

# Self-selection: The Roy model

Heidi L. Williams  
MIT 14.662  
Spring 2015

## Outline

- (1) Preliminaries: Overview of 14.662, Part II
- (2) A model of self-selection: The Roy model ([Roy, 1951](#))
- (3) Application: Immigration  
([Chiswick, 1978](#); [Borjas, 1985, 1987](#); [Abramitzky, Boustan and Eriksson, 2012, 2014](#))
- (4) Application: Health care ([Chandra and Staiger, 2007](#))
- (5) Application: Redistribution ([Abramitzky, 2009](#))

## 1 Preliminaries: Overview of 14.662, Part II

In their 2000 *Handbook of Income Distribution* review, [Neal and Rosen \(2000\)](#) discuss empirical regularities that have motivated theories of the distribution of labor earnings: for example, earnings distributions tend to be skewed to the right, and mean earnings tend to differ greatly across groups defined by occupation, education, experience and other observed traits (such as gender and race). A variety of models have been proposed as frameworks to explain these types of facts. For example, several theories of earnings distributions were covered in 14.661 and Part I of 14.662, such as human capital models and Rosen-style superstar models.

We are going to start off Part II of 14.662 by covering four additional classes of theoretical models with implications for earnings distributions: the Roy model, the compensating differentials model, discrimination models, and models of rent-sharing. For each class of models, we will discuss recent empirical papers applying these models to questions in labor economics as well as applied microeconomics more generally. From there we will cover three closely related topics - management practices, intergenerational mobility and early life determinants of long-run outcomes - which speak to other empirically important determinants of the distribution of labor earnings.<sup>1</sup>

---

<sup>1</sup>Almond and Currie's 2011 *Handbook of Labor Economics* chapter ([Almond and Currie, 2011](#)) motivates the increase in economics research on the latter two of these topics over the past few decades by saying: "*Child and family characteristics measured at school entry do as much to explain future outcomes as factors that labor economists have more traditionally focused on, such as years of education.*"

## 2 A model of self-selection: The Roy model

Roy (1951) analyzes the impact of self-selection in occupational choice on the income distribution. Roy motivates his analysis by saying that his contemporaries implicitly assumed that the distribution of incomes is arbitrary - “developed by the process of historical accident.” In contrast, the core of Roy’s model is to ask how the distribution of earnings is affected if individuals purposively select their occupation. Roy’s paper is definitely worth reading, but the key characteristics of the model are somewhat difficult to wade through given the verbal (no mathematical notation) style of the text. Instead, I’ll walk through the (formally identical) Borjas (1987) version of the Roy model, which is a ‘standard’ formalization of the Roy model that it is important for you to be comfortable with.

Roy’s original model was based on two occupations – (rabbit) hunting and fishing. The goal was to understand self-selection: will the individuals best suited for fishing choose to fish? Will the individuals best suited for hunting choose to hunt? The core idea of the model is to take seriously the idea that - in a market economy - individuals will not randomly sort themselves across the two occupations. In markets where non-random sorting is important, comparing (for example) the wage gap between hunters and fishermen will reflect not only a “real” difference in potential earnings (that would exist even if individuals *were* randomly distributed across occupations), but will also be a function of which individuals select into hunting and fishing.

This type of self-selection comes up as an issue in nearly every sub-field of economics. I will focus on three applications of the Roy model: immigration decisions, geographic variation in physician practice style, and redistribution.

## 3 Application: Immigration

Borjas (1987)’s application of the Roy model was motivated by wanting to explain heterogeneity in earnings across immigrants and natives in the US, with a focus on the self-selection induced by the migration decision. His model is written from the perspective of an immigrant who is thinking of migrating from her home (non-US) country to the US. The idea is that individuals compare their potential income in the US with their income in their home country, and make their migration decision based on this income differential (net of migration costs). This type of decision rule induces self-selection which then gives empirically testable predictions: if the US has higher returns to skill than the home country (higher income inequality), then migrants will be disproportionately drawn from the top of the home country’s skill distribution; in contrast, if the US has lower returns to skill than the home country (lower income inequality), then migrants will be disproportionately drawn from the bottom of the home country’s skill distribution.

Before digging into the specifics of the model, I want to start by laying out some of the intellectual history of this literature in order to give you some context for the Roy model. Borjas (1999) provides a survey of the literature on the economic analysis of immigration, and focuses attention on two questions: first, why do some people move? And second, what happens when they do? You covered the second question in 14.661 - covering papers such as Card (1990)

and [Borjas \(2003\)](#) that looked at how an inflow of immigrants affects the labor market outcomes of natives. In contrast, we here focus on the first question - looking at why people move, and in particular looking at the skill composition of immigrants (which is important for interpreting the question of how immigrants fare in the labor market relative to natives). Of course, these two questions are closely linked, because the economic impact of immigration depends on the skill distributions of immigrants and natives.

In order to appreciate the contribution of the Roy model to our understanding of this first question - why some people move - it is useful to have a concrete sense of how people thought about this question *before* the Borjas paper. In a well-known contribution, [Chiswick \(1978\)](#) noted: “*Economic theory suggests that migration in response to economic incentives is generally more profitable for the more able and more highly motivated,*” and then has a footnote outlining a simple model that generates that prediction. A key assumption underlying the Chiswick conclusion is that ability has the same effect on earnings in both the origin and destination countries. The Roy model relaxes this assumption: as we will see, a core insight of the [Borjas \(1987\)](#) application of the Roy model is that what type of selection we expect to see (that is, immigrants being positively or negatively selected) critically depends on the correlation between the value of ability in the home country and the value of ability in the destination country. Hence, in the Roy model, self-selection will not always imply that immigrants are the most able individuals from the home country.

### 3.1 [Chiswick \(1978\)](#) and [Borjas \(1985\)](#): Assimilation

Prior to the [Borjas \(1987\)](#) paper, there was an intellectual exchange between Barry Chiswick and George Borjas over how to interpret observed earnings differences between natives and immigrants, which motivated Borjas’s later application of the Roy model.

[Chiswick \(1978\)](#) was interested in estimating how “time spent in the US” affects the earnings of immigrants – that is, how the wages of immigrants change as they accumulate additional years of residence in the US. He investigated this question by estimating cross-sectional Mincer-style human capital earnings function equations that included a variable for “years since migration” to ask how that variable affected the earnings of immigrants. Luckily for him, for the first time since 1930, the (at the time, recently released) 1970 US Census asked a question about what year immigrants arrived in the US - enabling this analysis. Chiswick’s conclusions were thus based on a cross-sectional comparison of different cohorts at the point in time when the 1970 Census was taken.

In the 1970 Census data, Chiswick estimated regressions like the following:

$$\ln(\text{earnings}_i) = \mathbf{X}'_i\theta + \delta I_i + \alpha_1 I_i \text{Years}_i + \alpha_2 I_i \text{Years}_i^2 + \epsilon_i$$

where  $\mathbf{X}'_i$  includes covariates such as education and potential experience (calculated as age less education minus 5),  $I_i$  is an indicator for foreign-born, and  $\text{Years}_i$  represents years since migration. His main estimates are presented in Table 2:

TABLE 2  
REGRESSION ANALYSIS OF EARNINGS FOR NATIVE- AND  
FOREIGN-BORN ADULT WHITE MEN, 1970

	NATIVE BORN	NATIVE AND FOREIGN BORN		FOREIGN BORN	
	(1)	(2)	(3)	(4)	(5)
EDUC .....	.07154 (53.78)	.07058 (55.68)	.07004 (55.18)	.07164 (54.11)	.05740 (12.93)
$T$ .....	.03167 (22.99)	.03050 (22.86)	.03071 (22.99)	.03097 (23.10)	.02028 (3.47)
$T^2$ .....	-.00052 (-20.77)	-.00049 (-20.45)	-.00050 (-20.78)	-.00051 (-20.93)	-.00031 (-3.18)
LN WW .....	1.10335 (81.75)	1.10326 (84.78)	1.10169 (84.70)	1.10111 (84.67)	1.07151 (21.97)
RURALEQ1 ...	-.17222 (-20.28)	-.16970 (-20.25)	-.17080 (-20.39)	-.16915 (-20.18)	-.05821 (-1.13)
SOUTHEQ1 ...	-.12090 (-14.17)	-.12620 (-15.01)	-.12530 (-14.91)	-.12389 (-14.74)	-.21587 (-4.38)
NOTMSP .....	-.30647 (-27.76)	-.31078 (-28.97)	-.30947 (-28.86)	-.30874 (-28.79)	-.34498 (-7.66)
FOR .....	*	.02951 (1.75)	-.16359 (-4.32)	.00990 (0.18)	*
(FOR) (YSM)	*	*	.01461 (3.98)	.01555 (4.23)	.01500 (3.87)
(FOR) (YSM2)	*	*	-.00016 (-2.47)	-.00018 (-2.79)	-.00019 (-2.82)
(FOR) (EDUC)	*	*	*	-.01619 (-4.23)	*
CONSTANT ...	-1.03646	-1.01537	-1.00016	-1.02156	-.78891
Observations ( $N$ ) .....	34,321	36,245	36,245	36,245	1,924
$R$ .....	.55423	.55455	.55533	.55564	.58194
$R^2$ .....	.30717	.30753	.30839	.30873	.33866
Standard error	.70900	.71008	.70966	.70949	.71676

SOURCE.—U.S. Bureau of the Census 1972.

NOTE.— $t$ -ratios in parentheses; dependent variable: natural logarithm of earnings in hundreds of dollars.  
\* Variable not entered.

© 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/>.

Based on this analysis, Chiswick drew two conclusions:

1. The experience-earnings profile of immigrants is steeper than the experience-earnings profile of natives with the same measured skills. Chiswick focuses on deriving this conclusion based on the estimated coefficients in Column 2 and 5 of Table 2, evaluated at 10 years of experience ( $T = 10$ ) and 5 years of residency in the US ( $YSM = 5$ ).<sup>2</sup> For natives (Column 2), he calculates that  $(\beta_T = 0.03050) + 2 \cdot (\beta_{T^2} = -0.00049) \cdot (T = 10)$  implies a return to an additional year of experience of  $\approx 2.07\%$ . For immigrants (Column 5), he calculates that  $(\beta_T = 0.02028) + 2 \cdot (\beta_{T^2} = -0.00031) \cdot (T = 10) + (\beta_{YSM} = 0.01500) + 2 \cdot (\beta_{YSM^2} = -0.00019) \cdot (YSM = 5)$  implies a return to an additional year of experience in the US of  $\approx 2.718\%$ . Hence, he concluded that the return to experience for immigrants (2.718%) is steeper than for natives (2.07%). Chiswick interpreted this fact in a human capital framework in which immigrants and natives differ in the nature and financing of post-schooling

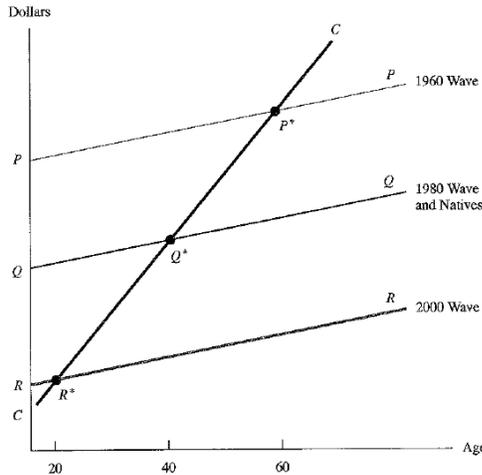
<sup>2</sup>Side note: it seems more natural to base this comparison on the estimated coefficients in Column 1 and 5? I am here following the discussion in his paper.

training (“larger worker-financed investments” by immigrants, because of the expectation of greater job mobility).

2. The experience-earnings profile of immigrants crosses the experience-earnings profile of natives about 10-15 years after immigration. Chiswick focuses on deriving this conclusion based on the estimated coefficients in Column 3 of Table 2, holding constant schooling and total labor market experience. For various years since migration ( $YSM$ ), he calculates the difference in earnings between natives and immigrants as  $(\beta_{FOR} = -0.16359) + (\beta_{YSM} = 0.01461) \cdot (YSM) + (\beta_{YSM^2} = -0.00016) \cdot (YSM^2)$ . For  $YSM = 10$ , the predicted percent difference in earnings between natives and the foreign born is  $\approx -3.349\%$ ; for  $YSM = 15$ , this is  $\approx +1.956\%$ ; hence he concluded that the immigrant experience-earnings profile crossed that of natives between 10 and 15 years after immigration. Chiswick interpreted this fact as evidence of self-selection in migration in favor of “high ability, highly motivated workers, and workers with low discount rates for human capital investments.”

However, because the 1970 Census is a single cross-section, the “years since migration” variable may confound two effects: (1) a true “assimilation” effect; and (2) fixed quality differences across immigrant cohorts. Borjas (1985) discusses this problem, and Figure 8-5 in Borjas’s *Labor Economics* text (Fifth Edition, p. 333) illustrates an example of how the type of cross-sectional analysis presented in Chiswick (1978) can erroneously estimate patterns in the age-earnings profile that in fact may be driven by fixed differences across cohorts.

**FIGURE 8-5 Cohort Effects and the Immigrant Age-Earnings Profile**  
 The typical person migrating in 1960 is skilled and has age-earnings profile  $PP$ ; the 2000 immigrant is unskilled and has age-earnings profile  $RR$ ; the 1980 immigrant has the same skills as the typical native and has age-earnings profile  $QQ$ . Suppose all immigrants arrive at age 20. The 2000 census cross section reports the wages of immigrants who have just arrived (point  $R^*$ ); the wage of immigrants who arrived in 1980 when they are 40 years old (point  $Q^*$ ); and the wage of immigrants who arrived in 1960 when they are 60 years old (point  $P^*$ ). The cross-sectional age-earnings profile erroneously suggests that immigrant earnings grow faster than those of natives.



© McGraw-Hill Professional Publishing. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <http://ocw.mit.edu/help/faq-fair-use/>.

The quality of immigrant cohorts – in terms of their earnings – could change over time as a function of, *e.g.*, changes in immigration policies (such as policy changes that emphasize or de-emphasize skills as a criteria for admission). These effects are indistinguishable in the 1970 Census because (year of migration) + (years in US) = 1970. Stated differently, the Chiswick

cross-sectional earnings function approach encountered a version of the (now) well-known problem that it is impossible to separately identify age and cohort effects in a single cross-section such as the 1970 census.

[Borjas \(1985\)](#) realized that progress can be made (using a presumably-hot-off-the-press “test version of Stata” that he thanks someone for in the acknowledgements) by using repeated cross-section or longitudinal data. [Borjas \(1985\)](#) took advantage of the (at the time, recently released) 1980 US Census which again asked a question about what year immigrants arrived in the US. Borjas’s contribution was to combine the 1970 and 1980 US Census data to examine how well Chiswick’s cross-sectional predictions about earnings growth predicted the actual earnings growth experienced by specific immigrant cohorts during the period 1970-1980.

In order to identify both the assimilation effect and the cohort indicators while also controlling for Census year indicators, a restriction must be imposed. Borjas’s analysis imposed the restriction that time-specific shocks have the same effect on log earnings of natives and immigrants. In a pooled sample of native-born and foreign-born individuals, this effectively uses native-born individuals to estimate the Census year indicators. The implicit assumption is that factors that are fixed within Census year have the same effect on log earnings of natives and immigrants. For factors like inflation, that assumption seems reasonable. However, it may be that other year-specific factors – such as business cycle variation – have differential effects on natives and immigrants.

Substantively, whereas Chiswick had concluded - based on his cross-sectional analysis - that immigrants adapt quite rapidly into the US labor market, Borjas’s analysis of earnings within immigrant cohorts suggests a different conclusion. He finds relatively slower rates of earnings growth for most immigrant groups (that is, slower than what is predicted by Chiswick-style cross-section regressions), implying a decline in the quality of immigrant cohorts in recent decades.

In terms of big-picture take-aways from this literature, I want to stress two things. First is the methodological point on age-time-cohort effects: the impossibility of identifying age, time, and cohort effects in a linear model comes up in a variety of contexts, and is a useful question to have in the back of your mind when reading papers, attending seminars, and working on your own research. For example, Dave Molitor’s MIT dissertation looked at physician practice patterns controlling for calendar year fixed effects and “cohort” (year of medical school graduation) fixed effects, omitting age fixed effects ([Molitor, 2011](#)). Second is how the substantive conclusions of this Chiswick-Borjas exchange relate to Borjas’s later Roy model application. [Chiswick \(1978\)](#) had interpreted the fact that the experience-earnings profile of immigrants crosses that of natives as evidence of self-selection in migration in favor of “high ability, highly motivated workers, and workers with low discount rates for human capital investments.” [Borjas \(1985\)](#) clarified that this fact could instead reflect cohort effects, which then raises the question of how cohort effects (specifically, cohort “quality”) relate to self-selection. This question provides the starting point for [Borjas \(1987\)](#)’s application of the Roy model.

### 3.2 Abramitzky, Boustan and Eriksson (2014): Assimilation

Abramitzky, Boustan and Eriksson (2014) re-examine this assimilation question using data on European immigrants to the US labor market during the Age of Mass Migration (1850-1913). They motivate their interest in this time period by noting that just as today, contemporaries were concerned about the ability of migrants to assimilate into the US economy. Fears about immigrant assimilation encouraged Congress to convene a special commission in 1907, which concluded that immigrants - particularly from southern and eastern Europe - would be unable to assimilate. This report provided fuel for legislators to subsequently restrict immigrant entry via a literacy test (in 1917) and quotas (in 1924). After that report was published, subsequent analyses suggested that - contrary to the commission's report - immigrants caught up with the native-born after 10 to 20 years in the US. However, all of these studies compared earnings in a single cross-section, and hence are subject to the Borjas (1985) critique of changes in immigrant cohort quality over time, as well as to a critique about selective return migration.

As stressed by Borjas (1985), the concern about changes in immigrant cohort quality over time can be addressed with repeated cross-sections on arrival cohorts. However, even with repeat cross-sections inferences on migrant assimilation may still be inaccurate because of selective return migration: if temporary migrants have lower skills or exert less effort, compositional changes in repeated cross sections will generate the appearance of wage growth within cohorts over time as lower-earning migrants return to Europe.

Abramitzky, Boustan and Eriksson (2014) address this concern by constructing a novel panel data set that follows native-born workers and immigrants from 16 sending countries through the US censuses of 1900, 1910, and 1920. They match individuals over time by first and last name, age, and country or state of birth. Because these censuses do not contain individual information about wages or income, they assign individuals the median income in their reported occupation.

They start by comparing occupation (as a proxy for labor market earnings) of native-born and immigrant workers, as a function of variables measuring time spent in the US (with native-born as the omitted group), indicators for year and country of origin, and age controls in pooled data (that is, omitting arrival cohort indicator variables) from the 1900, 1910, and 1920 censuses (which they refer to as the “cross-section” model). They then add arrival cohort indicators (which they refer to as the “repeated cross-section” model, because it follows arrival cohorts across census years); comparing the cross-section and repeated cross-section allows them to infer how much of the earnings difference between natives and immigrants is attributable to differences in the quality of arrival cohorts. Finally, they compare the repeated cross-section model with a “panel” sample which follows individuals across census years. By comparing the estimates in the repeated cross-section and the panel, they can infer whether and to what extent return migrants were positively or negatively selected from the immigrant population. Specifically, if they observe more (less) convergence in the repeated cross-section relative to the panel, they will infer that temporary migrants are drawn from the lower (upper) end of the occupation-earnings distribution.

**Table 4.** In the cross section, new immigrants hold occupations that earn \$1,200 below natives of similar ages, and appear to completely close this gap over time (Column 1). When indicator variables for arrival cohorts are added in Column 2, the initial occupation earnings gap shrinks to \$400, and in the panel sample (Column 3) there is no initial earnings penalty – if anything, immigrants start out slightly ahead of natives. The immigrant-native earnings gaps in Columns 2 and 3 are statistically distinguishable for immigrants who arrived between 0-5 and 6-10 years ago, suggesting negative selection of return migrants to Europe.

TABLE 4  
ORDINARY LEAST SQUARES ESTIMATES, AGE-EARNINGS PROFILE FOR NATIVES AND  
FOREIGN-BORN, 1900–1920: 1950 OCCUPATION-BASED EARNINGS IN 2010 DOLLARS

RIGHT-HAND-SIDE VARIABLE	POOLED CROSS SECTION AND PANEL		
	CROSS SECTION (1)	Cross Section Coefficients (2)	Panel Coefficients (3)
0–5 years in US	–1,255.73 (143.44)	–384.49 (187.30)	293.51 (237.96)
6–10 years in US	–734.51 (147.44)	–2.89 (172.05)	467.64 (213.61)
11–20 years in US	–352.93 (131.27)	173.83 (134.02)	329.38 (150.49)
21–30 years in US	–294.87 (142.10)	128.44 (138.93)	74.34 (150.33)
30 years in US	22.41 (184.65)	155.77 (178.49)	231.90 (186.55)
Arrive 1891+	...	–739.18 (106.99)	–232.77 (160.58)
Native-born	...	...	–153.83 (176.14)
Observations	205,458	259,093	

NOTE.—See the table 1 note for sample restrictions. Columns report coefficients from estimation of eq. (1). Column 1 pools three cross sections (1900–1920); the regression in cols. 2 and 3 adds the matched panel sample. The coefficients in col. 2 are interactions between the right-hand-side variables listed and a dummy for being in the cross section, and col. 3 reports interactions between the right-hand-side variables and a dummy for being in the panel. The omitted category is native-born men in the cross section. Coefficients on age, census year dummies, and country-of-origin fixed effects are not shown.

© 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/>.

**Figure 2.** Figure 2 plots the differences in implied convergence in these three specifications. Here, it is even easier to see the patterns from Table 4.

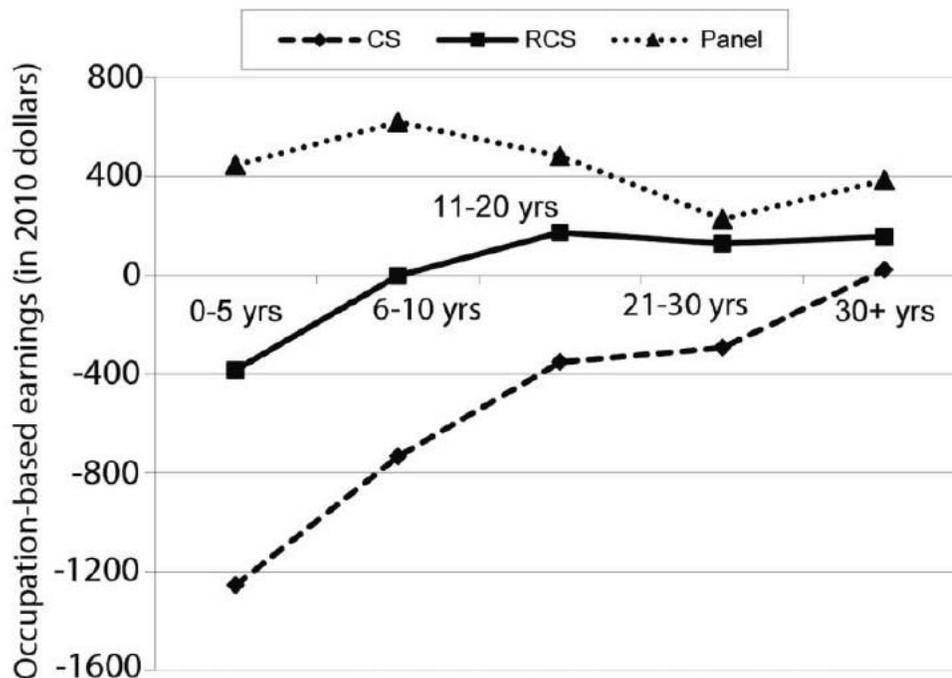


FIG. 2.—Convergence in occupation score between immigrants and native-born workers by time spent in the United States, cross-sectional and panel data, 1900–1920. The graph plots coefficients for years spent in the United States indicators in equation (1). Note that for the panel line, we subtract the native-born dummy from the years in the United States indicators (because the omitted category in that regression is natives in the panel sample). See table 4 for coefficients and standard errors.

© 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/>.

With the caveat that location choice within the US is endogenous, Table 6 (not shown) adds state fixed effects to these specifications, comparing immigrants and natives who live in the same states. This adjustment doubles the immigrant earnings penalty in the cross-section, and converts the small panel occupation earnings premium into an earnings penalty. While at best suggestive, this is consistent with immigrants achieving earnings parity with natives by moving to locations with a well-paid mix of occupations.

### 3.3 Borjas (1987): A model of self-selection

#### 3.3.1 Basic set-up of the model

Let country 0 denote the individual’s home country of origin, and country 1 denote the destination country (the US).<sup>3</sup> The wage of individual  $i$  in country  $j$  is:

$$\ln w_{ij} = \mu_j + \epsilon_{ij} \quad (1)$$

<sup>3</sup>Dahl (2002) develops a version of the Roy model in which potential movers have the option of moving to more than one potential destination.

where  $\epsilon_{i0}$  and  $\epsilon_{i1}$  are jointly normally distributed, and  $\sigma_{0,1}$  denotes  $cov(\epsilon_{i0}, \epsilon_{i1})$ :

$$\begin{pmatrix} \epsilon_{i0} \\ \epsilon_{i1} \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_{0,1} \\ \sigma_{0,1} & \sigma_1^2 \end{pmatrix} \right) \quad (2)$$

From here forward we drop the  $i$  subscripts; they were useful above just to clarify that these distributions describe agent  $i$ 's wages if she were in each of country 0 and country 1, but in practice we of course only observe one of those two wages. Borjas assumes that  $\mu_1$  also gives the earnings of the average native worker in the US. Borjas notes that we can think of these expressions as decomposing earnings into the part explained by observable characteristics such as age and completed education ( $\mu_0$  and  $\mu_1$ ) and a part due to unobserved characteristics ( $\epsilon_0$  and  $\epsilon_1$ ).

Let  $\rho_{0,1}$  denote the correlation coefficient of  $\epsilon_0$  and  $\epsilon_1$ , which represents the correlation of productive ability in the home country with productive ability in the US:

$$\rho_{0,1} = \frac{cov(\epsilon_0, \epsilon_1)}{\sigma_0 \sigma_1} \quad (3)$$

$$= \frac{\sigma_{0,1}}{\sigma_0 \sigma_1} \quad (4)$$

Let  $C$  denote the cost of moving. Borjas defines  $C = \pi w_0$  so that moving costs are expressed relative to the home country wage; expressing moving costs like this allows them to be neatly included in the  $\ln(w_o + C)$  expression below. An individual's decision of whether to migrate to the US is then determined by the sign of the following index function  $I$ :<sup>4</sup>

$$I = \ln \left( \frac{w_1}{w_o + C} \right) \quad (5)$$

$$= \ln(w_1) - \ln(w_o + C) \quad (6)$$

$$= \ln(w_1) - \ln(w_0 + \pi w_0) \quad (7)$$

$$= \ln(w_1) - \ln(w_0(1 + \pi)) \quad (8)$$

$$= \mu_1 + \epsilon_1 - \mu_0 - \epsilon_0 - \ln(1 + \pi) \quad (9)$$

$$\approx (\mu_1 - \mu_0 - \pi) + (\epsilon_1 - \epsilon_0) \quad (10)$$

Define  $v \equiv \epsilon_1 - \epsilon_0$ . Since migration occurs if  $I > 0$ , we can write the migration rate  $P$  as:

$$P = \Pr[\epsilon_1 - \epsilon_0 > -(\mu_1 - \mu_0 - \pi)] \quad (11)$$

$$= \Pr[v > -(\mu_1 - \mu_0 - \pi)] \quad (12)$$

$$= \Pr[v > (\mu_0 - \mu_1 + \pi)] \quad (13)$$

Define  $z = \frac{\mu_0 - \mu_1 + \pi}{\sigma_v}$ , and let  $\Phi$  denote the CDF of the standard normal distribution. Note that because  $v = \epsilon_1 - \epsilon_0$ , and because  $\epsilon_1$  and  $\epsilon_0$  are normally distributed with mean zero, we know that  $v$  is normally distributed with mean zero, and that  $\frac{v}{\sigma_v}$  follows a standard normal

---

<sup>4</sup>The approximation here is to approximate  $\ln(1 + \pi)$  as  $\approx \pi$ , based on the first-order Taylor approximation of  $f(x) = \ln(1 + x)$  around 0:  $f(a) + \frac{f'(a)}{1!}(x - a)|_{a=0} = \ln(1 + 0) + \frac{1}{1+0} \frac{1}{1!}(x - 0) = x$ .

distribution. Therefore, we can re-write the above equation as:

$$P = \Pr \left[ \frac{v}{\sigma_v} > \frac{\mu_0 - \mu_1 + \pi}{\sigma_v} \right] \quad (14)$$

$$= 1 - \Pr \left[ \frac{v}{\sigma_v} \leq \frac{\mu_0 - \mu_1 + \pi}{\sigma_v} \right] \quad (15)$$

$$= 1 - \Phi \left( \frac{\mu_0 - \mu_1 + \pi}{\sigma_v} \right) \quad (16)$$

$$= 1 - \Phi(z) \quad (17)$$

For higher values of  $z$ ,  $P$  is lower - implying migration is less likely. The migration rate  $P$  is increasing in mean US wages ( $\frac{\partial P}{\partial \mu_1} > 0$ ), decreasing in mean home country wages ( $\frac{\partial P}{\partial \mu_0} < 0$ ), and decreasing in moving costs ( $\frac{\partial P}{\partial \pi} < 0$ ). Borjas assumes that  $P < 1$ , so that at least part of the population in the country of origin is better off not migrating. He also assumes  $\mu_1 \approx \mu_0$ .

### 3.3.2 Useful facts

Deriving Borjas's expressions for self-selection requires applying some properties of the normal distribution and a version of the law of iterated expectations; in case anyone is rusty:

Property 1. If a vector of random variables  $X \sim N(\mu, \Sigma)$ , then  $AX + b \sim N(A\mu + b, A\Sigma A')$ .<sup>5</sup>

Property 2. If  $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N \left( \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_{X,Y} \\ \sigma_{X,Y} & \sigma_y^2 \end{pmatrix} \right)$ , then  $(Y|X = x) \sim N \left( \mu_y + \rho_{X,Y} \left( \frac{\sigma_y}{\sigma_x} \right) (x - \mu_x), \sigma_y^2 (1 - \rho_{X,Y}^2) \right)$ .<sup>6</sup>

Property 3. For any non-stochastic function  $f(\cdot)$  and  $X = f(W)$ ,  $E(Y|X) = E(E(Y|W)|X)$ .<sup>7</sup>

Property 4. Let  $\phi(z)$  and  $\Phi(z)$  denote the PDF and CDF of the standard normal distribution, respectively. If  $\frac{v}{\sigma_v} \sim N(0, 1)$ , then  $E \left( \frac{v}{\sigma_v} \mid \frac{v}{\sigma_v} > z \right) = \frac{\phi(z)}{1 - \Phi(z)}$ ; we refer to this expression as the Inverse Mills Ratio. Because  $\phi(z) = \phi(-z)$  and  $1 - \Phi(z) = \Phi(-z)$ , we can also write the Inverse Mills Ratio as  $\lambda(z) = \frac{\phi(-z)}{\Phi(-z)}$ .<sup>8</sup>

### 3.3.3 Analyzing self-selection

In order to analyze self-selection, we want to derive expressions that let us compare  $E(\ln w_0 | I > 0)$  and  $E(\ln w_1 | I > 0)$  - that is, for individuals who immigrate we'd like to compare average log earnings in country 0 and average log earnings in country 1. Let's start with  $E(\ln w_0 | I > 0)$ , which can be re-written as follows:

<sup>5</sup>See, *e.g.*, page 198 of [Goldberger \(1991\)](#). You can prove this using the moment generating function.

<sup>6</sup>See, *e.g.*, page 175-177 of [Casella and Berger \(2001\)](#).

<sup>7</sup>This is a version of the law of iterated expectations. See, *e.g.*, page 31 of [Wooldridge \(2010\)](#).

<sup>8</sup>See, *e.g.*, page 672 of [Wooldridge \(2010\)](#).

$$E(\ln w_0 | I > 0) = E\left(\mu_0 + \epsilon_0 \left| \frac{v}{\sigma_v} > z \right.\right) \quad (18)$$

$$= \mu_0 + E\left(\epsilon_0 \left| \frac{v}{\sigma_v} > z \right.\right) \quad (19)$$

$$= \mu_0 + \sigma_0 E\left(\frac{\epsilon_0}{\sigma_0} \left| \frac{v}{\sigma_v} > z \right.\right) \quad (20)$$

The following steps are useful in deriving a simplified version of the  $E\left(\frac{\epsilon_0}{\sigma_0} \left| \frac{v}{\sigma_v} > z \right.\right)$  term:

1. Because  $\epsilon_0$  and  $\epsilon_1$  are jointly normally distributed, applying Property 1 you can show that  $\epsilon_0$  and  $v \equiv \epsilon_1 - \epsilon_0$  are jointly normally distributed:  $\begin{pmatrix} \epsilon_0 \\ \epsilon_1 - \epsilon_0 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_{0,1} - \sigma_0^2 \\ \sigma_{0,1} - \sigma_0^2 & \sigma_0^2 + \sigma_1^2 - 2\sigma_{0,1} \end{pmatrix}\right)$ .<sup>9</sup>
2. Given that  $\epsilon_0$  and  $v \equiv \epsilon_1 - \epsilon_0$  are jointly normally distributed, applying Property 2 you can show that  $E(\epsilon_0 | v) = \rho_{0,v} \left(\frac{\sigma_0}{\sigma_v}\right) v$ , where  $\rho_{0,v} = \frac{\sigma_{0,v}}{\sigma_0 \sigma_v}$ . Simplifying implies  $E(\epsilon_0 | v) = \frac{\sigma_{0,v}}{\sigma_0} v$ .
3. Applying Property 3, you can show that  $E\left(\frac{\epsilon_0}{\sigma_0} \left| \frac{v}{\sigma_v} > z \right.\right) = E\left(E\left(\frac{\epsilon_0}{\sigma_0} \left| \frac{v}{\sigma_v} \right.\right) \left| \frac{v}{\sigma_v} > z \right.\right)$ .<sup>10</sup>
4. Finally, it will be useful to have a simplified expression for  $E\left(\frac{\epsilon_0}{\sigma_0} \left| \frac{v}{\sigma_v} \right.\right)$ . Let  $s = \frac{v}{\sigma_v} \sim N(0, 1)$ . Applying Property 2, you can show that  $E(\epsilon_0 | s) = \frac{\sigma_{0,s}}{\sigma_s^2} s$ . Substituting  $\rho_{0,v} = \frac{\sigma_{0,v}}{\sigma_0 \sigma_v}$  gives:

$$E\left(\frac{\epsilon_0}{\sigma_0} \left| \frac{v}{\sigma_v} \right.\right) = \frac{1}{\sigma_0} E(\epsilon_0 | s) \quad (21)$$

$$= \frac{1}{\sigma_0} \frac{\sigma_{0,s}}{\sigma_s^2} s \quad (22)$$

$$= \frac{1}{\sigma_0} \frac{\frac{1}{\sigma_v} \text{cov}(v, \epsilon_0)}{1} \frac{v}{\sigma_v} \quad (23)$$

$$= \frac{\sigma_{0,v}}{\sigma_0 \sigma_v} \frac{v}{\sigma_v} \quad (24)$$

$$= \rho_{0,v} \frac{v}{\sigma_v} \quad (25)$$

<sup>9</sup>You can show this by letting  $A = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$ .

<sup>10</sup>You can show this by letting  $X = 1\left(\frac{v}{\sigma_v} > z\right)$  be a function of  $W = \frac{v}{\sigma_v}$ .

Returning to our derivation of  $E(\ln w_0|I > 0)$ , we now have:

$$E(\ln w_0|I > 0) = \mu_0 + \sigma_0 E\left(\frac{\epsilon_0}{\sigma_0} \middle| \frac{v}{\sigma_v} > z\right) \quad (26)$$

$$= \mu_0 + \sigma_0 E\left(E\left(\frac{\epsilon_0}{\sigma_0} \middle| \frac{v}{\sigma_v}\right) \middle| \frac{v}{\sigma_v} > z\right) \quad (27)$$

$$= \mu_0 + \sigma_0 E\left(\rho_{0,v} \frac{v}{\sigma_v} \middle| \frac{v}{\sigma_v} > z\right) \quad (28)$$

$$= \mu_0 + \sigma_0 \rho_{0,v} E\left(\frac{v}{\sigma_v} \middle| \frac{v}{\sigma_v} > z\right) \quad (29)$$

$$= \mu_0 + \sigma_0 \rho_{0,v} \left(\frac{\phi(z)}{1 - \Phi(z)}\right) \quad (30)$$

The last equality follows from Property 4. We can derive a similar expression for  $E(\ln w_1|I > 0)$ :

$$E(\ln w_1|I > 0) = E\left(\mu_1 + \epsilon_1 \middle| \frac{v}{\sigma_v} > z\right) \quad (31)$$

$$= \mu_1 + \sigma_1 \rho_{1,v} \left(\frac{\phi(z)}{1 - \Phi(z)}\right) \quad (32)$$

It will be useful to re-write these expressions for the expected wages of migrants in each country ( $E(\ln w_0|I > 0)$ ,  $E(\ln w_1|I > 0)$ ). Using that  $\sigma_{0,v} = \text{cov}(\epsilon_0, v) = E[\epsilon_0 \cdot (\epsilon_1 - \epsilon_0)] = \sigma_{0,1} - \sigma_0^2$ :

$$E(\ln w_0|I > 0) = \mu_0 + \sigma_0 \rho_{0,v} \left(\frac{\phi(z)}{1 - \Phi(z)}\right) \quad (33)$$

$$= \mu_0 + \sigma_0 \frac{\sigma_{0,v}}{\sigma_0 \sigma_v} \left(\frac{\phi(z)}{1 - \Phi(z)}\right) \quad (34)$$

$$= \mu_0 + \frac{\sigma_{0,v}}{\sigma_v} \left(\frac{\phi(z)}{1 - \Phi(z)}\right) \quad (35)$$

$$= \mu_0 + \frac{\sigma_{0,1} - \sigma_0^2}{\sigma_v} \left(\frac{\phi(z)}{1 - \Phi(z)}\right) \quad (36)$$

$$= \mu_0 + \frac{\sigma_0 \sigma_1}{\sigma_v} \left(\frac{\sigma_{0,1}}{\sigma_0 \sigma_1} - \frac{\sigma_0}{\sigma_1}\right) \left(\frac{\phi(z)}{1 - \Phi(z)}\right) \quad (37)$$

$$= \mu_0 + \frac{\sigma_0 \sigma_1}{\sigma_v} \left(\rho_{0,1} - \frac{\sigma_0}{\sigma_1}\right) \left(\frac{\phi(z)}{1 - \Phi(z)}\right) \quad (38)$$

Analogously for  $E(\ln w_1|I > 0)$ , substituting  $\sigma_{1,v} = \sigma_1^2 - \sigma_{0,1}$ , we have:

$$E(\ln w_1|I > 0) = \mu_1 + \sigma_1 \rho_{1,v} \left(\frac{\phi(z)}{1 - \Phi(z)}\right) \quad (39)$$

$$= \mu_1 + \frac{\sigma_0 \sigma_1}{\sigma_v} \left(\frac{\sigma_1}{\sigma_0} - \rho_{0,1}\right) \left(\frac{\phi(z)}{1 - \Phi(z)}\right) \quad (40)$$

Define  $\mu_j + Q_j$  as the expected wage of migrants in country  $j$ . In order to understand the position of migrants in the distribution of workers in each country (that is, whether migrants are positively or negatively selected), we want to know the signs of  $Q_0$  and  $Q_1$ :

$$Q_0 \equiv E(\epsilon_0|I > 0) = \frac{\sigma_0\sigma_1}{\sigma_v} \left( \rho_{0,1} - \frac{\sigma_0}{\sigma_1} \right) \left( \frac{\phi(z)}{1 - \Phi(z)} \right) \quad (41)$$

$$Q_1 \equiv E(\epsilon_1|I > 0) = \frac{\sigma_0\sigma_1}{\sigma_v} \left( \frac{\sigma_1}{\sigma_0} - \rho_{0,1} \right) \left( \frac{\phi(z)}{1 - \Phi(z)} \right) \quad (42)$$

### 3.3.4 Four cases of immigrant selection

1. **Positive selection:**  $Q_0 > 0$  and  $Q_1 > 0$ . Migrants are positively selected relative to either country's income distribution  $\iff \rho_{0,1} > \frac{\sigma_0}{\sigma_1}$ .<sup>11</sup> This requires a high correlation between the value of skills in countries 0 and 1, and that income is more dispersed in the US than in country 0. Borjas's example is high-skilled migration from Western Europe.
2. **Negative selection:**  $Q_0 < 0$  and  $Q_1 < 0$ . Migrants are negatively selected relative to either country's income distribution  $\iff \rho_{0,1} > \frac{\sigma_1}{\sigma_0}$ .<sup>12</sup> This requires a high correlation between the value of skills in countries 0 and 1, and that income is *less* dispersed in the US than in country 0. Borjas's example is the US social safety net drawing low-skilled immigrants from countries with less of a social safety net.
3. **Refugee selection:**  $Q_0 < 0$  and  $Q_1 > 0$ . Migrants are negatively selected relative to the home country income distribution, but fall in the top of the US income distribution  $\iff \rho_{0,1} < \min(\frac{\sigma_0}{\sigma_1}, \frac{\sigma_1}{\sigma_0})$ . This requires a low correlation between the value of skills in country 0 and in country 1. Borjas argues this may be the case for countries that have recently experienced a Communist takeover.
4. **No fourth case:**  $Q_0 > 0$  and  $Q_1 < 0$ . Mathematically, this case is ruled out because it would require  $\rho_{0,1} > 1$ .

### 3.3.5 A note on the joint normality assumption

As an econometrician, what you observe is individuals' migration decisions (whether they moved to the US or stayed in their home country), data on the US wages of migrants  $E(\ln w_1|I > 0)$ , and data on the home country wages of non-migrants  $E(\ln w_0|I > 0)$ . Given this data, we would like to know the joint distribution of  $\ln w_0$  and  $\ln w_1$  so that we can make statements about where migrants fall in the home and US country income distributions. Heckman and Honore (1990) show that the joint normality assumption in the original Roy model allows you to identify the joint distribution of  $\ln w_0$  and  $\ln w_1$  in a single cross section of data, but that if you relax this joint normality assumption the Roy model is no longer identified. French and Taber (2011) give some intuition for these results.

<sup>11</sup>Note that this implies  $\frac{\sigma_1}{\sigma_0} > 1$ , since  $\rho_{0,1} \leq 1$ .

<sup>12</sup>Note that this implies  $\frac{\sigma_0}{\sigma_1} > 1$ , since  $\rho_{0,1} \leq 1$ .

### 3.4 Testing the Roy model: [Abramitzky, Boustan and Eriksson \(2012\)](#)

Research investigating patterns of immigration in recent decades has generally provided mixed evidence for the Roy model. For example, [Chiquiar and Hanson \(2005\)](#) document evidence against negative selection of Mexican migrants (as would be predicted by the Roy model). I'm going to focus on discussing a (second) recent paper by Ran Abramitzky, Leah Boustan, and Katherine Eriksson that investigated the predictions of the Roy model during the Age of Mass Migration (1850-1913). Because the US maintained essentially open borders during this period, this setting is a natural one in which to test for patterns of self-selection free of the legal factors that have governed migration in more recent years.

[Abramitzky, Boustan and Eriksson \(2012\)](#) investigate whether migrants were positively or negatively selected from the European population, focusing on Norwegian migrants to the US. During the age of mass migration, Norway had a more unequal income distribution than did the US (as illustrated in Figure 1, and supported in the paper by other past work):

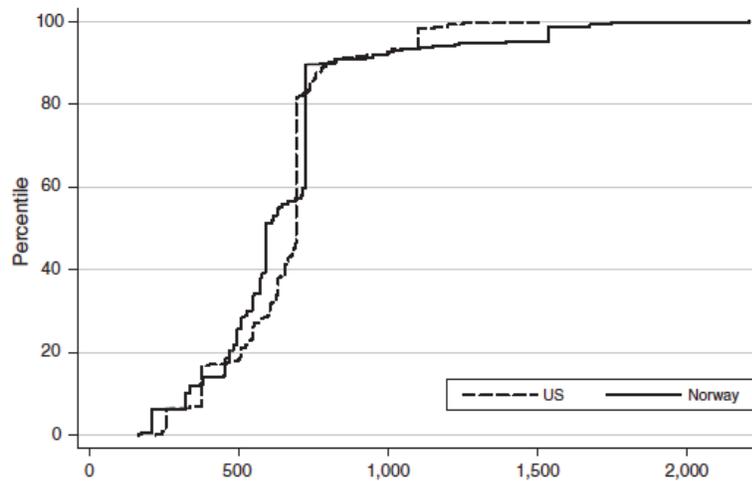


FIGURE 1. CUMULATIVE INCOME DISTRIBUTION FUNCTIONS IN THE UNITED STATES AND NORWAY IN 1900

*Notes:* US and Norwegian distributions contain all men aged 38 to 50 in the respective censuses of 1900. The x-axis is scaled in 1900 US dollars. Individuals are assigned the mean earnings for their occupation and are arrayed from lowest- to highest-paid occupations. The Norwegian distribution is rescaled to have the same mean as the US distribution (the actual Norwegian and US means are US\$(1900)350 and US\$(1900)643, respectively).

Courtesy of Ran Abramitzky, Leah Platt Boustan, Katherine Eriksson, and the American Economic Association. Used with permission.

Given this pattern that (at the time) income in Norway was more dispersed than was income in the US, the Roy model predicts that Norwegian migrants will be negatively selected. A key contribution of this paper is collecting the data necessary to test this prediction. They combine two fully digitized versions of the Norwegian censuses (1865 and 1900) with a dataset of all Norwegian-born men in the US in 1900 using the now-publicly-available census records from 1900 (the latter is drawn from the same data as they used in [Abramitzky, Boustan and Eriksson \(2014\)](#)). They match migrants and stayers in the US and Norwegian censuses of 1900 to birth families in 1865 based on name and age. The earnings-related outcome they observe is occupation, again as in [Abramitzky, Boustan and Eriksson \(2014\)](#).

Empirically, they document mixed evidence on the selection of rural-born Norwegian men, but strong evidence of negative selection among urban-born Norwegian men. That is, in the

urban sample men with poorer economic prospects in Norway were more likely to migrate to the US. They provide two pieces of direct evidence on this negative selection. First, they compare the occupational distributions of migrants and stayers. Figure 3 (Panel B, focusing on the urban sample) arrays occupations from lowest-paid to highest-paid according to the average US earnings in that occupation. Migrants are more likely to hold low-paying jobs, while men remaining in Norway exhibit an occupational distribution that is skewed towards higher-paying jobs. Figure 3 Panel A (not shown, focusing on the rural sample) suggests that men born in rural areas are employed in similar jobs in both countries.

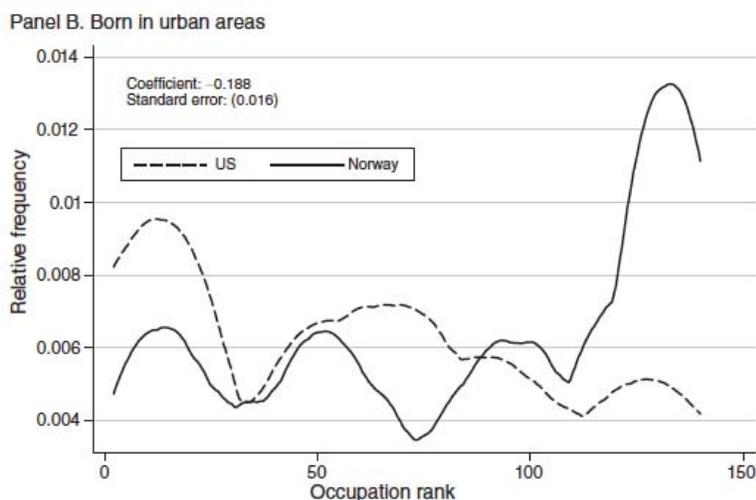


FIGURE 3. COMPARING THE OCCUPATIONAL DISTRIBUTIONS OF NORWEGIAN-BORN MEN IN THE UNITED STATES AND NORWAY IN 1900

*Notes:* Each figure presents the relative frequency of 144 earning categories (representing 189 distinct occupations) for Norwegian-born men in the United States and in Norway. All men are assigned the mean US earnings in their occupation. Men are divided by rural or urban place of birth. Farmers, the largest occupational category, is excluded from the figure for reasons of scale. We report coefficients and standard errors from OLS regressions of  $\ln(\text{earnings})$  on a dummy for living in the United States controlling for a quadratic in age.

Courtesy of Ran Abramitzky, Leah Platt Boustan, Katherine Eriksson, and the American Economic Association. Used with permission.

Second, they compare fathers of migrants and non-migrants. Table 4 compares the occupations, asset holdings, and property tax values of the heads of migrant and nonmigrant households in 1865. For both the urban and rural samples, this analysis provides evidence of negative selection. For example, in the urban sample heads of migrant households are 5.8 percentage points less likely to hold assets than are heads of non migrant households.

TABLE 4—ECONOMIC OUTCOMES OF HOUSEHOLD HEADS WITH MIGRANT AND NONMIGRANT SONS, 1865

Dependent variables	All households		Households with matched sons	
	Mean	Coefficient on migrant HH	Mean	Coefficient on migrant HH
<i>Panel A. Urban</i>				
Occ. > median	0.593	−0.001 (0.022)	0.583	−0.030 (0.042)
Occupational income	428.27	−27.501 (10.216)	440.47	−26.555 (20.798)
Assets	0.260	−0.030 (0.018)	0.252	−0.058 (0.035)
<i>N</i>		4,038		1,074
<i>Panel B. Rural</i>				
Occ. > median	0.608	0.008 (0.014)	0.577	−0.054 (0.032)
Occupational income	321.21	6.092 (3.847)	315.30	−9.077 (9.072)
Assets	0.665	−0.032 (0.012)	0.613	−0.035 (0.028)
Match tax records	0.130	−0.037 (0.009)	0.134	−0.040 (0.021)
Property tax bill	2.759	−0.372 (0.307)	2.821	0.044 (0.887)
<i>N</i>		12,177		2,499

*Notes:* Results for Match 1. The left-hand panel includes all households while the right-hand panel considers only households in which all sons can be matched to 1900. In each panel, sample means are reported in Column 1 and the coefficient and standard error from a regression of each dependent variable on an indicator for being the head of a migrant household is reported in Column 2. Regressions also control for a quadratic in head's age and a series of province dummies. We assign income levels to household heads using mean Norwegian earnings by occupation in 1900. Above-median occupations are those earning more than US\$(1900)311 (urban) and US\$(1900)393 (rural) per year. Assets is an indicator variable equal to 1 for men who own a business, own land, or are master craftsmen in an artisanal workshop. Tax records refers to the 1886 Land Registers. For fathers who match to the tax records, we report the value of the property tax bill in 1900 US dollars. The data is provided in speciedalers and marks. In 1875, one speciedaler was equal to four Norwegian kroner. The number of households that can be matched to the property tax records is reported below that variable.

Courtesy of Ran Abramitzky, Leah Platt Boustan, Katherine Eriksson, and the American Economic Association. Used with permission.

In the paper, they also infer the direction of migrant selection indirectly by comparing an OLS estimate of the return to migration (comparing the earnings of migrants with the earnings of stayers) with a family fixed effects estimate of the return to migration (comparing the earnings of migrants with earnings of their stayer brothers). If the OLS estimate measures the return to migration plus a selection term, and if migrants are negatively selected, then the OLS estimate will be smaller than the family fixed effect estimate. While the family fixed effect estimate is clearly not free from selection concerns, this analysis can inform the direction of across-household selection into migration. As shown in Table 3, this indirect method also provides evidence in favor of negative selection.

TABLE 3—OLS AND WITHIN-HOUSEHOLD ESTIMATES OF THE RETURN TO MIGRATION.  
HOUSEHOLDS WITH TWO OR MORE MEMBERS IN THE MATCHED SAMPLE

	Dependent variable = $\ln(\text{earnings})$ ; Coefficient on = 1 if migrant		
	Full sample, 1865	Rural, 1865	Urban, 1865
<i>Panel A. Unweighted</i>			
OLS	0.545 (0.027)	0.607 (0.034)	0.384 (0.044)
Within household	0.511 (0.035)	0.508 (0.045)	0.508 (0.057)
Chi-squared	1.49	7.47	8.31
<i>p</i> -value	0.2218	0.0063	0.0039
<i>N</i>	2,655	1,823	832
Number of migrant-stayer pairs	326	167	159
<i>Panel B. Weighted</i>			
OLS	0.586 (0.029)	0.609 (0.033)	0.443 (0.067)
Within household	0.542 (0.039)	0.529 (0.042)	0.561 (0.049)
Chi-squared	2.13	4.60	5.65
<i>p</i> -value	0.1441	0.0320	0.0175
<i>N</i>	2,241	1,666	306
Number of migrant-stayer pairs	269	140	129

*Notes:* Each cell contains coefficient estimates and standard errors from regressions of  $\ln(\text{earnings})$  on a dummy variable equal to one for individuals living in the United States in 1900. Regressions also include controls for age and age squared. In each panel, the first row conducts an OLS regression for the restricted sample of households that have at least two matched members in the dataset and the second row adds household fixed effects. Panel B contains results from regressions weighted to reflect the urban status (full sample only), asset holdings, and occupational distribution of fathers in the full population. We conduct chi-squared tests of the null hypothesis that the OLS and within-household coefficients are equal.

Courtesy of Ran Abramitzky, Leah Platt Boustan, Katherine Eriksson, and the American Economic Association. Used with permission.

Of course, even within households brothers can differ in unmeasured personal attributes, so it is difficult to interpret even the family fixed effect estimates as estimates of the returns to migration. In Appendix A of their paper, [Abramitzky, Boustan and Eriksson \(2012\)](#) document an analysis estimating the returns to migration using the gender composition of a man’s siblings and his place in the household birth order as instruments for migration. The idea is that both of these factors influence a man’s expectation of inheriting farmland in Norway and therefore his probability of migrating to the US; while the exclusion restrictions for these variables need not obviously be satisfied, the authors provide some evidence that the necessary assumptions appear reasonable in their context.

In terms of take-aways from this paper, there are three things that I would stress. First, this is a very recent paper that was providing new, interesting evidence testing the predictions of the Roy model. This is a classic question, but that doesn’t mean that there isn’t room for more good papers on it! Second, this paper as well as [Abramitzky, Boustan and Eriksson \(2014\)](#) highlight the value of looking for the ‘right’ empirical setting (testing for self-selection in migration during

a time period with open borders) and of constructing the ‘right’ data (the key tests of this paper are just tabulations of the data, but that’s because the authors did an enormous amount of work in order to construct data that would allow for transparent empirical tests). Finally, this paper as well as [Abramitzky, Boustan and Eriksson \(2014\)](#) are great examples of how economic history can overlap nicely with core questions in labor economics. There is a lot of exciting work being done in economic history that overlaps with labor and other applied micro fields, and I would encourage you to keep historical settings in mind as you are thinking about settings in which to investigate research questions of contemporary interest (in addition to what I think of as the more ‘traditional’ focus of economic history, which is shedding light on the long-run impacts of economic phenomena).

## 4 Application: Health care

Despite its origin in labor economics, the Roy model has been applied across a wide range of fields in economics. As an example, I’m going to talk in detail about an application of the Roy model from the field of health economics - [Chandra and Staiger \(2007\)](#). This is 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.

### 4.1 Brief background on geographic variation in medical expenditures

The earliest work I’m aware of that documented geographic variation in medicine is [Glover \(1938\)](#). Using a variety of data sources from England and Wales, Glover documented significant variation in small-area tonsillectomy rates. Despite looking for correlations with “any factor which might have some ætiological bearing on chronic tonsillitis and adenoidal growths - such factors for example as overcrowding and unemployment...not the slightest suggestion of correlation has been obtained.” Maybe not the regressions we would estimate today, but the start of a puzzle!

In his recent *Handbook of Health Economics* review, [Skinner \(2012\)](#) provides an overview of the economics and medical literatures that have subsequently documented variation in the use of medical care across “observably similar” patients. Perhaps the key reference in this area is the Dartmouth Atlas, which has used the Medicare claims data to document tremendous variation in expenditures across “hospital referral regions” (markets defined to include at least one hospital offering some key services such as cardiovascular treatments). While other market definitions and other datasets have been used (such as private claims data documenting the use of medical services in the under-65 non-Medicare population), many of the key ‘facts’ that have shaped peoples’ thinking in this area have come from the Dartmouth Atlas. The position taken by the [Skinner \(2012\)](#) and [Chandra and Staiger \(2007\)](#) is that explanations such as differences in income, patient preferences, and underlying health status don’t explain these variations. Other work has shown that adjusting for prices doesn’t hugely change these factors either - although price adjustments do matter in some places, like the Bronx and Manhattan ([Gottlieb et al.](#),

2010). Given that most research has examined the importance of one factor in isolation, we don't currently have a good sense of what share of the variation could be explained by these factors when taken together.

The consensus view from the Dartmouth Atlas is that this geographic variation in health spending is not associated with improved satisfaction, outcomes, or survival. This conclusion should be digested together with recent research that has suggested that - at least in some contexts - more spending *is* associated with better health outcomes (see, *e.g.*, Cutler (2005) and Joe Doyle's line of research). That said, for now let's take as a first fact that geographic variation in medical expenditures is not associated with improved health outcomes. This first fact is surprising in light of the second fact that many (not all) new medical technologies are shown in randomized clinical trials to be associated with improvements in health outcomes (if you are interested, ask me more about how this interacts with product entry regulation that differs across pharmaceuticals vs. medical procedures and medical devices).

Chandra and Staiger (2007) discuss how these two facts are often interpreted as evidence of diminishing returns. Randomized trials tend to be conducted on groups of patients most likely to benefit from treatment, whereas the lack of a cross-sectional relationship between spending and outcomes could be explained by a "flat of the curve" argument where physicians perform the intervention until the marginal return is zero. Chandra and Staiger outline three problems with this argument:

1. No explanation of why we observe geographic variation in the first place.
2. Still predicts a positive relationship between medical spending and patient outcomes unless all areas are in the range of zero or negative marginal benefits; this has never been documented in the literature.
3. Predicts that the marginal benefit from more intensive patient treatment should be lower in areas that treat more aggressively. However, the available evidence from US-Canada comparisons suggest the opposite: even though the US treats heart attacks more intensively than does Canada, the marginal benefit from intensive heart attack treatments appears to be larger in the US.

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.

## 4.2 Preliminaries: What motivates a model with productivity spillovers?

In the Chandra-Staiger model, patients can receive either of two alternative treatments: non-intensive management (medical management, denoted by subscript 1), and intensive intervention (surgery, denoted by subscript 2). Physicians choose the treatment option for each patient that maximizes utility on the basis of the expected survival rate ( $Survival_1$ ,  $Survival_2$ ) and

cost ( $\text{Cost}_1, \text{Cost}_2$ ). The productivity spillovers component of the model arises because survival and cost are positively related to the fraction of patients who receive the same treatment ( $P_1, P_2 = 1 - P_1$ ).<sup>13</sup>

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”). Some evidence exists for this, although there’s room for tightening the links on what implications this type of model would have (*e.g.* gradual vs. immediate adjustment to shocks, comparative statics across procedures where this should be more vs. less important).
2. Availability of support services. Some places have cardiac catheterization labs whereas other don’t, which *a priori* seems likely to be important. Obviously technology adoption is a choice variable, but this is one mechanism through which practice style may depend on the mix of patients in your area (excluding yourself).
3. Selective migration. Physicians who are more skilled at the intensive treatment may self-select into areas that treat more intensively. As we’ll discuss, this has different welfare implications than the first two stories.

### 4.3 Model

To the basic framework outlined above, add heterogeneity across patients that affects both expected survival and cost: some of this heterogeneity can be captured by observable patient characteristics ( $Z$ ); other factors ( $\epsilon$ ) are known to the patient and physician at the time of choosing treatment but are not observed by the econometrician. This is the Roy model component of the model: patients are sorted into the two treatments based on expected returns.

Putting these pieces together, for treatments  $i \in \{\text{nonintensive}, \text{intensive}\}$ , let the survival rate and cost associated with each treatment take the following forms:

$$\text{Survival}_i = \beta_i^s Z + \alpha_i^s P_i + \epsilon_i^s \text{ for } i = 1, 2 \quad (43)$$

$$\text{Cost}_i = \beta_i^c Z + \alpha_i^c P_i + \epsilon_i^c \text{ for } i = 1, 2 \quad (44)$$

Let  $\lambda$  represent the value of life (survival per dollar), capturing the trade-off being made by physicians/patients between improved survival and increased cost. Note that  $\lambda$  could be zero, if - for example, because of insurance - medical decisions are made without regard to the financial cost of treatment. We can then write the patient’s indirect utility  $U$  as:

$$U_i = \text{Survival}_i - \lambda \text{Cost}_i = \beta_i Z + \alpha_i P_i + \epsilon_i \text{ for } i = 1, 2 \quad (45)$$

---

<sup>13</sup>This parametrization of productivity spillovers has been used in other papers looking at network externalities, such as [Katz and Shapiro \(1985\)](#). Note that Chandra and Staiger focus on *geographic* spillovers, whereas in other contexts these spillovers can be market-wide.

where  $\beta_i = \beta_i^s - \lambda\beta_i^c$ ,  $\alpha_i = \alpha_i^s - \lambda\alpha_i^c$ , and  $\epsilon_i = \epsilon_i^s - \lambda\epsilon_i^c$ .  $\beta_i Z$  represents an index of how medically appropriate a given patient is for each treatment,<sup>14</sup>  $\alpha_i P_i$  captures the productivity spillover,<sup>15</sup> and  $\epsilon_i$  represents unobservables that influence survival and cost. An individual is treated intensively ( $i = 2$ ) if  $U_2 > U_1$  (that is, treatment choice maximizes patient utility, not accounting for the externalities that the treatment decision for that patient will have on the treatments for other patients):

$$\Pr\{\text{intensive treatment}\} = \Pr\{i = 2\} \quad (46)$$

$$= \Pr\{U_2 - U_1 > 0\} \quad (47)$$

$$= \Pr\{\beta_2 Z + \alpha_2 P_2 + \epsilon_2 - \beta_1 Z - \alpha_1(1 - P_2) - \epsilon_1 > 0\} \quad (48)$$

$$= \Pr\{P_2(\alpha_1 + \alpha_2) - \alpha_1 + (\beta_2 - \beta_1)Z > \epsilon_1 - \epsilon_2\} \quad (49)$$

$$= \Pr\{\alpha P_2 - \alpha_1 + \beta Z > \epsilon\} \quad (50)$$

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:

$$E[U_2 - U_1 | U_2 - U_1 > 0] = \beta Z + \alpha P_2 - \alpha_1 + E[\epsilon | U_2 - U_1 > 0] \quad (51)$$

Thus, patients receiving the intensive treatment will have a higher expected utility gain if they are more appropriate (higher  $\beta Z$ ) or if they live in a more intensive region (higher  $\alpha P_2$ ). One way of thinking about the intuition here is that patients are given the best care conditional on where you live, but marginal patients would be better off in an area with the other specialization.

#### 4.4 Equilibrium

In equilibrium, the fraction of patients choosing intensive treatment ( $P_2$ ) must match the demand equation for  $\Pr\{\text{intensive treatment}\}$  outlined above. Intuitively, this requires that the proportion of patients choosing intensive treatments must generate benefits (including the productivity spillovers) that are consistent with that proportion. Letting  $f(Z)$  represent the distribution of  $Z$  in the population, this implies the following equilibrium (fixed point) condition:

$$P_2 = \int_Z \Pr\{\alpha P_2 - \alpha_1 + \beta Z > \epsilon\} f(Z) dZ \quad (52)$$

$$= G(P_2) \quad (53)$$

<sup>14</sup>What you should have in mind for ‘appropriateness’ here is something like heterogeneity by age, where very intensive treatments are high risk for older (more frail) patients.

<sup>15</sup>That is, if there were no productivity spillovers, a patient’s utility from a treatment should be unaffected by how other patients are treated, in which case  $\alpha_i$  would be zero.

Variation across areas in the use of the intensive treatment can arise for two reasons, illustrated in Chandra-Staiger's Figure 1: multiple equilibria, or single equilibrium that are determined by small differences in patient characteristics. Distinguishing between these two cases matters for welfare.

In Figure 1(A), there are two stable equilibria: an intensive equilibrium in which most patients receive intensive treatment and the returns to doing so are high, and a nonintensive equilibrium in which so few patients receive intensive treatment that the returns to doing so are (relatively) low. Chandra-Staiger don't have predictions on what determines the choice among multiple equilibria.

In Figure 1(B), variation across areas instead arises because of differences in the distribution of patients across areas: if most patients in an area are appropriate for intensive treatments, the intensive equilibrium will arise, and vice versa. Productivity spillovers imply that even small differences in patient types across areas could be magnified into large equilibrium differences in specialization across areas.

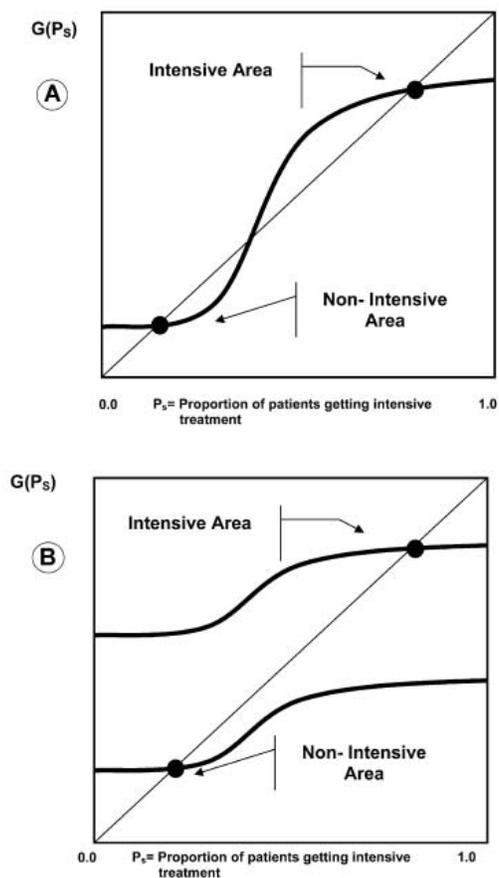


FIG. 1.—Characterizations of area variations: A, multiple equilibrium; B, single equilibrium.

© 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/>.

In Chandra-Staiger’s Figure 2, they ignore patient-level unobservables ( $\epsilon$ ) and plot patient utility as a function of appropriateness for intensive treatment ( $Z$ ) for each treatment option. This is really the key figure in the paper, so it’s worth making sure you understand what they are doing here. You can think of  $Z$  as a propensity score that predicts clinical appropriateness (say, as a function of age and comorbidities);  $Z$  is plotted on the  $x$ -axis, almost like the running variable in a regression discontinuity design. Patients further to the left on the  $x$ -axis are less appropriate for intensive treatment and patients further to the right on the  $x$ -axis are more appropriate for intensive treatment. Figure 2(a) plots patient utility on the  $y$ -axis for the two treatments, within a given area: less appropriate patients receive higher utility from nonintensive treatment, whereas more appropriate patients receive higher utility from intensive treatment. The gap between the intensive and nonintensive curves is greater for more appropriate patients, implying that the return to intensive treatment is higher for patients who are more appropriate for intensive treatment. Figure 2(b) plots these patient utility curves for two areas which differ in their treatment intensity, clarifying several predictions. First, patients less appropriate for the intensive treatment are worse off intensive areas (as a result of the productivity spillover). Second, intensive areas treat less appropriate patients on the margin. Third, patients more appropriate for the intensive treatment are better off in an intensive area (again, as a result of the productivity spillover).

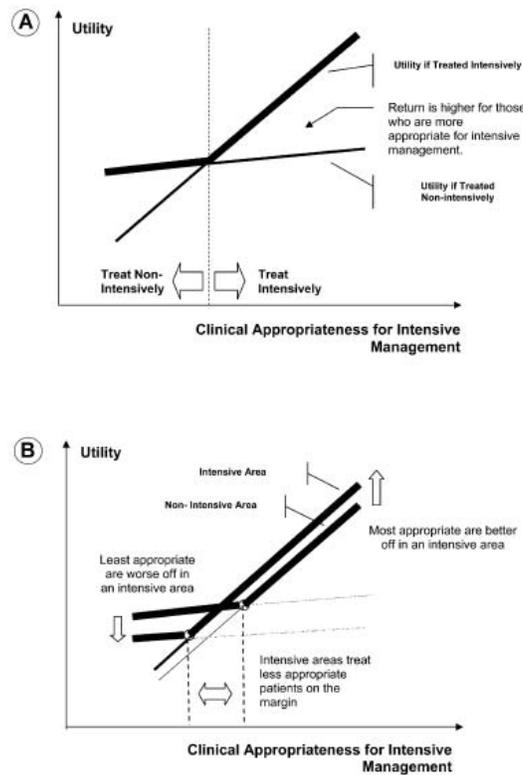


FIG. 2.—Graphical illustration of the Roy model with productivity spillovers. *A*, Relationship between two alternative ways to treat patients within an area. The production possibilities frontier describes the best treatment for a patient of given clinical appropriateness. The model predicts that the returns to intensive management are increasing in patients’ appropriateness for such interventions. *B*, Contrasts the care across two areas that differ in their surgical intensity. As a result of the productivity spillover, patients appropriate for intensive management are better off in the surgically intensive areas, whereas patients appropriate for nonintensive management are worse off in such areas.

In general, this paper has an excellent discussion of alternative models and their implications - some of which they are able to test empirically. See, *e.g.* the discussion of productivity differences (in the absence of spillovers) on p.114, the flat-of-the-curve model on p.115, and an alternative way of modeling productivity spillovers in Footnote 3.

#### 4.5 Welfare

In this model, spillovers imply that an increase in the share of patients receiving the intensive treatment has a positive externality on some patients (those receiving the intensive treatment) and a negative externality on others (those receiving the nonintensive treatment). Unsurprisingly, externalities suggest the equilibrium may not be optimal.

If all areas have identical patient distributions and variation in use of intensive treatments across areas arises from multiple equilibria, then the usual “area approach” of comparing patient survival and costs across areas can be used to determine the optimal equilibrium rate of intensive treatment: this view implies we can find the optimum, and if appropriate then cut spending without negatively impacting outcomes. However, if we are in the single equilibrium case and differences across areas arise due to differences in patient distributions, then we would reach the opposite conclusion: expected patient utility in more intensive regions would be raised by increasing intensive treatment rates above their equilibrium value. That is, in a single equilibrium world, from a welfare perspective there is *too little* area variation in treatment as long as the marginal patient ignores the externality she imposes on other patients. These different conclusions suggest that it is important to understand whether the observed geographic variations are the consequence of single or multiple equilibria.

#### 4.6 Data and estimation

As with much (too much?) of the health economics literature, Chandra and Staiger test their model in the context of heart attacks (‘acute myocardial infarctions,’ or AMIs). This context is convenient because this is a common condition with extensive data (Medicare claims and CCP chart data), a relatively high mortality rate (implying mortality is a meaningful outcome), and a limited role for patients to select providers (because of urgent nature of the condition).

There is a clear mapping from treatments in this market to the intensive/nonintensive procedures in the model: beta blockers - a form of medical management - as the nonintensive treatment, and cardiac catheterization - a marker for receiving angioplasty or bypass surgery - as the intensive treatment. (Although note that all patients should be prescribed beta blockers.) Chandra and Staiger use ‘standard’ market definitions - dividing the US into 306 ‘hospital referral regions’ (as in *The Dartmouth Atlas of Health Care*) - and document substantial variation in the use of the intensive and nonintensive treatments across these markets. Patients are assigned to an HRR based on their residence, not the hospital at which they receive treatment.

The data they use is the Health Care Financing Administration’s Health Care Quality Improvement Initiative Cooperative Cardiovascular Project (CCP) data (so-called ‘chart data’)

linked to Medicare insurance claims data. This chart data gives detailed patient observables that are taken down at the time of admission, which are vastly superior to the patient covariates included in normal claims data. The only other data I know of that contains a similar level of detail is cancer registry data (worth checking out).

Chandra and Staiger partition patients into groups based on their appropriateness for intensive treatments, and index these groups by  $k$ . Then for  $\text{Outcome}_{ijk} \in \{\text{Survival}_{ijk}, \text{Cost}_{ijk}\}$  for patient  $i$  in HRR  $j$ , their key estimating equations are of the following form:

$$\text{Outcome}_{ijk} = \beta_{0k} + \beta_{1k}\text{Intensive Treatment}_i + \mathbf{X}_i\Pi_k + u_{ijk} \quad (54)$$

Even conditional on the relatively detailed covariates available in the CCP chart data, there's a concern that intensive treatment will be administered to those who will benefit most from treatment, which would bias OLS estimates. Specifically, the 'usual' concern is that (unobservably) sicker patients will receive more health inputs, which would bias a correlation of health inputs and survival towards finding that health inputs are bad for your health. In regressions where this is a concern (which is some but not all of their regressions), Chandra and Staiger follow [McClellan et al. \(1994\)](#) in using 'differential distance' as an instrumental variable - defined as the distance between the patient's zip code of residence and the nearest catheterization hospital minus the distance to the nearest noncath hospital. This is the 'classic' health economics instrument and the standard IV checks on it have been presented in a number of past papers, which is one reason why only a limited number of checks are presented in the current paper.

The intuition we developed for the propensity score interpretation of the  $x$ -axis in Figure 2 is empirically implemented using  $\Pr(\text{Cardiac Cath}_{ij}) = \hat{G}(\theta_0 + \mathbf{X}_i\Phi)$  as a measure of clinical appropriateness for cardiac catheterization.

## 4.7 Results

**Table 1.** Patients with higher ‘cath propensity’ benefit more from intensive treatment: Column (3) of Panel B, above versus below median propensity IV estimates of 0.038 versus 0.002. The age cut is an alternative parametrization of cath propensity, because clinical guidelines recommend treating patients over age 80 nonintensively; the results in Panel C are similar. In both Panel B and Panel C, the larger IV estimates for more appropriate patients come from both higher survival and lower costs. Consistent with Roy model prediction.

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

SAMPLE	INSTRUMENTAL VARIABLE ESTIMATES OF		
	Impact of Cath		
	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 (.036)	9.086 (1.810)	.016 (.005)
B. By cath propensity:			
Above the median ( $N = 64,799$ )	.184 (.034)	4.793 (1.997)	.038 (.017)
Below the median ( $N = 65,096$ )	.035 (.083)	17.183 (3.204)	.002 (.005)
Difference	.149 (.090)	-12.39 (3.775)	.036 (.018)
C. By age:			
65–80 ( $N = 89,947$ )	.171 (.037)	6.993 (1.993)	.024 (.009)
Over 80 ( $N = 39,948$ )	.016 (.108)	16.026 (2.967)	.001 (.007)
Difference	.155 (.114)	-9.033 (3.574)	.023 (.011)

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.

**Table 2.** Split the sample by above versus below median values of the instrument. Instrument predicts cath (first stage, 48.9 - 42.8 = 6.1 percentage points) and survival (second stage, 67.6 - 66.7 = 0.9 percentage points) but does not predict differences in ‘predicted survival’ (basically a placebo check, 67.5 - 67.2 = 0.3 percentage points). Intuitively, the lack of a difference in predicted survival says that a summary measure of pre-period fixed patient factors that affect survival do not differ across individuals with above- vs. below-median values of the instrument, which is comforting. Obviously the hope is that this means that unobservable determinants of survival are also balanced across these groups.

TABLE 2  
RELATIONSHIP BETWEEN DIFFERENTIAL DISTANCE (DD) AND PROBABILITY OF CATHETERIZATION AND SURVIVAL, AND DIFFERENTIAL DISTANCE AND OBSERVABLE CHARACTERISTICS (%)

SAMPLE	30-DAY CATH RATE		ONE-YEAR SURVIVAL		ONE-YEAR PREDICTED SURVIVAL		30-DAY PREDICTED CATH RATE FOR PATIENTS GETTING CATH	
	DD Below Median	DD Above Median	DD Below Median	DD Above Median	DD Below Median	DD Above Median	DD Below Median	DD Above Median
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(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.

© 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/>.

**Table 3 and Figure 3.** Here, Chandra-Staiger implement a test in the spirit of Gruber et al. (1999). They use the patient-level propensity score variable as the *outcome*, and limit the sample to those patients who received cath. Their explanatory variable of interest is (log, risk-adjusted) HRR-level cath rates: if the coefficient on this variable is negative, that means that average appropriateness of patients receiving cath is lower in more intensive areas. Consistent with this, the first coefficient in Column (2) suggests a negative relationship. They illustrate this relationship graphically in Figure 3. Consistent with Roy model prediction.

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 (.002)	-.045 (.008)
Over age 80	.125 (.002)	.063 (.012)
Not eligible for cath using ACC/AHA guidelines	.028 (.001)	.010 (.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 AML 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.

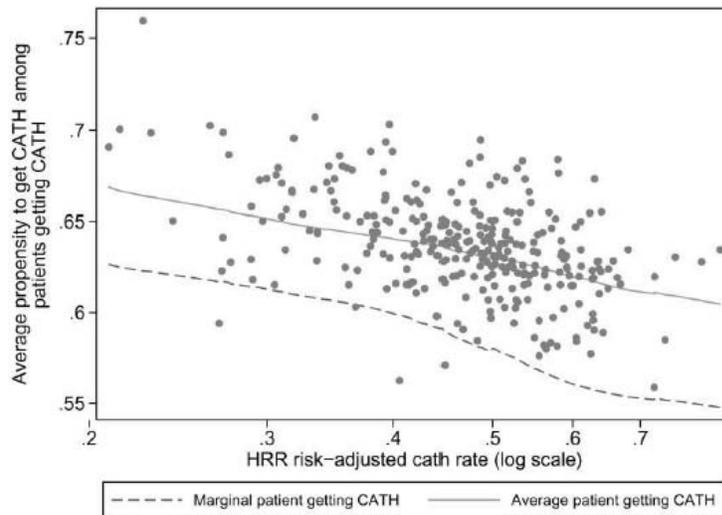


FIG. 3.—Relation between average patient and marginal patient receiving cardiac catheterization. For each of the 306 HRRs we graph the average propensity to receive cardiac catheterization (among patients who actually received it) against the log of the area risk-adjusted cath rate. Using local regression, we estimated the relationship between the average propensity and the risk-adjusted cath rate and the slope of this line at each point. These estimates were then used to plot the average (upper line) and marginal patient (lower line and estimated as the local difference in the average) receiving treatment.

**Table 4.** Quality of nonintensive treatment is worse in areas that treat more intensively: risk-adjusted beta blocker rate (a ‘good practice’ that is a standard measure of hospital quality) is negatively correlated with risk-adjusted cath rate at the HRR level (-0.31). Consistent with productivity spillovers.

TABLE 4  
HRR-LEVEL MEASURES OF INTENSIVE TREATMENT, MEDICAL MANAGEMENT, SUPPORT OF MEDICAL TREATMENT, AND DEMOGRAPHIC CHARACTERISTICS

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

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/>.

**Table 5.** Patents are more likely to receive cath if they live in areas where the average patient is more appropriate for cath (that is, they are regressing whether you get cath on the average propensity score for patients in your HRR): a one percentage point increase in the average propensity of patients in your HRR implies a 0.53 percentage point increase in the probability you receive cath. Consistent with productivity spillovers.

TABLE 5  
OLS RESTIMATES OF THE RELATIONSHIP BETWEEN PROBABILITY OF RECEIVING CATHETERIZATION AND HRR PATIENT CHARACTERISTICS  
(N = 138,873)

HRR-LEVEL INDEPENDENT VARIABLE	PROBABILITY OF RECEIVING CATHETERIZATION	
	(1)	(2)
Average propensity to get cath	.529 (.172)	.575 (.167)
Percent under age 65		.150 (.135)
Log of resident population		-.003 (.005)
Log of per capita income		.024 (.024)

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/>.

**Table 6.** Returns to intensive treatment are about three times higher in intensive areas relative to nonintensive areas (0.038 vs. 0.009) - opposite prediction from flat-of-the-curve model. Difference in IV estimates appear to be driven by differences in survival rather than differences in cost (Panel A). Consistent with productivity spillovers.

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

SAMPLE	INSTRUMENTAL VARIABLE ESTIMATES OF		
	Impact of Cath		
	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 (.061)	6.691 (3.510)	.038 (.021)
Below the median ( $N = 66,124$ )	.09 (.059)	9.835 (3.155)	.009 (.007)
Difference	.166 (.085)	-3.144 (4.720)	.029 (.022)
B. Patients above the median cath propensity:			
HRR risk-adjusted cath rate:			
Above the median ( $N = 32,388$ )	.271 (.064)	.347 (4.370)	.78 (9.820)
Below the median ( $N = 32,411$ )	.168 (.046)	4.962 (2.890)	.034 (.021)
C. Patients below the median cath propensity:			
HRR risk-adjusted cath rate:			
Above the median ( $N = 31,383$ )	.206 (.129)	16.21 (5.130)	.013 (.009)
Below the median ( $N = 33,713$ )	-.139 (.165)	22.064 (6.870)	-.006 (.007)

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/>.

**Table 7.** In intensive areas, patients more appropriate for cath are better off (0.052), and patients less appropriate for cath are worse off (-0.075). Even though on average higher spending doesn't translate into improved survival (Row A), this hides important heterogeneity. These results are really quite striking. Consistent with productivity spillovers.

TABLE 7  
RELATIONSHIP BETWEEN HRR CATHETERIZATION RATE, SURVIVAL, AND COSTS, BY CLINICAL APPROPRIATENESS FOR INTENSIVE MANAGEMENT

SAMPLE	OLS ESTIMATES OF THE RELATIONSHIP BETWEEN HRR RISK-ADJUSTED CATH RATE AND			
	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 (.019)	8.093 (1.410)	-.28 (.073)	.702 (.004)
B. By cath propensity:				
Top tercile ( $N = 46,287$ )	.052 (.019)	10.012 (1.439)	-.366 (.073)	.802 (.032)
Middle tercile ( $N = 46,295$ )	.03 (.030)	11.154 (1.784)	-.271 (.082)	.906 (.021)
Bottom tercile ( $N = 46,291$ )	-.075 (.028)	2.763 (1.612)	-.209 (.073)	.369 (.021)
Difference (top - bottom)	.127 (.034)	7.249 (2.161)	-.157 (.103)	.433 (.038)
C. By age:				
65-80 ( $N = 96,093$ )	.023 (.021)	9.616 (1.448)	-.311 (.072)	.775 (.012)
Over 80 ( $N = 42,780$ )	-.031 (.028)	4.738 (1.603)	-.215 (.080)	.531 (.022)
Difference (top - bottom)	.054 (.035)	4.878 (2.160)	-.096 (.108)	.244 (.025)
D. By AHA/ACC criterion:				
Ideal ( $N = 89,569$ )	.027 (.023)	9.845 (1.599)	-.302 (.076)	.769 (.010)
Appropriate ( $N = 31,800$ )	-.002 (.024)	6.174 (1.537)	-.282 (.080)	.752 (.026)
Not appropriate ( $N = 17,504$ )	-.08 (.040)	2.958 (1.511)	-.177 (.065)	.264 (.021)
Difference (top - bottom)	.107 (.046)	6.887 (2.200)	-.125 (.100)	.505 (.023)

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. 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/>.

**Table 8.** Tests their alternative model of productivity spillovers mentioned in Footnote 3.

TABLE 8  
RELATIONSHIP BETWEEN SURVIVAL AND ALTERNATIVE MEASURES OF HRR SIZE, BY CLINICAL APPROPRIATENESS OF PATIENT

HRR-LEVEL INDEPENDENT VARIABLE	ALL PATIENTS (1)	BY CATH PROPENSITY			ALL PATIENTS (5)	BY CATH PROPENSITY		
		Top Tercile (2)	Middle Tercile (3)	Bottom Tercile (4)		Top Tercile (6)	Middle Tercile (7)	Bottom Tercile (8)
Risk-adjusted cath rate	.006 (.019)	.052 (.019)	.031 (.030)	-.077 (.028)	.008 (.019)	.053 (.019)	.031 (.030)	-.075 (.028)
Log of resident population	-.002 (.002)	-.001 (.002)	.002 (.003)	-.007 (.003)				
Log of AMIs per hospital					.005 (.004)	.014 (.005)	.006 (.007)	-.007 (.006)
Observations	138,873	46,287	46,295	46,291	138,873	46,287	46,295	46,291

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. HRR surgical and medical 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.

© 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/>.

## 4.8 Take-aways

The ‘facts’ of geographic variation have been around for a long time. This paper had high impact by combining a simple/intuitive theoretical model with some careful empirical tests. From a research perspective, this is not your ‘standard’ IV paper – they have an instrument that looks valid and they use an IV strategy when needed, but they also devise tests of equilibrium predictions that can be tested via OLS. From a policy perspective, the paper has an important conclusion: it isn’t obvious that cutting spending in areas that treat intensively will be welfare-improving. Although some papers have documented evidence of knowledge spillovers, more careful thinking about the plausibility of the three proposed mechanisms for productivity spillovers (knowledge spillovers, availability of support services, selective migration) would be useful.

## 5 Application: Redistribution

Ran Abramitzky has a series of papers - and a book (in progress) - investigating the equality-incentives trade-off in the context of the Israeli kibbutz. Most kibbutzim were established in the 1930s and 1940s, and aimed to recreate egalitarian societies based on Marxist principles. Key features of kibbutzim included equal sharing in the distribution of income, no private property, and a non-cash economy. A large public finance literature has stressed that we would expect individuals to be geographically mobile in response to such redistributive policies, with high-ability individuals moving away from such arrangements and low-ability individuals choosing to enter such arrangements (generating adverse selection). This relates to the Roy model because positive self-selection of migrants is expected when the place of origin has lower returns to skill (more distribution) than the destination, while negative self-selection is expected when the place of origin has higher returns to skill.

[Abramitzky \(2009\)](#) tests these ideas in the context of the Israeli kibbutzim - specifically, whether their redistributive policies encourage the exit of more productive individuals and encourage the entry of less productive individuals. As in his later work on immigration in the US ([Abramitzky, Boustan and Eriksson, 2012, 2014](#)), he is able to do this by taking advantage of a longitudinal data set of individuals linked across population censuses that allows him to compare migrants and non-migrants.

### 5.1 Data and estimation

He uses data on a random representative sample of individuals linked between the 1983 and 1995 Israeli Censuses of Population (a linkage done by the Israeli Central Bureau of Statistics). These data identify individuals who live in “a cooperative rural settlement, in which production, marketing, and consumption are organized in a cooperative manner,” which Abramitzky uses to identify kibbutz members.

He focuses on three subsamples:

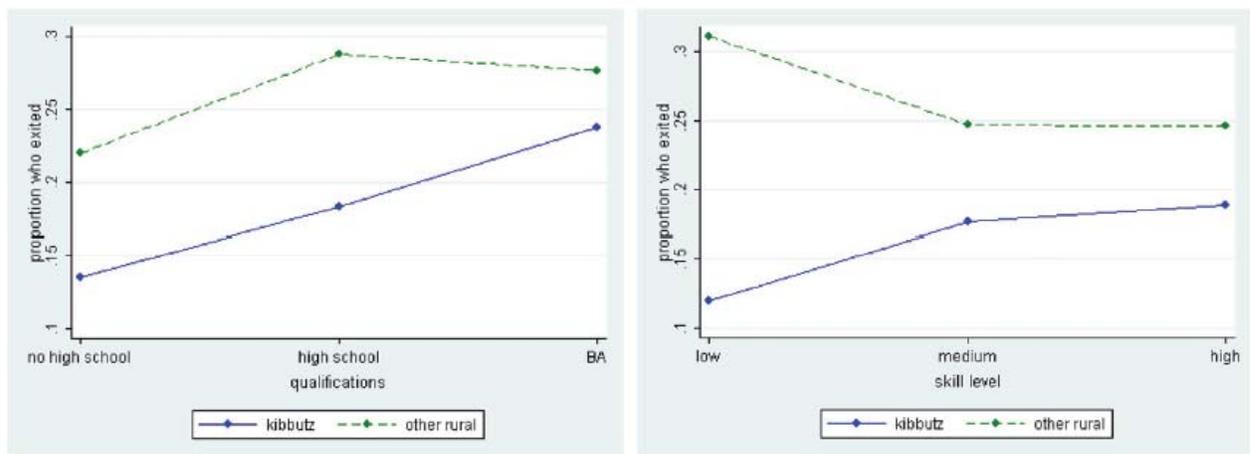
1. 1983 kibbutz members and other rural residents also observed in 1995. This sample allows him to compare kibbutz-to-city migrants both with kibbutz members who stayed in their kibbutz and with other rural-to-city migrants (the idea being that other rural locations do not engage in intensive redistribution, and provide a rough counterfactual for how rural-to-city migration rates might vary by skill level in the absence of differences in redistributive policies).
2. City residents observed in 1995, including individuals who migrated from the kibbutz and from other rural areas between 1983 and 1995. This sample allows him to analyze the 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 migrant to the kibbutz or other rural localities between 1983 and 1995. This sample allows him to compare the pre-entry earnings of city-to-kibbutz migrants with the earnings of city stayers and city-to-other rural migrants.

He focuses on Jewish individuals between the ages of 21 and 54 in 1983 (ages of 33 and 66 in 1995). 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%. He notes that entry is low in part because kibbutzim are well aware of the tendency of low-ability individuals to apply; they engage in centralized screening to mitigate adverse selection. Note that this makes it more difficult to document negative selection in entry, because actual entrants are probably less negatively selected than applicants.

## 5.2 Testing for positive selection in exit

To test for positive selection in exit from kibbutzim, he examines individuals who lived in a kibbutz in 1983 and either stayed or left by 1995, and compares the skill levels of movers with those of stayers. He also compares this skill bias in moving from kibbutzim with the skill bias in moving from other rural locations.

**Figure 1.** Figure 1 illustrates the key results on 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.

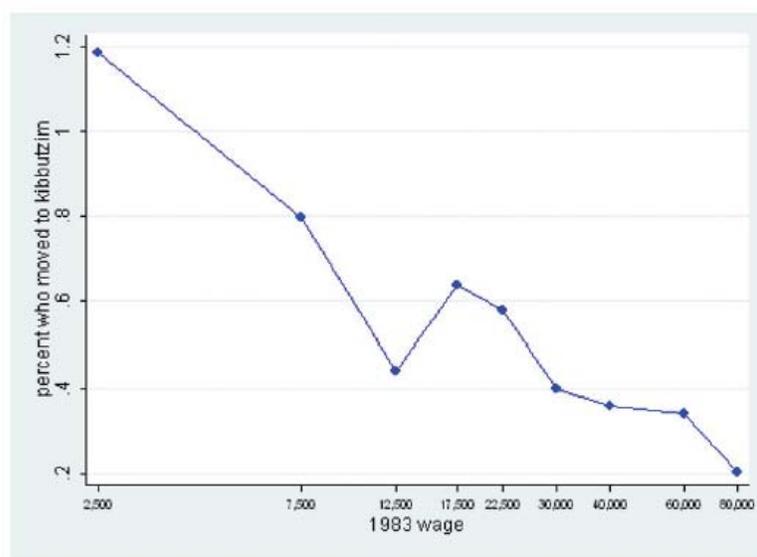


**Fig. 1.** Exit from kibbutzim and other rural areas, 1983–1995. Notes: The left hand panel shows the proportion of kibbutz members (solid line) and individuals from other rural areas (dashed line) who moved to the city between 1983 and 1995 by level of qualifications in 1983. The right hand panel shows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

Courtesy of Elsevier, Inc. <http://www.sciencedirect.com>. Used with permission.

### 5.3 Testing for negative selection in entry

**Figure 3.** Figure 3 illustrates the key results on 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.

Courtesy of Elsevier, Inc. <http://www.sciencedirect.com>. Used with permission.

While the paper of course explores the patterns from Figures 1 and 3 in regressions (which, among other things, allow for statistical inference), a great aspect of this paper is that the key ideas are clearly illustrated in these two simple figures. Ran's broader research agenda on the kibbutz uses this institution 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.

## 6 Other applications

I put four papers with additional applications of the Roy model on the syllabus: the [Willis and Rosen \(1979\)](#) application to college attendance (using the NBER-Thorndike data), the [Borjas \(2002\)](#) application to public employment (using Census and CPS data), the [Kirkeboen, Leuven and Mogstad \(2014\)](#) application to fields of study, and the [Rothschild and Scheuer \(2013\)](#) application to models of optimal taxation. All are interesting and worth reading.

## References

- Abramitzky, Ran**, “The effect of redistribution on migration: Evidence from the Israeli Kibbutz,” *Journal of Public Economics*, 2009, *93* (5-6), 498–511.
- , **Leah Boustan**, and **Katherine Eriksson**, “Europes tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration,” *American Economic Review*, 2012, *102* (5), 1832–1856.
- , —, and —, “A nation of immigrants: Assimilation and economic outcomes in the age of mass migration,” *Journal of Political Economy*, 2014, *122* (3), 467–717.
- Almond, Douglas and Janet Currie**, “Human capital development before age five,” in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics Volume 4B*, 2011, pp. 1487–1541.
- Borjas, George**, “Assimilation, changes in cohort quality, and the earnings of immigrants,” *Journal of Labor Economics*, 1985, *3* (4), 463–489.
- , “Self-selection and the earnings of immigrants,” *American Economic Review*, 1987, *77* (4), 531–553.
- , “The economic analysis of immigration,” in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics, Volume 3, Part 1*, 1999, pp. 1697–1760.
- , “The wage structure and the sorting of workers into the public sector,” in John Donahue and Joseph Nye, eds., *For the People: Can We Fix Public Service?*, 2002, pp. 29–54.
- , “The labor demand curve *is* downward sloping: Reexamining the impact of immigration on the labor market,” *Quarterly Journal of Economics*, 2003, *118* (4), 1335–1374.
- Card, David**, “The impact of the Mariel boatlift on the Miami labor market,” *Industrial and Labor Relations Review*, 1990, *43* (2), 245–257.
- Casella, George and Roger Berger**, *Statistical Inference, Second Edition*, Duxbury Press, 2001.
- Chandra, Amitabh and Douglas Staiger**, “Productivity spillovers in health care: Evidence from the treatment of heart attacks,” *Journal of Political Economy*, 2007, *115* (1), 103–140.
- Chiquiar, Daniel and Gordon Hanson**, “International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States,” *Journal of Political Economy*, 2005, *113* (2), 239–281.
- Chiswick, Barry**, “The effect of Americanization on the earnings of foreign-born men,” *Journal of Political Economy*, 1978, *86* (5), 897–921.
- Cutler, David**, *Your Money or Your Life: Strong Medicine for America’s Health Care System*, Oxford University Press, 2005.
- Dahl, Gordon**, “Mobility and the return to education: Testing a Roy model with multiple markets,” *Econometrica*, 2002, *70* (6), 2367–2420.
- French, Eric and Christopher Taber**, “Identification of models of the labor market,” in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics, Volume 4A*, 2011, pp. 537–617.
- Glover, J. Allison**, “The incidence of tonsillectomy in school children,” *Proceedings of the Royal Society of Medicine*, 1938, *31* (10), 1219–1236.
- Goldberger, Arthur**, *A Course in Econometrics*, Harvard University Press, 1991.
- Gottlieb, Daniel, Weiping Zhou, Yunjie Song, Kathryn Gilman Andrews, Jonathan Skinner, and Jason Sutherland**, “Prices don’t drive regional Medicare spending variations,” *Health Affairs*, 2010, *29* (3), 537–543.

- Gruber, Jonathan, Philip Levine, and Douglas Staiger**, “Abortion legalization and child living circumstances: Who is the ‘marginal child’?,” *Quarterly Journal of Economics*, 1999, 114 (1), 263–291.
- Heckman, James and Bo Honore**, “The empirical content of the Roy model,” *Econometrica*, 1990, 58 (5), 1121–1149.
- Katz, Michael and Carl Shapiro**, “Network externalities, competition, and compatibility,” *American Economic Review*, 1985, 75 (3), 424–440.
- Kirkeboen, Lars, Edwin Leuven, and Magne Mogstad**, “Field of study, earnings, and self-selection,” 2014. NBER working paper #20816.
- McClellan, Mark, Barbara McNeil, and Joseph Newhouse**, “Does more intensive treatment of acute myocardial infarction reduce mortality? Analysis using instrumental variables,” *Journal of the American Medical Association*, 1994, 272 (11), 859–866.
- Molitor, David**, “The evolution of physician practice styles: Evidence from cardiologist migration,” 2011. MIT mimeo.
- Neal, Derek and Sherwin Rosen**, “Theories of the distribution of earnings,” in Anthony Atkinson and François Bourguignon, eds., *Handbook of Income Distribution, Volume 1*, 2000, pp. 379–427.
- Rothschild, Casey and Florian Scheuer**, “Redistributive taxation in the Roy model,” *Quarterly Journal of Economics*, 2013, 128 (2), 623–668.
- Roy, Andrew**, “Some thoughts on the distribution of earnings,” *Oxford Economic Papers*, 1951, 3 (2), 135–146.
- Skinner, Jonathan**, “Causes and consequences of regional variations in health care,” in Mark Pauly, Thomas McGuire, and Pedro Pita Barros, eds., *Handbook of Health Economics Volume 2*, 2012, pp. 45–93.
- Willis, Robert and Sherwin Rosen**, “Education and self-selection,” *Journal of Political Economy*, 1979, 87 (5 (Part 2: Education and Income Distribution)), S7–S36.
- Wooldridge, Jeffrey**, *Econometric Analysis of Cross Section and Panel Data: Second Edition*, MIT Press, 2010.

MIT OpenCourseWare  
<http://ocw.mit.edu>

14.662 Labor Economics II  
Spring 2015

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.