Geography, Labor Movement, and Academic Knowledge Diffusion

David Jinkins*
Penn State University

August 27, 2015

Abstract

This paper measures the effect of geography and labor movement on the diffusion of a new idea. Empirically, the idea I will study is a particular academic research paper, and diffusion is measured by the spread of citations of this paper. Using a panel data set of academic economists, I develop and estimate a dynamic model of location choice in which an idea is more likely to be encountered when nearby colleagues already know about it. I find that for ex-post interested academics (as evidenced by multiple lifetime citations of the research paper), additional lifetime departmental exposure to a citer correlates with a 2% increase in the probability of citation in the next year, and the addition of a nearby (but potentially non-departmental) citer correlates with an 7% increase in the probability of citation in the next year. ADD MOBILITY RESULT

1 Introduction

People spread knowledge as they move from place to place. Firms include non-compete clauses in contracts with employees to prevent them from taking information on business practices to competitors. Governments encourage international exchange with programs such as the Erasmus program in Europe or the Fulbright program in the United States. Berkeley Astrophysicist Frank Shu wrote in 2002 that "Taiwan is a small country, and cannot develop every kind of technology by itself. Some people must go abroad to learn the latest developments and then bring them back." (Taiwan Panorama, 2002)

This paper measures the role of geography and movement between firms on the speed at which knowledge spreads. One way to think about this question is in terms of the European Union. At

*I thank Jim Tybout for his guidance, encouragement, and support at many stages of this project. I also thank Jonathan Eaton for his substantial advice and guidance. In addition this paper benefited from comments by Bulat Gafarov, Paul Grieco, Kala Krishna, Konstantin Kucheryavvy, Peter Newberry, Mark Roberts, Felix Tintelnot, Neil Wallace, Daniel Xu, and Steve Yeaple, as well as participants in the Penn State trade and macro reading groups. All remaining errors are my own.

one time it was difficult for a German to take a job in the UK. EU regulations on the movement of labor make it easier for a worker to move from Berlin to London. How much faster today do new ideas developed in Germany spread to the UK? If easing labor mobility restrictions leads to a significant increase in the speed of knowledge diffusion, then governments should take the spread of technology into account when designing immigration law. To some degree they already do − recent US immigration reform proposals have been explicit about preference for high-skill workers.

I develop a model of movement among departments and the diffusion of knowledge. The diffusion of knowledge is taken to be a stochastic process. The probability of learning about a new idea depends only on the fraction of nearby peers who already know about it, and fixed, potentially unobservable characteristics of a worker. The second part of the model is movement between departments. Moving is costly, payoffs also depend on permanent characteristics and unobservables, and an academic moves to maximize expected lifetime utility.

The model is estimated using academic citations, a sort of paper trail left behind by ideas, as well as observed movement of academics between departments. I construct a new panel data set of academics moving between departments in the United States using data from the citation database Web of Knowledge. Not only do I have information on the diffusion of citations through the network of American academics, I also have information on the workplace of an academic each time he publishes. The data set is large, containing thousands of authors, hundred of departments, and information on more than one hundred thousand academic papers.

I estimate that if 5% of the coworkers of an academic know about a new paper, he is around 50% more likely to learn about the paper in the next year than he would be if none of his coworkers knew about it. If we counterfactually increase mobility by reducing the cost of moving, we expect that within a few years after a paper is published the fraction of departments housing an employee who knows about the new paper will grow by up to 18%, the coefficient of variation between departments in the fraction of workers who know about the paper will fall by as much as 12%, and there is an as much as a 1.5% increase in the fraction of academics who have heard about the new paper. The size of the effect depends on how much we reduce the movement costs.

In a calibration using the estimates from the baseline domestic model, I analyze the effect of Chinese scholars visiting the United States on the diffusion of knowledge of a new American paper among Chinese academics and departments. Visits significantly increase diffusion. This result is driven by the relative ease of learning about the new paper in the United States, as well as the strong effect of coworker knowledge in facilitating learning.

The key challenge in estimating the structural model is endogenous sorting. That many people in a department cite a paper soon after it is published can be explained either by peer learning or by common interests. Since academics choose to work together based on mutual interests, a model which ignores sorting will overestimate the effect of learning from coworkers. The identification problem here is similar to the well-known difficulty in estimating peer effects on test scores in the education literature and peer effects on productivity in the labor literature.

The structural model developed in this paper allows for sorting on fixed unobservables. If we assume that the unobservables which jointly affect sorting and citing are fixed during the estimation period, the model is identified by moves between departments and time-series variation in citations. Put simply, we can compare the citation behavior of academics in a department before and after someone moves in or out to make inference about peer learning. Versions of this assumption are common in the structural spillover literature. For instance, when measuring productivity spillovers of supermarket cashiers, Mas and Moretti (2009) assume that the scheduling of workers with different levels of ability is unrelated to transient changes in the productivity of other workers in the shift, except through a spillover effect.[1] In order to measure peer spillover on test scores, Arcidiacono, Foster, Goodpaster, and Kinsler (2012) assume that either the fundamental ability of a student is fixed over time as he is observed taking different classes, or his ability grows in a deterministic manner.

But still the potential confounding effect of unobserved serially correlated shocks remains. To mitigate problems arising from such unobserved shocks, the estimation also utilizes a source of exogenous variation – variation which affects location choices, but does not affect the diffusion of knowledge except through its effect on location choice. There is a shock to movement into and out of public universities created by the oil price jump and subsequent recession of 1990-1991. Some states were largely unaffected by the crisis, while other states had serious budget shortfalls. Newspaper articles from the period document a number of state schools implementing hiring freezes in the Spring of 1991. I show that in my data, 1991 budget deficits have a statistically significant negative effect on net moves into state schools in 1991, even when university fixed effect, year fixed effects, and university specific trends are controlled for. A probit model using budget deficit as an instrument finds an effect of peers on learning of the same order of magnitude as the estimate in the structural model.

My research adds to the empirical literature on geography and knowledge diffusion.[2] Several reduced-form papers in this literature find evidence that workers take knowledge with them as they move between firms (Almeida and Kogut, 1999; Oettl and Agrawal, 2008; Poole, 2013).[3] This paper makes two contributions to the existing literature. First, I explicitly develop and estimate both a diffusion process for knowledge and a forward-looking inter-firm movement problem for workers. This structural approach allows me to go beyond testing for a knowledge spillover as done in previous work, and analyze how counterfactual changes in the barriers to movement between firms affect the knowledge diffusion process.[4] I also add to the literature by constructing a new data set tailored to addressing questions about inter-firm movement and knowledge diffusion. While

[1] A challenge to Mas and Moretti would be that during high volume periods low productivity cashiers must work harder, and managers schedule more productive workers. Mas and Moretti do a number of tests to check for this and other identification hurdles.

[2] (Jaffe, Trajtenberg, and Henderson, 1993) is the classic citation, see (Breschi and Lissoni, 2001) for a somewhat dated survey. Social networks are also important for knowledge diffusion. Conley and Udry (2010) show that social networks in Ghana were important for the diffusion of technology related to pineapple growing.

[3] (Almeida and Kogut, 1999) estimate stronger geographical spillovers in locales with more movement between firms. Oettl and Agrawal (2008) find that the year after an engineer who once worked at a foreign firm appears in Canada, Canadian firms are more likely to cite that foreign firm in patents. Poole (2013) shows that when a Brazilian worker moves from a multinational firm to a domestic firm, the wages of the other workers at the domestic firm rise.

[4] In Appendix A I show that my data is consistent with the knowledge localization literature. Using several reduced-form methods I can reject the null hypothesis of no peer effect.

patents have been used to infer the location of workers (Oettl and Agrawal, 2008), the relative frequency of academic publication allows me to construct a more accurate measure of the set of academics in a department in any given year.[5] An accurate measure of worker location is crucial for the estimation of the model in this paper, as well as any model which aims at measuring a worker peer effect in knowledge diffusion.

The structural model developed here can be thought of as combining recent work from two literatures. The knowledge diffusion process was motivated by the treatment of disease spread between and within households in recent epidemiology literature (Cauchemez, Carrat, Viboud, Valleron, and Boelle, 2004). That the spread of innovation is similar to the spread of infectious disease is not a new insight. Ken Arrow made such an observation in 1969, for instance.[6] Empirical models of the spread of disease often focus on the diffusion path of a particular outbreak. Early work on the diffusion of technology such as Griliches (1957) or Rogers (1962) similarly studied the empirical diffusion curves of narrowly defined technologies. More recent work by economists has focused on aggregate growth and diffusion models (Caballero and Jaffe, 1993; Kortum, 1997; Eaton and Kortum, 1999; Lucas, 2008; Chor and Lai, 2013).

The worker location choice model I develop builds on recent work by Kennan and Walker (2011) on American interstate migration. The model gives workers a chance to change locations each period by incurring a moving cost. Kennan and Walker's forward-looking dynamic discrete choice framework allows me to capture two important features of the data. First, many academics move more than once in their careers – earlier migration literature such as Dahl (2002) allowed only a single migration decision. The discrete choice model also allows me to capture choice among many locations, a feature that is not present in much of the macro literature on repeat and return migration (Dos Santos and Postel-Vinay, 2003; Dustmann, 2003).

Figure 1 provides some motivation for the model I develop below. The data in the figure is a large pool of economics papers (originating papers), and the papers which cite them (citing papers). The horizontal axis is time since an originating paper was published, and the vertical axis is the percentage of its citing papers which have an author sharing an affiliation with an author of the originating paper. Any citing paper sharing an author with its originating paper is excluded. Whatever is causing cites to largely come from own department just after a paper is published, it dies away over time. This picture suggests that the diffusion of knowledge depends in some way on physical proximity.

In what follows, I will describe the main model, then discuss data and estimation. In the estimation section I will discuss the identification strategy, describe the source of exogenous variation, and discuss the actual implementation of the estimation routines. Following that is a

[5] One paper using such a matched patent data set found an average of "a little over one" lifetime patents per inventor (Kim, Lee, and Marschke, 2009). Compare this to an average of 4.7 lifetime publications per academic in my full data set, 14.1 lifetime publications in the estimation sample of around four thousand economists who worked in one of the top 100 US departments in the years 1987-1994, and 22.9 average lifetime publications for the subset of those economists who started and ended in different departments.

[6] 'Although mass media plays an important role in alerting individuals to the possibility of an innovation, it seems to be personal contact that is most relevant in leading to its adoption. Thus, the diffusion of an innovation becomes a process formally akin to the spread of an infectious disease.' -Ken Arrow, 1969

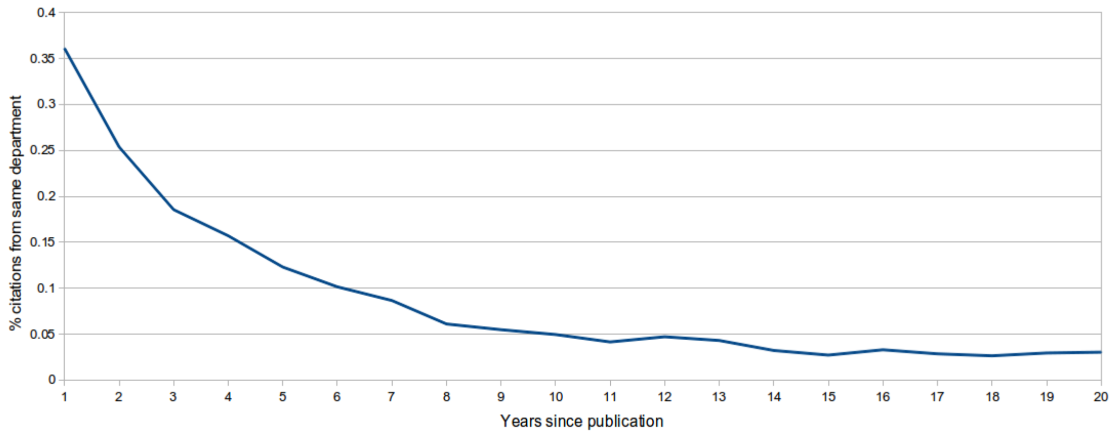

Figure 1: Citing paper location sharing over time.

results section, and a counterfactual section, and the cross-country calibration. Penultimately, several alternative specifications are estimated, and the models are simulated and checked against data. Finally, the results of a reduced-form probit model with a similar message to the main structural model are presented.

2 Data Description

2.1 General Data Construction

In this section I describe with some generality how the data used in all exercises in this paper were collected and constructed. Section 2.2 describes the specific construction of variables used in the estimation of the structural model described above.

An academic's current place of employment is listed under the byline on academic papers. The Thomson-Reuters Web of Knowledge, a citation database, records affiliation for each academic on each available paper. I use the python web scraping library BeautifulSoup to download citation data for more than one hundred thousand economics articles from the Web of Knowledge, and then use affiliation data to construct a panel of economists moving between departments.[7]

In addition to direct information on each economics paper, I also collected data on all papers from any discipline which cite either the most cited hundred economics papers, or that cite any economics paper published in 1980 or 2005.[8] This is a large, rich data set, containing thousands of economists, hundreds of departments from around the world, and more than one hundred thousand

[7] Agrawal, McHale, and Oettl (2013) constructs a similar panel of academics also from Thomson-Reuters Web of Knowledge, but uses evolutionary biologists rather than economists.

[8] In another project using this data, I am looking at the effect of the internet on diffusion rates. I choose 1980 because it is well before the internet era, and 2005 because it is well after.

papers and citation records. More information on how the data was cleaned and filtered is in Appendix C.

In several exercises I use the field of an economist. I construct a field for each economist using data from IDEAS, based on which curated mailing lists the work of an academic is mainly distributed in. This way of classifying field is not original to me – it is currently an experimental classification system on IDEAS itself.

An economist can simultaneously work in many areas. What I will refer to as a field is a 91-dimensional unit vector describing research area. This is a fine disaggregation scheme. For instance, someone doing trade and operations research will correctly have a different field vector from someone doing trade and public economics. If the work of an economist is distributed in the IDEAS development mailing list as well as the game theory mailing list, he will have a field vector of 89 zeros, with $\frac{1}{\sqrt{2}}$ in the dimensions corresponding to game theory and development.

Journal fields are constructed using the JEL field rankings in Barrett, Olia, and Bailey (2000). To get a field for each paper, journal and academic fields are combined and normalized. Field construction is described in detail in the data appendix, Appendix C.

2.2 Construction of Variables

In this section I describe the construction of variables specific to the estimation of the structural model described in Section 4. The model is estimated using first citation times of a single paper: Michael Jensen's 1986 *American Economic Review* piece "Agency costs of free cash flow, corporate finance, and takeovers". Estimating the structural model on citations of a single paper allows me to use a binary field, which greatly reduces the complexity of the department choice problem. I can also focus on data from a relatively small number of years around the time the paper was published. Finally, using a single paper allows me to keep the parameter space small. Papers have widely varying citation trajectories, and most papers have very few citations.[9]

Jensen's 1986 *American Economic Review* piece is one of the most highly cited papers in my data set, giving me many observations of citation times. The paper was published just before Jensen ended his joint appointment with the University of Rochester where he spent the first 20 years of his career, and permanently moved to the Harvard Business School. I use the Jensen paper because most of the other highly cited papers are in the field of econometrics. The most highly cited econometrics papers are those which become widely used by applied economists. For example, two of the most cited papers in my dataset are Heckman (1979) and White (1980). For these papers, field is a poor measure of interest. The Jensen paper, on the other hand, is still more likely to be cited by economists working in contract theory or business economics.

The binary academic specific field f_i is set to one if an academic works in either of the Jensen fields: "Contract Theory and Applications" or "Business Economics". The department field F_d

[9] As a robustness check, I reestimate the entire model using Grossman and Hart (1986), another influential paper published in 1986. In Appendix I I present a comparison of the reestimated results to the baseline results, and find essentially no difference.

is the mean fraction of academics in the department in the Jensen field, averaged over all years in my sample. To create the quality of an academic, I first calculate his mean coauthor-adjusted lifetime citations per published paper. To reduce the dimensionality of the problem, I then partition academics into equally-sized low and high quality groups.[10] I assign the high-quality group $q_i = 1$, and the low quality group $q_i = 0$. Department quality Q_d is based on the REPEC ranking of US departments, with departments assigned equally spaced values of $Q_d \in [0, 1]$.

I use data from the 104 American departments ranked in the top 25% of US departments by REPEC. Data from lower-ranked departments is available, but noisy because, economists at low ranked departments publish relatively rarely. I observe the location of an economist only when his work appears in a journal. To give some idea about what is excluded, the three lowest ranked included universities are Clark University, the Georgia Institute of Technology, and the University of New Mexico. I drop all economists who never worked at any of the 104 departments in my dataset. If an economist in my dataset spent some years at a department not included, I classify his department in those years as "other". I estimate the model on data for the eight years beginning in 1987, the year after Jensen's paper was published. Tables 1 and 2 contain summary statistics for the data I use in estimation.

	Obs Number	In Field	Citers	Moves	Cits / Pap, avg	Cits / Pap, sd
Academicss	3876	150	122	679	26	29

Table 1: Academic summary statistics

	Obs Number	1987 med size	1994 med size	Field, (avg, sd)	Avg cits / pap, (avg, sd)
Departments	104	16	24	(0.03,0.06)	(20,9)

Table 2: Department summary statistics

3 Reduced Form Evidence

3.1 Peer Learning

Before describing the structural model, I present some motivating reduced-form evidence. The unit of obervation in the following analysis is an academic, and we will be interested in how long it takes an academic to cite the Jensen paper conditional on a number of covariates. The natural setup for this type of problem is survival analysis. Table 3 presents the results of a series of logistic models, reporting odds ratios. We are most interested in the odds ratios of two regressors: department citers and total exposure. Department citers is the number of other people in one's department who have cited a paper. Total exposure is the sum of department citers for each year

[10] One can imagine several ways to measure the quality of an academic based on publications and citations. One alternative would be a simple count of published papers per year. Another would be total citations per year. While the choice of quality metric will change the ranking of academics to some degree, I believe different metrics will lead to similar aggregated high and low quality groups.

since an academic has been in the data set. Included in the controls are the lifetime citations of the paper, whether the academic has published in business or contract theory journals, department size and lagged department size. There are both year and duration fixed effects, with duration fixed effects controlling for how long an academic has been observed working. The lifetime citations of the paper are meant to proxy for interest.

We see that lagged department citers goes in the correct direction, but is not significant. Lagged total exposure is also in the right direction and is significant at the 10% level. The reported coefficient implies that additional exposure to a colleague who has cited the paper leads to a 4.3% higher odds ratio of citing in the next period. When both lagged department citers and lagged total exposure are included in Model (3), we see that department citers is in the wrong direction and still not statistically significant. I take this as an indication that previous exposure matters as much single period lagged exposure. That multiple lags matter is perhaps not surprising in light of the time it takes to publish a paper. Finally, in Model (4) we see that it is really the interaction between lifetime paper citations and lagged total exposure that matters. That is, those who are interested in the paper are the ones likely to learn about it from colleagues.

Taking these results into consideration, the structural model will use total exposure as the measure of peer influence, and unobservable interest will play a role both in the likelihood that an academic cites the paper, and also in deciding when to move between departments.

4 Model

4.1 Learning and citation

Time is discrete. There is a finite number A of academics partitioned at any particular time into D departments. Each academic is endowed with a quality q_i and either the same field as the new paper ($f_i = 1$) or another field ($f_i = 0$). Each academic also has an unobserved latent type $h_i \sim \mathcal{N}(0, \sigma^2)$, which captures his field-specific skill in discovering new research.

At time 0 a new paper is written in a particular field. If an academic is *potentially interested* in the paper, he is susceptible to learning about its existence. I assume that potential interest in a paper is independently drawn once for each academic from a Bernoulli distribution with success probability dependent upon field f_i: $\gamma_{f_i} \in \{\gamma_0, \gamma_1\}$. The probability of a potentially interested academic learning about the paper depends on the fraction of other academics in his department who have already learned about the paper, his observable characteristics, and his unobserved latent type. Upon learning about the paper, an interested academic immediately cites it.[11]

More formally, the probability that a potentially interested academic who has not yet learned about a paper learns about it at time t is given by the logit:[12]

[11]There is an extension in Section 8 in which the model is estimated with a deterministic one-year publication lag. If we are willing to additionally assume that an academic can not transfer knowledge until he actually publishes something citing the new paper, it would be feasible to make publication lags random as well.

[12]At first glance, this looks like the typical dynamic logit, but it is simpler. A dynamic logit has state dependence.

	(1)	(2)	(3)	(4)	(5)
	Odds Ratio	Odds Ratio	Odds Ratio	OLS	Odds Ratio
Lag. Total Experience	1.039**		1.215***	0.000380*	1.038*
	(0.0468)		(0.000320)	(0.0665)	(0.0669)
Lag. No. of Citing Colleagues		1.079	0.866		
		(0.161)	(0.493)		
Lag. No. of Cit. Coll. x Lag. Tot. Exp.			0.984*		
			(0.0761)		
Lifetime Jensen 1986 Citations			2.077***		
			(6.20e-08)		
Contract Theory/Business Econ			2.188*		2.706***
			(0.0769)		(0.00346)
Department Size			1.033**	-0.000288	
			(0.0242)	(0.186)	
Lag. Department Size			0.958***	-3.12e-05	
			(0.00486)	(0.856)	
Year					2.464***
					(0.000309)
Year^2					0.919***
					(0.00178)
Duration					0.784
					(0.192)
Duration^2					1.004
					(0.881)
Constant	0.00119***	0.00119***	4.05e-05***	0.0155*	0.00112***
	(0)	(0)	(7.30e-07)	(0.0820)	(0)
Observations	23,216	23,216	23,216	23,216	23,216
Duration FE	YES	YES	YES	YES	NO
Year FE	YES	YES	YES	YES	NO
Department x Individual FE	NO	NO	NO	YES	NO

Robust pval in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All standard errors are clustered at the department and individual level.

Table 3: Duration model, hazard of citing Jensen (1986)

$$\frac{e^{\alpha+\beta K(t-1)+\psi t+h_i}}{1+e^{\alpha+\beta K(t-1)+\psi t+h_i}} \tag{1}$$

$K(t-1)$ is the number of colleagues who have cited the paper in each year up to $t-1$. Equation (1) is the only place in the model where β, the main parameter of interest, appears. The parameter β measures the direct effect of the knowledge of colleagues on own learning. The latent type h_i can either increase or decrease the probability of citing, depending on its sign.

4.2 Dynamic Department Choice Problem

In this section, I develop a model of labor movement between departments. It is necessary to model movement because the counterfactual exercises we are interested in involve reducing the cost of moving between departments. Explicitly modeling movement disciplines the way that movement patterns change when mobility is increased. In addition, treating moves as random would bias the estimates of the learning and citing parameters in Section 4.1.

In the model, an academic decides in which department to work in order to maximize discounted lifetime expected utility. If the academic chooses to move, he must pay a movement cost. The model is a dynamic discrete choice model, similar in spirit to recent work by Kennan and Walker (2011) on interstate migration.[13] Department choice is the only decision in the model.

Let $\mathbf{X}_i$ be the vector of personal characteristics of academic i: field f_i, quality q_i, and latent type h_i. If academic i works at department d in period t, he gets period random utility:

$$u_{i,t}(d) = W(d, \mathbf{X}_i) + \varepsilon_{i,d,t} \tag{2}$$

Period utility is a department-specific, time-invariant payoff, plus a time-varying preference shock. The preference shock $\varepsilon_{i,d,t}$ is distributed IID Type 1 Extreme Value. The current-period payoff to working at department d can be split into two parts:

$$W(d, \mathbf{X}_i) = w_v(\mathbf{X}_i) + w(d, \mathbf{X}_i) \tag{3}$$

The first part of the period payoff w_v depends only on personal characteristics like quality and field. This component is the same at any department. The second part of the payoff w is department

The econometrician needs to estimate the extent that the outcome today depends on the outcome yesterday. In the model here, once an academic learns about a paper, he needs not learn about it a second time. Learning about a paper is like contracting a chronic disease – one time is enough.

[13] Since knowledge diffusion is the main focus of this paper, for tractability the location choice model developed here is simpler than that in Kennan and Walker. In particular, I assume that movement costs are the same for all department pairs. Dahl (2002) is an alternative for estimating the migration decision between many possible locations. In the Dahl model, however, migration decisions are taken only once in the lifetime of a worker. Since my data contains repeated migration observations, a version of the dynamic Kennan and Walker model is more appropriate.

specific. It depends on time-invariant measure of department field F_d, and time-invariant measure of department quality Q_d, both of which interact with individual field, quality, and latent type.[14]:

$$\ln w(d, \mathbf{X}_i) = \xi_0 + \xi_q q_i Q_d + \xi_f f_i F_d + \xi_h h_i F_d \tag{4}$$

Latent type h_i is interacted with department field F_d because those with high skill in discovering new research value having colleagues in the same field differently than those with low skill.

Movement costs $\mathcal{C}$ must be paid each time an academic changes departments.[15] Saving is not allowed. Agents choose departments to maximize discounted lifetime expected utility. The value function below is net of the non-department-specific payoff component w_v.[16] The set of departments is $\mathcal{D}$. Suppress the permanent characteristic vectors $\mathbf{X}_i$ and write the recursive value function as:

$$\begin{aligned} V(d) &= \rho \mathbb{E}_\varepsilon \big[\max\{V(d') + w(d') + \varepsilon_{i,d',t} - \mathbb{1}_{\{d' \neq d\}}\ \mathcal{C}\}_{d' \in \mathcal{D}} \big] \\ &= \rho\gamma_e + \rho \ln \left(\sum_{d' \in \mathcal{D}} e^{V(d') + w(d') - \mathbb{1}_{\{d' \neq d\}}\ \mathcal{C}} \right) \end{aligned} \tag{5}$$

The substitution of the expectation of the maximum of Type I Extreme Value errors in (5) follows Rust (1987), and is derived in Appendix E. This value function is defined on D departments for each type $\mathbf{X}_i$, with $\gamma_e \approx 0.577$ being the Euler-Mascheroni constant. I show in Appendix D that the natural operator on (5) is a contraction mapping. We can use the value function to get the probability of moving from department d to another department d':

$$Pr(d, d') = \frac{e^{V(d') + w(d') - \mathbb{1}_{\{d' \neq d\}} \mathcal{C}}}{\sum_{d'' \in \mathcal{D}} e^{V(d'') + w(d'') - \mathbb{1}_{\{d'' \neq d\}} \mathcal{C}}} \tag{6}$$

4.3 Summary

The model can be thought of as consisting of two parts. The learning and citing part is a stochastic process governed by (1). Only the observable department citation fraction $K(d, t-1)$ varies over time.[17] The part of the model governing movement between firms is a dynamic discrete choice problem, characterized by the value function (5) and the utility function (2). Solving the value function results in transition probabilities which depend on fixed observable and unobservable

[14] In practice, department field is the mean annual fraction of academics in the field working at the department. Academic quality will be lifetime mean citations per paper, and department quality is the mean annual average quality of academics working at the department during the sample period. Details are contained in Section 2.

[15] In principle, movement costs could depend on interactions between department and the observable characteristics in the payoff equation, with some exclusion for identification. The current specification forces all sorting to go through interactions in the payoff equation. I suspect that substitution patterns would not be much more rich in a specification with characteristic-dependent movement costs, so to save parameters I estimate the simpler model.

[16] In particular, if we add $\frac{\rho}{1-\rho} w_v$ to the left hand side of (5), and add $w_v + \frac{\rho}{1-\rho} w_v$ to every appearance of $V(.)$ on the right hand side, the new terms cancel out and the equation remains the same.

[17] See Section 8 for an extension in which there is a national knowledge spillover as well.

characteristics. The link between the two parts of the model is the latent type h_i, which affects both payoffs and learning probabilities. While latent type is unobserved, it is assumed to be fixed over time. This extreme form of serial correlation will be discussed further in Section 5.1.

4.4 Initial conditions

In the model described above, latent type h_i is assumed to be independently randomly distributed. There is, however, an interaction in payoffs between unobserved latent type and department field. This interaction will induce sorting before my sample period, so if I take latent type to be randomly distributed I will get inconsistent estimates. Put simply, an academic is more likely to be of high latent type if he is first observed at a department with a high fraction of workers in the field of the new paper. To mitigate this problem, I assume that the mean of the distribution of latent type depends upon the department observed in an academic's first year. Here $F_i^{(1)}$ denotes the field fraction and $Q_i^{(1)}$ the quality of the observed first department of academic i:

$$h_i = \phi_Q Q_i^{(1)} + \phi_F F_i^{(1)} + h_i^*, \qquad h_i^* \sim \mathcal{N}(0, \sigma^2) \tag{7}$$

Any level effect in (7) will be absorbed by α in the learning probability equation (1), and the size of parameter ξ_l in period payoff equation (4). The quantities $F_i^{(1)}$ and $Q_i^{(1)}$ depend only on the *initial* observed department of an academic, while the department field F_d and quality Q_d entering into (4) depend on the current location of the academic which may change from year to year.

4.5 Likelihoods

My data describe a set of academic economists over time, and the citations of a particular paper over time. For academic i, the key variables are the (possibly empty) year of academic i's first citation of the new paper $C_i \in 1, 2, \ldots, T \cup \emptyset$, and a (possibly empty) department for academic i in each year $M_{i,t}$. Collect into sets $\mathbf{C} = \{C_i\}_{i\in A}$, $\mathbf{M} = \{M_{i,t}\}_{i\in A, t\in 1,\ldots,T}$, and $\mathbf{M}_i = \{M_{i,t}\}_{t\in 1,\ldots,T}$. As before, let $\mathbf{X}_i$ be all individual characteristics, both observable and unobservable. Let $\mathbf{X}_{o,i}$ denote only observable individual characteristics, and let $\mathbf{X}_o = \{\mathbf{X}_{o,i}\}_{i\in A}$. Let H denote the mean-zero normal CDF with variance σ^2, i.e. the distribution of h_i^* as in (7). Let $\mathfrak{D}$ be the set of departments. Department fields F_d and qualities Q_d are contained in the vector $\mathbf{Z} = \{F_d, Q_d\}_{d\in\mathfrak{D}}$. Suppose that we have calculated the transitions $Pr(d, d'|\mathbf{X}_i, \mathbf{Z}, \theta)$ from value function iteration. We can consider the likelihood for each individual separately. The likelihood for academic i is:

$$\begin{aligned} Pr(C_i, \mathbf{M}_i|\mathbf{X}_{o,i}, \mathbf{Z}, \theta) &= \int Pr(C_i, \mathbf{M}_i|h_i, \mathbf{X}_{o,i}, \mathbf{Z}, \theta) dH \\ &= \int Pr(C_i, \mathbf{M}_i|\mathbf{X}_i, \mathbf{Z}, \theta) dH \end{aligned} \tag{8}$$

We can split the integrand in (8) into multiplicative terms:

$$Pr(C_i, \mathbf{M}_i | \mathbf{X}_i, \mathbf{Z}, \theta) = Pr(C_i | \mathbf{X}_i, \mathbf{M}_i, \theta) Pr(\mathbf{M}_i | \mathbf{X}_i, \mathbf{Z}, \theta) \tag{9}$$

The second part comes directly from the transitions derived from the value function iteration. We construct it by multiplying probabilities of observed moves:[18]

$$Pr(\mathbf{M}_i | \mathbf{X}_i, \mathbf{Z}, \theta) = \prod_{t=1}^{T-1} Pr(M_{i,t}, M_{i,t+1} | \mathbf{X}_i, \mathbf{Z}, \theta) \tag{10}$$

The first part is a little more complicated. Set $M_{i,0} = \emptyset$ for notational convenience:

$$\begin{aligned} Pr(C_i | \mathbf{X}_i, \mathbf{M}_i, \theta) = & \left[(1 - \gamma_{f_i}) + \gamma_{f_i} \prod_{t=1}^{T} \left(1 - 1_{\{M_{i,t} \neq \emptyset\}} \frac{e^{\alpha + \beta K(M_{i,t}, t-1) + h_i}}{1 + e^{\alpha + \beta K(M_{i,t}, t-1) + h_i}} \right) \right]^{1_{\{C_i = \emptyset\}}} \\ \times & \left[\gamma_{f_i} \prod_{t=0}^{C_i - 1} \left(1 - 1_{\{M_{i,t} \neq \emptyset\}} \frac{e^{\alpha + \beta K(M_{i,t}, t-1) + h_i}}{1 + e^{\alpha + \beta K(M_{i,t}, t-1) + h_i}} \right) \frac{e^{\alpha + \beta K(M_{i,C_i}, C_i - 1) + h_i}}{1 + e^{\alpha + \beta K(M_{i,C_i}, C_i - 1) + h_i}} \right]^{1_{\{C_i \neq \emptyset\}}} \end{aligned} \tag{11}$$

The two big multiplied terms in (11) reflect the difference between those I observe citing the paper and those I do not. If I observe that an academic cited the paper ($C_i \neq \emptyset$), he must have been interested in it. If the academic didn't cite the paper ($C_i = \emptyset$), then either he was not interested, or he would have been interested but did not hear about the paper during the years in my data. In the top term, the $(1 - \gamma_{f_i})$ is the probability γ_{f_i} that the academic was not interested. The second term on the top line is the probability that the academic would have been interested had he heard about the paper, multiplied by the probability that he did not hear about the paper. The indicator function $1_{\{M_{i,t} \neq \emptyset\}}$ eliminates years when an academic is not in the data set. The bottom line is simply the probability that an academic was interested γ_{f_i} multiplied by the probability that the academic did not hear about the paper until the year he did, and then multiplied by the probability that he did hear about the paper in the year he cited it.

Combining all the academics, the total likelihood is then:

$$Pr(\mathbf{C}, \mathbf{M} | \mathbf{X}_o, \mathbf{Z}, \theta) = \prod_i Pr(C_i, \mathbf{M}_i | \mathbf{X}_{o,i}, \mathbf{Z}, \theta) \tag{12}$$

5 Estimation

5.1 Identification and causality

The main parameter of interest is β in (1), the impact of colleagues on own learning about new ideas. It governs not only peer-learning directly, but it also reflects the importance of movement

[18] Entry and exit are treated as exogenous. If $M_{i,t} = \emptyset$ or $M_{i,t+1} = \emptyset$, then $Pr(M_{i,t}, M_{i,t+1} | \mathbf{X}_i, \mathbf{Z}, \theta) = 1$.

between departments. A high β implies that a knowledgeable colleague makes an academic much more likely to learn about the new paper.

The three common peer effect identification challenges are endogenous sorting, correlated effects, and the reflection problem. The peer effect in my model works with a lag, that is citation probabilities are affected only by lagged colleague knowledge. There is a clear direction of causality implied by time, so the reflection problem is not an issue.

Endogenous sorting and correlated shocks remain a challenge. Academics might sort into departments based on unobservables. An academic may cite earlier because his colleagues have already learned about a paper, or it could could just be that he is working with people interested in similar things. Even if he had been at a different department, he would have been among the early citers.

The model developed above allows the citing probability to be influenced by unobserved fixed individual characteristics h_i, and allows for sorting on these unobserved characteristics. Even if academics sort into departments based on time-invariant unobservables, identification is possible using moves between departments and time series variation. For example, suppose that an academic who has cited the Jensen paper moves from Cornell to Penn State. I can observe citing behavior at Penn State before the academic arrives, and citing behavior at Cornell after he leaves. If all characteristics of Penn State and Cornell academics are fixed, then the change in citing behavior can be used to infer β. In the language of an experiment, the control group is Penn State academics just before the new colleague arrives, and the treatment group is Penn State academics after the colleague arrives. As mentioned in the introduction, the assumption that fixed effects are time-invariant is common in the structural spillover literature, especially the non-experimental labor literature on peer effects in school classrooms (Betts and Zau, 2004; Arcidiacono et al., 2012; Burke and Sass, 2013). [19]

What is not in the model is serially-correlated, time-varying unobserved individual or correlated shocks. If such persistent time varying shocks cause a group of people to sort together and subsequently begin citing each other papers, the baseline structural model will overestimate learning from colleagues. In the current setting, however, ignoring serially correlated shocks is unlikely to seriously bias the estimates. While research interests can change over the lifetime of an academic, there is a strong lock-in effect due to the high fixed costs of reaching the research frontier in an unfamiliar area. Substantial change in research focus takes place at most several times in a career, and the model is estimated on only eight years of data.

If serially correlated shocks were important, however, to identify the causal effect of coworkers on learning one would need an exogenous shock which affects the location choice of an academic, but does not affect citations. I use the US recession of 1990-1991 to induce exogenous movement. Some states were hit particularly hard by the crisis, and some state schools were forced to implement

[19] There is an even larger labor literature on the value-added effect of teachers on student achievement. This literature also needs to deal with endogenous sorting, and uses student fixed effects when possible (see Harris and Sass (2011) for a recent example). In this context, Rothstein (2010) finds evidence that student fixed effects are not sufficient to control for endogenous sorting.

temporary hiring freezes. I will argue that these hiring decisions by state schools in 1991 induced exogenous changes in movement patterns between departments, but did not affect citing behavior.

In the baseline structural model, I include the shocks as a temporary source of variation in payoff. I describe how I do this in detail in the next section. I also run an endogenous probit, using state budget deficits as an instrument, with the exogeneity assumption that the 1991 recession affected movement choices but did not affect citing behavior.[20] The estimated coefficient on β in the reduced-form instrumental variable exercise is similar in size to the estimate in the structural estimation.

As for the other parameters, first consider the citing probability (1). The variance in citing frequencies net of the β coworker effect will identify the dispersion of latent type, and the level of citation frequencies net of β identifies α. Since the dispersion of latent type is identified from (1), the parameter ξ_h along with the other payoff parameters ξ are identified by observed department move choices in the data, i.e. substitution patterns. The cost of movement $\mathcal{C}$ is identified by the frequency of moves.

5.2 Exogenous variation: Economic malaise of 1990-1991

Induced partly by an oil price shock caused by the Iraqi invasion of Kuwait, the United States went through an economic recession from July, 1990 to March, 1991. The effects of the downturn differed by state (Washington Post, 1991; Moore, 1991). In several of the hardest hit states, public universities implemented hiring freezes for various lengths of time (LA Times, 1991; Money Magazine, 1991; **?**). I use data on state budget deficits in fiscal year 1991 to proxy for temporary, unanticipated hiring reductions at public universities in 1991 (Gold, 1995).

Let b_d be the 1991 budget deficit divided by total state expenditures in the state of public university d. Several linear regressions show that the 1991 economic downturn induced observable variation in movement patterns. The unit of observation is a department-year, and the dependent variable is net moves into a department. The independent variable we care about is dum91bd, a 1991 dummy multiplied by budget shortfall b_d. The regression is performed on 104 departments with seven years of observation starting in 1986, which is similar to the data cut I use in the structural exercise. Table 4 reports results.

Some states had 1991 budget deficits b_d as large as 15 and 20%, implying one to two fewer net in-moves into public universities compared with a typical year.

In the structural model, I make use of the shock to state budgets by assuming that in 1991 payoffs (wages) are suddenly and temporarily shocked so that:

$$w_{1991}(d, \mathbf{X}_i) = e^{-\xi_{ex} b_d p_d} w(d, \mathbf{X}_i) \tag{13}$$

[20] There is also a quasi-experimental labor literature on peer-effects in the classroom employing instrumental variable methods to deal with endogenous sorting and other identification issues (Kang, 2007; De Giorgi, Pellizzari, and Redaelli, 2010).

	net in-moves	net in-moves	net in-moves	net in-moves
dum91bd	-8.568**	-12.280**	-6.784**	-7.055**
	(4.24)	(4.93)	(2.88)	(2.84)
year dummies	no	yes	yes	yes
dep dummies	no	no	yes	yes
dep dummies × year	no	no	no	yes
Obs	617	617	617	617
R^2	0.01	0.02	0.77	0.82

Table 4: The net in-move effect of state budget shortfalls

Here p_d is a dummy set to 1 if d is a public university. Since this payoff cut is sudden and temporary, it does not affect expectations in the value function. What this means for the estimation is that for the single year 1991 transition probabilities between departments in the movement likelihood (10) are given by (14) rather than the original transition probabilities (6).[21]

$$Pr_{1991}(d, d') = \frac{e^{V(d') + w_{1991}(d') - \mathbb{1}_{\{d' \neq d\}} \mathcal{C}}}{\sum_{d'' \in \mathcal{D}} e^{V(d'') + w_{1991}(d'') - \mathbb{1}_{\{d'' \neq d\}} \mathcal{C}}} \tag{14}$$

5.3 Implementation

I estimate the twelve parameters in the likelihood function (12) using Bayesian inference and Markov Chain Monte Carlo (MCMC). Description of the priors are contained in Table 5. The priors are mostly designed to be proper but relatively uninformative. For parameters which a priori fall anywhere on the real line I use the normal distribution centered at zero with variance 100, and for parameters which are a priori non-negative I use the exponential distribution with parameter 300.[22] There are weakly informative priors are on the interest parameters γ because I can observe whether the academics in my eight-year sample cited the Jensen paper anytime up to 2012. If an academic has not cited the influential Jensen paper 25 years after it was published, it is probably not because he has yet to hear about it. About a third of people in Jensen's field ultimately cite the paper, so I assign a beta distribution with parameters 1 and 2 to the field interest probability γ_1. The standard deviation of the prior is 0.24. About 6% of people not in the field ultimately cite the paper, so I assign to the non-field interest probability γ_0 a beta distribution with parameters 1/8 and 2, which gives a standard deviation of 0.13.

[21] See Section 9 for an alternative reduced-form analysis using 1991 budget deficits as an instrument.

[22] The exponential prior has been used for a similar purpose in the epidemiology literature (Cauchemez et al., 2004). In earlier versions of the paper I used mostly improper diffuse priors (for parameters including the peer effect β) and ended up with very similar posteriors in the baseline model. Appendix K compares the baseline model with an estimated version in which β's prior is diffuse and finds almost no qualitative difference. In the model extensions, some results are sensitive to the choice of prior, in particular the version in which I allow all parameters in the knowledge diffusion process to depend upon observed field. There are too few people in Jensen's field to provide strong evidence on so many parameters. Since there are theoretical reasons to expect that colleague knowledge should not cause less learning, I use the exponential prior for all peer effects.

	Prior
α	Norm(0,100)
β	Exp(300)
γ_F	Beta(1,2)
γ_{NF}	Beta(0.125,2)
ξ_f	Norm(0,100)
ξ_l	Norm(0,100)
ξ_q	Norm(0,100)
$\mathcal{C}$	Exp(300)
ϕ_F	Norm(0,100)
ϕ_Q	Norm(0,100)
σ	Exp(300)
ξ_{ex}	Exp(300)

Table 5: Priors

Recall that I partition academics into two quality groups, field is binary, and I use four points to approximate the one-dimensional numerical integral over h_i. Thus, in each iteration of the estimation routine there are 16 independent value functions to solve, each on a space of 104 departments.

The MCMC employed for the estimation is a random walk Metropolis algorithm with an adaptive proposal distribution. The art in MCMC is choosing efficient proposals. In the plain random-walk metropolis method, a proposal is just mean-zero Gaussian random noise ϵ added to the current parameter set. It is difficult to determine an efficient covariance structure for ϵ a priori. If the jumps are large or in unlikely directions, then the proposal is accepted too rarely and it takes a long time to move around the posterior. On the other hand, if jumps are too small, then the proposal is almost always accepted and the routine must be run a long time to spend enough time in the high probability areas of posterior distribution.

To get an efficient covariance structure, I employ the adaptive algorithm suggested by Haario, Saksman, and Tamminen (2001). In every step of the algorithm the empirical covariance structure of many previously accepted parameters is calculated. The random noise for the next proposal is then drawn from a mean-zero Gaussian distribution with the calculated empirical covariance structure. Haario et al. (2001) show that this algorithm will asymptotically approach the efficient covariance structure. Parameters are updated block by block, with only a single block being updated in each step. There are three blocks: parameters related to learning and citing ($\{\alpha, \beta, \gamma\}$), parameters related to moving ($\{\xi, \psi, \lambda_o\}$), and latent type parameters ($\{\phi, \sigma\}$). The covariance structures are updated for each of the parameter blocks separately.

The MCMC routine is implemented in python 2.71, making heavy use of the excellent pandas (panel data analysis) library as well as the python multiprocessing library. Each time I estimate the model, I ran 10 separate chains using Penn State Research Computing resources. The first half of each chain is discarded as a burn-in. Appendix F contains convergence diagnostics for the

MCMC chains as well as mixing plots. Running many chains in parallel allows for implementation of the Gelman-Rubin convergence criterion (Gelman and Rubin, 1992). Without going into detail, the criterion tests for the similarity of the separate chains in terms of mean and variance of each parameter. If each of the chains is 'indistinguishable' from the other chains after a burn in, then we say that draws from the chains are independent draws from the posterior distribution. All parameters in the estimation routine pass the Gelman-Rubin test.

6 Results

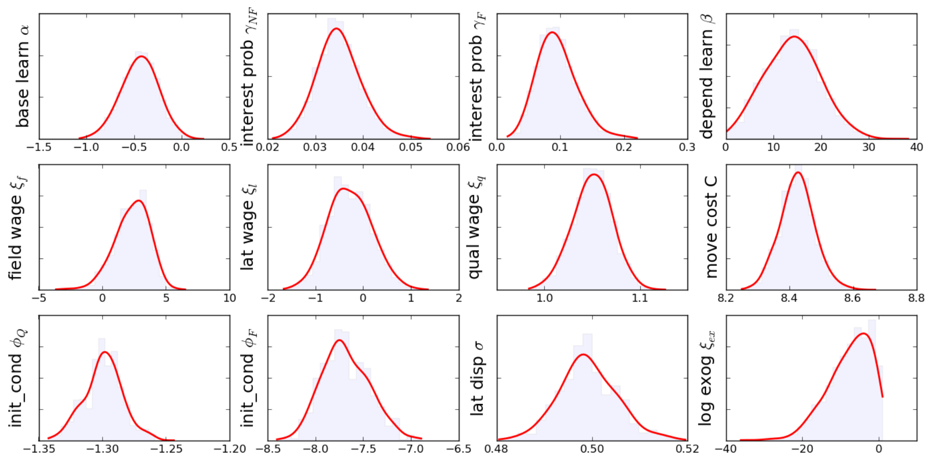

Figure 2: Posterior distributions

	mean	std	25%	50%	75%
α	-0.447	0.179	-0.571	-0.440	-0.322
β	14.128	5.850	9.801	14.138	18.244
γ_F	0.035	0.004	0.031	0.034	0.038
γ_{NF}	0.094	0.031	0.072	0.092	0.112
ξ_f	2.208	1.374	1.419	2.367	3.216
ξ_l	-0.293	0.411	-0.577	-0.289	-0.019
ξ_q	1.050	0.021	1.035	1.049	1.065
$\mathcal{C}$	8.426	0.052	8.387	8.426	8.462
ϕ_Q	-1.302	0.014	-1.312	-1.302	-1.293
ϕ_F	-7.603	0.230	-7.754	-7.611	-7.443
σ	0.499	0.005	0.495	0.499	0.503
ξ_{ex}	0.919	0.739	0.343	0.756	1.312

Table 6: Posterior moments

Posterior moments for the twelve estimated parameters are listed in Table 6, and parameter posterior distribution kernel densities and histograms are plotted in Figure 2. The first row of Figure 2 contains the posterior distributions of the base learning parameter α, the dependent learning parameter β, the interest probability γ_{NF} of those not in Jensen's field and γ_F of those in Jensen's field. The relative magnitudes and signs of the interest parameters are in line with what one might expect. The expected interest probability is a little more than twice as high for academics in Jensen's field. The main parameter of interest β is relatively large and positive, reflecting the importance of colleagues knowledge on own learning.

Interpreting the magnitude of the raw citation parameters is difficult. Figure 3 presents the percent change in annual learning probability from an increase from 0% to 5% of colleagues knowing about a new paper. The size of the effect depends on the latent type. The histogram in in 3 shows typical latent types in the data, which are below zero because of the initial condition equation described in Section 4.4. For the latent types in the data, an increase from 0% to 5% coworker knowledge raises annual learning probabilities by 35-60%.

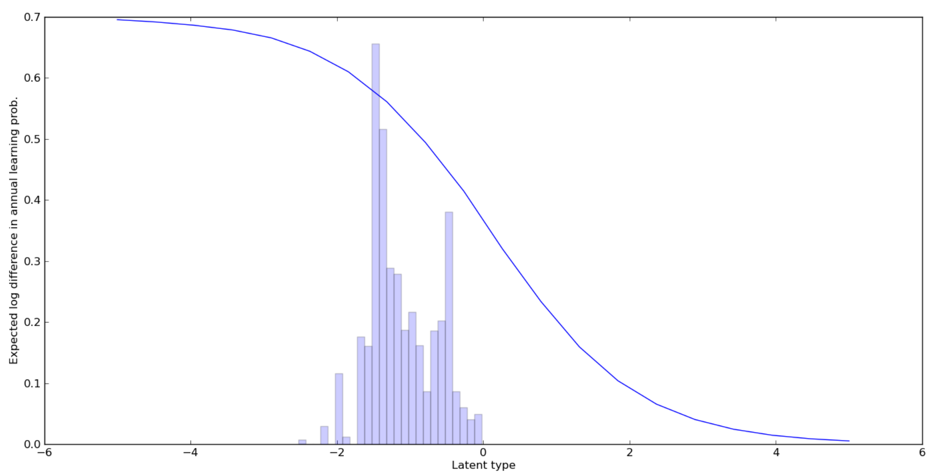

Figure 3: Annual learning probability percent increase, 0% to 5% coworker knowledge of new paper

The bottom two rows of Figure 2 contains posterior distributions for the moving parameters. The field and quality wage interaction coefficients ξ have a positive sign in expectation, meaning that we expect people to sort towards own type in both field and quality dimensions. The coefficient on the field interaction, however, is estimated without much precision and very well may be close to zero or negative. The movement cost parameter is large relative to the wage interactions, providing a strong disincentive to moving. The 1991 payoff effect ξ_{ex} reflects the extent that wages in affected states dropped to generate moving patterns in the data. The mean value is 0.87, which implies that a public school in a state with a budget shortfall of 10% would see a 1991 payoff drop of about $e^{-0.919\times 0.1} \approx 9\%$.[23]

[23]The entry and exit processes of academics is not modeled here, and the size of economics departments has been

The latent type distribution parameter posteriors are located in the bottom row of Figure 2. The first column relates to the initial mean of the distribution of latent type h_i. ϕ_F and ϕ_Q are both negative, so that a department with high field fraction and quality has lower average initial latent type values. This result is consistent with the sorting implied by the negative coefficient on the wage interaction between latent type and field ξ_l. The standard deviation of latent type is estimated to be a bit less than one.

6.1 Intuition for counterfactuals

Before I get to the results of the counterfactual exercise, first I present some intuition using a toy model. The goal in this section is to show that we should expect an increase in movement between departments to make them more similar in terms of knowledge fractions, as well as increase aggregate diffusion. Consider a simple continuous-time theoretical model of diffusion. Let there be a single firm with a continuum of workers in which there is a hazard of learning about a new idea given by:

$$\lambda(t) = \alpha + \beta S(t) \tag{15}$$

S(t) is the share of people in the firm who know about the new idea at time t. As in the empirical model developed above, it is easier to learn about a new paper as more people come to know about it. Suppose that a new innovation is developed at time zero. We can describe the evolution of S by the 2nd-order differential equation:

$$\frac{dS(t)}{dt} = (\alpha + \beta S(t))\,(1 - S(t)) \tag{16}$$

Solving for S gives:

$$S(t) = \frac{\alpha e^{(\alpha+\beta)t} - \alpha}{\alpha e^{(\alpha+\beta)t} + \beta} \tag{17}$$

This is the logistic curve, which has long been used to model the spread of innovations.

Now consider two symmetric firms in which the innovation is spreading independently as above. If the firms are exactly the same, movement will not have any effect on knowledge spread. Suppose instead that one firm, the leading firm, gets a head start learning about the new innovation. The second firm, the lagging firm, begins to learn about the innovation only after some time. This situation is illustrated in Figure 4. The horizontal axis is time since the beginning of learning about the innovation, and the vertical axis is share of people who know about the new idea. The leading firm is farther up the logistic diffusion curve.

growing rapidly over the last thirty years. The model therefore has little to say about what distribution of academics over departments we should expect to see in the data. Even so, Appendix H contains some discussion of the long-run distribution of academics across departments implied by the estimation results, and compares this distribution with what is observed in the data.

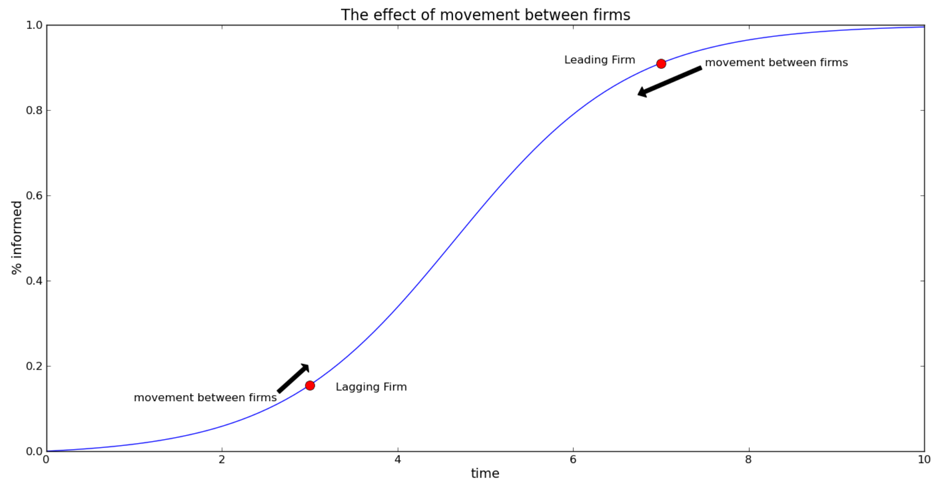

Figure 4: Movement between firms on a diffusion curve

Randomly swap workers between firms. This will raise the share of people in the lagging firm which know about the idea, and lower the share at the leading firm, moving the two firms closer to each other on the diffusion curve. The steeper the section a firm is on, the faster knowledge is diffusing within the firm. In the illustrated case movement between firms will speed up diffusion because both firms will be pulled onto a steeper part of the diffusion curve.

If both firms were on the initial convex part of the diffusion curve, however, one might expect that aggregate diffusion could be slowed down by movement.[24] This is not so. From (16), the effect of an additional worker becoming informed on the rate of diffusion is given by:

$$\frac{dS'(t)}{dS(t)} = -\left(\alpha + \beta S(t)\right) + \beta\left(1 - S(t)\right) \tag{18}$$

The first term on the RHS says that now there are less uniformed workers to learn the new idea, which slows the change in $S(t)$. The second term says that the remaining uninformed workers are more likely to learn about the new idea, which increases the change in $S(t)$. Suppose that a particular time the leading firm has $S(t) = s_h$ and the lagging firm has $S(t) = s_l$, with $s_h > s_l$. Then the change in aggregate diffusion resulting from an informed-uninformed worker swap is given by:

$$-\left(\alpha + \beta s_l\right) + \left(\alpha + \beta s_h\right) + \beta\left(1 - s_l\right) - \beta\left(1 - s_h\right) = 2\beta(s_h - s_l) \tag{19}$$

This expression is positive, so the effect of marginal worker movement on aggregate diffusion is positive. [25]

[24] An earlier draft of this paper made such an informal argument.

[25] If $\beta = 0$ so that learning does not depend on coworkers, then (17) reduces to the exponential distribution.

Even though the model developed in this section is just a toy, the logic goes through to the empirical model developed above. Movement between firms should make firms both more similar in terms of knowledge shares and increase aggregate diffusion.

6.2 Counterfactual results

The counterfactuals in this section involve varying the movement cost parameter $\mathcal{C}$. I begin by drawing a set of parameters from the estimated posterior distribution, and then simulate the model using the academics in my dataset and the ergodic distribution of academics over departments. I draw a department, latent type, interest for each academic using the estimated parameters. Results are generated for four values of $\mathcal{C}$: the full estimated cost parameter, 70% of the parameter, 50% of the parameter, and totally shutting down the cost of moving between departments. In the estimation data, about 3-4% of academics move each year. 70% of the cost parameter is chosen because it induces 9-10% of academics to move each year.

We will focus on three statistics to characterize the generated data: the percentage of academics who have cited the paper, the percentage of departments housing someone who cited the paper, and the coefficient of variation over departments in fraction of members who have cited the paper. Figure 5 plots the expected evolution of those three statistics in the different scenarios. Figure 6 plots the posterior expectation of the log difference between counterfactual statistics and simulated data statistics. The bottom row of Figure 6 is the log difference between the data offer rate and half of the offer rate. In both figures, the dotted lines are the 90% confidence intervals on the expectation of the posterior distribution.

As expected, more mobility increases the fraction of departments employing at least one person who has cited the paper, and the reduces variation in knowledge fractions between departments.

In this case, mixing between firms has no effect on the diffusion rate, as one would expect. To see this, consider two firms both of size one, one at t_1 on the diffusion curve, and the other at t_2. The aggregate diffusion rate is $z^* = \alpha e^{\alpha t_1} + \alpha e^{\alpha t_2}$. Now combine the two firms. The knowledge share at the combined firm is:

$$y^* = \frac{(1 - e^{-\alpha t_1}) + (1 - e^{-\alpha t_1})}{2} = 1 - \frac{z^*}{2\alpha}$$

Find the appropriate time argument associated with share y^* on the diffusion curve:

$$\begin{aligned} y^* &= 1 - e^{-\alpha t^*} \\ 1 - \frac{z^*}{2\alpha} &= 1 - e^{-\alpha t^*} \\ t^* &= \frac{\ln(\frac{z^*}{2\alpha})}{-\alpha} \end{aligned}$$

Finally get the new diffusion rate (the new firm has population two):

$$2\alpha e^{-\alpha t^*} = 2\alpha e^{\ln(\frac{z^*}{2\alpha})} = z^*$$

As expected, combining the firms has no effect on the aggregate diffusion rate. The only way that movement can effect aggregate diffusion is through the dependence parameter β.

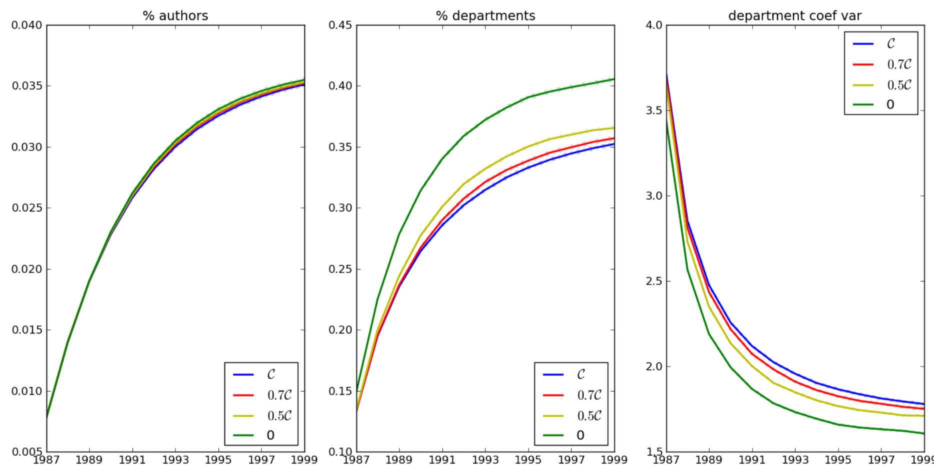

Figure 5: Counterfactual statistics, posterior expectations

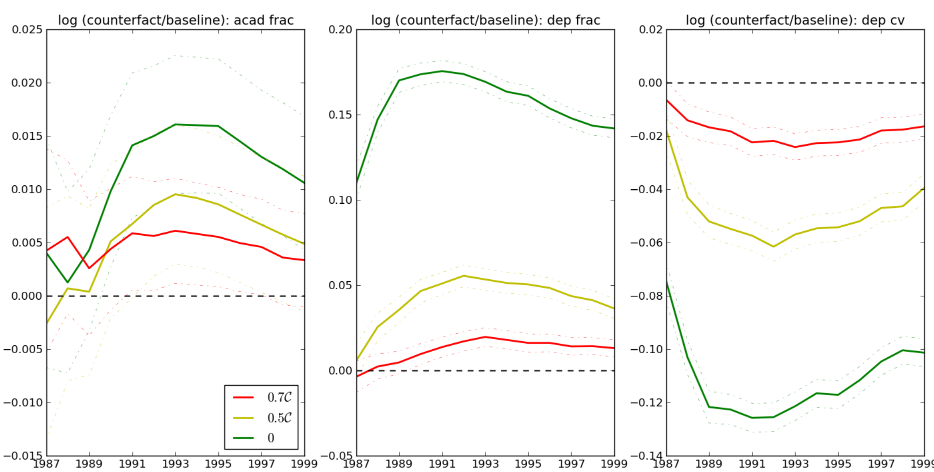

Figure 6: log change in posterior expectations

If we compare the no cost to the benchmark case, then within several years after the idea begins diffusing, we expect 15-20% more departments to house at least one person who knows about the paper. Likewise, we expect a 12% lower coefficient of variation between departments. The less dramatic counterfactuals push diffusion in the same direction, although the effects are smaller. We expect that 1.5% more academics will have heard about a new paper seven years after it is published in the no cost counterfactual.

All of the log difference plots exhibit a U shape because in the model ideas eventually diffuse completely. As time goes to infinity, all potentially interested academics learn about the new paper.

As expected, increasing movement between departments speeds up knowledge diffusion.

7 Cross-country diffusion

This section calibrates a model of international knowledge diffusion using information from the estimated structural model. In particular, I show that a small increase in movement of Chinese academics between China and the United States can significantly increase the diffusion of foreign knowledge in China. In recent years, Chinese scholars have been spending more time as visitors in the United States. In my time at Penn State, our department has housed several visiting Chinese researchers, and a faculty member told me that he frequently receives emails from Chinese economists asking to pay their own way to visit.

As mentioned above, an academic only appears in my data when he publishes in one of the journals tracked by the Web of Knowledge. The data is very sparse for China and other developing countries. Using outside data, I calibrate the structure of the Chinese academic labor market. I assume that the 117 participating universities in a Chinese government program for improving higher education make up the universe of active research universities.[26] Based on my impression from clicking through department websites, I further assume that each department has on average twenty active research faculty. Since I have no information on department or academic quality or field, I assume that departments and academics are homogeneous.

As in the United States, I model academics in China moving between departments, and learning about new papers. Since departments and academics are homogeneous, moves are random. I assume that each Chinese academic has a 2.2% chance of changing his domestic affiliation each year, matching the observed movement rate in the American data.

The probability of learning about a new paper in China is still given by (1), but with different parameter values than in the United States:

$$\frac{e^{\alpha_c+\beta_c K(d,t-1)+h_i}}{1+e^{\alpha_c+\beta_c K(d,t-1)+h_i}} \tag{20}$$

As Chinese moves are random, for simplicity I will assume that all Chinese academics have the same latent type equal to the average latent type of Americans. The first Chinese citation of the Jensen paper in my data is from Hong Kong in 1995, and then from mainland China in 1997. Assuming that Chinese academics have the same probability of interest in the Jensen paper as American academics not in Jensen's field ($\gamma_c = \gamma_0$), I set the Chinese base learning hazard α_c so that the first cite is expected nine years after publication. I calibrate the Chinese dependent hazard β_c so that the increase in learning probability from no coworker knowledge of a new paper to 5% is the same as in the estimated domestic structural model for the average latent type.[27]

[26] I am referring to the 221 Program. For more information on this program see Lixu (2004). I found the list of universities on Wikipedia.

[27] To be clear, I set α_c to satisfy:

Everything related to American academics works exactly as in the domestic structural model. New in this model is international movement of Chinese academics. With annual probability λ_c, a Chinese academic visits a random American department for one year. While in the United States, a potentially interested visitor will learn about the new paper with the American probability (1). I simulate the model with three values of λ_c: 0, 0.01, and 0.02, and 1100 random draws of parameters from the posterior distribution.

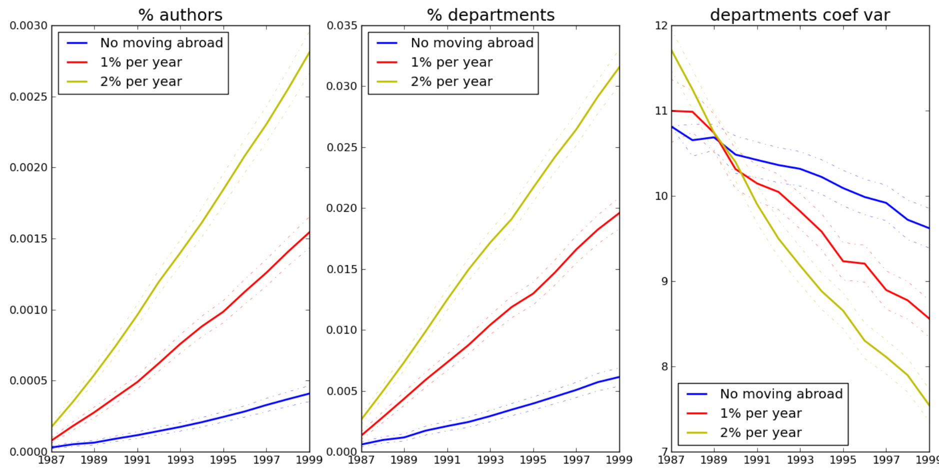

Figure 7: Expected Chinese knowledge diffusion

Figure 7 plots for China the expected evolution of the three statistics we focused on in the counterfactual section above: the percentage of academics who know about the paper, the percentage of departments housing at least one academic who knows about the paper, and the coefficient of variation over departments in fraction of informed academics. International exchange directly increases the fraction of informed Chinese academics because it is easier to learn about the new paper abroad. There is a slight convexity in the left-hand panel plotting the percentage of Chinese academics who have learned about the paper by a given year. There are two causes for the convexity. The first is that it is getting easier over time to learn about the paper in the United States, so an academic is more likely to learn while visiting abroad. This effect dies after the first few years because knowledge about the paper spreads quickly in the United States. Secondary transmission causes the convexity of the line in later years. It is hard to discover a new American paper alone in China, but the knowledge of coworkers greatly facilitates learning.

The calibration contained in this section is rough and suggestive, but it underlines an important

$$\frac{1}{9} = \alpha_c * \gamma_c * |A_c| \tag{21}$$

Here $|A_c|$ is the number of Chinese academics. I set β_c to satisfy the following equality:

$$\frac{e^{\alpha+\beta 0.05+\bar{h}}}{1+e^{\alpha+\beta 0.05+\bar{h}}} - \frac{e^{\alpha+\bar{h}}}{1+e^{\alpha+\bar{h}}} = \frac{e^{\alpha_c+\beta_c 0.05+\bar{h}}}{1+e^{\alpha_c+\beta_c 0.05+\bar{h}}} - \frac{e^{\alpha_c+\bar{h}}}{1+e^{\alpha_c+\bar{h}}} \tag{22}$$

direction for future research. Past research has pointed to large welfare gains from reduction in barriers to migration (Clemens, 2011). Typically this line of research does not consider migrants as vectors for technology diffusion. If migrants can move knowledge between places, then not only will welfare gains to additional migration be larger than the previous literature has estimated, but distribution of welfare gains will change as source country workers benefit from the knowledge of return migrants. More rigorous estimation of international labor movement and knowledge diffusion is a natural next step in this project.

8 Extensions

In this section, three extensions to the basic structural model are presented. In the first extension, I add a national dependent probability to the learning specification (1) above. This model captures a time effect. As more people learn about a new idea, an academic is more likely to run into someone who knows about the idea at a conference or seminar. This model weakens the importance of location, as anyone learning about the paper anywhere increases learning probabilities for all other academics. In (23), $K(n, t-1)$ is the aggregate percentage of academics who cited the paper by $t-1$.

$$\frac{e^{\alpha+\beta K(d,t-1)+\beta_n K(n,t-1)+h_i}}{1+e^{\alpha+\beta K(d,t-1)+\beta_n K(n,t-1)+h_i}} \tag{23}$$

In a second extension, the diffusion process parameters are all allowed to depend on field. That is, if an academic is in the Jensen field, we rewrite (1) all with f subscripts as in (24).

$$\frac{e^{\alpha_f+\beta_f K(d,t-1)+h_i}}{1+e^{\alpha_f+\beta_f K(d,t-1)+h_i}} \tag{24}$$

The last extension is a simple publication lag. I assume that if we observe a cite in, say, 1991, the academic actually learned about the paper in 1990. To maintain comparability with the other model specifications, I maintain the assumption that an academic cannot learn about the Jensen paper until 1987, the year after it was published. To estimate the publication lag extension, I pool the three observed 1987 cites in with the observed 1988 cites.

Priors are the same relatively uninformative priors used in the baseline model. Table 7 compares the expectations of posteriors for the baseline model and extensions.[28] In the all extensions, the movement related parameters at the bottom of Table 7 are similar to those in the baseline model.

First consider the national dependence specification. The department level β is about half of the size of that estimated in the baseline model, and the national parameter β_n is large. The national β_n is, of course, multiplied by very small numbers since relatively few people ever cite the Jensen paper overall. The posterior for interest levels γ and base learning parameter α are similar to those in the baseline model.

[28] Estimated posterior kernel densities for all extensions can be found in Appendix J.

param	baseline		field-dep		nation		lag	
α	-0.447	(0.179)	-0.574	(0.204)	-0.955	(0.184)	-0.252	(0.159)
α_f			0.170	(0.488)				
β	14.128	(5.850)	16.991	(5.881)	6.537	(4.654)	10.418	(4.295)
β_f			78.846	(71.576)				
β_n					81.257	(16.031)		
γ_{nf}	0.035	(0.004)	0.036	(0.005)	0.027	(0.003)	0.028	(0.003)
γ_f	0.094	(0.031)	0.078	(0.032)	0.071	(0.022)	0.084	(0.026)
ξ_f	2.208	(1.374)	2.310	(1.379)	2.347	(1.434)	2.302	(1.379)
ξ_l	-0.293	(0.411)	-0.269	(0.425)	-0.082	(0.421)	-0.095	(0.423)
ξ_q	1.050	(0.021)	1.051	(0.020)	1.084	(0.019)	1.084	(0.018)
ϕ_Q	-1.302	(0.014)	-1.302	(0.017)	-1.297	(0.015)	-1.298	(0.016)
ϕ_F	-7.603	(0.230)	-7.627	(0.265)	-7.629	(0.224)	-7.606	(0.234)
σ	0.499	(0.005)	0.499	(0.005)	0.496	(0.005)	0.496	(0.005)
$\mathcal{C}$	8.426	(0.052)	8.425	(0.049)	8.651	(0.049)	8.655	(0.049)
ξ_{ex}	0.919	(0.739)	0.001	(0.026)	0.095	(0.282)	0.082	(0.245)

Table 7: Posterior expectations, extensions versus baseline

As for the field-specific parameter model, the posteriors for those not in the Jensen field are similar to the baseline model. relatively small sample size causes the parameters for those in Jensen's field to be estimated with less accuracy. The base learning parameter α and the dependent learning parameter β are both higher for those in the field.

The publication lag extension looks fairly similar to the baseline model. The dependent learning parameter β is a bit lower than in the baseline, and the base learning parameter α is a bit higher. This is due to the model trying to match the larger number of 1987 citers. Since there is no colleague knowledge at that point, base learning must be ratcheted up to rationalize learning.

8.1 Model Checking

This section uses the baseline model as well as the three extensions to simulate data, and then compares statistics of the simulated data to the same statistics of the observed data. We will check the models on same three dimensions: the diffusion of citations among academics, the diffusion of citations between departments, and the coefficient of variation across departments of percentage of academics who have cited the paper. For each exercise, 2000 vectors of parameter values are drawn from the posterior distributions, and the model is simulated at each parameter value. The movement posteriors in the regional and field dependence extensions are nearly identical to those in the baseline model. In order to increase computation speed, we simulate moves out of the baseline model, and then simulate idea diffusion using the baseline and extended models separately. Both moves and citation times are simulated separately for the publication lag model.

Figure 8 contains posterior means for the three model scenarios and data. When interpreting

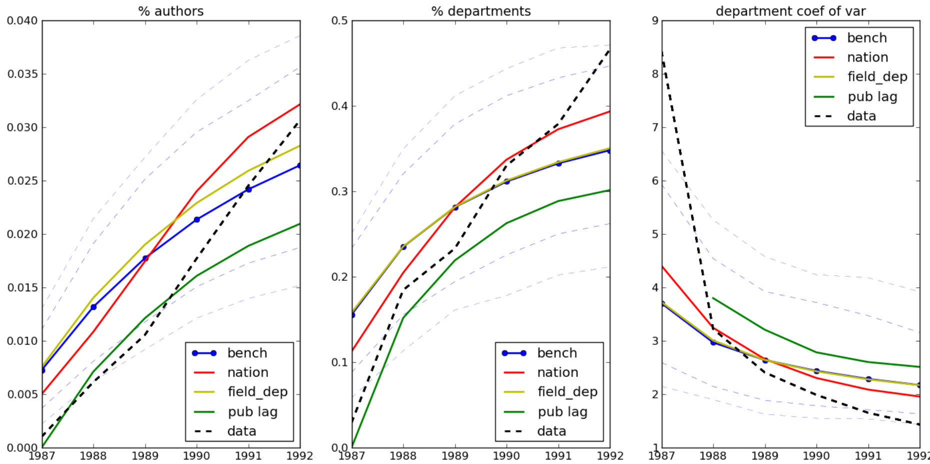

Figure 8: Model checking, simulations vs data

this exercise, the reader should keep in mind that we are comparing means of many simulations to the data, which should be thought of as a single random realization. We have only 122 first citations in the data, and most of these come toward the end of the data period. There is sizable random variation in the simulated trajectories. To make this point, the 95% and 99% credible intervals of the baseline model are included as faint lines in Figure 8. Save for the first year, the data is always within the 99% credible interval of the baseline model.

Except for the publication lag model, all models overestimate the number of citers in the first year in the raw data. In the raw data, there are only three citing academics in the first year. As time goes by, the levels in the data and in the models become more similar. Of the three models, the publication lag model does the best in the first few years, but misses the data in the final years. The national dependence model fits the qualitative slope of the data the best, but its level is too high for the entire simulation period. The benchmark and field-specific parameter models display very similar behavior, and are generally in between the baseline and the publication lag model in level.

9 Probit Model

This section estimates an alternative reduced-form model for knowledge diffusion. The probit model developed here is for citing in 1992, and uses 1991 budget deficits as an instrument for knowledge fractions. The idea is that budget cuts affect movement and substitution patterns across departments. To give an example, suppose that Berkeley was planning on hiring a junior faculty member in contract theory in 1991, but couldn't because there was a hiring freeze. The junior contract theorist who had already cited Jensen's paper and would have gone to Berkeley instead went to NYU. The exogeneity condition is that a budget cut only affects citation probability

	cit92 b/se	cit92 b/se	cit92 b/se	cit92__93 b/se	cit92__93 b/se	cit92__93 b/se
probit eq 1						
kfrac	26.952*	37.604*	40.753**	26.362**	38.034**	41.238**
	(14.78)	(19.84)	(16.95)	(11.68)	(15.97)	(18.75)
field		0.000	0.000		0.023	0.029
		(.)	(.)		(0.18)	(0.12)
qual		0.001	0.001		0.001	0.001
		(0.00)	(0.00)		(0.00)	(0.00)
dep_qual		-0.048	-0.053		-0.052*	-0.070*
		(0.04)	(0.04)		(0.03)	(0.04)
dep_field		1.996	1.275		1.366	1.151
		(4.21)	(5.64)		(3.22)	(4.58)
public			0.060			-0.010
			(0.30)			(0.32)
dep_size						0.006
						(0.01)
_cons	-2.513***	-1.012	-0.491	-2.319***	-0.735	-0.277
	(0.64)	(2.77)	(3.84)	(0.47)	(2.08)	(3.65)
probit eq 2	kfrac	kfrac	kfrac	kfrac	kfrac	kfrac
bd	-0.061***	-0.023	-0.014	-0.061***	-0.023	-0.011
	(0.01)	(0.05)	(0.06)	(0.01)	(0.05)	(0.06)
field		0.000	0.000		-0.001	-0.001
		(.)	(.)		(0.00)	(0.00)
qual		-0.000	-0.000		-0.000*	-0.000*
		(0.00)	(0.00)		(0.00)	(0.00)
dep_qual		0.001***	0.001***		0.001***	0.002***
		(0.00)	(0.00)		(0.00)	(0.00)
dep_field		-0.005	-0.004		-0.005	-0.011
		(0.08)	(0.08)		(0.08)	(0.08)
public			-0.001			-0.000
			(0.01)			(0.01)
dep_size						-0.000
						(0.00)
_cons	0.024***	-0.009	-0.008	0.024***	-0.010	-0.009
	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
CD Wald F	31.41	5.66	0.74	16.38	5.66	0.74
obs	2940	2845	2845	2940	2940	2940

Table 8: Probit model

through its affect on knowledge fractions at departments. Let C_i^{1992} be a dummy which is one if academic i cited the Jensen paper for the first time in 1992, let $\mathbf{X}_{o,i}$ be a vector of observed characteristics, let $b_{d,i}$ be the budget shortfall in the state of the (public university) department in which academic i worked in 1991, and let K_i, the knowledge fraction, be the fraction of coworkers who have cited Jensen before 1992. The observations are all academics who have not cited the Jensen paper as of 1992. The natural probit specification is:

$$C_i^{1992} = 1_{\{\beta_X \mathbf{X}_{o,i} + \beta_K K_i + \varepsilon_i > 0\}} \tag{25}$$

And:

$$K_i = \Gamma_X \mathbf{X}_{o,i} + \Gamma_{b_d} b_{d,i} + v_i \tag{26}$$

Here we assume that v_i and ε_i are jointly normal and correlated. I estimate the probit twice using the ivprobit function in STATA, once using 1992 citers as described above and once assuming the budget deficit effect lasted for two years, with the dependent variable being a dummy for either a 1992 or a 1993 first cite. The right-hand-side variables are the 1991 analogues of the quantities in the structural model. Field is a dummy which is one if an academic has the field of contract theory or business economics. Department field fraction is the mean field value of academics in the department in 1991. Quality is mean lifetime citations per paper, and department quality is the mean quality in the department in 1991. Department size is just the number of authors in the department in 1991, and public is a dummy for public universities.

The probit results are contained in Table 8. Field is omitted in the second and third models because it is a perfect predictor of not citing for the first time in 1991. Department size is omitted in the third model because the likelihood would not converge with it included. All standard errors are clustered at the department level. The direction of the kfrac (K_i) coefficient is significant and in the expected direction in all models. Some of the Cragg-Donald F Statistics are lower than ideal. The rule of thumb from Staiger and Stock (1997) is that this statistic should be greater than ten. The instrument may be weak in some specifications. A back of the envelope calculation using the models with only K_i indicates that if 5% of coworkers know about a new paper rather than 0% of coworkers, citing probability is higher by about $\Phi(-1.15) - \Phi(-2.5) \approx 12\%$. Appendix A contains two additional reduced-form exercises which provide evidence on the importance of location on knowledge diffusion.

10 Conclusion

This paper develops a model of movement and the diffusion of knowledge between firms. The model is estimated on data from a panel of academics and the diffusion citations. Both the main structural model and a reduced-form exercises show that physical proximity facilitates learning about a new idea. In a counterfactual section, I find that increased worker mobility speeds the diffusion of knowledge between locations, reduces the dispersion in fraction of informed workers across firms, and has a positive effect on the total diffusion of ideas across workers. In a calibrated exercise describing Chinese scholarly visits to the United States, I find that the international movement of workers can have a large effect on domestic idea diffusion in a developing country.

There are several directions in which to develop this research. One is to examine the effect of the internet on the diffusion of ideas. My data span the early 1980's when there was no internet to the present. The speed at which citations diffuse in the data should be informative about how the internet has affected idea diffusion. A second and stickier direction is to explicitly model serially correlated shocks which affect both sorting and citing. Recent research by Arcidiacono and Miller (2011) in estimating dynamic discrete choice models with serially correlated, unobserved state variables might prove useful for such an exercise. Finally, as mentioned in the section on Chinese migration, using a similar model to rigorously estimate the effect of labor migration on international technology diffusion is a natural next step.

Appendices

A Additional evidence on location and idea diffusion

This section contains several independent empirical exercises which support the finding of the structural model, that physical location is an important part of knowledge diffusion. In particular, the two exercises below show that that those in the same department as an author learn about his new work first.

A.1 Those nearby are the first to learn about new ideas

The first exercise follows Jaffe et al. (1993). Suppose we have a paper C which cites another paper O, and another paper R in the same field as C. I compare the probability that any author of O shares a department with any author of C, to the probability that any author of O shares a department with any author of R. If the papers O and C are more likely to come from the same department, we will take it as evidence that face-to-face contact spreads ideas. The exercise will show that being nearby is important in the first few years after a paper is published, but after some time location no longer matters. The time element suggests that, just like infectious disease, ideas diffuse over time.

There three kinds of papers in the exercise: originating papers (O), citing papers (C), and reference papers (R). An originating paper is where the analysis starts. To be concrete, let the set of originating papers be all economics papers published in 1980. A citing paper is any economics paper which cites an originating paper. To each citing paper, a reference paper is matched. A reference paper is published in the same year as a citing paper, and shares a similar field. Recall that a paper's field is a unit vector. To choose a reference paper, I consider all papers published in the same year as a citing paper, and choose the paper with a field vector closest to the citing paper in the Euclidean sense. Any citing paper which shares an author with its originating paper, a self-cite, is dropped.

Each paper is associated with the departments of its authors. For each year after the publication of an originating paper, I calculate the fraction of citing papers which share a location with their originating papers, and the fraction of reference papers which share a location with their originating papers. In Figure 9, I use all papers published in 1980, all papers published in 2005, and the most cited 100 economics papers as originating papers. The blue line is the percentage of citing papers which share a department with the originating papers, and the red line is the percentage of reference papers which share a department with the originating papers. The lighter lines are 5% confidence intervals.

Citing papers are initially more likely to come from the same department compared to reference papers, but the effect fades over time. A little under a decade after a paper is published those outside the originating department are just as likely as those in the department to cite the originating

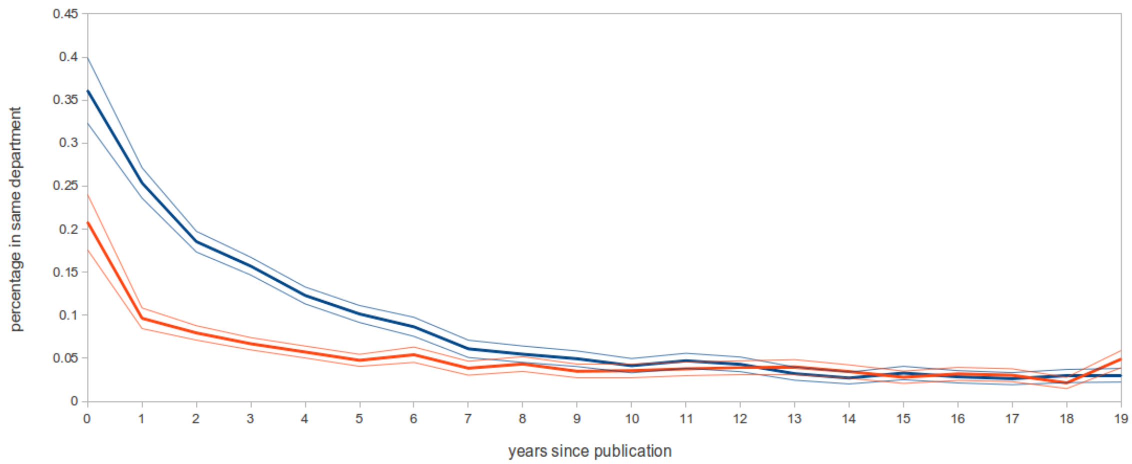

Figure 9: Results of Jaffe exercise, Dark Blue = Citing Papers, Light Red = Reference Papers

paper. Like a disease, it takes knowledge of a paper some time to diffuse outside the originating department. Similar results using patent data were recorded in Jaffe et al. (1993).

A.2 Dealing with endogenous sorting

The Jaffe exercise above recently came under attack. Thompson and Fox-Kean (2005) redid the original exercise with more detailed patent data, and nearly all of the important results in the original were overturned. Applied to my situation, the Thompson critique is that field does not adequately control for the endogenous sorting of academics with similar interests into the same department. The reason I observe departmental colleagues citing new research first may just be that coworkers are the most interested in each other's work. Controlling for field as I did in the Jaffe exercise above mitigates this problem, but will not eliminate it completely. In Appendix B a simple related model shows that bias will make spillovers look stronger than they are.

In order to deal with possible bias, in a second exercise I control for research interest more carefully and use an instrumental variable. As in the structural exercise above, I will use the work of Michael Jensen. I downloaded information on all the citations of all of his papers, and collected the number of total times each economist cited any of Jensen's papers as of 2013. An economist who cites Jensen more often is more interested in Jensen's research. I then choose a single Jensen paper, as above his 1986 *American Economic Review* piece "Agency costs of free cash flow, corporate finance, and takeovers".

I record the year of first citation for each of the 669 academics who ever cited the 1986 paper. My hypothesis is that, conditioning on interest in Jensen's work, those who ever worked at Harvard will cite Jensen's paper earlier. Let CY_i be the year that author i first cited the 1986 Jensen paper. $Harv_i$ is a dummy which is set to 1 if author i ever worked at Harvard. $Jcits_i$ is the total number of times author i ever cited any of Jensen's papers. I run the following regression:

$$CY_i = \beta_0 + \beta_H Harv_i + \beta_J Jcits_i + \beta \mathbf{X}_i + \varepsilon_i \tag{27}$$

Here $\mathbf{X}_i$ contains characteristics of author i such as dummies for the first year i was observed in my data.

	Citation year	Citation year	Citation year	Citation year
Worked at Harvard	-2.38**	-2.13*	-2.04**	-27.93**
Total Jensen cites	–	-0.13***	-0.10***	-0.05
First year dummies	no	no	yes	yes
Only first cohort	no	yes	no	no
Instrumented	no	no	no	yes
IV First Stage				Worked at Harvard
Quality				.001***
Total Jensen cites				.002
First year dummies				yes
Obs	669	438	669	669
R^2	0.01	0.03	0.11	–

Table 9: Effect of working at Harvard on year of Jensen citation

Table 9 contains the estimation results from (27). The only first cohort model uses only authors who were working in 1986. The correlations described in the table support the hypothesis that location matters. In the simple OLS models of the first three columns, those who worked at Harvard with Jensen cited him around two years earlier than others on average, and for every other citation of Jensen's other work an author cited Jensen's new paper on average a month or so earlier.

The instrument used in the last column is the quality of an academic measured by mean coauthor-adjusted citations per published paper. The exogeneity assumption is that quality is correlated with working at Harvard, but quality only affects the timing of citing Jensen's paper through its effect on location. The instrumented effect of working at Harvard is both very strong and statistically significant. I ran placebo tests on everything in the regression table using a Princeton, Berkeley, and University of Pennsylvania dummy rather than a Harvard dummy. The coefficients on the placebo dummies were never statistically significant.

B Sorting and Bias

Sorting of academics into departments is not random. People tend to work alongside others with similar interests. The econometric challenge in this paper is sorting out how working together affects citing behavior and how having similar interests affects citing behavior. To motivate the difficulty, suppose the hazard of citing a paper is observed, and given by:

$$\lambda_{it} = \beta_0 + \beta_1 \text{dep\_frac}_{it} + \beta_2 \text{interest}_i + \varepsilon_{it} \tag{28}$$

In words, the hazard of citing a paper depends on the fraction of colleagues who know about the paper as well as personal interest in the topic. Suppose that instead of estimating the true model above via OLS, we estimated:

$$\lambda_{it} = \gamma_0 + \gamma_1 \text{dep\_frac}_{it} + \varepsilon_{it} \tag{29}$$

It is a standard result that the asymptotic expected value for the estimator $\hat{\gamma_1}$ can be written:

$$\mathbb{E}[\hat{\gamma_1}] = \beta_1 + \rho_{\{\text{dep\_frac,interest}\}} \sqrt{\frac{\sigma^2_{\text{interest}}}{\sigma^2_{\text{dep\_frac}}}} \tag{30}$$

ρ is the Pearson correlation coefficient, and here we expect it to be positive. This specification, then, leads to an overestimation of the effect of being in the same department on the hazard of citing a paper. In most of this paper, I deal with similar bias by using a dummy for working in the same field as a paper. Suppose that I now estimate the following model:

$$\lambda_{it} = \delta_0 + \delta_1 \text{dep\_frac}_{it} + \delta_2 \text{field} + \varepsilon_{it} \tag{31}$$

Now we can write the asymptotic expected value of the estimator $\hat{\delta_1}$ as (Hanushek and Jackson 1977):

$$\mathbb{E}[\hat{\delta_1}] = \beta_1 + \frac{\rho_{\{\text{dep\_frac,interest}\}} - \rho_{\{\text{dep\_frac,field}\}}\rho_{\{\text{field,interest}\}}}{1 - \rho^2_{\{\text{dep\_frac,field}\}}} \sqrt{\frac{\sigma^2_{\text{interest}}}{\sigma^2_{\text{dep\_frac}}}} \tag{32}$$

If field and interest are perfectly correlated, then the second term is eliminated and the estimator is no longer biased. The less well field acts as a proxy for interested, the more biased the estimator will be. As long as we believe that field and interest are positively correlated, the estimate of β_1 is biased upward.

Loosely speaking, this intuition goes through for all the exercises in the paper. The second exercise in the reduced form section is an exception. In that exercise I construct an interest variable more detailed than field, and also use an instrument to help with identification.

C Data construction

C.1 ISI Web of Knowledge

My primary data source is the Thomson-Reuters Web of Knowledge (http://thomsonreuters.com/web-of-knowledge/). The Web of Knowledge is a citation database including conference proceedings, journal articles, books, and patents. The Web of Knowledge is similar to other citation databases such as Google Scholar. One difference is that Google Scholar indexes working papers from a variety of sources, while the Web of Knowledge tracks only published papers. For my purposes, the most important distinction is that in the Web of Knowledge, there is a uniform page for each paper containing summary details such as academic names and affiliations. Google Scholar links to outside web pages which each have different information formats.

Using the python library beautiful soup, I scraped data from the Web of Knowledge. The program started with a list of all papers classified as economics papers by the Web of Knowledge (a distinction based on the journal the paper was published in), and clicked through the link to each paper one at a time. Detailed information on the paper was then recorded on my hard disk. In particular, I recorded the following information for each paper:

1. Academic names
2. Academic affiliations
3. Paper Title
4. Journal Title
5. Number of citing papers

In addition to getting this information for the most cited 100,000 economics papers, I also recorded the same information for every paper (not necessarily economics) which either cited an economics paper published in 1980, cited an economics paper published in 2005, or cited one of the one hundred most cited economics papers of all time.

I cleaned and processed the data using a large data processing tool called OpenRefine. I used the cleaned data to link academics to departments. There were several difficulties in doing this. The first is that I dropped all information about a department except the name of the university. While university names are recorded fairly consistently in the database, department information is not. One affiliation might list Harvard University, Economics Dept. Another might list Harvard University, Department of Economics, and still another might give no department information at all. The upshot is that I conflate every Harvard department together, so that the Harvard Business School, the Kennedy School of Government, and the Economics Department are all considered to be the same location for my purposes.

A second, similar problem is in recording the names of academics. First and middle names are sometimes completely recorded, and sometimes only initials are given. I dealt with this by dropping all first and middle names except the first initial of the first name. This certainly causes problems with common names, especially with Chinese names like Li. If two Li's have the same first initials, they will be conflated in my data.

Another difficulty is with academics who have several affiliations within a given year. Many economists list the National Bureau of Economic Research as a second affiliation, for instance. I dropped all NBER affiliations, and dealt with other affiliation problems case by case.

C.2 Constructing Fields

C.2.1 IDEAS Fields

In the main structural exercise we need fields for each academic. In some of the additional exercises reported in Section A, we need fields for papers as well. My main source for this information is IDEAS, a database of economists hosted by the St. Louis Federal Reserve Bank (http://ideas.repec.org/). Ideas allows economists to register themselves, report affiliation, and report current working papers and publications. Using this information, IDEAS ranks economists and institutions along a number of dimensions. Registering with IDEAS is voluntary, and some 37,000 economists have registered.

IDEAS classifies economists into field based on something called the NEP mailing lists. NEP, for New Economics Papers, curates new articles appearing on ideas into 91 different categories. Each category is curated by a particular economist, and over time people take turns being curator. Every so often, in my experience about once a week, an email in each category is sent out listing new papers. IDEAS puts academics into categories based on which mailing list distributes their papers. If either 5 of an economists papers have been included in a particular mailing list, or at least 25% of all of an economists papers have been included, then the economist is deemed to be working in the field of the mailing list.

The IDEAS website maintains a list of economists classified in this way (http://ideas.repec.org/i/e.html). An economist can be classified as working in any number of fields, at least any number of to 91. I again used python and beautiful soup to record every economist affiliation on IDEAS. This amounts to about 30,000 economists. In the structural section, when I say that an economist has the same field as the Jensen paper, I will mean that IDEAS lists him as working in either the field of contract theory, or business economics.

C.2.2 Journal Fields

In some exercises in this study, a paper field is required as well. To get a field for each paper, I combine the academic field described above with a journal field. For the journal field, I use the classification of Barrett et al. (2000). These classifications are JEL field, which is different than the NEP field I have for academics. Using the JEL classification descriptions from the JEL website, I linked the 91 NEP fields to the JEL fields the journals were classified into. I first used fuzzy matching on words in the field descriptions, and then went through and hand corrected odd matches.

To construct the field for each paper, I added together the 91 x 1 field vector of all of the academics, then added the 91 x 1 field vector of the journal the paper was published in. I then normalized to that the resulting vector is a unit vector. Distance between paper fields is then the Euclidean distance between 91 x 1 field vectors.

C.3 Simplifications

As mentioned in the main body of the paper, I cut out all but the top 104 departments when performing the structural estimation. Academics at lower ranked departments publish more rarely, so that I do not observe them very often making the affiliation data noisy. Table 10 at the end of this file is a sample of included departments and fields.

All the data collection program files will be available on my website soon.

D Verifying contraction mapping

Let T be the operator on the space of bounded functions f on the space $1, 2, \ldots, D$. In particular, let T be defined as follows:

$$(Tf)(d) = \rho \ln \left(\sum_{d' \in \mathcal{D}} e^{f(d') + w(d') - \mathbb{1}_{\{d' \neq d\}} \mathcal{C}} \right) \tag{33}$$

We will verify that T is a contraction mapping using Blackwell's sufficient conditions. Recall Theorem 3.3 from Stokey (1989):

Theorem 3.3 (Blackwell's sufficient conditions for a contraction): Let $X \subseteq \mathbb{R}^l$, and let $B(X)$ be a space of bounded functions $f : X \to \mathbb{R}$, with the sup norm. Let $T : B(X) \to B(X)$ be an operator satisfying:

a. (monotonicity) $f, g \in B(X)$, and $f(x) \leq g(x)$, for all $x \in X$, implies $(Tf)(x) \leq (Tg)(x)$, for all x in X.
b. (discounting) there exists some $\beta \in (0, 1)$ such that $[T(f + a)](x) \leq (Tf)(x) + \beta a$, for all $f \in B(X)$, $a \geq 0$, $x \in X$.

In (33), monotonicity is immediate. Discounting is almost immediate as well. Take $a \geq 0$, and let $f(x) \in B(X)$:

$$\begin{aligned} [T(f + a)](d) &= \rho \ln \left(\sum_{d'} \left(e^{f(d') + w(d') - \mathbb{1}_{\{d' \neq d\}} \mathcal{C} + a} \right) \right) \\ &= \rho \ln \left(e^a \sum_{d'} \left(e^{f(d') + w(d') - \mathbb{1}_{\{d' \neq d\}} \mathcal{C}} \right) \right) \\ &= (Tf)(d) + \rho a \end{aligned} \tag{34}$$

Both monotonicity and discounting hold, so by Blackwell's sufficient conditions the operator T is a contraction.

DEPT	QUAL	JENS. FIELD	NEP ABREV	FIELD NAME
UNIV MICHIGAN	0.884	0.035	NEP-ACC	Accounting & Auditing
PURDUE UNIV	0.490	0.0036	NEP-AFR	Africa
PENN STATE UNIV	0.663	0.067	NEP-AGE	Economics of Ageing
HARVARD UNIV	1.0	0.014	NEP-AGR	Agricultural Economics
UNIV PENN	0.903	0.041	NEP-ARA	Arab World
CORNELL UNIV	0.817	0.016	NEP-BAN	Banking
UNIV ROCHESTER	0.355	0.086	NEP-BEC	Business Economics
MIT	0.980	0.086	NEP-CBA	Central Banking
UNIV MARYLAND	0.798	0.060	NEP-CBE	Cognitive & Behavioural Economics
STANFORD UNIV	0.932	0.028	NEP-CDM	Collective Decision-Making
UNIV DELAWARE	0.346	0.0	NEP-CFN	Corporate Finance
CARNEGIE MELLON UNIV	0.625	0.047	NEP-CIS	Confederation of Independent States
YALE UNIV	0.923	0.026	NEP-CMP	Computational Economics
PRINCETON UNIV	0.971	0.051	NEP-CNA	China
UNIV ILLINOIS	0.557	0.031	NEP-COM	Industrial Competition
BOSTON UNIV	0.913	0.012	NEP-CSE	Economics of Strategic Management
UNIV N CAROLINA	0.403	0.0	NEP-CTA	Contract Theory & Applications
...				
UNIV NEW MEXICO	0.038	0.140	NEP-ORE	Operations Research
CUNY HUNTER COLL	0.105	0.0	NEP-PBE	Public Economics
UNIV MIAMI	0.086	0.0	NEP-PKE	Post Keynesian Economics
TUFTS UNIV	0.538	0.0	NEP-POL	Positive Political Economics
BRIGHAM YOUNG UNIV	0.336	0.0	NEP-PPM	Project, Program & Portfolio M
UNIV CALIF IRVINE	0.644	0.0	NEP-PUB	Public Finance
UNIV HAWAII MANOA	0.144	0.0	NEP-REG	Regulation
EMORY UNIV	0.326	0.0	NEP-RES	Resource Economics
UNIV CALIF SAN DIEGO	0.855	0.056	NEP-RMG	Risk Management
WELLESLEY COLL	0.432	0.257	NEP-SBM	Small Business Management
CUNY	0.375	0.0	NEP-SEA	South East Asia
DREXEL UNIV	0.067	0.0	NEP-SOC	Social Norms & Social Capital
MIDDLEBURY COLL	0.182	0.0	NEP-SOG	Sociology of Economics anagent
SANTA CLARA UNIV	0.115	0.0	NEP-SPO	Sports & Economics
UNIV CALIF SANTA CRUZ	0.615	0.0	NEP-TID	Technology & Industrial Dynami
SUNY ALBANY	0.394	0.0	NEP-TRA	Transition Economics
TULANE UNIV	0.317	0.192	NEP-TRE	Transport Economics
APPALACHIAN STATE UNIV	0.269	0.0	NEP-TUR	Tourism Economics
RENSSELAER POLYTECH INST	0.298	0.0	NEP-UPT	Utility Models & Prospect Theo
CHAPMAN UNIV	0.548	0.0	NEP-URE	Urban & Real Estate Economics

Table 10: Selected departments and fields.

E Deriving Emax expectation

In this appendix, I show that if X and Y are constants, and ε_1 and ε_2 are distributed IID Gumbel (0,1), then:[29]

$$\mathbb{E}\left[\max\{X+\varepsilon_1, Y+\varepsilon_2\}\right] = \gamma_e + \ln\left(e^X + e^Y\right) \tag{35}$$

The CDF of the Gumbel distribution is $F(z) = e^{-e^{-z}}$, so we can write the distribution of the maximum in (35) as:

$$G(z) = e^{-\left(e^{X-z}+e^{Y-z}\right)} \tag{36}$$

Take the derivative with respect to z to get the PDF:

$$g(z) = \left(e^{X-z}+e^{Y-z}\right)e^{-\left(e^{X-z}+e^{Y-z}\right)} \tag{37}$$

Now we can rewrite the LHS of (35) as:

$$\int_{-\infty}^{\infty} zg(z)dz \tag{38}$$

Use the change of variable $t = e^{X-z}+e^{Y-z}$ (the trick is that $\ln(t) = \ln\left(e^X+e^Y\right) - z$):

$$\int_0^{\infty}\left(\ln\left(e^X+e^Y\right) - \ln(t)\right)e^{-t}dt = \ln\left(e^X+e^Y\right) + \gamma_e \tag{39}$$

The substitution of γ_e is an identity.[30]

F MCMC diagnostics

This appendix presents mixing plots and diagnostics from the parallel MCMC estimation. Eyeballing the plots in Figure 10, everything looks good. Eyeballing, however, is not very reliable. Table 11 presents Gelman-Rubin convergence criterion results. Recall that to pass the Gelman-Rubin convergence criterion, the test value must be below 1.1. All parameters pass the Gelman-Rubin test.

[29] My Russian officemates tell me that this is just trivial probability theory. Even so, I saw the result in several papers which all referenced Rust (1987), but it is not derived there either! I leave the derivation here for future puzzled American graduate students.

[30] This identity can be found in both the Wikipedia and Wolfram Mathworld articles on the Euler-Mascheroni constant. The Wolfram article references the textbook Whittaker and Watson (1996) after a list of identities involving γ_e, but I was not able to find this particular identity there.

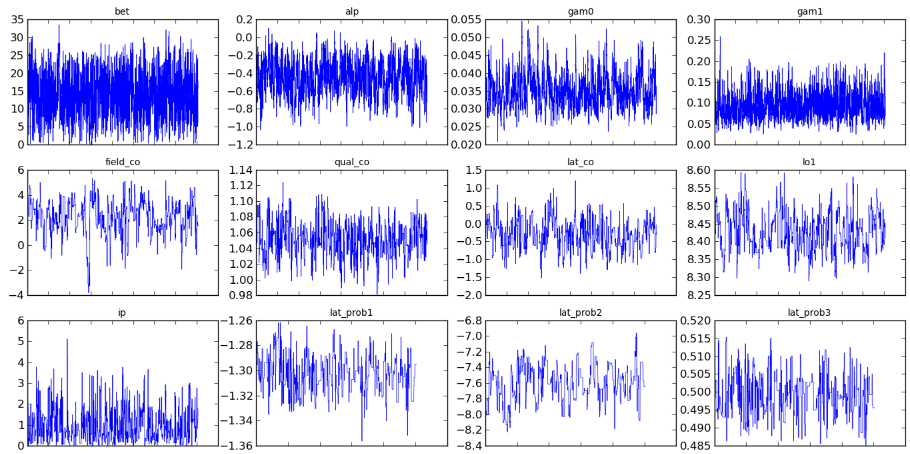

Figure 10: Mixing plots

parameter	GR criterion
α	1.013
β	1.003
γ_F	1.019
γ_{NF}	1.005
ξ_f	1.016
ξ_l	1.008
ξ_q	1.029
ϕ_Q	1.070
ϕ_F	1.034
σ	1.073
$\mathcal{C}$	1.010
ξ_{ex}	1.013

Table 11: Gelman-Rubin Test

G Patent vs Academic Citations

Citations are footprints left behind by ideas moving between brains. If an academic uses an old idea in a new paper, he cites it. At least since the late 1980's, researchers have used citation data to measure the spread of ideas, although more often the citations have been of patents rather than academic citations (Griliches, 1998). Some researchers have been explicit about why they prefer patent data. (Jaffe and Trajtenberg, 2002) write: "Academics may cite a friend (or neighbor) just to be nice, since the price of doing so is infinitesimal, or even negative if a longer list of references is perceived as making the research look more thorough. An inventor who did the same is in effect leaving money lying on the table: if those citations are included in the final patent, the inventor has reduced the scope of her monopoly." Few understand the language of *quid pro quo* better than academics, but the problem of undeserved citation is no less severe in patenting. It isn't clear to me how citing an irrelevant patent hurts an inventor. In fact, the value of a patent is related to the number of other patents which cite it, so there is an incentive for inciting there as well.

In some ways academic citations are better than patent citations at measuring knowledge diffusion. The rules for who is listed as inventor on a patent are as complicated as a rocket schematic. In a 1972 legal opinion, Judge Newcomer of the Eastern Pennsylvania Circuit Court reflected on the meaning of inventorship:[31]

> [Joint inventorship] is one of the muddiest concepts in the muddy metaphysics of the patent law. On the one hand, it is reasonably clear that a person who has merely followed instructions of another in performing experiments is not a co-inventor... To claim inventorship... perhaps one need not be able to point to a specific component as one's sole idea, but one must be able to say that without his contribution to the final conception, it would have been less – less efficient, less simple, less economical, less something of benefit.

Due to difficulties with inventorship, the patent literature has been confined to studying the flow of ideas between firms. Since academic citations follow clear norms for authorship, I can use my data to understand how ideas spread within firms as well. The bar for authorship does differ between academic fields. In economics coauthors can usually be counted on one (invisible) hand, while it is not unusual for papers published in Nature to have more than 100 authors (Greene, 2007).

A second advantage of using academic citation data is seeing the dogs that did not bark. My panel includes academics who did not cite a paper, often departments full of non-citers. In this respect academic citation data is even richer than most epidemiology data. The epidemiologists from whom I derive my model work with data on only households with at least one influenza infection. If I were using patent citation data, I would not have information on the pool of people who might have cited a patent. Information on non-citers will make my estimates more precise.

H Discussion of Long-Run Behavior of Movement Model

In this section I simulate the long-run distribution of academics over departments in the baseline structural model, and compare to the empirical distribution in the data. Using the 1994 distribution of academics over departments as the initial distribution, I simulated the model for 1000 years. The first 500 years were thrown out as a burn-in.

The top row of Figure 11 contains a histogram of the data department sizes and a histogram of simulated mean department sizes. The histograms are qualitatively similar, although there are

[31] 352 F. Supp. 1357; 1972 U.S. Dist. LEXIS 10602; 176 U.S.P.Q. (BNA) 361. LexisNexis Academic. Web. Date Accessed: 2013/07/30.

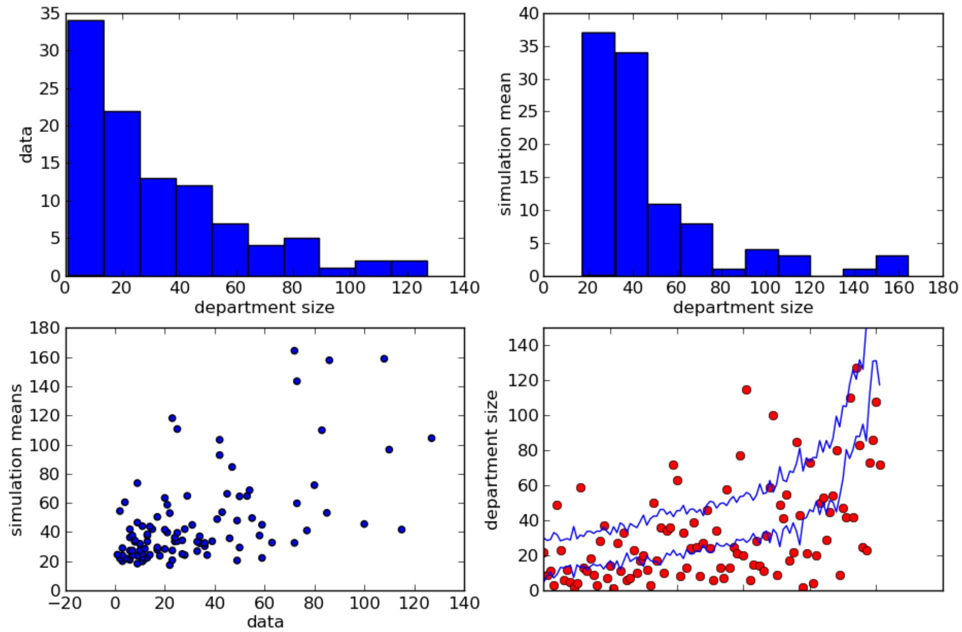

Figure 11: TL: Histogram of data, TR: Histogram of simulation means, BL: Scatter plot data vs. simulation means, BR: Two standard deviations of simulation variation against data.

no very small departments in the simulation. The scatter plot in the lower-left panel compares department sizes in the data to simulated mean department sizes. While there is positive relationship between the data and the simulation means, the correlation is far from perfect. The Pearson correlation coefficient is 0.39.

Even in the long-run, department sizes fluctuate as opportunities to move stochastically arise. The bottom right panel of Figure 11 plots bounds of two standard deviations around the simulated mean department size (the two lines), as well as the 1994 data (the dots). Departments are ordered on the x-axis according to size of simulated mean department size. The simulation cannot account for the smallest department sizes observed in the data. It is not unexpected that the long-run department sizes implied for the model differ somewhat from the department size distribution observed in 1994. The model developed above has nothing to say about entry or exit, and during this period the economics profession is growing rapidly.

I Robustness Check: Estimating with an Alternative Paper

In this appendix I present a table comparing the baseline Jensen estimates to a re-estimation using Grossman and Hart (1986). This is a slightly older version of the model in the baseline section. As can be seen in Table 13, the estimates are almost identical.

	Baseline		Grossman Hart	
	mean	std	mean	std
α	-0.651	0.163	-0.456	0.176
β	19.272	6.402	17.394	6.069
γ_F	0.089	0.028	0.089	0.004
γ_{NF}	0.034	0.004	0.034	0.029
ξ_f	0.973	0.910	1.100	0.881
ξ_l	0.666	0.182	0.675	0.174
ξ_q	0.404	0.031	0.403	0.032
λ_o	0.045	0.002	0.045	0.002
ϕ_F	-21.586	0.282	-21.575	0.281
σ	0.820	0.009	0.820	0.008
ξ_{ex}	0.266	1.276	0.225	0.827

Table 12: Posterior moment comparison

J Posterior Kernal Densities for Alternative Models

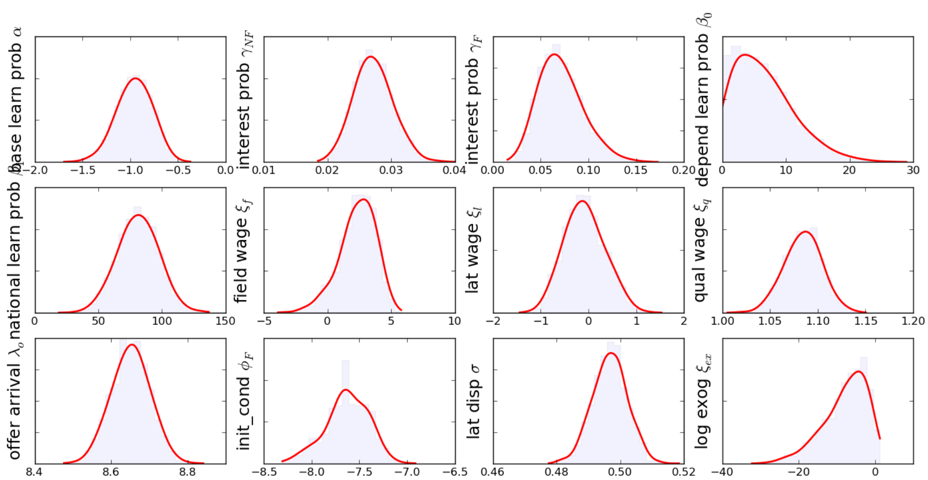

Figure 12: Posteriors for national model

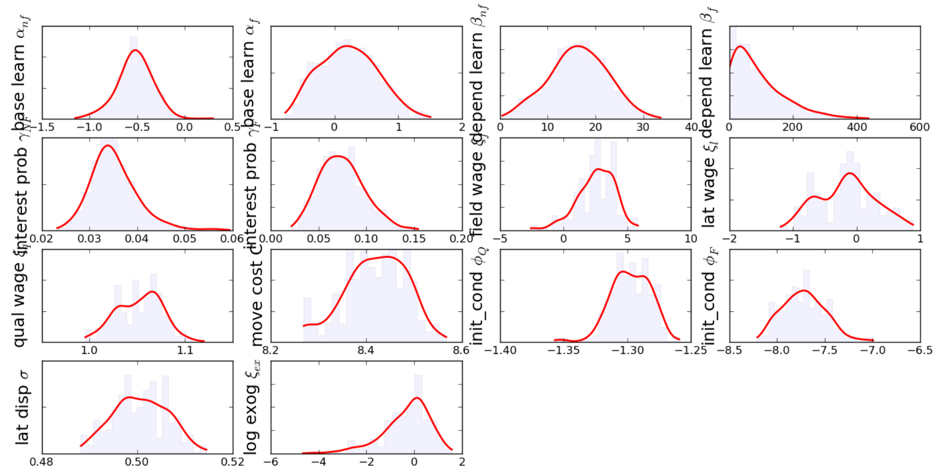

Figure 13: Posteriors for field-specific model

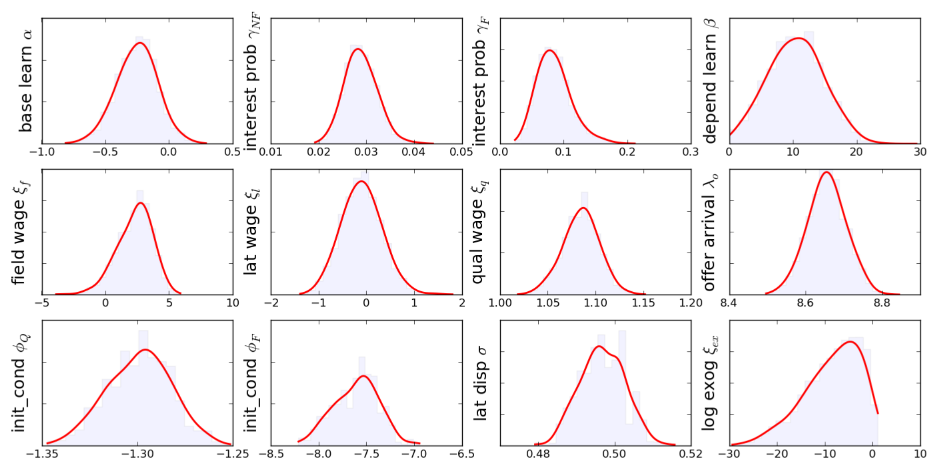

Figure 14: Posteriors for publication lag model

K Comparison of baseline priors on β

	Exponential (300)		Diffuse	
	mean	std	mean	std
α	-0.447	0.179	-0.628	0.178
β	14.128	5.850	15.992	6.466
γ_F	0.035	0.004	0.034	0.004
γ_{NF}	0.094	0.031	0.091	0.030
ξ_f	2.208	1.374	2.246	1.557
ξ_l	-0.293	0.411	-0.061	0.428
ξ_q	1.050	0.021	1.084	0.019
$\mathcal{C}$	8.426	0.052	8.655	0.052
ϕ_Q	-1.302	0.014	-1.300	0.015
ϕ_F	-7.603	0.230	-7.588	0.227
σ	0.499	0.005	0.496	0.005
ξ_{ex}	0.919	0.739	0.021	0.126

Table 13: Posterior moment comparison, alternate priors

References

Ajay Agrawal, John McHale, and Alexander Oettl. Why stars matter. *mimeo*, 2013.

Paul Almeida and Bruce Kogut. Localization of knowledge and the mobility of engineers in regional networks. *Management science*, 45(7):905–917, 1999.

Peter Arcidiacono and Robert A Miller. Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica*, 79(6):1823–1867, 2011.

Peter Arcidiacono, Gigi Foster, Natalie Goodpaster, and Josh Kinsler. Estimating spillovers using panel data, with an application to the classroom. *Quantitative Economics*, 3(3):421–470, 2012.

Christopher B Barrett, Aliakbar Olia, and Dee Von Bailey. Subdiscipline-specific journal rankings: whither applied economics? *Applied Economics*, 32(2):239–252, 2000.

Julian R Betts and Andrew Zau. Peer groups and academic achievement: Panel evidence from administrative data. *mimeo*, 2004.

Stefano Breschi and Francesco Lissoni. Knowledge spillovers and local innovation systems: a critical survey. *Industrial and corporate change*, 10(4):975–1005, 2001.

Mary A Burke and Tim R Sass. Classroom peer effects and student achievement. *Journal of Labor Economics*, 31(1):51–82, 2013.

Ricardo J Caballero and Adam B Jaffe. How high are the giants' shoulders: An empirical assessment of knowledge spillovers and creative destruction in a model of economic growth. In *NBER Macroeconomics Annual 1993, Volume 8*, pages 15–86. MIT press, 1993.

S Cauchemez, F Carrat, C Viboud, AJ Valleron, and PY Boelle. A bayesian mcmc approach to study transmission of influenza: application to household longitudinal data. *Statistics in medicine*, 23(22):3469–3487, 2004.

Davin Chor and Edwin L.C. Lai. Cumulative innovation, growth, and welfare improving patent policy. *mimeo*, 2013.

Michael A Clemens. Economics and emigration: Trillion-dollar bills on the sidewalk? *The Journal of Economic Perspectives*, 25(3):83–106, 2011.

Timothy G Conley and Christopher R Udry. Learning about a new technology: Pineapple in ghana. *The American Economic Review*, pages 35–69, 2010.

Gordon B Dahl. Mobility and the return to education: Testing a roy model with multiple markets. *Econometrica*, 70(6):2367–2420, 2002.

Giacomo De Giorgi, Michele Pellizzari, and Silvia Redaelli. Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics*, pages 241–275, 2010.

Manon Domingues Dos Santos and Fabien Postel-Vinay. Migration as a source of growth: the perspective of a developing country. *Journal of Population Economics*, 16(1):161–175, 2003.

Christian Dustmann. Return migration, wage differentials, and the optimal migration duration. *European Economic Review*, 47(2):353–369, 2003.

Jonathan Eaton and Samuel Kortum. International technology diffusion: Theory and measurement. *International Economic Review*, 40(3):537–570, 1999.

Andrew Gelman and Donald B Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, pages 457–472, 1992.

Steven D Gold. *The fiscal crisis of the states: Lessons for the future*, volume 10. Georgetown University Press, 1995.

Mott Greene. The demise of the lone author. *Nature*, 450(7173):1165–1165, 2007.

Zvi Griliches. Hybrid corn: An exploration in the economics of technological change. *Econometrica, Journal of the Econometric Society*, pages 501–522, 1957.

Zvi Griliches. Patent statistics as economic indicators: a survey. In *R&D and productivity: the econometric evidence*, pages 287–343. University of Chicago Press, 1998.

Sanford J Grossman and Oliver D Hart. The costs and benefits of ownership: A theory of vertical and lateral integration. *The Journal of Political Economy*, pages 691–719, 1986.

Heikki Haario, Eero Saksman, and Johanna Tamminen. An adaptive metropolis algorithm. *Bernoulli*, pages 223–242, 2001.

Douglas N Harris and Tim R Sass. Teacher training, teacher quality and student achievement. *Journal of public economics*, 95(7):798–812, 2011.

James J Heckman. Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, pages 153–161, 1979.

Adam B Jaffe and Manuel Trajtenberg. *Patents, citations, and innovations: A window on the knowledge economy.* MIT press, 2002.

Adam B Jaffe, Manuel Trajtenberg, and Rebecca Henderson. Geographic localization of knowledge spillovers as evidenced by patent citations. *the Quarterly journal of Economics*, 108(3):577–598, 1993.

Changhui Kang. Classroom peer effects and academic achievement: Quasi-randomization evidence from south korea. *Journal of Urban Economics*, 61(3):458–495, 2007.

John Kennan and James R Walker. The effect of expected income on individual migration decisions. *Econometrica*, 79(1):211–251, 2011.

Jinyoung Kim, Sangjoon John Lee, and Gerald Marschke. International knowledge flows: evidence from an inventor-firm matched data set. In *Science and Engineering Careers in the United States: An Analysis of Markets and Employment*, pages 321–348. University of Chicago Press, 2009.

S. Kortum. Research, patenting, and technological change. *Econometrica*, 65(6):1389–1419, 1997.

LA Times. 173 at UCI accept 'golden handshake' : Budget: The campus hopes to make up half of a projected usd 8-million shortfall through early retirements. April 12 1991.

Li Lixu. China's higher education reform 1998–2003: A summary. *Asia Pacific Education Review*, 5(1):14–22, 2004.

R.E. Lucas. Ideas and growth. *Economica*, 76(301):1–19, 2008.

Alexandre Mas and Enrico Moretti. Peers at work. *The American Economic Review*, 99(1):112–145, 2009.

Money Magazine. Paying the bills the budget crunch hits the campuses: with colleges hurting for money and cutting back on faculty and services, you must be extra careful when shopping for a school. September 10 1991.

Stephen Moore. State spending splurge: The real story behind the fiscal crisis in state government. 1991.

Alexander Oettl and Ajay Agrawal. International labor mobility and knowledge flow externalities. *Journal of International Business Studies*, 39(8):1242–1260, 2008.

Jennifer P Poole. Knowledge transfers from multinational to domestic firms: evidence from worker mobility. *Review of Economics and Statistics*, 95(2):393–406, 2013.

Everett M Rogers. *Diffusion of innovations.* Simon and Schuster, 1962.

Jesse Rothstein. Teacher quality in educational production: Tracking, decay, and student achievement. *The Quarterly Journal of Economics*, 125(1):175–214, 2010.

John Rust. Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica: Journal of the Econometric Society*, pages 999–1033, 1987.

Douglas Staiger and James H Stock. Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557–586, 1997.

Nancy L Stokey. *Recursive methods in economic dynamics.* Harvard University Press, 1989.

Taiwan Panorama. Don't taiwanese want to study abroad anymore? September 2002.

Peter Thompson and Melanie Fox-Kean. Patent citations and the geography of knowledge spillovers: A reassessment. *American Economic Review*, pages 450–460, 2005.

Washington Post. Public colleges in most states cut budgets this year; recession brings end to decade of growth, survey finds; further reductions expected. April 7 1991.

Halbert White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, pages 817–838, 1980.

Edmund Taylor Whittaker and George Neville Watson. *A course of modern analysis*. Cambridge university press, 1996.