54.9</td>
<td>74.3</td>
<td>73.5</td>
</tr>
<tr>
<td>Over 80 (N = 39,961)</td>
<td>20.3</td>
<td>16.5</td>
<td>52.1</td>
<td>52.1</td>
</tr>
</tbody>
</table>

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 CCI 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.
Test in the spirit of Gruber et al. (1999)

- Sample: patients receiving cath
- Patient appropriateness = f(log, risk-adjusted HRR cath rate)
- Negative: average patient appropriateness lower in more intensive areas
- Consistent with Roy model
Results: Table 4

- Quality of non-intensive treatment (beta blockers) worse in intensive areas: -0.31
- Consistent with productivity spillovers

<table>
<thead>
<tr>
<th>HRR Indicator</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>10th Percentile</th>
<th>90th Percentile</th>
<th>Correlation with HRR Cath Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measures of intensive treatment:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk-adjusted 30-day cath rate</td>
<td>46.3%</td>
<td>9.1%</td>
<td>34.5%</td>
<td>58.3%</td>
<td>1.00</td>
</tr>
<tr>
<td>Risk-adjusted 30-day PTCA rate</td>
<td>17.7%</td>
<td>5.1%</td>
<td>11.3%</td>
<td>23.6%</td>
<td>.81</td>
</tr>
<tr>
<td>Risk-adjusted 30-day CABG rate</td>
<td>13.4%</td>
<td>2.9%</td>
<td>10.2%</td>
<td>16.9%</td>
<td>.51</td>
</tr>
<tr>
<td>Risk-adjusted 12-hour PTCA rate</td>
<td>2.7%</td>
<td>2.6%</td>
<td>.6%</td>
<td>5.8%</td>
<td>.52</td>
</tr>
<tr>
<td>Measures of quality of medical management:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk-adjusted beta-blocker rate</td>
<td>45.6%</td>
<td>9.5%</td>
<td>34.2%</td>
<td>58.3%</td>
<td>-.31</td>
</tr>
<tr>
<td>Support for intensive treatment:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiovascular surgeons per 100,000</td>
<td>1.06</td>
<td>.27</td>
<td>.70</td>
<td>1.40</td>
<td>.33</td>
</tr>
<tr>
<td>Cath labs per 10,000</td>
<td>2.40</td>
<td>.76</td>
<td>1.50</td>
<td>3.30</td>
<td>.39</td>
</tr>
<tr>
<td>Demographic characteristics:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log of resident population</td>
<td>13.96</td>
<td>.89</td>
<td>12.72</td>
<td>15.18</td>
<td>-.05</td>
</tr>
<tr>
<td>Log of per capita income</td>
<td>9.55</td>
<td>.20</td>
<td>9.31</td>
<td>9.85</td>
<td>.02</td>
</tr>
<tr>
<td>Percent college graduates</td>
<td>19.3%</td>
<td>5.5%</td>
<td>13.1%</td>
<td>26.6%</td>
<td>-.05</td>
</tr>
</tbody>
</table>

Note: HRR surgical and medical intensity rates are computed as the risk-adjusted fixed effects from a patient-level regression of the receipt of cath or beta-blockers on HRR fixed effects and CCP risk adjusters.

© The University of Chicago Press. All rights reserved. This content is excluded from our Creative Commons license. For more information, see http://ocw.mit.edu/help/faq-fair-use/.
Results: Table 5

- Cath = f(average appropriateness of patients in your HRR)
- 1pp increase in average propensity of patients in your HRR
  ⇒ 0.53pp increase in the probability you receive cath
- Consistent with productivity spillovers

<table>
<thead>
<tr>
<th>HRR-Level Independent Variable</th>
<th>Probability of Receiving Catheterization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Average propensity to get cath</td>
<td>.529</td>
</tr>
<tr>
<td></td>
<td>(.172)</td>
</tr>
<tr>
<td>Percent under age 65</td>
<td>.150</td>
</tr>
<tr>
<td>Log of resident population</td>
<td>-.003</td>
</tr>
<tr>
<td>Log of per capita income</td>
<td>.024</td>
</tr>
</tbody>
</table>

Note.—The table reports OLS estimates of the relationship between a patient receiving catheterization and the average appropriateness for catheterization in an HRR. Regressions control for patient risk adjusters, and standard errors are clustered at the level of HRRs.

© The University of Chicago Press. All rights reserved. This content is excluded from our Creative Commons license. For more information, see [http://ocw.mit.edu/help/faq-fair-use/](http://ocw.mit.edu/help/faq-fair-use/).
Results: Table 6

- IV: outcome = f(cath), survival = f(spending) (like Table 1, but area)
- Returns to treat$^2$ higher in intensive areas (0.038 vs. 0.009); opposite prediction from “flat of the curve” model
- Difference in IV from survival, not costs
- Consistent with productivity spillovers

**TABLE 6**

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

Note.—HRR intensity rates are computed as the risk-adjusted fixed effects from a patient-level regression of the receipt of cath on HRR fixed effects and CCP risk adjusters. Differential distance (measured as the distance between the patient’s zip code of residence and the nearest catheterization hospital minus the distance to the nearest hospital) is the instrument. Each model includes all the CCP risk adjusters, and the standard errors are clustered at the level of each HRR.

© The University of Chicago Press. All rights reserved. This content is excluded from our Creative Commons license. For more information, see [http://ocw.mit.edu/help/faq-fair-use/](http://ocw.mit.edu/help/faq-fair-use/).
Results: Table 7

- **OLS**: outcome = f(HRR-level cath rate)
- **On average**, no return to spending; hides important heterogeneity
- **Intensive areas**: appropriate patients better off (0.052), less appropriate patients worse off (-0.075)
- **REALLY striking**
- **Consistent with productivity spillovers**

<table>
<thead>
<tr>
<th>Sample Description</th>
<th>One-Year Survival (1)</th>
<th>One-Year Cost ($1,000s) (2)</th>
<th>Beta-Blocker in Hospital (3)</th>
<th>Catheterization within 30 Days (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. All patients (N = 138,875)</td>
<td>.007 (.019)</td>
<td>8.093 (1.410)</td>
<td>-.28 (.073)</td>
<td>.702 (.004)</td>
</tr>
<tr>
<td>B. By cath propensity:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top tercile (N = 46,287)</td>
<td>.052 (.019)</td>
<td>10.012 (1.439)</td>
<td>-.366 (.073)</td>
<td>.802 (.032)</td>
</tr>
<tr>
<td>Middle tercile (N = 46,295)</td>
<td>.03 (.030)</td>
<td>11.154 (1.784)</td>
<td>-.271 (.082)</td>
<td>.906 (.021)</td>
</tr>
<tr>
<td>Bottom tercile (N = 46,291)</td>
<td>-.075 (.028)</td>
<td>2.763 (1.612)</td>
<td>-.209 (.073)</td>
<td>.369 (.021)</td>
</tr>
<tr>
<td>Difference (top – bottom)</td>
<td>.127 (.034)</td>
<td>7.249 (2.161)</td>
<td>-.157 (.103)</td>
<td>.433 (.038)</td>
</tr>
<tr>
<td>C. By age:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65–80 (N = 96,093)</td>
<td>.023 (.021)</td>
<td>9.616 (1.448)</td>
<td>-.311 (.072)</td>
<td>.775 (.012)</td>
</tr>
<tr>
<td>Over 80 (N = 42,780)</td>
<td>-.031 (.028)</td>
<td>4.738 (1.603)</td>
<td>-.215 (.080)</td>
<td>.531 (.022)</td>
</tr>
<tr>
<td>Difference (top – bottom)</td>
<td>.054 (.035)</td>
<td>4.878 (2.160)</td>
<td>-.096 (.108)</td>
<td>.244 (.025)</td>
</tr>
<tr>
<td>D. By AHA/ACC criterion:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal (N = 89,569)</td>
<td>.027 (.023)</td>
<td>9.845 (1.599)</td>
<td>-.302 (.076)</td>
<td>.769 (.010)</td>
</tr>
<tr>
<td>Appropriate (N = 31,800)</td>
<td>-.002 (.024)</td>
<td>6.174 (1.537)</td>
<td>-.282 (.080)</td>
<td>.752 (.026)</td>
</tr>
<tr>
<td>Not appropriate (N = 17,504)</td>
<td>-.08 (.040)</td>
<td>2.958 (1.511)</td>
<td>-.177 (.065)</td>
<td>.264 (.021)</td>
</tr>
<tr>
<td>Difference (top – bottom)</td>
<td>.107 (.046)</td>
<td>6.887 (2.200)</td>
<td>-.125 (.100)</td>
<td>.505 (.023)</td>
</tr>
</tbody>
</table>

**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 adjustors. HRR surgical and medical intensity rates are computed as the risk-adjusted fixed effects from a patient-level regression of the receipt of cath or beta-blockers on HRR fixed effects and CCP risk adjustors.

© The University of Chicago Press. All rights reserved. This content is excluded from our Creative Commons license. For more information, see [http://ocw.mit.edu/help/faq-fair-use/](http://ocw.mit.edu/help/faq-fair-use/).
Take-aways

- ‘Facts’ of geographic variation have been around a long time
  - Health economists’ favorite puzzle: tremendous variation in medical spending across observationally similar patients
  - Caveat: other industries...

- High impact paper: simple model, careful empirics

- Not your ‘standard’ IV paper

- Not much on mechanisms for productivity spillovers

- Very policy relevant, but pretty silent on welfare
  - Is high spending evidence of overuse?
  - Can’t infer overuse if productivity is heterogeneous
  - Chandra-Staiger (2011):
    “Expertise, overuse, and underuse in health care”
Expertise, overuse, and underuse in health care

- Variation in treatment intensity across hospitals due to:
  1. Greater benefits of treatment (“expertise”)
  2. Withholding of treatment (“underuse”)
  3. Providing harmful treatment (“overuse”)

- Model:
  - Expected benefit from treatment: \( B_{ih} = \alpha_h + X_{ih}\beta + \epsilon_{ih} \)
  - Expertise: \( \alpha_h \)
  - Hospital treats if \( B_{ih} \) exceeds hospital-specific threshold \( \tau_h \)
  - \( E(B_{ih}|treat_{ih} = 1) = \alpha_h + X_{ih}\beta + E(\epsilon_{ih}| - \epsilon_{ih} < X_{ih}\beta + \alpha_h - \tau_h) \)

- Tentative conclusion: Expertise varies widely, lots of overuse

- Won’t go through empirics, just an example
Different model/example: Prostate cancer

- Two treatment options: surgery, watchful waiting
- Assume PSA score is a perfect risk adjuster
  - Different from Chandra-Staiger: No uncertainty
- Patients as good as randomly allocated to providers
- Providers use a threshold rule

- Frame as a regression discontinuity (RD) framework
  - Goal: Clarify how parameters are identified in the data

- Want to identify:
  - $\alpha_h$: Hospital “expertise”
  - $\tau_h$: Look for evidence of overuse/underuse ($\tau_h \geq 0$)
Pr(1-yr survival)

Note: Drawn such that higher PSA score patients benefit more from the treatment.

hospital A

PSA score

no treatment

(“intensive”) 0 4 10 ("less intensive")

RD estimate tells you $\tau_A$:
- Optimal for patient if $\tau_A = 0$
- Underuse if $\tau_A > 0$: patients with positive benefits not being treated (as drawn)
- Overuse if $\tau_A < 0$: patients with negative benefits are being treated
What can we learn if hospitals switch at the same point and have different $\tau$'s?
- Both A, A' switch at PSA = 4
  - $\tau_A > \tau_{A'}$
  => hospital A more productive at PSA = 4
  True in both RD and Chandra-Staiger frameworks
What can we learn if hospitals switch at different points and have the same $\tau$'s?
- Hospital A switches at PSA = 4
- Hospital B switches at PSA = 10
- $\tau_A = \tau_B$

$\Rightarrow$ Hospital A more productive in Chandra-Staiger

In RD framework, no basis for inference

Same tau, less appropriate $X$'s $\Rightarrow$ higher $\alpha_h$
What can we learn if hospitals switch at the same point and have the same $\tau$’s?
- Both $A, A'$ switch at PSA = 4
- $\tau_A = \tau_{A'}$
=> hospitals $A, A'$ equally productive at PSA = 4
True in both RD and Chandra-Staiger frameworks
Note: slopes could be different?
Connecting this to the Chandra-Staiger model

- Identification of $\tau_h$ is clearer than is identification of $\alpha_h$
- In RD framework, $\alpha_h$ is identified by comparing realized returns across hospitals using the same threshold
- How is $\alpha_h$ identified in Chandra-Staiger?
  - Component of treatment returns common across patients in $h$
  - Is this how we would ideally model expertise?
    - What if good at treating easy but not complicated patients?
    - Does this complicate decomposition exercise?
  - Estimation of $\alpha_h$ is very important, because $\alpha_h$ pins down where hospital A’s optimal threshold is if currently $\tau_A \neq 0$
    - Key input into welfare analysis with counterfactuals
    - Exactly what RD can’t tell us: Almond et al. (2010) example
Add’l application: Welfare effects of fixed thresholds

- Treatment guidelines based only on patient characteristics
  - Care for newborns: e.g. birthweight \( \leq 1250g \)
  - Prostate cancer antigen (PSA) score: e.g. PSA = 4
  - Hypertension: e.g. systolic blood pressure \( \geq 160mm \)

- Implicit assumption: all variation is driven by \( \tau_h \), not \( \alpha_h \)

- Criticisms of guidelines usually focus on patient heterogeneity

- But: heterogenous productivity \( \Rightarrow \) different optimal thresholds

- Chandra-Staiger model could be applied to estimate the welfare gains and losses from uniform treatment guidelines
  - Net welfare effect ambiguous
    - Welfare gain from limiting overuse or avoiding underuse
    - Welfare loss from heterogeneous hospitals using fixed threshold
  - Could estimate policy counterfactuals, conduct welfare analysis
1 Roy model application #1: Health care

2 Roy model application #2: Redistribution

3 Looking ahead
Abramitsky (2009)

- Series of papers - and a book (in progress) - investigating the equality-incentives trade-off in the context of the Israeli kibbutz
- Key features:
  1. Equal sharing in the distribution of income
  2. No private property
  3. Non-cash economy
- PF literature: expect mobility in response to redistributive policies
- Roy model:
  - Positive self-selection of migrants expected when place of origin has lower returns to skill (more redistribution) than destination
  - Negative self-selection of migrants expected when place of origin has higher returns to skill (less redistribution) than destination
- Abramitsky (2009) tests these ideas in context of Israeli kibbutzim
  - As in work on US immigration, takes advantage of a new longitudinal data set of individuals linked across population censuses
Data

- Random representative sample of individuals linked between 1983 and 1995 Israeli Censuses of Population
  - Censuses identify individuals who live in “a cooperative rural settlement, in which production, marketing, and consumption are organized in a cooperative manner” (kibbutz members)
- Focuses on Jewish individuals between the ages of 21 and 54 in 1983 (ages of 33 and 66 in 1995)
### Three subsamples

1. **1983 kibbutz members and other rural residents also observed in 1995**
   - Compare kibbutz-to-city migrants both with kibbutz members who stayed in their kibbutz and with other rural-to-city migrants

2. **City residents observed in 1995, including individuals who migrated from the kibbutz and from other rural areas between 1983 and 1995**
   - Analyze earnings of kibbutz-to-city migrants in the city labor market compared with earnings of city natives and other rural-to-city migrants

3. **City residents observed in 1983, including individuals who would migrate to kibbutz or other rural localities between 1983 and 1995**
   - Compare the pre-entry earnings of city-to-kibbutz migrants with the earnings of city stayers and city-to-other rural migrants
Summary statistics on entry and exit

- A total of 343 out of the 1577 individuals in the sample who lived in a kibbutz in 1983 left the kibbutz between 1983 and 1995, over 20%
- A total of 90 out of the 16,789 individuals in the sample who lived outside of kibbutzim in 1983 (with non-missing earnings) entered a kibbutz in this period, around 0.5%
  - Note: low in part because screening mitigates adverse selection
  - Makes it harder to document negative selection
Testing for positive selection in exit

More educated members and those with higher skilled occupations are more likely to leave kibbutzim, and this skill bias in out-migration is stronger in kibbutzim than in other rural localities. These results suggest a positive selection away from redistribution.

![Graphs showing exit proportions](http://www.sciencedirect.com)

*Fig. 1. Exit from kibbutzim and other rural areas, 1983–1995. Notes: The left hand panel shows the proportion of kibbutz members (solid line) and individuals from other rural areas (dashed line) who moved to the city between 1983 and 1995 by level of qualifications in 1983. The right hand panel shows the same, but broken down by the skill level of the member’s occupation in 1983.*

Testing for negative selection in entry

Individuals with lower wages are more likely to enter a kibbutz.

Fig. 3. Entry to kibbutzim from cities by wages. Notes: This figure shows the proportion of people living in cities in 1983 who entered kibbutzim between 1983 and 1995, broken down by wage categories in 1983. The numbers on the x-axis are plotted on a log scale.

Take-aways

- Paper formalizes these results, but I like that the key ideas are clearly illustrated in these two simple figures.
- Broader research agenda uses the kibbutz as a laboratory for understanding how one form of intensive redistribution was able to survive over time; many kibbutzim eventually moved away from full equal sharing to something closer to capitalism and taxation.
1. Roy model application #1: Health care

2. Roy model application #2: Redistribution

3. Looking ahead
Looking ahead

Equalizing wage differentials

For next Wednesday: Please read Goldin-Katz (forthcoming, JOLE)