The Pennsylvania State University
The Graduate School
College of the Liberal Arts

Essays in Status, Trade, and Knowledge Diffusion

A Dissertation in
Economics
by
David C. Jinkins

Submitted in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

August 2014

The dissertation of David C. Jinkins was reviewed and approved[*] by the following:

James R. Tybout
Professor of Economics
The Pennsylvania State University
Dissertation Adviser and Chair of Committee

Jonathan Eaton
Professor of Economics
Brown University

Edward Green
Professor of Economics
The Pennsylvania State University

Gordon De Jong
Distinguished Emeritus Professor of Sociology and Demography
The Pennsylvania State University

Robert C. Marshall
Distinguished Professor of Economics
The Pennsylvania State University
Head of the Department of Economics

[*]Signatures are on file in the Graduate School.

Dissertation adviser: Professor James R. Tybout David C. Jinkins

Essays in Status, Trade, and Knowledge Diffusion

Abstract

Each of the three chapters in this dissertation is based on an empirical research paper. While the topic of each chapter is different, they are linked by methodology. Each chapter develops a structural model of economic interaction, applies econometric techniques to estimate model parameters from data, and then uses the estimated model for policy analysis.

In the first chapter, I modify a recent theoretical model of conspicuous consumption to empirically measure the importance of peer beliefs to Americans and Chinese. In the model, a consumer cares not only about the direct utility she receives from consumption, but also about the way her consumption pattern affects her peer group's belief about her well-being. I estimate the model on household budget surveys using an EM algorithm. According to model estimates, a Chinese consumer cares 20% more than an American consumer about peer beliefs. I use the estimated model to evaluate the welfare effect of the 1990-2002 American luxury tax on automobiles. The luxury tax benefited nearly all Americans a small amount, but hurt the small fraction of consumers who love automobiles the most.

The second chapter, a joint work with my adviser and others, seeks to understand the way Colombian and American firms interact in U.S. Customs data. After documenting patterns in the data, we develop an estimable empirical model in which heterogeneous sellers engage in costly search for buyers. Through meeting buyers, a firm gradually learns about the appeal of its product in the market, which affects its incentive to search for more buyers. Fit using indirect inference, the model both replicates key patterns in the customs data and allows us to quantify several types of trade costs, including the search costs of identifying potential clients and the costs of maintaining business relationships with existing clients. We also estimate the effect of previous exporting activity on the costs of meeting new clients, and to characterize the cumulative effects of learning on a firm's search intensity. Finally, we use our fitted model to explore the effects of these trade costs and learning effects on aggregate export dynamics.

The third chapter measures the effect of mobility between workplaces on the speed at which new ideas diffuse. Using a new panel data set linking academics to departments and citations, I develop and estimate a dynamic model of location choice in which an idea is more likely to be encountered when colleagues already know about it. Several exercises indicate that coworker knowledge significantly affects the probability of learning about a new idea. Counterfactual exercises show that labor mobility increases the speed at which new ideas spread between locations, makes locations more uniform in the fraction of people who know about a new idea, and raises the percentage of people who know about a new idea at a given time. A calibration using results from my baseline estimation indicates that international movement of workers can have a large effect on diffusion of knowledge into a developing country.

Contents

Listing of figures

List of Tables

Acknowledgments

No graduate student is an island. I am deeply indebted to my adviser Jim Tybout for patience with my terrible whiteboard handwriting in countless meetings, for both giving me excellent advice and the freedom to make my own decisions, and for support of many kinds. Straight is the gate and narrow is the way through graduate school, and I was lucky to have Jim as a guide.

I am grateful to Jonathan Eaton for his thorough comments on everything from barely coherent research notes to my polished job-market paper, for his guidance in choosing a research topic, and for his advice and training over the last five years.

My research has greatly benefited from discussions with Kala Krishna, Ed Green, Paul Grieco, Steve Yeaple, Neil Wallace, Peter Newberry, Venky Venkateswaran, Mark Roberts, Russell Cooper, Michal Fabinger, Gordon De Jong, John Riew, and other Penn State faculty.

My understanding of economic research was tempered in the oven that is Kern 6th floor and tested over pitchers of beer at Cafe 210 West with my friends and colleagues Yu Awaya, Xue Bai, Chun-Ting Chen, Mohammad Davoodalhoseini, Bernardo Diaz, Pelin Dogruer, Alex Fakos, Farid Farrokhi, Bulat Gafarov, Gustavo Gudino, Yoske Igarashi, Konstantin Kucheryavyy, Ahmad Lashkaripour, Joosung Lee, Shengyu Li, Jicheng Liu, Gary Lyn, Wataru Nozawa, Bruno Salcedo, Yelena Sheveleva, Wiroy Shin, Bruno Sultanum, Felix Tintelnot, Marlon Williams, Hoon Sik Yang, and many others.

My parents Mark and Ann Jinkins were generous enough to continue calling me even after repeatedly suffering through fumbling descriptions of my *next* great idea. I hope they know how seriously I take their advice, and how they have shaped my life.

I am indebted to the National Science Foundation for supporting research which contributed to the first and second chapters of my dissertation. Funding for research contributing to the first chapter was provided through the NSF East Asia and Pacific Summer Institute Program.

My companions Petek and Aydin have put up with a moody husband and father, many lonely dinners, and a pile of shoes which made our foyer look like we were running a mosque. Graduate school was an adventure, and I am grateful that they came along.

To Petek and Aydin
You have made all the difference

When a country is run well, it is shameful to be poor.
When a country is run poorly, it is shameful to be wealthy.

Analects of Confucious, Chapter I

1

Conspicuous Consumption in the United States and China

I WEAR a Seiko automatic watch. Over the course of a month, it picks up about five minutes. I knew it would do this before I bought it from reading online reviews, but even so I purchased it for about $100 a few years ago. At the time, I could have picked up a much less expensive digital Casio from Wal-Mart which would have run more reliably, been easier to read, and been more water resistant. On just about any measure of watch performance the Casio would have outrun the Seiko, and yet there is the relatively expensive Seiko on my wrist. Why did I buy that watch? Why did you buy yours?

When buying a car or a suit, a consumer considers how her social group will view the new purchase. In addition to evidence from personal introspection, I will cite some papers below which find reduced form evidence of such consumer behavior. This paper adds to the empirical literature on conspicuous consumption by developing and estimating a partial-equilibrium heterogeneous-agent structural model in which a consumer's peers infer his wealth after observing a subset of his purchases. Inference about welfare by his peer group causes a consumer to distort his consumption toward the purchase of visible goods.

The structural model I develop closely follows recent reduced-form empirical work on conspicuous consumption.[51] In order to estimate the model, I use a survey on the relative visibility of different categories of goods and household-level consumption expenditure data. As it is used to calculate purchasing power, expenditure data is available for many countries and time periods. This is primarily a *measurement* paper.

I estimate the model separately using American and Chinese consumption expenditure data. The estimated model fits the data very well. I find that the Chinese consumers care 20% more than American consumers about peer group beliefs. Using the estimated model, I find that the 1990-2002 American luxury tax on automobiles had a small but positive welfare effect on all but around 2 in 10,000 American households. The households hurt by the tax were the extreme Jeremy Clarkson fans, gearheads that get a large amount of pleasure from automobile purchases.

In this paper as well as the literature I am following, people consider peer group belief an end in itself. I put peer group belief about welfare directly into the utility function. Some might argue that people only care about peer group beliefs as the means to an ultimate consumptive end–wearing a nice watch makes people trust you more, so you are more likely to get a loan or secure a business deal. I am sympathetic to this point of view, and this sort of signaling is doubtless going on to some degree. The two points of view about peer beliefs are complementary. From a long perspective, our brains might have been selected to care about peer group beliefs precisely because good standing makes successful reproduction more likely. In this case, the utils we get from positive peer group beliefs are an evolutionary rule of thumb.[78]

There are several strands of empirical literature that support the presence of a social component in the utility function. Consider the ultimatum game in which one player proposes a split of a sum of money, and the other player decides whether to accept or reject. If the second player accepts, the money is allocated according to the split. If the second player rejects, neither player gets anything. There is a long and robust experimental literature showing that if people only care about immediate monetary payoffs, the splits they propose are too fair. Researchers have been careful to pair subjects who do not know each other and are unlikely to have interaction after the experiment, and the result still holds. One explanation is that there is some sort of social component in the utility function.[40,17] A second defense comes from the literature on self-reported happiness and relative wealth. Luttmer[65] finds that relative wealth compared with neighbors has a robust positive correlation with self-reported happiness, controlling for absolute wealth level. On the face of it, it seems hard to explain this fact without some sort of social component in the utility function. If you are not convinced that there is a fundamental social belief component in the utility function, then you can think of this paper as estimating a reduced form of a more complicated dynamic game.

This is not the first paper to take an empirical look at conspicuous consumption.[16,24,69,70] My paper borrows both data and functional forms from Heffetz[51], who conducts a telephone survey in the United States to determine the visibility of consumption goods. Heffetz analyzes household budget survey data, and finds evidence that the relatively visible goods identified by the survey are being used as a means to signal wealth. To my knowledge, the only other structural estimation of a utility function including conspicuous consumption is Perez-Truglia[75]. Perez-Truglia follows earlier literature in using a two-good functional form, and a variety of specifications for how non-market goods like status enter utility. My specification below differs from Perez-Truglia's in a few important ways. Some cosmetic differences include that I allow for individual level preference heterogeneity and estimate a many good utility function. Any good can be used for signaling in my model, while in Perez-Truglia's model cars and clothes are the visible goods. More substantively, while Perez-Truglia is focused on the provision of unobservable non-market goods (status), I assume that society cares only about an individual's unobservable welfare. This allows me to consider peer-group beliefs as an equilibrium outcome, rather than assume a functional form for the provision of a non-market good. In his framework, Perez-Truglia finds that a tax on luxury goods can create welfare gains significantly larger than those I find in my estimation.

Using a structural estimation, I can examine both the absolute and relative magnitude of the motive for conspicuous consumption, and I can measure the welfare gains from an sales tax on visible good categories. A well-designed excise tax can raise nearly everyone's welfare.

If wealth were directly observable by the peer group, there would be no reason to distort consumption towards visible goods and welfare would be higher than in the incomplete information world. One way to get people closer to the complete information allocation is to raise the price of the visible good, and then redistribute the proceeds of the tax. Loosely speaking, the rich are better off because they distort consumption less, and the poor are better off because they are getting a subsidy from the rich. If people care deeply about peer group belief, then the welfare gains from this sort of tax can be large.*

There is a relatively large and old related literature estimating what are known as interdependent preferences. Beginning with James Duesenberry's 1949 doctoral thesis,[36] researchers have theorized that the consumption of neighbors affects own demand. A typical econometric model in this literature lets household demand parameters depend linearly on the average of the consumption of a reference group. A relationship between neighbor consumption and own consumption is taken to mean that preferences are interdependent. The literature, however, does not take a stand on why consumption neighborhood consumption should be linked in this particular way.

1.1 An Empirical Model of Conspicuous Consumption

There is a finite set of goods G. Each good has an exogenous price p_g. There is a continuum of consumers I. For each consumer, nature draws a wealth w_i, a preference type $\boldsymbol{\gamma}_i$, and an observation type $t_i \in G$. A consumer decides how to allocate his wealth to goods in order to maximize his utility. Following earlier theoretical literature,[51,53] I assume a consumer's utility

*Signaling distortions are particularly worrying when considering the economic lives of the poor. A recent study reports that in parts of India, the *median* household making under a dollar a day spends 10% of its income on festivals–this while 43% of such households did not have enough to eat throughout the year.[11]

function consists of two additively separable parts.

$$U(\mathbf{C}_i, \boldsymbol{\gamma}_i, t_i) = (1-\alpha)u(C_i, \boldsymbol{\gamma}_i) + \alpha\, u(C_b(c_{t_i}, \boldsymbol{\gamma}_i, t_i), \boldsymbol{\gamma}_i) \tag{1.1}$$

The first term on the right-hand side of (1.1) is a fundamental utility $u : \mathbb{R}^I_+ \to \mathbb{R}$. Fundamental utility describes the pleasure a consumer gets directly from consuming a bundle of goods. The second term is the belief of a consumer's peer group over his utility. Peer group belief over the utility level of consumer i is based on his expenditure on good category t_i. C_b maps consumption of the observable good, observation type, and preference type to the unobservable full consumption vector. The preference type and observation type of consumer i are known to his peer group.[†]

1.1.1 Equilibrium Concept

An equilibrium is a social belief function C_b and a vector-valued consumption function C on (W, Γ, G) such that:

1. For each consumer type $(w_i, \gamma_i, t_i),\ C(w_i, \gamma_i, t_i)$ solves the consumer's problem.

2. For each consumer types $(w_i, \gamma_i, t_i),\ C(w_i, \gamma_i, t_i) = C_b(c_{t_i}(w_i, \gamma_i, t_i), \gamma_i, t_i)$.

The first condition says that a consumer chooses an optimum consumption bundle, and the second condition says that Consumer i's peer group learns his true type.

[†]The peer-group infers the one-dimensional wealth of a consumer from the one-dimensional observed consumption choice of the observable good. If I allow for more than one observed good, then one-dimensional would be inferred from multi-dimensional consumption. As in a typical multi-dimensional screening model, the equilibrium will be driven by beliefs off the equilibrium path and there will be many possible equilibria.

1.1.2 Specializing to Cobb-Douglas

Let the fundamental utility function be Cobb-Douglas:

$$u(\mathbf{C}, \boldsymbol{\gamma}) = \sum_{g=1}^{G} \gamma_g \ln(c_g)$$

The model can then be written as a generalization of the Heffetz model to many goods and preference heterogeneity.[‡] In what follows I drop subscripts for Consumer i to simplify notation. Let $t \in G$ be Consumer i's observation type, and let c_t^* be Consumer i's equilibrium consumption of the visible good. Equilibrium demand for good $g \neq t$ conditional on spending on the visible good is the standard Cobb-Douglas constant expenditure share:

$$p_g c_g^* = \gamma_g \left(\sum_{j \neq t} \gamma_j \right)^{-1} (w - p_t c_t^*) \tag{1.2}$$

Using the demands, we can write the utility function as a function of visible good consumption.

$$U(c_t) = (1-\alpha)\left(\hat{\gamma} \ln\left(w - p_t c_t\right) + \gamma_t \ln\left(c_t\right)\right) + \alpha\left(\gamma_t \ln\left(s(c_t)\right) + \gamma_t \ln\left(c_t\right)\right) + \zeta(\mathbf{p}, \boldsymbol{\gamma}) \tag{1.3}$$

Here $\hat{\gamma} = \sum_{g \neq t} \gamma_g$ and $\zeta(\mathbf{p}, \boldsymbol{\gamma})$ is a constant which depends only on utility parameters and prices. The single-valued function $s(c_t)$ is the belief of the peer group about spending on non-visible goods $w - p_t c_t$.

Consumer i maximizes utility function (1.3) subject to his budget constraint. The first

[‡]In the Heffetz version, there are only two goods, one visible and the other invisible to society. In my version, there is one visible good for each observation type, and all the other goods are invisible.

order condition for an interior solution to his problem can be written:

$$s'(c_t^*) = \frac{1}{\alpha}\left((1-\alpha)\,p_t - \frac{\gamma_t}{\hat{\gamma}}\frac{s(c_t^*)}{c_t^*}\right) \tag{1.4}$$

This differential equation has the solution:

$$s(c_t^*) = \frac{\hat{\gamma}\,(1-\alpha)}{\gamma_t + \alpha\hat{\gamma}} p_t c_t^* + \frac{\hat{\gamma}\alpha}{\gamma_t + \alpha\hat{\gamma}} \underline{\mathrm{W}} \frac{p_t c_t^*}{p_t \underline{\mathrm{c}}}^{-\frac{\gamma_t}{\alpha\hat{\gamma}}} \tag{1.5}$$

The constant in the solution (1.5) is pinned down because the lowest possible wealth type $\underline{\mathrm{W}} > 0$ has no reason to signal in a separating equilibrium. His expenditure on the visible good $\underline{\mathrm{c}}$ is the fraction $\gamma_t/\sum_j \gamma_j$ of his wealth. As one might expect, the function s is jointly homothetic in c_t and $\underline{\mathrm{W}}$.

Define equilibrium expenditure share on the visible good category $r = p_t c_t^*/w$, the ratio $\gamma = \gamma_t/\hat{\gamma}$, and the ratio of minimum wealth to own wealth $\hat{w} = \underline{\mathrm{W}}/w$. Substituting in for the s function and dividing by wealth, we have a simplified equilibrium condition:

$$(1-r)(1+\frac{\gamma}{\alpha}) = \frac{(1-\alpha)}{\alpha} r + \left(r\left(1+\gamma^{-1}\right)\right)^{-\frac{\gamma}{\alpha}} \hat{w}^{1+\frac{\gamma}{\alpha}} \tag{1.6}$$

1.2 Description of Data and Sources

This project requires two types of data. We need household-level consumer expenditure data, and we need information about how visible different good categories are relative to each other. Household expenditure data is widely available from national statistical agencies. Information on the visibility of different good categories is taken from a survey conducted in Heffetz[51].

1.2.1 Household Expenditures

American household expenditure data is taken from the National Bureau of Economic Research.[73] This data set is publicly available, and features a large random sample of American household consumption decisions for selected years between 1981 and 2002. In addition to detailed information on household income and expenditures, the NBER data set contains demographic data on household members such as age, race, sex, and location. There are 47 good categories available in the NBER data set. Following Heffetz[51] exactly,[§] I aggregate into 29 expenditure categories. The NBER data set contains 160,617 household observations across 18 years.

Households display widely varying consumption behavior. Figure 1.1 is a scatter plot the 2001 log budget shares by log expenditures. Representative household models in the literature such as those by Heffetz and Ireland cannot replicate this heterogeneity.[¶] The heterogeneous preference model estimated in this paper can potentially match the noise observed in the data.

For the Chinese household expenditures, I use publicly available data from the Chinese Household Income Project (CHIP).[62] Like the American household expenditure data, the CHIP data is comprised of repeated cross-sections of Chinese households. In this study I use urban households surveyed in 1995 and 2002 for a total of 13,767 observations. I use 14 good categories which correspond to aggregates of those in the American household expenditure survey. Table 1.1 details the link between the American and Chinese expenditure data.

[§] Heffetz was kind enough to give me his STATA code.

[¶] Heffetz[51] contains a discussion of this issue.

[‖] Air, Gas, Cmn, Cin

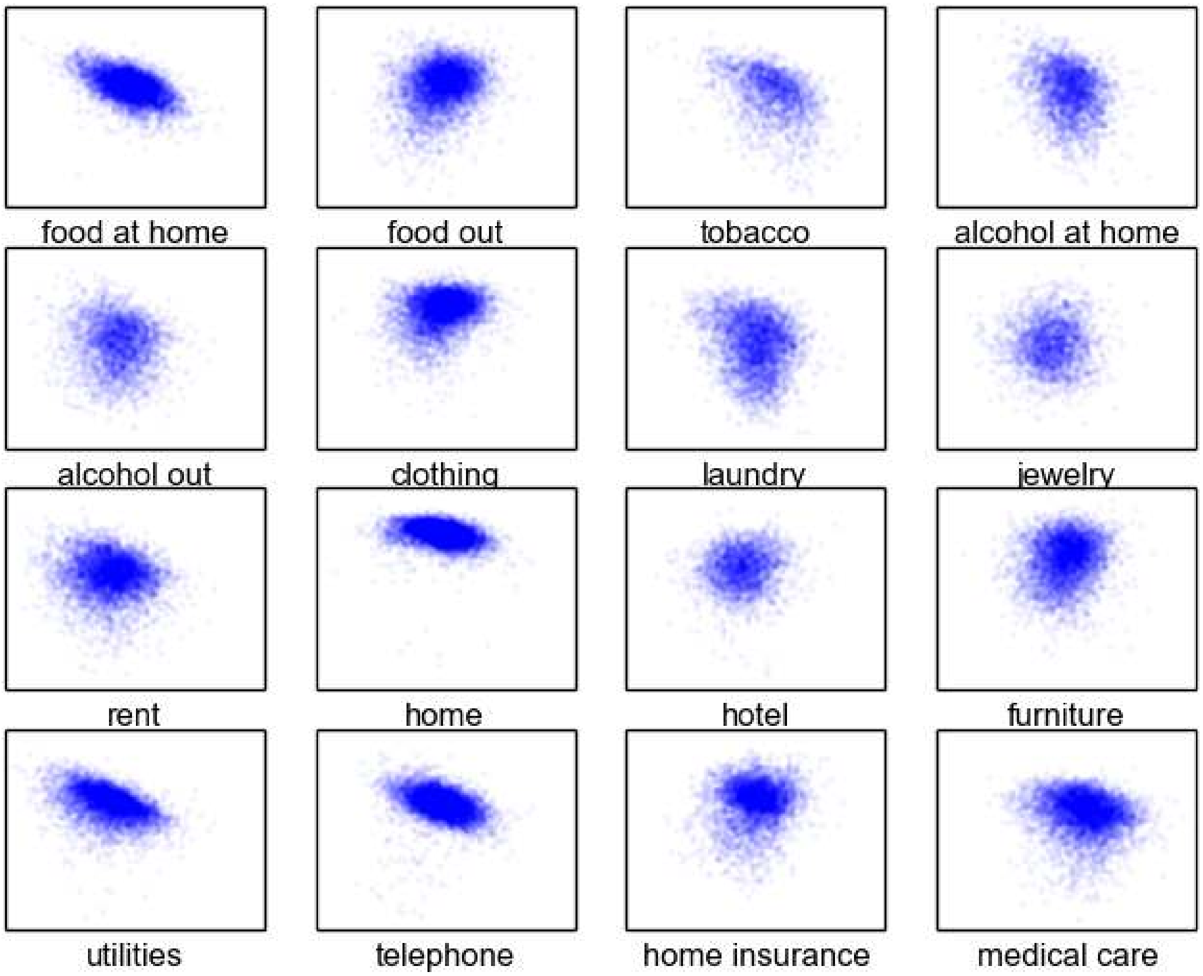


Figure 1.1: Log expenditure shares (y) by log expenditure (x)

US Cat	1995 Chn Cat	2002 Chn Cat	Chn Cat Name
Fdh,Fdo	h27	e1-e152-e153	Food-Cig.-Alcohol
Alh,Alo	h30-h31	e153	Alcohol
Cig	h31	e152	Cigarettes
Bks	h37	f631	Textbooks
Edu	h38 to h42	f63-f631	Education-Textbooks
Bus,Car[‖]	h44	f514	Transportation
Utl	h45 to h46	f72	Water,Elec.,Fuel
Tel	h47	f522	Communication
Clo,Jwl	h32	f2	Clothes
Ot1,Ot2	h33	f6-f63	Entertain.
Fur,Lry,Brb	h34,h36	f3	HomeEquip.,Facil.,
Med,Lin	h48	f4	Health
Hom,Htl	h64	f71	Housing
Fee,Cha	h35	f8	Misc.Goods

Table 1.1: Chinese Consumption Categories

1.2.2 Visibility Indexes

Data about the relative visibility of goods is taken from Heffetz[51]. Heffetz bases the index on randomized telephone surveys conducted in the United States in several waves around 2004. Heffetz asked respondents how long it would take them to notice if a new acquaintance similar to themselves spent more than average on a particular good category. Respondents chose from five time periods ranging from almost immediately to almost never. Basic demographics were also recorded for respondents.

From the survey responses, Heffetz creates indexes (called vindexes) between zero and one for each category of goods by averaging over survey results. A higher index implies that a good category is more conspicuous. A result of this aggregation methodology is that the index is cardinal rather than ordinal. Two goods with similar index values are similar in visibility. Details on the implementation of the survey and calculation of the index are available in the original paper. Table A.1 in the appendix presents vindex survey data.

I do not have a vindex equivalent for China, so I use the aggregated American vindex data for the Chinese estimation. Since there are fewer good categories in the Chinese data, I collapse the American vindex by taking the mean over aggregated good categories.

1.3 Discussion of Model Identification

We are interested in α, the weight given to the peer-belief part of the utility function. The key identification issue is that, for a fixed α, *any* consumption bundle can be rationalized by a particular set of utility function parameters $\boldsymbol{\gamma}_i$. In order to separate preferences and conspicuous consumption, we need to take a stand on how utility parameters might be distributed. One natural assumption is that most people's preferences are broadly similar, while a few people have atypical preferences. To operationalize this idea, I assume that preferences for each household and each good category are independently drawn from lognormal distributions. In addition, to rationalize zero expenditure in a good categories I assume that with some probability a consumer doesn't derive any pleasure from consumption of a particular category ($\gamma_{ig} = 0$).

A second challenge is that the Cobb-Douglass base utility assumption implies that there are no luxury or inferior goods. Absent any conspicuous consumption, expenditure shares are constant as household wealth increases. Figure 1.1 shows that expenditure shares are changing on average as household wealth increases. The combination of Cobb-Douglass utility and changing expenditure shares in principle identifies my model.

The Cobb-Douglass assumption is too strong, however. I want to allow a good like "food at home" to be inferior even without conspicuous consumption effects. To do this, I allow the location of the distribution of utility parameters to drift as a function of wealth. In particular, the location parameter $\hat{\mu}_g(w_i)$ of the lognormal distribution for good category g is given by (1.7).

$$\hat{\mu}_g(w_i) = \psi_g \ln\left(\frac{w_i}{\underline{\mathrm{W}}}\right) + \mu_g \tag{1.7}$$

This 'money-in-the-utility-function' specification is somewhat ad hoc, but it allows us to keep the simple equilibrium condition (1.6) as well as allowing for rich evolution of expenditure shares with wealth. This distribution of utility parameters also breaks simple identification of my model from the correlation of household expenditure shares and wealth.

In order to regain identification, I use differences in observed vindexes across demographics. I assume that all utility parameters $\boldsymbol{\gamma}_i$ are drawn out of the same distribution, but observation types t_i are drawn with probability weighted by an individual's demographic specific vindex. The size of differences in average consumption between demographic groups, are then informative about the weight α of peer group beliefs in the utility function.

I have only a single Chinese demographic category, so when estimating Chinese preference parameters I cannot use an identification strategy based on differences in demographic groups. In the Chinese estimation, I take the ψ_g's in equation (1.7) as data from the American estimation. This assumption implies that luxury and inferior good categories are the same in both China and the United States. Deviations from Chinese expenditure share trends along with vindex probabilities identify α.

1.4 Estimation Procedure

In order to estimate the parameter of interest α, we must jointly estimate the observation type of each household and four preference distribution parameters for each good category.

This is a large problem, so I split the estimation into two steps by using a 'hard' expectation maximization algorithm. In the first step (maximization), I condition on the observation type of each household and update α and preference distribution parameters. In the second stage, I take α and the preference distribution parameters as given and find the most likely observation type of each household (expectation). The algorithm stops when there is no change in α.

1.4.1 Maximization: Updating α and Preference Distribution Parameters

In the maximization step, I condition the likelihood function on the observation type t_i of each household and update α and lognormal preference distribution parameters μ_g, σ_g, wealth-scaling parameter ψ_g, and a zero probability z_g. Given α, the preference parameters $\boldsymbol{\gamma}_i$ of each household can be calculated using observed consumption shares. Once we have preference parameters for each household, we can analytically calculate the most likely log-normal distribution and zero parameters. The outer structure of the maximization step is to let a numerical optimizer maximize the conditional likelihood over α, and to treat the likelihood-maximizing preference parameters as functions of α.

1.4.1.1 Recovering Household Preference Parameters Given α

Taking observation type t_i and α as given, there is a mapping from observed consumption shares directly to household preference parameters. Consider a household of observed wealth type w, observed consumption vector C, and observation type t. Rearranging (1.2), γ_g for

$g \neq t$ are given by :

$$p_g c_g = \frac{\gamma_g}{\sum_{g \neq t} \gamma_g} (w - p_t c_t) \tag{1.8}$$

$$\gamma_g = \frac{p_g c_g}{(w - p_t c_t)} \sum_{g \neq t} \gamma_g \tag{1.9}$$

$$\gamma_g = \frac{p_g c_g}{(w - p_t c_t)} \tag{1.10}$$

We can solve for the 28 non-observation type γ_g's up to a scaling factor $\sum_{g \neq t} \gamma_g = 1$. Using (1.10) and the equilibrium condition (1.4) we can then solve for γ_t. Unfortunately, (1.4) is non-linear and in principle needs to be solved numerically for each household. To decrease estimation time, in practice I solve (1.4) on a 1000 point grid of visible consumption shares and wealths, and then linearly interpolate to find household specific γ_t's.

1.4.1.2 Updating Preference Distribution Parameters

Given α, we have now recovered $\boldsymbol{\gamma}_i$ for each household. The most likely zero probability z_g^* for good category g is the fraction of zero γ_{ig}'s:

$$z_g^* = \frac{1}{\|I\|} \sum_i \mathbf{1}_{\gamma_{ig}=0}$$

Let an upper bar denote sample means over non-zero γ_i's, and let m_i refer to normalized income, $m_i = w_i / \underline{\mathrm{W}}$. The other likelihood-maximizing preference parameters are:

$$
\begin{aligned}
\psi_g^* &= \frac{\hat{\text{cov}}(\ln m, \ln \gamma)}{\hat{\text{var}}(\ln m)} \\
\mu_g^* &= \overline{\ln \gamma} - \psi_g^* \overline{\ln m} \\
\sigma_g^{2*} &= \overline{\left(\ln \gamma - t_g^* \ln m - \mu_g^*\right)^2}
\end{aligned}
\tag{1.11}
$$

1.4.1.3 Full Conditional Likelihood Function

I have shown how, given observation types, it is straight-forward to calculate preference parameters and likelihood maximizing preference distribution parameters as a function of α. Let ϕ be the log-normal probability density function. The maximization step conditional log-likelihood function is given in (1.12). All preference parameters and preference distribution parameters are implicitly functions of α.

$$
l^1(\alpha) = \sum_{ig} \left(\mathbf{1}_{\{\gamma_{ig}=0\}} \ln(z_g) + \mathbf{1}_{\{\gamma_{ig}\neq 0\}} \left(\ln(1-z_g) + \ln \phi(\gamma_{ig}, m_i | \mu_g, \sigma_g, t_g)\right)\right) \tag{1.12}
$$

Likelihood (1.12) is the objective function used by the numerical solver in the search for α. This completes the characterization of the maximization step in the algorithm.

1.4.2 Expectation: Updating Observation Type t_i

Given the utility weight of social beliefs α and a set of preference distribution parameters, we find the most likely observation type for each household. Now preference parameters γ_{ig} are a function of observation type t and are calculated exactly as in Section 1.4.1.1. $\mathbf{v}_i$ is the household-specific vector of observation type probabilities. Household i's (unnormalized) probability of being observation type $t \in G$ is given by (1.13).

$$l_i^2(t) = \ln(v_{it}) + \sum_g \left(\mathbf{1}_{\{\gamma_{ig}=0\}} \ln(z_g) + \mathbf{1}_{\{\gamma_{ig}\neq 0\}} \left(\ln(1 - z_g) + \ln \phi(\gamma_{ig}, m_i | \mu_g, \sigma_g, t_g) \right) \right) \tag{1.13}$$

For each household, I assign the observation type giving the highest probability.

In practice, in the United States I have visibility indexes for eight different types of households. One dimension of differentiation is the age of the survey respondent (over/under age 40). The other dimension of differentiation is region in the United States (Northeast, Midwest, West, and South). The visibility probabilities are taken directly from Heffetz and normalized so that they sum to one. Table A.2 in the appendix characterizes observation-type probability distributions for the demographic groups.

In the American data, $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$, $\mathbf{z}$, and $\mathbf{t}$ are each 28 dimensional vectors. Adding α gives us a total of 118 parameters to be estimated, giving me more than 1000 observations per parameter. In the Chinese version of the model, the preference heterogeneity vectors are 14 dimensions and I take the $\mathbf{t}$ vector as data from the American estimation, so there are 43 parameters to be estimated. The Chinese estimation has over 300 observations per parameter.

1.5 Results and an Application to an American Luxury Tax

Chinese care about 20% more than Americans about social beliefs. The weight of social beliefs α in American utility is 0.027 with standard error 1×10^{-4}. In Chinese utility, the weight of social beliefs is 0.033 with standard error 0.001. Standard errors are bootstrapped by repeatedly redrawing from the data and reestimating the model. All estimated parameters are presented in Appendix A.2.

The model is capable of simulating data similar to the real data set. In Figure 1.2 is a scatter plot of simulated US data, made in the same way as and superimposed on top of the scatter plots of the actual US data in figures 1.1.

The estimation also does well fitting observation types. The observation type distribution (for a particular demographic) should be the same as the vindex probability distribution. Figure 1.3 is a scatter plot of the vindex probabilities and the estimated observation type densities. Each point is labeled with the relevant good category, and the colors represent different demographic types (region and age). Although there is not a perfect correlation between vindex probabilities and observation type frequencies, there is clearly a trend in the right direction. The model misses the most on good categories “car” and “jewelry”. I suspect the problem is that these are durable goods, so that a single year of expenditure is a poor reflection of average expenditure in those categories.

1.5.1 Policy Analysis: Luxury Tax

In the model developed above, a consumer distorts his full-information utility-maximizing consumption bundle in order to signal his wealth. The signal is on expenditures, however, not on physical goods. In principle, a social planner could impose a sales tax on a highly visible good category in order to reduce physical consumption. In the real world, such a tax is known as a luxury tax. In this section I consider the welfare implications of one such tax scheme, an American luxury tax on automobiles.

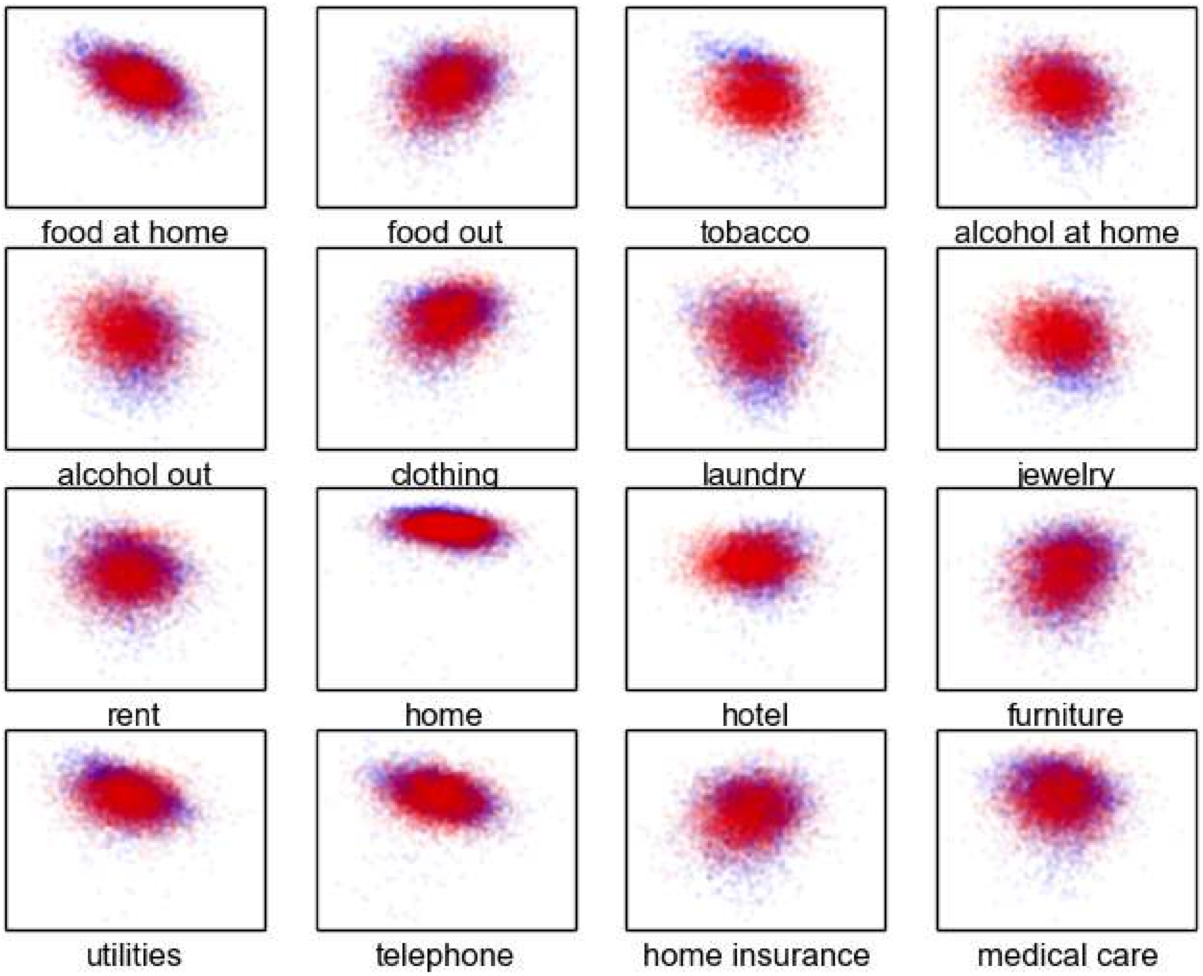


Figure 1.2: Log expenditure shares (y) by log expenditure (x), sim=red, dat=blue

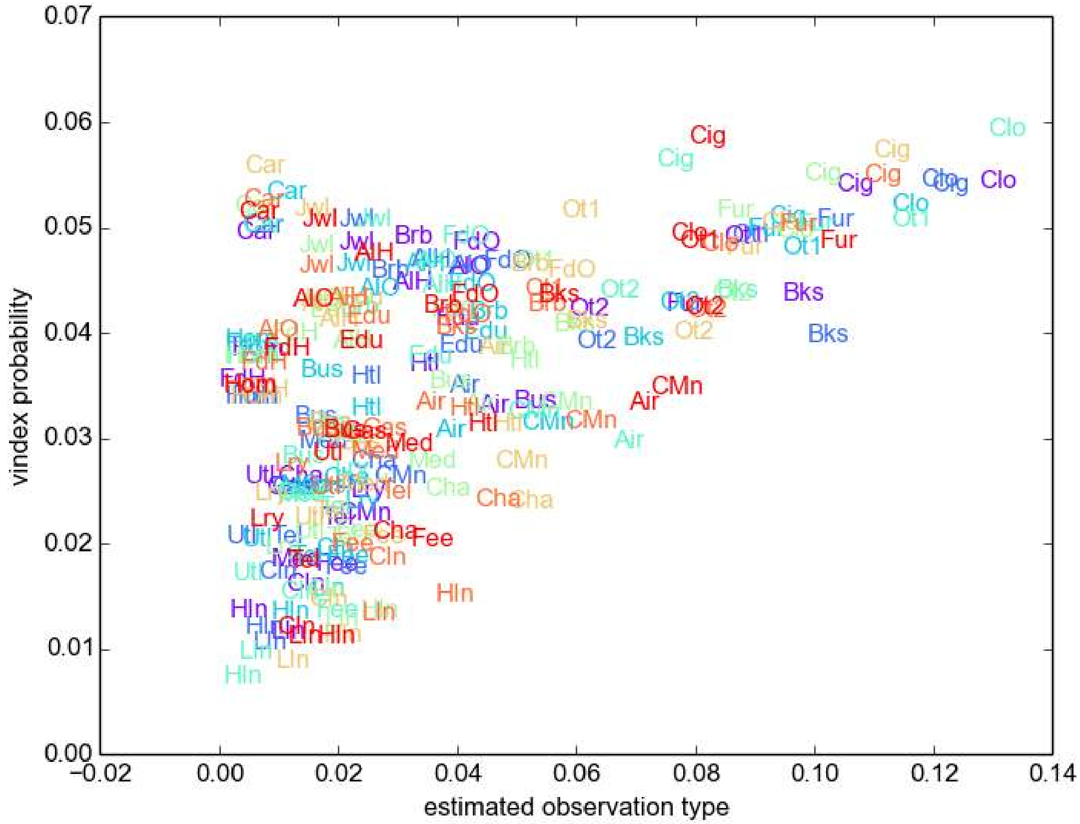


Figure 1.3: Estimated observation type frequencies vindex probabilities, by demographic

1.5.1.1 Application: Welfare Effect of US Automotive Luxury Taxes

In 1990, President George H.W. Bush signed the Omnibus Budget Reconciliation Act into law.[**] The OBRA contained a provision for a luxury tax on automobiles, as well as jewelry, furs, yachts, and personal aircraft. The tax on autos was 10% of the price exceeding $30,000. As one might imagine, the luxury tax did not go over well at campaign fundraisers and was repealed in 1993 for all goods except automobiles.[††] Congress finally scrapped the auto tax in 2002.

In this section, I measure the welfare effects of a 10% tax on automobiles, redistributed lump sum as a proportion of wealth. Redistributing the tax proportionally to wealth conveniently abstracts from the welfare effect of a transfer from the rich to the poor. In addition, taxes redistributed this way change neither the individual nor aggregate fraction of wealth optimally allocated to any particular good category, as relative wealth remains unchanged.

My luxury tax will be 10% of spending on automobiles. Let $\tau = 0.1/1.1$ be the fraction of spending on autos taken by the government, let s be the share by which the government increases wealth levels, let l_i be the equilibrium fraction of luxuries in consumer i's total expenditures, and let L be the aggregate fraction of spending on luxuries. Condition (1.14) balances the budget.

[**] Some readers might remember that this act proved television to be a poor medium for lip-reading.

[††] A cynical political realist might observe that luxury vehicles are often imported from Europe.

$$
\begin{aligned}
(1+s)\tau\sum_i w_i l_i &= s\sum_i w_i \\
s &= \frac{\tau\sum_i w_i l_i}{s\sum_i w_i - \tau\sum_i w_i l_i} \\
s &= \frac{\tau L}{1-\tau L}
\end{aligned}
\tag{1.14}
$$

Welfare change under the tax scheme is as in (1.15).

$$
\Delta u_i = \sum_{g\in G}\gamma_{ig}\ln(1+s) + \sum_{g\in l}\gamma_{ig}\ln(1-\tau) \tag{1.15}
$$

It can be shown that $\ln(1+s)+\ln(1-\tau)<0$, so it is impossible to have a truly Pareto tax scheme. That is, it is always possible that some unlucky consumer will draw all zero γ_{ig}'s in non-luxury good categories, ensuring he will be harmed by luxury taxes. We can, however, potentially design taxes which benefit all but a vanishingly small fraction of consumers.

The relationship between α and the tax scheme here is through the link from the tax level to government subsidies. Fixing preference parameters and the tax level τ, the higher α the higher government subsidies s to consumers.

Figure 1.4 displays a histogram of welfare changes resulting from a 10% luxury tax, calculated for one million American households simulated using estimated model parameters from Section 1.5. About 0.02%, or two in 10,000 households are harmed by the auto luxury tax. The vast majority of households benefit from the automobile luxury tax. In contrast, a similar 10% sales tax on food at home harms 90% of households.

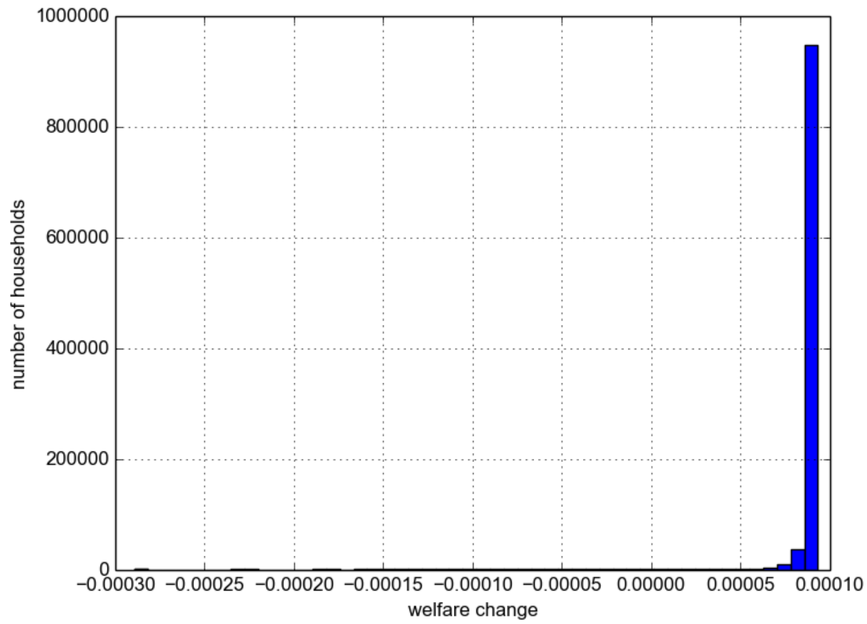


Figure 1.4: Histogram of welfare changes from a 10% luxury auto tax

1.6 Summary

This chapter develops a structural conspicuous consumption model with preference heterogeneity estimable from widely available consumption expenditure data. In an application, I show how the estimated model can be used to measure the welfare implications of a tax on luxury goods.

The results of the estimation show that:

1. Peer group belief plays a small role in overall consumption decisions. American and Chinese consumers value peer group belief under five percent as much as they value the direct utility from consumption.

2. Chinese consumers value peer group belief 20% more than American consumers.

3. Simple luxury taxes can lead to small welfare gains for nearly all households.

The strongest assumption in the model is that a household's peer group sees only consumption expenditures on one good category. While a single-dimensional signal generates the unique and simple equilibrium solution to the model, it is clearly counterfactual. In the real world, one's peer group sees a full, noisy vector of consumption expenditures. An earlier version of this paper had a model with this feature, but estimation involved numerically calculating a thirty dimensional integral for each consumer for each parameter trial. Future research might focus on relaxing this stark assumption about the observability of consumption.

Isn't it a pleasure to study and practice what you have learned? Isn't it also bliss when friends visit from distant places?

Analects of Confucious, Chapter I

2

A Search and Learning Model of Export Dynamics

with Jonathan Eaton, Marcela Eslava, C.J. Krizan, and James R. Tybout

RESEARCH ON EXPORTING has been digging deeper into microeconomics data to understand the barriers that producers face in entering foreign markets and their implications for export dynamics. Firm-level datasets have provided insights first into the costs of exporting at all, and then, as data became available, to penetrating individual markets. We take this analysis one step forward by examining exporters' relationships with individual buyers in a market, both descriptively and through the lens of a dynamic model.

2.0.1 Scope

We begin by summarizing patterns in a decade's worth of data on individual merchandise shipments from Colombia to the United States. First, following work by some of the authors,[38] we review patterns of entry into the U.S. market of individual Colombian exporters across different cohorts. We note that most new exporters drop out of the U.S. market within a year, but those who survive this shakedown period have much lower exit rates in the future. Indeed, surviving members of new cohorts tend to expand their sales very rapidly, causing their market shares to grow as they mature. After a decade, nearly a quarter of total Colombian exports to the U.S. originate from firms that were not supplying the U.S. market at the beginning of the period.

We then look at relationships between buyers and sellers. Colombian firms which export to the U.S. ship at least once per year to an average of 1.3 U.S. clients. In contrast, U.S. firms place at least one order per year with an average of 2.2 Colombian suppliers if they deal with Colombian firms at all. Overall, the distribution of U.S. clients across Colombian exporters is very nearly Pareto, with a handful of large sellers accounting for a substantial fraction of total shipments. Most buyer-seller matches are short-lived, lasting less than two years, on average. Matches are even less durable if they begin with a small initial shipment. But enough exporters gain buyers each period that the ergodic distribution implied by the transitions and by entry replicates closely the distribution in the cross section.

Finally, we develop a model that is consistent with these facts. It is based on the conjecture that firms' exporting behavior reflects search and learning processes in a foreign market. That is, producers who are interested in a particular market devote resources to identifying potential buyers there. When they find one, they learn something (receive a noisy signal) about the appeal of their products in this market. Taking stock of the available information, these firms update their beliefs concerning the scope for export profits, and they adjust the intensity of their search efforts accordingly, seeking to maximize their expected profit streams. At the same time, firms manage their portfolio of existing clients, investing in their profitable business relationships and letting the others expire. These features of the model are not only motivated by the exporting patterns observed in the data, but also by the exporting strategies documented by a series of interviews with Colombian exporters[33]. Interviewed exporters described engaging in costly strategies both to search for new clients and to maintain existing relationships alive. They also frequently mentioned learning from previous relationships about the appeal of their products in a particular market, and using that information to adjust their searching behavior.

Fit to our data on shipments and business relationships, the model quantifies the role of several frictions in shaping firm-level export dynamics. We estimate that for non-exporters, the cost of maintaining low-level searches for clients in the U.S. is small, amounting to $1,405 per year for an expected yield of one potential client every two years. However, search costs are very convex in buyer arrival hazards, rising to $51,471 for an expected yield of one potential client per year. Both of these figures describe the search costs for a firm that has not yet established a successful business relationship abroad. But network effect are very important. We estimate that after the first relationship is formed, search costs for one client every two years drop to $106, and $3,898 for one client per year. Finally, once a successful match is formed, we estimate that it costs exporters $2,855 dollars per shipment to maintain the relationship. As a benchmark, the Doing Business project of the World Bank estimates that procedures required to export a one-container shipment cost $1,745 in Colombia in 2005. Even when a seller pays the fixed cost, her relationship dissolves with probability 0.27 per year for exogenous reasons.

In addition to trade costs, the model quantifies the effects of learning on exporter behavior. We estimate that on average, only 1 in 5 potential buyers that an exporter meets will be interested in forming a business relationship. However, this success rate varies substantially across sellers, so they adjust their search intensities dramatically as they form opinions concerning the scope of the market for their particular product. A typical firm which has met four potential buyers will choose a match hazard of 1.35 (new clients per year) if all of its encounters have led to successful business relationships, while it will choose a hazard of 0.22 if each encounter has been a failure.

This learning process, in combination with the various trade costs mentioned above, induces frictions and irreversibilities in export responses to market-wide shocks. We conclude our analysis with some experiments that quantify their implications for export dynamics. A 20 percent reduction in the cost of searching for new clients leads to an increase in total exports of around 5 percent, which takes some time to kick in. Increased exports are mostly explained by the entry of new sellers into exporting, and to a lesser extent by an increase in the mean number of clients per seller. In turn, a decrease of 20 percent in the per-shipment fixed cost leads to a much more marked increase in both the number of exporters and the mean number of clients, and also to an increase in mean sales per client. The latter occurs despite the entry into exporting of less productive sellers, and is explained by increased search by the more productive firms.

2.0.2 Relation to literature

While we look at the evolution of firms' sales in a particular market, our analysis is related to the literature on the dynamics of firm size in general. The model explains the size distribution of firm sales through two interacting mechanisms. One, as in Melitz[68], Bernard et al.[13], Luttmer[66], and Irarrazabal and Opromolla[52], is firm efficiency: More efficient firms sell more to a given set of buyers by having a lower price or a higher quality product. A second is that some firms have larger networks of buyers than others, as in Jackson and Rogers[54] or Chaney[23].

Investments in building a client base constitute a type of sunk cost, so our model also relates to the export hysteresis literature, where firms pay a one-shot start-up cost to break into new markets.[10,29,32,3,4] But unlike these formulations, our sunk costs are incurred on the client margin rather than the country margin, and they pay off in terms of market knowledge and reputation as well as revenue streams. These features of our model allow us to explain why new exporters who don't exit tend to rapidly expand, and why established exporters' sales are relatively stable. They also explain why many firms export for short periods on a very small scale.

Our formulation is also related to the two-period learning models developed by Rauch and Watson[77] and Albornoz et al.[2]. In the former, importers experiment with foreign suppliers by placing trial orders with them, and they gain access to a supplier network if they establish a successful business relationship. In the latter, firms choose to experiment in markets with low entry costs in order to learn about their product's appeal elsewhere. Like our model, these formulations provide interpretations for the fact that when new exporters survive, their exports tend to grow rapidly.*

Finally, in allowing firms to attract more buyers by incurring greater costs, our analysis relates to Drozd and Nosal[35] and Arkolakis[8 9]. By positing that firms face marketing costs that are convex in the number of foreign clients they service, Arkolakis also accounts for small-scale exporters and the age-dependence of export growth rates. However, since all exporting relationships last a single period in his models and learning is absent, Arkolakis's models do not explain the irreversibilities observed in firms' exporting behavior, nor do they speak to the duration of matches.

2.1 Firm-Level Trade: Transaction Level Evidence

2.1.1 Data

The empirical motivation for our model comes from a comprehensive data set that describes all imports by buyers in the United States from Colombian exporters (as well as other origins) during the period 1992-2009. The source is the U.S. Census Bureau's Longitudinal Foreign Trade Transactions Database (LFTTD). Each record includes a date, the US dollar value of the product shipped, a 6-digit harmonized system product code, a quantity index, and, critically, ID codes for both sellers and buyers. These IDs allow us to identify the formation and dissolution of business relationships between individual buyers in the U.S. and sellers in

*Ruhl and Willis[81] also note this pattern in plant-level export data and show that market entry costs are insufficient to explain it.

Colombia, hereafter referred to as "matches."†

To identify foreign exporters, the U.S. import transactions records include a manufacturer's identification code.‡ This field is an amalgamation of the manufacturer's country, company name, street address, and city. Anecdotal information from customs brokers indicates that commonly used software constructs it automatically as the name and address information is entered in other fields. So this variable is sensitive to differences in the way exporters' names and addresses are recorded as they pass through customs, and shipments from the same exporter can appear to originate from distinct Colombian firms. To gauge the importance of this problem, we have conducted various checks on the matches that are based on this variable; these are explained in the appendix.

We limit our analysis to transactions between non-affiliated trade partners, and we consider only imports of manufactured goods. The latter restriction notably excludes oil and coffee exports, which constitute the bulk of trade between the two countries and are dominated by a few Colombian sellers.§ Our final data set of manufacturing transactions spans the years 1992-2009. It contains 26,625 unique Colombian exporters, 12,921 unique U.S. importers, and 42,767 unique trading pairs. Value data have been deflated to 1992 prices using the U.S. CPI. Since we exclude a number of large HS codes from our data, as well as affiliated trade, and because we also lose information due to disclosure restrictions, the total value covered

†There are two ways to track U.S. importers in the LFTTD: Employment Identification Numbers (EINs) and the firm identifiers in the Longitudinal Business Database ("alphas"). Though an EIN does not necessarily identify a complete firm, it is unique to a firm, and there is one associated with every import transaction. Alphas map to entire firms, but the match rate between trade transactions and alphas is only about 80 percent[14]. To maximize the coverage of our sample, we use Employment Identification Numbers (EIN) to identify U.S. buyers.

‡This variable is based on Block 13 of CBP form 7501, the import declaration form and customs brokers are required to input the data.

§Colombian commercialization of coffee is centralized to an important degree by the National Federation of Coffee Growers. A few players also dominate oil exports.

by our data is not comparable to total Colombian exports to the U.S. Table B.1 in appendix B.1 compares patterns in our sample to patterns in official aggregates from both the U.S. and Colombia.

In addition to U.S. customs records, we use establishment level survey data from Colombia's national statistics agency (Departmento Administrativo Nacional de Estadistica, or DANE). These data provide annual information on the sales volumes, exports, and other characteristics of all Colombian manufacturing plants with at least 10 workers. Because they have been widely analyzed, we do not discuss summary statistics for this data set herein. Later, however, when estimating our search and learning model, we use such statistics to characterize the size distribution of Colombian firms, the fraction of Colombian plants that export and, among these firms, the relationship between exports and domestic sales.

2.1.2 Exports and Exporters

Following Brooks[19] and Eaton et al.[38], Tables 2.1-2.3 provide various annual measures of Colombian exports of manufactured goods to the United States for the years 1992-2009.[¶] Each column follows an exporting cohort—i.e., a group of firms that began exporting in a particular year—from the year of its appearance through time. The tables report number of exporters, total exports, and exports per firm, respectively. Note that, since we don't know the history of firms before 1992, the 1992 "cohort" consists of all firms present that year, regardless of when they began exporting; given re-entry. This implies that the first few cohorts are in general overestimated in terms of their initial size. Nonetheless, the patterns highlighted below apply also to the most recent cohorts.

[¶] Similar tables for Colombian exports of all goods and to all destinations appear in Eaton et al.[38].

year	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	total
1992	2,232																		2,232
1993	823	1,235																	2,058
1994	583	330	1,160																2,073
1995	440	213	339	953															1,945
1996	372	163	178	255	899														1,867
1997	321	128	133	170	248	877													1,877
1998	268	104	124	132	153	256	893												1,930
1999	232	85	87	114	117	187	262	1,026											2,110
2000	203	85	79	91	103	136	170	344	1,372										2,583
2001	187	70	65	79	85	109	145	229	389	1,251									2,609
2002	173	64	62	72	68	88	112	171	242	399	1,373								2,824
2003	165	51	58	62	62	77	86	140	185	301	440	1,719							3,346
2004	150	52	41	53	63	76	80	132	164	223	327	616	1,768						3,745
2005	140	52	47	39	54	77	69	115	145	196	235	398	661	1,902					4,130
2006	122	46	44	39	44	71	65	110	131	157	168	308	410	564	1,896				4,175
2007	113	37	39	31	42	55	48	91	101	132	156	240	305	365	548	1,681			3,984
2008	93	29	30	24	38	50	45	74	90	117	130	184	198	230	331	447	1,455		3,565
2009	80	25	28	24	28	40	39	60	72	88	97	145	175	157	230	248	386	1,378	3,300

Table 2.1: Number of Exporting Firms, by Entry Cohort

year	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	total
1992	469																		469
1993	352	83																	435
1994	336	83	92																510
1995	313	75	102	58															549
1996	256	67	62	40	60														484
1997	247	84	43	41	48	119													581
1998	225	49	42	36	45	131	63												590
1999	207	51	49	41	39	197	74	81											739
2000	180	53	55	37	51	102	53	158	109										799
2001	150	22	51	41	28	57	36	80	101	111									677
2002	124	23	47	34	27	28	23	45	65	83	40								538
2003	147	42	51	31	42	24	22	37	71	107	50	78							702
2004	156	43	53	19	57	21	23	42	78	106	60	107	90						855
2005	150	22	75	17	52	18	23	43	78	80	58	81	75	84					855
2006	117	31	52	14	64	43	17	38	61	79	32	51	52	112	78				838
2007	103	7	18	11	67	58	19	30	28	64	22	35	33	66	67	62			689
2008	95	6	9	8	33	37	17	33	26	34	20	31	37	54	42	53	57		591
2009	68	22	7	6	13	24	10	23	16	16	14	22	41	25	39	37	36	64	485

Table 2.2: Value of Exports, by Entry Cohort (millions of $US)

year	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	pooled
1992	210																		210
1993	428	67																	211
1994	576	251	79																246
1995	712	353	300	61															282
1996	687	411	346	158	67														259
1997	771	652	321	241	192	136													310
1998	839	468	339	269	297	510	71												306
1999	893	601	561	361	336	1,054	281	79											350
2000	885	623	697	407	496	750	313	460	80										309
2001	801	316	783	519	329	521	251	350	259	89									260
2002	716	353	757	473	399	318	207	260	268	207	29								191
2003	891	827	870	493	677	315	257	260	385	355	114	46							210
2004	1,039	828	1,281	358	900	281	291	318	478	476	183	174	51						228
2005	1,071	413	1,593	444	967	231	326	375	535	408	248	204	113	44					207
2006	958	675	1,177	356	1,448	605	256	341	464	505	188	165	126	198	41				201
2007	915	175	466	357	1,606	1,048	391	327	278	481	140	145	108	181	123	37			173
2008	1,023	208	283	341	860	747	379	443	289	287	153	166	186	236	125	120	39		166
2009	855	864	262	266	478	607	255	389	221	176	143	152	235	162	169	151	93	47	147

Table 2.3: Exports per Firm, by Entry Cohort (thousands of $US)

Consider Table 2.1 first. Naturally, each cohort's membership falls as it matures. But note that there is especially high attrition the first year, with more than 60 percent of firms dropping out. Conditional on making it to the second year, the survival probability is much higher, however, with an attrition rate around 40 percent the second year, and further declines occur thereafter. Thus, in terms of numbers, the most recent cohort is always larger than any previous one. Firms that were exporting to the United States in 1992 account for less than five percent of the firms exporting to the United States towards the end of the sample.

Table 2.2 shows that the rapid initial decline in its membership is not followed by a similar collapse of the total sales of a cohort. The decline in number of firms per cohort along with their relatively stable total sales means, of course, that sales per firm are growing substantially From the first to the second year of any cohort average sales more than double (Table 2.3).

2.1.3 Evidence on Buyer-Seller Matches

We next use the data to characterize the buyer-seller matches that took place during 1992-2009.

2.1.3.1 Monogamous and Polygamous Matches

The number of Colombian exporters appearing in our sample grew from 2,232 in 1992 to 3,300 in 2009, a growth of 2 percent per annum, while the number of U.S. importing firms grew by 3 percent per annum (Table 2.4). The number of Colombian exporter-U.S. importer pairs (representing at least one transaction between them in a year) also grew at an annual rate of 2 percent. Roughly 80 percent of matches are monogamous in the sense that the buyer deals with only one Colombian exporter and the exporter ships to only one buyer in the United States. However, since the remainder of the matches are polygamous, the average

Colombian exporter was involved in relationships with around 1.3 U.S. firms while the average U.S. buyer was involved with around 2.3 Colombian firms. Both figures declined slightly over the period.

Year	Colombian Sellers	U.S. Importers	Pairs
1992	2,232	1,190	3,087
1993	2,058	1,183	2,824
1994	2,073	1,212	2,810
1995	1,945	1,173	2,588
1996	1,867	1,191	2,490
1997	1,877	1,208	2,480
1998	1,930	1,191	2,495
1999	2,110	1,386	2,793
2000	2,583	1,661	3,411
2001	2,609	1,698	3,483
2002	2,824	1,826	3,733
2003	3,346	2,110	4,483
2004	3,745	2,296	5,071
2005	4,130	2,457	5,552
2006	4,175	2,471	5,607
2007	3,984	2,343	5,307
2008	3,565	2,221	4,751
2009	3,300	2,079	4,467

Table 2.4: Size of Data Set

2.1.3.2 Transition Probabilities

Like exporting stints (Table 2.1), most matches are short-lived. Of the 3,087 buyer-seller matches that existed at the beginning of the period, 70 percent didn't make it to 1993. But, of those that made it into the next year, almost 50 percent made it into the next year. Similarly, of the relationships that existed in 2005, 57 percent started that year but of those that started before, 37 percent had been around at least three years before. Of the 3,210 matches identified in 1992, less than 25 endure (are present every year) throughout the period.

Table 2.5 reports the probability with which a Colombian firm participating in certain number of relationships with buyers transits into a different number of relationships the following year. (Confidentiality restrictions prevent us from reporting numbers for cells that are too sparsely populated.) This table reports the annual average for 1992-2009 across all industries. A firm that stops exporting but re-appears as an exporter sometime later in our sample period is considered to have gone ”dormant”, while those exporters that drop to zero foreign sales for the extent of our sample are considered to have gone ”out” of exporting. Those that have never been observed to export constitute the pool of potential entrants.

Among first-time exporters, 93.2 percent sell to only one firm. Of these, 62 percent don’t export the next year, and only about six percent go on to establish a larger number of relationships. For firms with three relationships in a year, about twelve percent enter into a larger number of relationships the next year. Hence there is an enormous amount of churning at the lower end. Even for firms with a large number of relationships the most likely outcome is to have fewer the next year.

We can ask what this pattern of entry and growth implies about the ergodic distribution of relationships. If we assume that entrants in a year replace exiting firms, the ergodic distribution implied by this transition matrix is given by Table 2.6.

For purposes of comparison, the year-specific average share of Colombian firms in each group is reported as well. Note that the ergodic distribution implied by the transition matrix is very close to the cross-sectional distribution in the data, suggesting that over the period we observe the process has been quite stationary. Interestingly, both distributions are very nearly

t \t+1	Out	Dormant	1	2	3	4	5	6-10	11+
Out	.	.	0.932	0.055	0.009	0.002	0.001	0.001	0.000
Dormant	.	.	0.876	0.100	0.015	0.008	.	.	0.000
1	0.539	0.080	0.321	0.048	0.010	0.002	.	0.001	.
2	0.194	0.077	0.375	0.241	.	0.024	0.009	0.004	.
3	0.090	0.042	0.220	0.271	0.210	0.092	.	0.027	.
4	0.059	.	0.129	0.216	0.215	0.184	0.083	0.095	.
5	.	.	0.095	0.184	0.181	0.181	0.126	0.178	.
6-10	.	.	0.039	0.073	0.089	0.123	0.157	0.419	0.073
11+	.	0.000	0.000	0.000	.	.	.	0.432	0.526

Table 2.5: Transition Probabilities, Number of Clients

	1	2	3	4	5	6-10	11+
Erg Distribution	0.792	0.112	0.031	0.016	0.009	0.022	0.016
Data	0.778	0.116	0.043	0.021	0.011	.	.

Table 2.6: Ergodic Client Distribution Implied by Transitions

Pareto, reflecting the coexistence of many small scale exporters with a few "super-exporters."

2.1.3.3 Match Maturation

The survival probability of new matches increases with initial sales volume. Table 2.7 sorts observations on matches according to their size in their first year of existence and reports year-to-year separation rates. In addition to the very low survival rates, two patterns stand out. First, those matches that begin with sales in the top quartile among all new matches are more likely to survive than matches that begin with smaller sales volumes. Second, survival probabilities improve after the initial year.

	1 year	2 years	3 years	4 years	5+ years
Quartile 1	82.9	63.2	57.3	55.0	49.7
Quartile 2	75.6	58.4	49.4	46.8	43.7
Quartile 3	67.7	52.1	44.6	40.8	37.6
Quartile 4	52.1	44.5	40.3	39.2	36.7

Table 2.7: Separation Rates, by Age of Match and Initial Sales

Further features of the match maturation process are evident in Figure 2.1, which shows the log of annual sales per match, broken down by initial size quartile. For each size quartile, matches are further distinguished according to their life span: less than one, 1 to 2 years, and so forth. And for each cluster of bars, the left-most bar corresponds to sales in the initial year of the match's existence, the next bar corresponds to sales during the second year of the match's existence, and so forth.

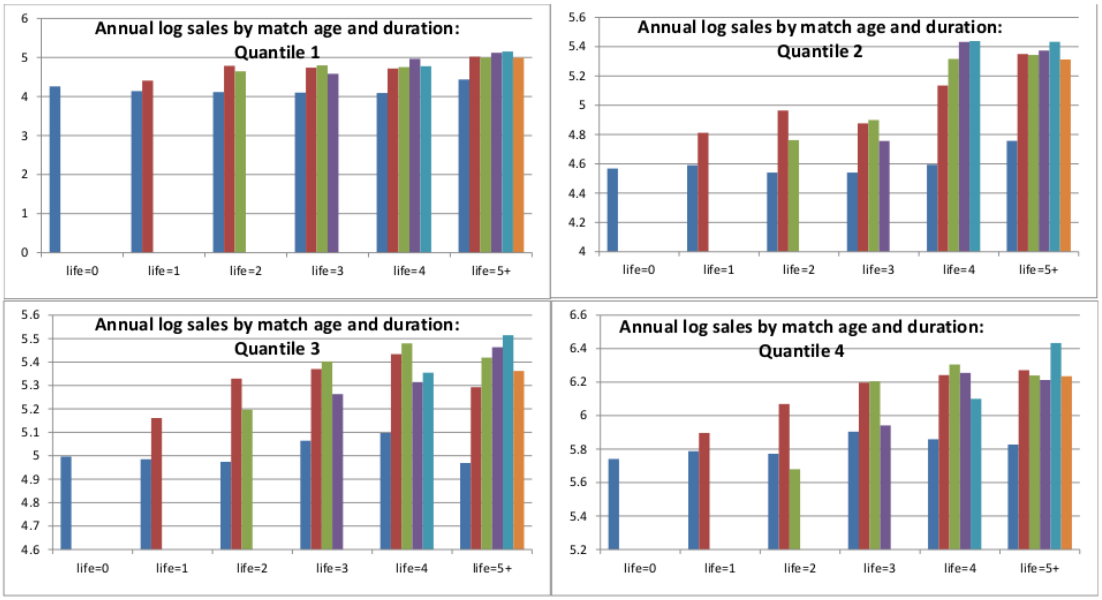


Figure 2.1: Average log annual sales per match, by initial size quartile

The first message of these graphs is that initial sales are a good predictor of sales in subsequent years, conditioning on survival. Those matches with first-year sales in the smallest quartile systematically generated the lowest annual sales in subsequent years, and more generally, first-year sales are monotonically related to annual sales in subsequent years. Second, sales tend to jump from the first to the second year, in large part simply because observations on a match's first year correspond to less than a full calendar year. (There is an analogous effect at work in the final year of a match's life.) Looking at complete-year observations reveals a tendency for annual sales to grow among matches that start small and survive, but no such tendency among matches that start in the largest quartile. Finally, looking across matches with different life spans, those that survive more years tend to have higher sales in all (full) years than matches that fail relatively quickly. This pattern is robust across matches in the different quartiles for initial sales.

2.1.3.4 Number of Clients and Sales per Client

Finally, firms that are successful at building a large client base also manage to sell relatively large amounts to each client. To summarize this relationship we fit the following regression:

$$\ln \overline{R}_{jt} = \phi_0^r + \phi_1^r \ln(n_{jt}^c) + \phi_2^r \ln(n_{jt}^c)^2 + \epsilon_{jt}^r$$

Here $\overline{R}_{jt}$ is exporter j's average revenue per client in year t, and n_{jt}^c is the number of clients who received shipments from j during the same year. The regression implies $\overline{R}$ is an increasing concave function of n^c: $\widehat{\phi_1^r} = 2.67$; $\widehat{\phi_2^r} = -0.14$.

2.2 A Model of Exporting at the Transactions Level

We now develop a model of exporter behavior consistent with the patterns reviewed above. Buyer-seller relationships form and disband at irregular intervals. Similarly, export shipments are discrete events distributed unevenly through time. To capture these features of the data, and to allow agents to update their behavior each time their circumstances change, we formulate our model in continuous time, treating all of the exogenous processes in our model as Markov jump processes.

Explaining the evolution of a firm's exports and domestic sales requires modeling both its sales to existing buyers and the evolution of its portfolio of clients. We can treat these two components sequentially. We first consider the relationship between a seller and an individual buyer. Having characterized the seller's profits from a relationship with an individual buyer, we then turn to her learning about the popularity of her product, i.e., the chance that a potential buyers likes her product. Finally, we characterize her search for buyers.

2.2.1 A Seller-Buyer Relationship

This section characterizes the profit streams that sellers generate from successful business relationships. The expressions we develop here describe relationships between domestic firms and foreign buyers, but with appropriate relabeling of market-wide variables they apply equally to relationships between domestic firms and domestic buyers.

2.2.1.1 Profits from a Single Shipment

Several features of our model are standard. First, at any time t seller j can hire workers at a wage w_t in real local currency units, each of whom can produce $\varphi_j \in \{\varphi^1, .., \varphi^{N_\varphi}\}$ units of output.[‖] Hence seller j's unit cost in local currency is w_t/φ_j. If she sells at price p_{jt} in foreign currency her unit profit in local currency is

$$p_{jt}/e_t - w_t/\varphi_j, \tag{2.1}$$

where e_t is the exchange rate. Second, goods markets are monopolistically competitive and each producer supplies a unique differentiated product.

Once buyer i has agreed to form a business relationship with seller j, he periodically places sales orders with j. For j, an order from i that arrives at time t generates revenue:

$$X_{ijt} = \left(\frac{p_{jt}}{P_t}\right)^{1-\eta} y_{ijt}\overline{X}_t, \tag{2.2}$$

where $\eta > 1$ is buyers' elasticity of demand, p_{jt} is the price of seller j's product, $\overline{X}_t$ is the average spending level among all potential foreign buyers, P_t is the relevant price index for all competing products in the foreign market, and $y_{ijt} \in \{y^1, .., y^{N_y}\}$ is a time-varying demand

[‖]We treat φ as time-invariant to facilitate model identification. Other sources of idiosyncratic temporal variation in sales will be discussed shortly.

shifter idiosyncratic to the ij relationship.[**]

For simplicity, and to keep the analysis as close as possible to other heterogeneous firm models, we assume that the seller posts a non-negotiable price, charging the optimal markup over unit cost:[††]

$$p_{jt} = \frac{\eta}{\eta - 1} \frac{e_t w_t}{\varphi_j} \tag{2.3}$$

By (2.1), (2.2), and (2.3), an order from buyer i at time t therefore generates the following profits for seller j:

$$\pi_{ijt} = \frac{1}{\eta} \frac{\overline{X}_t}{e_t} \left(\frac{e_t w_t \eta/(\eta - 1)}{\varphi_j P_t} \right)^{1-\eta} y_{ijt}.$$

We can combine all the macroeconomic variables affecting the profit of any seller from this source selling in this destination, along with constants, as:

$$x_t = \frac{1}{\eta} \frac{\overline{X}_t}{e} \left(\frac{e_t w_t \eta/(\eta - 1)}{P_t} \right)^{1-\eta},$$

where $x \in \{x^1, .., x^{N_x}\}$ is general to all potential buyers in the foreign market. Suppressing subscripts on state variables, this allows us to write the profits from a sale as:

$$\pi_{\varphi}(x, y) = x\varphi^{\eta-1} y, \tag{2.4}$$

In what follows, (2.4) is all we take from our specification of preferences and pricing behav-

[**]Not all buyers necessarily face the same range of goods and hence the same aggregate price index P. We treat idiosyncratic components of the price index as P as reflected in y_{ijt}.

[††]An alternative specification would introduce bilateral bargaining between buyer and seller.

ior into the dynamic analysis. Any set of assumptions that deliver this simple multiplicative expression for a firm's profit from a sale would serve us equally well.

2.2.1.2 Relationship Dynamics

At any point in time, each seller maintains business relationships with an endogenous number of buyers. These relationships form as a consequence of a search process that will be characterized in Section 2.2.3, and they dissolve for several reasons. First, there is a constant exogenous hazard δ that any particular relationship will terminate, which could be due to the demise of the buyer or the buyer no longer finding the seller's product useful. Second, after each sale to a particular buyer, the seller evaluates whether it is worth sustaining her relationship with him. Doing so keeps the possibility of future sales to him alive, but it also means paying the fixed costs F of maintaining the account, providing technical support, and maintaining client-specific product adjustments.[‡‡]

When deciding whether to maintain a particular business relationship, the seller knows her own type, φ, the macro state, x and profits from the current sale, $\pi_\varphi(x, y)$ to the buyer in question. She can therefore infer this buyer's current y value and calculate the value of her relationship with him to be:

$$\widetilde{\pi}_\varphi(x, y) = \pi_\varphi(x, y) + \max\left\{\widehat{\pi}_\varphi(x, y) - F, 0\right\}.$$

Here $\widehat{\pi}_\varphi(x, y)$ is the expected value of continuing a relationship that is currently in state (x, y). Clearly the seller terminates this relationship if $\widehat{\pi}_\varphi(x, y) < F$.

If a seller pays F to keep a relationship active, and if the relationship does not end anyway

[‡‡]For instance, Colombian producers of construction materials interviewed for a related project[33] referred that it is frequent for foreign buyers to request adjustments in the specifications of products or packages. In turn, these require adjustments in the production process that are costly to maintain.

for exogenous reasons, one of several events will next affect it: with hazard λ^b the buyer will place another order, with hazard $q^X_{xx'}$ x will jump to some new market-wide state $x' \neq x$, or with hazard $q^Y_{yy'}$ y will jump to some new buyer-specific shock $y' \neq y$.* Let τ_b be the random time that elapses until one of these events occurs. Given that x and y are Markov jump processes, τ_b is distributed exponentially with parameter $\lambda^b + \lambda^X_x + \lambda^Y_y$, where

$$\lambda^X_x = \sum_{x' \neq x} q^X_{xx'} \tag{2.5}$$

and

$$\lambda^Y_y = \sum_{y' \neq y} q^Y_{yy'}, \tag{2.6}$$

are the hazards of transiting from x to any $x' \neq x$, and from y to any $y' \neq y$, respectively. Then assuming the seller has a discount factor ρ, the continuation value $\widehat{\pi}_\varphi(x, y)$ solves the Bellman equation:

$$\begin{aligned}
\widehat{\pi}_\varphi(x, y) &= \mathbf{E}_{\tau_b}\left[e^{-(\rho+\delta)\tau_b}\frac{1}{\lambda^b + \lambda^X_x + \lambda^Y_y}\left(\sum_{x' \neq x} q^X_{xx'}\widehat{\pi}_\varphi(x', y) + \sum_{y' \neq y} q^Y_{yy'}\widehat{\pi}_\varphi(x, y') + \lambda^b \widetilde{\pi}_\varphi(x, y)\right)\right] \\
&= \frac{1}{\rho + \delta + \lambda^b + \lambda^X_x + \lambda^Y_y}\left(\sum_{x' \neq x} q^X_{xx'}\widehat{\pi}_\varphi(x', y) + \sum_{y' \neq y} q^Y_{yy'}\widehat{\pi}_\varphi(x, y') + \lambda^b \widetilde{\pi}_\varphi(x, y)\right)
\end{aligned}$$

Before a seller has met her next buyer, she does not know what state y this buyer will happen to be in. So when choosing her search intensity for new business relationships, she must base her decisions on the ex ante expected pay-off to forming a new business relationship.

*Since sales in the data are discrete events rather than flows, we model the buyer's purchases accordingly. We think of the buyer not as making use of the products continually but in discrete spurts. For example, the buyer might be a producer of a product that it makes in batches. At the completion of each batch it buys inputs for the next batch.

Given the market state x, a type-φ seller calculates this expected value as:

$$\widetilde{\pi}_{\varphi}(x) = \sum_{s} \Pr(y^s)\widetilde{\pi}_{\varphi}(x, y).$$

where $\Pr(y^s)$ is the probability that a randomly selected buyer is currently in state $y^s \in \{y^1, .., y^{N_y}\}$.[†]

For the purposes of the search model that follows, all that matters about an individual relationship is $\widetilde{\pi}_{\varphi}(x)$, and this object can be estimated directly from data on the revenue streams generated by matches. Nonetheless, the history of a seller's interactions with a given buyer affects its overall sales trajectory and hence matters for our characterization of aggregate export dynamics.

Hereafter, we will denote the expected value of a relationship with a foreign buyer by $\widetilde{\pi}^f_{\varphi}(x)$ and the expected value of a relationship with a home market buyer by $\widetilde{\pi}^h_{\varphi}(x)$. These two objects are calculated in the same way, but since expenditure levels ($\overline{X}_t$) and price indices (P_t) differ across markets, and no exchange rate factor e is necessary for domestic profit calculations, each has its own process for the market-wide state variable, x. These market-wide demand shifters are denoted x^f and x^h below.

2.2.2 Learning about Product Appeal

Sellers conduct market-specific searches for buyers. When searching in market $m \in \{h, f\}$, each recognizes that some fraction $\theta^m \in [0, 1]$ of the potential buyers she meets there will be willing to do business with her. An encounter with one of these willing buyers generates an expected profit stream worth $\widetilde{\pi}^m_{\varphi,x}$, while an encounter with any of the remaining potential

[†]Here we take the probabilities $\Pr(y^m)$ to be the ergodic distribution of y implied by the transition hazards $q^Y_{yy'}$. We could assume that the distribution at the time of the first purchase is different from the ergodic one.

buyers does not generate a sale then or subsequently.

Each seller's θ^h and θ^f values are drawn before she has met any clients. These draws remain fixed through time, inducing permanent cross-market differences in her product's popularity. All θ^m draws are independently beta-distributed across sellers and markets:

$$b(\theta^m|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} (\theta^m)^{\alpha-1} (1-\theta^m)^{\beta-1}, \ m \in \{h, f\},$$

where $\Gamma(\phi) = \int_0^\infty z^{\phi-1} e^{-z} dz$ is the gamma function (needed to ensure that the distribution has the proper limits). However, the independence of θ^h and θ^f does not mean sellers' domestic and foreign sales are likewise independent. Rather, cross-market correlation in sales will be induced by the firm type φ, which can be viewed as capturing aspects of product appeal that are common to both markets.[‡]

Sellers are presumed to have already met many potential customers in the domestic market, and thus to have learned their θ^h draws. But sellers typically have far less experience abroad, so we allow them to still be learning about their θ^f draws. Specifically, each seller recognizes that for any given θ^f, the probability a random sample of n potential foreign buyers will yield a customers is binomially distributed:

$$q\left[a|n,\theta^f\right] = \binom{n}{a} \left[\theta^f\right]^a \left[1-\theta^f\right]^{n-a}.$$

So after she has met n potential buyers abroad, a of whom were willing to buy her product,

[‡]The firm effect is similarly interpreted to reflect both productive efficiency and product appeal in Melitz[68] and many other papers based on CES demand systems. However in the present context, the global aspects of product appeal captured by φ are qualitatively distinct from the market-specific product appeal effects captured by θ. The former determines the amount of a product each buyer purchases, given that he is interested, while the latter determines what fraction of potential buyers are willing to place orders with the seller, should they happen to meet her.

a seller's posterior beliefs about her θ^f draw are distributed:

$$p(\theta^f|a,n) \propto q\left[a|n,\theta^f\right] \cdot b(\theta^f|\alpha,\beta)$$

where the factor of proportionality is the inverse of the integral of the right-hand side over the support of θ^f. Since the beta distribution is the conjugate prior for the binomial, a firm's expected success rate after a successes in n trials has a convenient closed-form representation:

$$\overline{\theta}^f_{a,n} = E\left[\theta^f|a,n\right] = \int_0^1 \theta p(\theta|a,n)d\theta = \frac{a+\alpha}{n+\alpha+\beta}. \tag{2.7}$$

This posterior mean converges to $p\lim\left(\frac{a}{n}\right) = \theta^f$ as n gets large.

2.2.3 Searching for Buyers

To complete our model we now consider sellers' search intensities in each market. Each seller continuously chooses the market-specific hazard s^m, $m \in \{h, f\}$, with which she encounters a potential buyer, recognizing that this involves the instantaneous flow cost $c(s^m, a)$, where $c(s^m, a)$ is increasing and convex in s^m.[§] Whether $c(s^m, a)$ increases or decreases in the number of successful matches, a, depends upon the relative strength of several forces and will be left for the data to determine. Costs might fall with a because encounters with interested buyers increase the seller's visibility and enhance her opportunities to meet additional

[§]Interviews conducted with Colombian exporters revealed a variety of activities firms pursue to meet potential buyers abroad[33]. Ranked roughly in terms of decreasing cost, these included maintaining a foreign sales office; paying the exports promotion office to organize visits with prospective clients abroad, and sending their sales representatives to those visits; sending sales representatives abroad to visit potential clients on their own; attending trade fairs; paying a researcher to search the web for foreign firms that purchase products similar to their own; paying browsers to ensure that their site appear near the top of a search for their product type; maintaining a web site in English. Interviewees also reported that relatively low-cost activities, such as traveling to trade fairs, or translating their websites to English, led to relationships with one or two clients every few years. Establishing a larger network of clients required much more costly activities.

potential buyers. Alternatively, costs might rise if the pool of easy-to-reach buyers becomes "fished out," as in Arkolakis[8].

We can now describe optimal search behavior, beginning with the foreign market. Recall that when the foreign market state is x^f, a type-φ seller expects the value of a new business relationship will be $\widetilde{\pi}_{\varphi}^f(x^f)$. Further, she believes the next match will yield such a relationship with probability $\overline{\theta}_{a,n}^f$. Combined with search cost function $c(s^f, a)$ and the jump process for x^f, these objects imply sellers' optimal search policy abroad.

To characterize this policy, let τ_s^f be the random time until the next foreign search event, which could be either a change in the market-wide state x^f or an encounter with a potential buyer. Then, suppressing market superscripts, the optimal search intensity s for a type-φ firm with foreign market search history (a, n) solves the following the Bellman equation:

$$V_{\varphi}(a,n,x) = \max_{s} \mathbf{E}_{\tau_s} \left[-c(s,a) \int_0^{\tau_s} e^{-\rho t} dt + \frac{e^{-\rho \tau_s}}{s + \lambda_x^X} \cdot \left(\sum_{x' \neq x} q_{xx'}^X V_{\varphi,}(a,n,x') \right.\right.$$
$$\left.\left. + s \left[\overline{\theta}_{a,n} (\widetilde{\pi}_{\varphi}(x) + V_{\varphi}(a+1, n+1, x) + (1 - \overline{\theta}_{a,n}) V_{\varphi}(a, n+1, x) \right] \right) \right]$$

(Recall that λ_x^X is given by (2.5).) Taking expectations over τ_s yields:

$$V_{\varphi}(a,n,x) = \max_{s} \frac{1}{\rho + s + \lambda_x^X} \left[-c(s,a) + \sum_{x' \neq x} q_{xx'}^X V_{\varphi,}(a,n,x') \right. \tag{2.8}$$
$$\left. + s \left\{ \overline{\theta}_{a,n} \left[\widetilde{\pi}_{\varphi}(x) + V_{\varphi}(a+1, n+1, x) \right] + (1 - \overline{\theta}_{a,n}) V_{\varphi}(a, n+1, x) \right\} \right]$$

Applying the multiplication rule for differentiation and using expression (2.8) for $V_{\varphi}(a,n,x)$, the optimal search intensity s^* satisfies:

$$\frac{\partial c(s^*, a)}{\partial s} = \overline{\theta}_{a,n}\left[\widetilde{\pi}_{\varphi}(x) + V_{\varphi}(a+1, n+1, x)\right] + (1-\overline{\theta}_{a,n})V_{\varphi}(a, n+1, x) - V_{\varphi}(a, n, x) \tag{2.9}$$

That is, the marginal cost of search must equal the expected marginal benefit of a match, which includes the expected value of the associated profit stream, $\overline{\theta}_{a,n}\widetilde{\pi}_{\varphi}(x)$, and the expected value of the information generated.

Now consider the home market. Since we assume sellers have already learned their true success rates at home, θ^h, new encounters do not influence expectations, and we need not condition the value function or the expected success rate on search histories. Again suppressing market superscripts, the Bellman equation collapses to:

$$V_{\varphi}(x) = \max_{s} \frac{1}{\rho + \lambda_x^X}\left[-c(s, a) + \sum_{x' \neq x} q_{xx'}^X V_{\varphi}(x') + s\theta_j \widetilde{\pi}_{\varphi}(x)\right]$$

and the first-order condition is simply:

$$\frac{\partial c(s^*, a)}{\partial s} = \theta_j \widetilde{\pi}_{\varphi}(x).$$

The marginal cost of search equals the expected profit from a successful relationship times the probability of success.

2.3 An Empirical Version of the Model

2.3.1 The Search Cost Function

To implement our model empirically, we impose additional structure in several respects. First, we specify a functional form for our search cost function. Generalizing Arkolakis[8] to allow for network effects, we write these costs as:

$$c(s,a) = \kappa_0 \frac{(1+s)^{(1+1/\kappa_1)} - 1}{(1+a)^{\gamma(1+1/\kappa_1)}(1+1/\kappa_1)}. \tag{2.10}$$

Several properties of this function merit note. First, marginal costs fall at a rate determined by γ with the number of successful matches a seller has already made, so $\gamma > 0$ implies "network" effects and $\gamma < 0$ implies "congestion" effects.[¶] Second, a seller who is not searching in a particular market incurs no search cost: $c(0,a) = 0$. Third, given the cumulative number of successful matches, a, the marginal cost of search increases with λ^s at a rate inversely related to $\kappa_1 : c_s(s,a) = \kappa_0(1+s)^{1/\kappa_1}/(1+a)^{\gamma(1+1/\kappa_1)}$. Finally, network effects endure, even if a firm is not actively searching.

2.3.2 Processes for exogenous state variables

Next we impose more structure on the exogenous state variables, φ, x^h, x^f, y^h and y^f. All are assumed to have zero means in logs, and the net effect of these normalizations is undone by introducing scalars Π^h and Π^f into the home and foreign profit functions, respectively:

$$\begin{aligned} \pi^f_\varphi(x^f, y^f) &= \Pi^f x^f \varphi^{\eta-1} y^f, \\ \pi^h_\varphi(x^h, y^h) &= \Pi^h x^f \varphi^{\eta-1} y^f \end{aligned}$$

More substantively, we impose that the cross-firm distribution of φ is log normal with standard deviation σ_φ, and we treat all of the Markov jump processes (x^h, y^h, x^f, y^f) as independent Ehrenfest diffusion processes. The idiosyncratic match shocks, y^f and y^h, are as-

[¶] To contain the dimensionality of the computational problem we solve, we assume that firms with more than a^* buyers have (i) exhausted their learning effects, and (ii) reap no additional network effects at the margin from further matches. We choose a^* to exceed the observed maximum a for 99 percent of sellers in the foreign (United States) market. Also, we set $a = a^*$ for all sellers in their home (Colombian) market.

sumed to share the same distribution, but we allow the x^f and x^h processes to differ. Among other things, the latter accommodates the fact that the exchange rate affects aggregate demand and price indices in the two markets differently.

Any variable z generated by an Ehrenfest process can be discretized into $2g+1$ possible values, $g \in I^+ : z \in \{-g\Delta, -(g-1)\Delta, .., 0, .., (g-1)\Delta, g\Delta\}$. Further, it jumps to a new value with hazard λ_z, and given that a jump occurs, it goes to z' according to:

$$z' = \begin{cases} z+\Delta \\ z-\Delta \\ \text{other} \end{cases} \text{with probability} \begin{cases} \frac{1}{2}\left(1-\frac{z}{g\triangle}\right) \\ \frac{1}{2}\left(1+\frac{z}{g\triangle}\right) \\ 0 \end{cases} .$$

Thus, given a grid size g, the intensity matrices $Q^X = \left\{q_{ij}^X\right\}_{i,j=1,N^X}$ and $Q^Y = \left\{q_{ij}^Y\right\}_{i,j=1,N^Y}$ that were introduced in section 3.1 are each block-diagonal and characterized by a single parameter, Δ.

2.4 Estimation

2.4.1 Stage 1: Estimating Observable Jump Processes

Shimer[83] shows that if z follows a continuous time Ehrenfest diffusion process, it asymptotes to an Ornstein-Uhlenbeck process with mean zero as the fineness of the grid increases:[‖]

$$dz = -\mu z dt + \sigma dW.$$

Here $\mu = \lambda_z/g$, $\sigma = \sqrt{\lambda_z}\Delta$, and W follows a Weiner process. Accordingly, since it is possible to observe proxies for x^f and x^h, these can be viewed as discrete time observations

[‖]Specifically, replacing the parameter vector (λ, g, Δ) with $(\lambda/\epsilon, g/\epsilon, \Delta\sqrt{\epsilon})$, $\epsilon > 0$, leaves the autocorrelation parameter μ and the instantaneous variance parameter σ unchanged. But as $\epsilon \to 0$, the innovation dW approaches normal.

	Parameter	*value*
home macro state jump hazard	λ^{x_h}	1.200
foreign macro state jump hazard	λ^{x_f}	1.215
home macro state jump size	Δ^{x_h}	0.003
foreign macro state jump size	Δ^{x_f}	0.053

Table 2.8: Market-wide Demand Shifters

on underlying Ornstein-Uhlenbeck processes, and the parameters of these processes can be econometrically estimated. Then, given μ and σ, estimates of Δ and λ for these processes can be inferred.

Measuring x^f as real expenditures on manufacturing goods in the U.S., and measuring x^h as real expenditures on manufacturing goods in Colombia, we obtain the results reported in Table 2.8.[**] They imply that x^f and x^h both jump 1.2 times per year, on average. However, jumps in the U.S. market tend to be much larger, essentially because they reflect movements in the real exchange rate as well as movement in dollar-denominated expenditures.

2.4.2 Stage 2: Indirect inference

Our data are relatively uninformative about the rate of time discount ρ and the demand elasticity η, so we do not attempt to estimate either one. For the former we follow convention and assume $\rho = 0.05$. For the latter, following many previous trade papers, we fix the demand elasticity at $\eta = 5$. All of the remaining parameters we estimate using the method of indirect inference[43]. These parameters include the exogenous match separation hazard (δ),

[**] Our foreign market size measure is the OECD time series on American GDP in 'Industry, including energy' adding imports and subtracting net exports of manufactures. Our home market size measure is real Colombian expenditures on manufacturing goods, taken from DANE. We converted all of the data used for the estimation into real 1992 US dollars, deflating nominal US dollars with the consumer price index available on the US Bureau of Labor Statistic website. We used an official Colombian Peso - US Dollar exchange rate time series downloaded from the Central Bank of Colombia to convert Pesos to nominal US Dollars

the market size scalars ($\Pi^h, \Pi^{f\cdot}$), the fixed costs of maintaining a match (F), the parameters of the product appeal distributions (α,β), the dispersion of the productivity distribution (σ_φ), the jump hazards for the idiosyncratic buyer shocks (λ_y), the hazard rate for shipments (λ_b), the network/congestion parameter (γ), the cost function convexity parameter (κ_1), and the cost function scaling parameter (κ_0). For notational convenience we hereafter collect these parameters in the vector Λ :

$$\Lambda = \left(\Pi^h, \Pi^{f\cdot}, \delta, F, \alpha, \beta, \sigma_\varphi, \lambda_y, \lambda_b, \gamma, \kappa_0, \kappa_1\right)$$

We seek the value of Λ that allows our model to replicate the features of the transactions-level data summarized in Section 2.1 above. In addition to the joint distribution of home and foreign sales across firms, these include the distribution of clients across exporters, the probabilities than a particular exporter will move up or down in this distribution, given its current position, the hazard that a given match will end, given its current age and size, the survival rates of exporting cohorts as they mature, and the distribution of shipment frequencies across matches.

The sample statistics that we use as a basis for inference are listed in Table 2.9. These same statistics are also repeatedly constructed using data simulated with the model at alternative candidate values for Λ. The method of indirect inference amounts to choosing the Λ value that minimizes a metric of the distance between sample and simulated statistics.[††]

[††]More precisely, our estimator for Λ is:

$$\widehat{\Lambda} = \arg\min \left[\widehat{M} - M_S(\Lambda)\right]' \widehat{W}^{-1} \left[\widehat{M} - M_S(\Lambda)\right]$$

where $\widehat{M}$ is the vector of data-based statistics listed in the right-most column of Table 2.9, $M_S(\Lambda)$ is their counterpart based on S simulations of our model at candidate vector Λ, and $\widehat{W}$ is a compatible matrix with $\widehat{se}(\widehat{M})$ on its diagonal and zeros elsewhere. These standard errors are constructed using the sample data. In

Data feature	Summary method	Statistics ($\widehat{M}$)
Distribution of home and foreign sales	OLS cross-plant regression: $\ln X_{jt}^{f} = \phi_0^{hf} + \phi_1^{hf} \ln X_{jt}^{h} + \epsilon_{jt}^{hf}$	$\widehat{\phi}_0^{hf}, \widehat{\phi}_1^{hf}, s\widehat{e}(\epsilon^{hf})$
	Cross-plant moments	$\widehat{E}(1_{X_{jt}^f>0}), \widehat{E}(\ln X_{jt}^f \vert X_{jt}^f > 0),$
	Standard deviation of foreign sales	$se(\ln X_{jt}^f)$
Distribution of clients across exporters, $\Phi(n^c)$	OLS regression for $n^c \in I^+$: $\ln[1-\Phi(n^c)] = \phi^c \ln(n^c) + \epsilon^n$	$\widehat{\phi}^c, s\widehat{e}(\epsilon^{n^c})$
Sales per client given number of clients	OLS cross-match regression: $\ln X_{ijt}^f = \phi_0^r + \phi_1^r \ln(n_{jt}^c) + \phi_2^r \ln(n_{jt}^c)^2 + \epsilon^m$	$\widehat{\phi}_0^r, \widehat{\phi}_1^r, \widehat{\phi}_2^r, s\widehat{e}(\epsilon^r)$
Autoregression, log domestic sales	$\phi_0^h + \phi_1^h \ln X_{jt-1}^h + \epsilon^h$	$\widehat{\phi}_1^h, s\widehat{e}(\epsilon^h)$
Transition probabilities, number of clients (n_{jt}^c)	Cross-plant year-to-year average transition rates	$\widehat{P}[n_{jt+1}^c = m \vert n_{jt}^c = k],$ $m, k = 0, 1, 2, 3+$
Match death hazards, given match age (A^m)	Cross-match average year-to-year death rates, given age	$\widehat{E}[1_{X_{ijt}^f=0} \vert X_{ijt-1}^f > 0, A_{ijt-1}^m],$ $A_{ijt-1}^m = 0, 1, 2, 3, 4+$
Exporter exit hazard by cohort age (A^c)	Cross exporter average exit rate, given years exporting	$\widehat{E}[1_{X_{jt}^f=0} \vert A_{jt}^c],$ $A_{jt}^c = 0, 1, 2, 3, 4+$
Cohort-specific exports per plant	Cross-exporter mean log exports	$\widehat{E}(\ln X_{jt}^f \vert A_{jt}^c),$ $A_{jt}^c = 1, 2, 3, 4+$
Match-specific shipments per year (n_{ijt}^s)	Cross-match mean shipments per year	$\widehat{E}\,(n^s)$ (trimmed)
Autoregression, match-specific sales (X_{ijt}^f)	$\ln X_{ijt}^f = \widehat{\beta}_0^f + \widehat{\beta}_1^f \ln X_{ijt-1}^f + \epsilon_{ijt}^f$	$\widehat{\beta}_1^f, s\widehat{e}(\epsilon^f)$
Match death prob. and match sales	$1_{X_{ijt}^f=0} = \widehat{\beta}_0^d + \widehat{\beta}_{\text{1st year}}^d 1_{A_{ijt-1}^m=0} + \widehat{\beta}_{\text{lsales}}^d \ln X_{ijt-1}^f + \epsilon_{ijt}^d$	$\widehat{\beta}_0^d, \widehat{\beta}_{\text{1st year}}^d, \widehat{\beta}_{\text{lsales}}^d, s\widehat{e}(\epsilon^d)$

Variable definitions:
n_{jt}^c: number of foreign clients served by firm j in year t
$\Phi(n^c)$: cumulative frequency distribution of number of foreign clients in population of exporters
A_{ijt}^m : age of match (in years) between seller j and foreign buyer i in year t
A_{jt}^c: number of consecutive years exporter j has made at least one shipment abroad.

Table 2.9: Statistics used for Indirect Inference

While there is no exact mapping between the statistics in the last column of Table 2.9 and the parameters we wish to estimate, it is possible to comment in general terms on sources of identification. First, several parameters are closely associated with sample means. Specifically, the profit function scaling parameters Π^h and Π^f are identified by average levels of sales in each market, $E(\ln X_{jt}^h)$ and $E(\ln X_{jt}^f)$, given market participation, as well as the fraction of firms that export, $E(1_{X_{jt}^f>0})$. And the shipment hazard λ_b is closely related to the average number of shipments per year, $\widehat{E}\left(n^s\right)$.

Second, the match-specific shock hazard, λ^y, the exogenous match separation hazard, δ, and the fixed costs of maintaining a match, F, are key determinants of the persistence and dispersion in client-specific sales trajectories. Accordingly, key statistics that help to identify these parameters include estimates of autoregressions for match-specific sales $\widehat{\beta}_1^f$, $\widehat{se}(\epsilon^f)$, match death hazards by age of match, $\widehat{E}[1_{X_{ijt}^f=0}|X_{ijt-1}^f > 0, A_{ijt-1}^m]$, and parameters of the regression relating match death hazards to match size: $\widehat{\beta}_0^d$, $\widehat{\beta}_{\text{1st year}}^d$, $\widehat{\beta}_{\text{lsales}}^d$, $\widehat{se}(\epsilon^d)$. Since the fixed costs of sustaining a match are incurred after each shipment, the difference in separation hazards between the first and all subsequent years helps to distinguish F from δ. Also, in the absence of shocks to market-wide conditions (x) or idiosyncratic buyer demands (y), all matches would survive A periods with probably $(1-\delta)^A$. Accordingly, the rate at which hazard rates decline with match age is informative about δ. Further identification comes from the fact that δ affects all firms equally, while the effect of F declines as $\widehat{\pi}_{\varphi,x}$ increases. This makes the association between shipment size and match longevity informative regarding the

addition to giving the greatest weight to those statistics that are most precisely estimated, W^{-1} serves to eliminate units of measurement as a factor in determining the fit. The efficient GMM estimator of ϕ would use $E\left[\widehat{M}-E(\widehat{M})\right]\left[\widehat{M}-E(\widehat{M})\right]'$ (adjusted for simulation error in $M(\Lambda)$) as its weighting matrix. But since our data on establishments and matches come from several sources, it computationally infeasible for us to construct this set of weights. Our weighting matrix yields a consistent estimator, provided that our model is properly specified.

importance of F.

Third, the θ distribution parameters, α and β, determine the cross-firm joint distribution of success rates in home and foreign markets and, similarly, the dispersion in firm types σ_φ helps determine the cross-distribution of domestic and foreign sales. The combined effects of these parameters is reflected in the means, variances, and covariances of foreign and domestic sales, which are implied in turn by $\widehat{\phi}_1^{hf}$, $\widehat{se}(\epsilon^{hf})$, and $se(\ln X_{jt}^f)$. Similarly, the cross-firm distribution of numbers of foreign clients, summarized by $\widehat{\phi}^c$ and $\widehat{se}(\epsilon^{n^c})$, responds to (α,β). This distribution also responds to σ_φ, since the firm effects φ strongly influence search intensities. But the role of the firm effects φ is distinct from that of the popularity indices θ^f and θ^h because φ induces correlation in sales across markets. This correlation, which is implied by $\widehat{E}(\ln X_{jt}^f|A_{jt}^c)$, $\widehat{\phi}_1^{hf}$, $\widehat{se}(\epsilon^{hf})$, and $se(\ln X_{jt}^f)$, helps to isolate the variance in firm effects, σ_φ.

Finally, the marginal cost of search and its sensitivity to previous matches are determined by γ, κ_0, and κ_1. Match rates, transition probabilities for numbers of clients, $\widehat{P}[n_{jt+1} = m|n_{jt} = k]$, and the client distribution are informative about the convexity of the matching cost function. Accordingly, $\widehat{\beta}^c$ and $\widehat{se}(\epsilon^n)$ are useful in their identification. Differences in match arrival rates among firms that have made many versus few matches help to distinguish the convexity parameter β from the network effect parameter, γ. And importantly, the shape of the client-per-seller distribution is informative about network effects, since these effects critically impact the ability of firms to sustain large client bases, and thus affect the "fatness" of the right-hand tail.

	Parameter	value	std. error
rate of exogenous separation	δ	0.267	0.001
domestic market size	Π^h	11.344	0.017
foreign market size	Π^f	10.675	0.017
fixed cost	F	7.957	0.018
First θ distribution parameter	α	0.716	0.007
Second θ distribution parameter	β	3.161	0.029
demand shock jump hazard	λ_y	0.532	0.001
demand shock jump size	Δ^y	0.087	0.001
shipment order arrival hazard	λ_b	8.836	0.006
std. deviation, log firm type	σ_φ	0.650	0.002
network effect parameter	γ	0.298	0.001
search cost function curvature parameter	κ_1	0.087	0.001
search cost function scale parameter	κ_0	111.499	0.512

Table 2.10: Parameters Estimated using indirect inference (Λ)

2.4.3 Parameter estimates

Table 2.11 reports estimates based on the data moments described in the previous subsection. Data-based estimates of these moments, $\widehat{M}$, are reported and juxtaposed with their simulated counterparts, $M_S(\Lambda)$, in Table 2.11.[‡‡] The Euclidean distance between these two vectors divided by the length of the latter vector is 0.118.

[‡‡]The share exporters, the coefficient of log foreign sales on log domestic sales, and the AR1 coefficient for log domestic sales in Table 2.11 are obtained from a combination of the Colombian Annual Manufacturing Survey (AMS) and the administrative records of exports transactions. The data used cover 1993-2007. Exports from administrative records are merged into the AMS using firm identifiers. This is done because the AMS has no export information for 1993-1999, and because the dynamics of aggregate exports reported in the EAM starting in 2004 differ substantially from aggregate reports from other sources.

No. clients (n^c)	Data	Model	Share of firms exporting	Data	Model
$\widehat{P}[n^c_{jt+1}=0 \mid n^c_{jt}=1]$	0.618	0.534	$\widehat{E}(1_{X^f_{jt}>0})$	0.299	0.351
$\widehat{P}[n^c_{jt+1}=1 \mid n^c_{jt}=1]$	0.321	0.358			
$\widehat{P}[n^c_{jt+1}=2 \mid n^c_{jt}=1]$	0.048	0.082	Log foreign sales on		
$\widehat{P}[n^c_{jt+1}\geq 3 \mid n^c_{jt}=1]$	0.013	0.024	log domestic sales	Data	Model
$\widehat{P}[n^c_{jt+1}=0 \mid n^c_{jt}=2]$	0.271	0.260			
$\widehat{P}[n^c_{jt+1}=1 \mid n^c_{jt}=2]$	0.375	0.321	$\widehat{\beta}^{hf}_1$	0.727	0.515
$\widehat{P}[n^c_{jt+1}=2 \mid n^c_{jt}=2]$	0.241	0.281	$\widehat{se}(\epsilon^{hf})$	2.167	1.424
$\widehat{P}[n^c_{jt+1}\geq 3 \mid n^c_{jt}=2]$	0.113	0.135			
Match death hazards	Data	Model	Exporter exit hazards	Data	Model
$\widehat{E}[1_{X^f_{ijt}=0} \mid X^f_{ijt-1}>0, A^m_{ijt-1}=0]$	0.694	0.857	$\widehat{E}[1_{X^f_{jt}=0} \mid A^c_{jt-1}=0]$	0.709	0.748
$\widehat{E}[1_{X^f_{ijt}=0} \mid X^f_{ijt-1}>0, A^m_{ijt-1}=1]$	0.515	0.329	$\widehat{E}[1_{X^f_{jt}=0} \mid A^c_{jt-1}=1]$	0.383	0.099
$\widehat{E}[1_{X^f_{ijt}=0} \mid X^f_{ijt-1}>0, A^m_{ijt-1}=2]$	0.450	0.304	$\widehat{E}[1_{X^f_{jt}=0} \mid A^c_{jt-1}=2]$	0.300	0.121
$\widehat{E}[1_{X^f_{ijt}=0} \mid X^f_{ijt-1}>0, A^m_{ijt-1}=3]$	0.424	0.281	$\widehat{E}[1_{X^f_{jt}=0} \mid A^c_{jt-1}=3]$	0.263	0.055
$\widehat{E}[1_{X^f_{ijt}=0} \mid X^f_{ijt-1}>0, A^m_{ijt-1}=4]$	0.389	0.305	$\widehat{E}[1_{X^f_{jt}=0} \mid A^c_{jt-1}=4]$	0.293	0.100
Log sales per client on			Log sales per exporter		
client no. regression	Data	Model	by cohort age	Data	Model
$\widehat{\beta}^m_1$	2.677	0.842	$\widehat{E}(\ln X^f_{jt} \mid A^c_{jt}=0)$	8.960	9.306
$\widehat{\beta}^m_2$	-0.143	0.042	$\widehat{E}(\ln X^f_{jt} \mid A^c_{jt}=1)$	10.018	10.806
$\widehat{se}(\epsilon^m)$	2.180	1.622	$\widehat{E}(\ln X^f_{jt} \mid A^c_{jt}=2)$	10.231	10.755
Client number inverse			$\widehat{E}(\ln X^f_{jt} \mid A^c_{jt}=3)$	10.369	10.679
CDF regression	Data	Model	$\widehat{E}(\ln X^f_{jt} \mid A^c_{jt}\geq 4)$	10.473	10.669
$\widehat{\beta}_1^{\,c}$	-1.667	-1.587	Log dom. sales autoreg.	Data	Model
$\widehat{\beta}_2^{\,c}$	-0.097	-0.280	$\widehat{\beta}^h_1$	0.976	0.896
$\widehat{se}(\epsilon^{n^c})$	0.066	0.128	$\widehat{se}(\epsilon^h)$	0.462	0.683
Match shipments per year	Data	Model	Log match sale autoreg.	Data	Model
$\widehat{E}\,(n^s)$	4.824	3.770	$\widehat{\beta}^f_1$	0.811	0.613
Match death prob regression	Data	Model	$\beta^f_{\text{1st year}}$	0.233	0.370
$\widehat{\beta}^d_0$	1.174	1.640	$\widehat{se}(\epsilon^f)$	1.124	0.503
$\widehat{\beta}^d_{\text{1st year}}$	0.166	0.203			
$\widehat{\beta}^d_{\text{lsales}}$	-0.070	-0.100			
$\widehat{se}(\epsilon^d)$	0.453	0.395			

Transition probabilities do not exactly match those in Table 2.5 because of different confidentiality restrictions.

Table 2.11: Data-based and simulated statistics ($\widehat{M}$ and $M_S(\Lambda)$)

2.5 Analysis of Results

2.5.1 Fitting the Moments

Comparing the data-based moments to their simulated counterparts in Table 2.11, one finds the model does a reasonably good job of explaining the patterns we discussed in Section 2.1 above. In particular, the simulated transition probabilities for numbers of clients are close to the data, as are the match death hazards, the relationship between exit rates and cohort age, and the relationship between average exports and cohort age. The model also qualitatively (but less accurately) captures the concentration of exporters at the low end of the client count distribution and the tendency for average sales per client to co-vary positively with number of clients. Finally the model also captures the positive association between domestic and foreign sales.

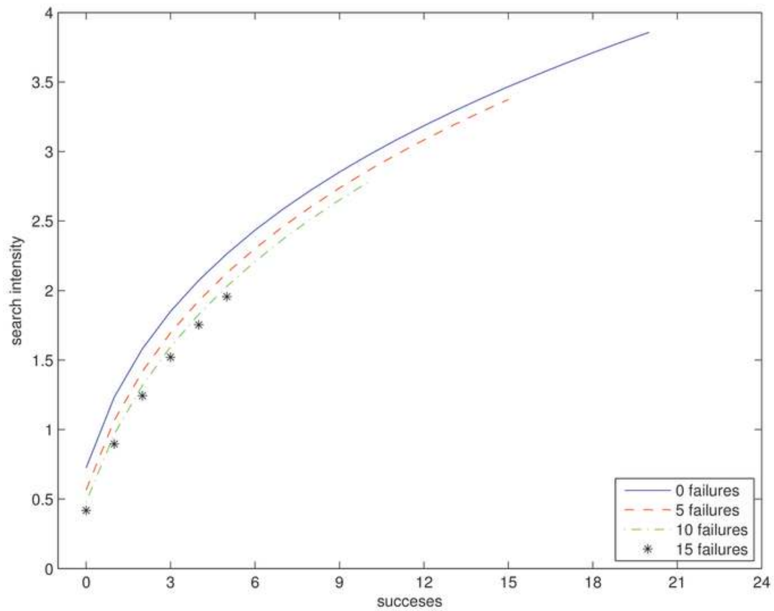
2.5.2 Interpreting the coefficients

Several immediate implications of the coefficient estimates merit note. First, although mature matches fail with probabilities exceeding 40 percent (Table 2.7), we estimate that the exogenous failure rate is only $\delta = 0.27$. Thus idiosyncratic shocks to buyer-seller matches appear to play a significant role in match survival. Second, the fixed per-shipment costs of sustaining a match are roughly $F = \exp(7.957) =$ \$US 2,855, about 70 percent higher than the per shipment costs of regulations by 2005, according to the Doing Business report. Third, the unconditional average success rate with potential U.S. buyers is $\alpha/(\alpha+\beta) \approx 0.184$, so less than one-fifth of the buyers that Colombian exporters meet are interested in establishing a business relationship. Fourth, however, success rates vary across exporters with standard deviation $\sqrt{\alpha\beta/\left[(\alpha+\beta)^2(\alpha+\beta+1)\right]} \approx 0.176$, so some firms have much higher success rates than others, and this creates considerable scope for learning. Fifth, network effects are extremely important. After a successful matches, search costs at any given s have fallen by the factor $(1+a)^{-\gamma(1+1/\kappa_1)}$ relative to the costs faced by a new exporter. Thus, for example, when a seller achieves her first successful match, her search costs for any given arrival hazard drop to 8 percent of their pre-match level, and after three success matches, they drop to 2 percent. Finally, there is considerable convexity in the search cost function $(1+1/\kappa_1 = 12.49)$, so holding the number of successful matches constant, intensifying the search process is very costly. This is how the model explains the fact that 80 percent of exporters have a single client.

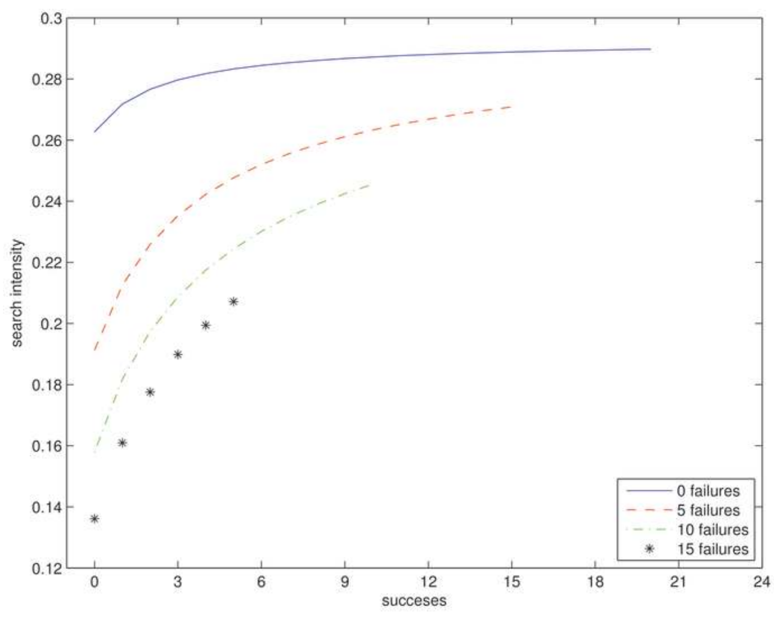
What are the combined implications of these estimates for sellers' search policy? Figure 2.2a below shows search intensity (s^f) as a function of number of successes (a) and failures ($n-a$), taking expectations over marketwide shocks (x) and productivity shocks (φ). For any given number of previous failures, search intensity is increasing in the number of previous successes. This reflects the fact that successes build a network and thus reduce the cost of making future matches. It is also clear that the effect of a successful match has the most dramatic effect on search intensity when firms have little experience. Partly this is due to the fact that early successes contain the most information, and thus move priors relatively more.

2.5.3 Restricted versions of the model

To explore identification of the learning effects and the reputation effects in our model, we consider two alternative specifications. The first, which we call the *no-learning* model, treats firms as knowing their exact θ^f draws, even before they acquire any experience in export markets. This specification involves the same set of parameters, none of which are constrained, so it isn't a nested version of the benchmark model. Rather it replaces one characterization of



(a) baseline



(b) no network

Figure 2.2: Search policy functions by match history

	Parameter	benchmark (Λ)	no learning (Λ^{NL})	no network (Λ^{NN})
rate of exogenous separation	δ	0.267	0.516	0.119
domestic market size	Π^h	11.344	12.670	10.884
foreign market size	Π^f	10.675	12.245	10.321
fixed cost	F	7.957	10.238	8.539
First θ distribution parameter	α	0.716	0.512	1.807
Second θ distribution parameter	β	3.161	0.351	0.963
demand shock jump hazard	λ_y	0.532	0.713	1.581
demand shock jump size	Δ^y	0.087	0.060	0.087
shipment order arrival hazard	λ_b	8.836	10.028	10.347
std. deviation, log firm type	σ_φ	0.650	1.268	1.355
network effect parameter	γ	0.298	0.112	0
search cost function curvature parameter	κ_1	0.087	0.0348	0.057
search cost function scale parameter	κ_0	111.499	234.764	175.953
fit metric	D	9.97 e+04	2.155 e+05	1.17 e+05
fit metric, no weighting	$\widetilde{D}$	0.117	0.182	0.143

Table 2.12: Parameter Estimates for Alternative Models

beliefs with another. The second alternative, which we call the *no-network* model, is nested by the benchmark model. It shuts down reputation effects by imposing $\gamma = 0$, but it retains the benchmark assumption that firms must learn their θ^f draws through experience. Both alternative models are calibrated to the same statistics we use for our benchmark model. The resulting parameter estimates and the associated fit metrics are reported in Table 2.12. Below we discuss the ability of each to fit the data.

2.5.3.1 No Learning

Other things held fixed, the elimination of learning effects makes the rapid turnover of novice exporters less likely, both by discouraging inexperienced low-θ^f firms from exploring foreign markets and by eliminating learning-based exit. Shutting down learning effects also means that high-θ^f firms do not intensify their search efforts as they receive positive feedback about their product appeal.

With these mechanisms inoperative, the no-learning model must use other means to explain the rapid turnover of new exporters and the rapid expansion of sales per surviving exporter as young cohorts mature. To accomplish the former, lower productivity firms are induced to participate in export markets by a rightward shift in the θ^f distribution and higher values for Π^f and λ_b, while match failure rates and market exit rates are sustained by higher values for F, δ, and λ_y (Table 2.12, column 3 versus column 2).* To get sales per exporter growing with cohort age, the no-learning model relies more heavily on selection effects. Low productivity firms are enticed into the market by the bigger Π^f value and the higher average popularity of their products. But these firms tend to end their matches as soon as the fixed costs (F) come due, which–being relatively large–ensures that the surviving exporters have substantially higher sales. The relatively large value of λ_y also helps to generate growth in match sales conditioned on match survival, since buyers who draw negative shocks tend to fail, while matches with positive shocks tend to survive. Finally, the no-learning model facilitates new exporter growth by reducing the convexity of the search cost function, κ_1.

While these parameter adjustments help the no-learning model qualitatively match patterns of exporter turnover and growth, the model's overall fit metric is much worse than that of the benchmark model (Table 2.12, lower panel). The reason is that the no-learning model badly overstates the share of firms that export (Table B.2 in appendix B.2), severely understates the persistence in match-specific sales, given match continuation, overstates the rela-

*Recall that $E(\theta^f) = \alpha/(\alpha+\beta)$ and $var(\theta^j) = \alpha\beta/\left[(\alpha+\beta+1)(\alpha+\beta)^2\right]$.

tionship between sales per client and number of clients, and fails to match the Pareto shape of the cross exporter client distribution

2.5.3.2 No Network Effect

Network effects mean that sellers with a history of successful matches face relatively low search costs, given search intensity. This allows firms with popular products to build larger customer bases than the sharply convex search cost function would have otherwise allowed, and thereby helps the benchmark model match the Pareto distribution of clients across sellers.

To determine the importance of this feature of the model, we set $\gamma = 0$ and re-estimated the remaining parameters, obtaining the no-network estimates reported in Table 2.12. Without network effects, the model moves part way toward matching the Pareto shape by reducing the convexity of the search cost function, κ_1. But this is an imperfect fix because all exporters are equally affected by κ_1, not just the larger ones. Accordingly, various other adjustments occur, including a modest increase in F, a rightward shift in the θ distribution, an increase in the variance of φ, and an increase in the jump hazard for buyer shocks, λ_y. Interestingly, these adjustments are qualitatively similar to those that occurred when we shut down learning effects. Here, however, market sizes Π^f and Π^h shrink a bit rather than expand.

Despite these adjustments, the no-network model does significantly worse than the benchmark model (Table 2.12, bottom panel). In particular, the client distribution is far from Pareto, reflecting the model's inability to explain the existence of very large exporters (Table B.2 in appendix B.2). The no-network model also overstates the fraction of firms that export and the average exports of surviving firms after the first year. Finally, it makes the correlation between domestic and foreign sales far too weak, and the log sales-per-client distribution far too non-linear in the log of the number of clients.

The inability of the no-network model to generate a set of super-exporters can be traced back to the search policy function this model delivers. Figure 2.2b summarizes its properties. Note that learning effects appear to be relatively important for the first several clients, but unlike in figure 2.2b, the policy function quickly flattens out as successes accumulate. So, within the general structure of our search and learning framework, sustained growth in search intensity among relatively established exporters cannot be sustained without network effects. Note also the very different scales between Figures 2.2a and 2.2b, indicating much lower search intensities when the network effect is not present.

2.5.4 Counterfactual experiments

It remains to use our model to explore the export dynamics in a search and learning world with network effects. These experiments will reveal the extent to which learning and network effects create deviations from the export path one would expect in a frictionless setting with the same market-wide shocks and idiosyncratic processes for buyer and seller shocks.

We graph three experiments in Figures 2.3-2.5 below. Each figure has separate panels decomposing aggregate exports into number of exporters, mean per-client exports, and mean number of clients. In Figure 2.3, we reduce the scalar κ_0 in the search cost function by 20% percent. In Figure 2.4, we decrease the fixed cost of maintaining a client relationship F by 20%, and in Figure 2.5, we reduce the size of foreign market jumps Δx_f by 20% percent. For all experiments, the shock takes place in 2002 and is unanticipated and permanent. The red line represents the time path that would have been observed in the absence of the shock, and the dashed blue line reflects the time path induced by the shock. We use the same draws for all stochastic processes, with and without the parameter change, so these changes are the only reason that the blue line differs from the red line after 2002. In all exercises, we take the market-wide demand shifters x_f and x_h from the data.

While the shock takes place in 2002, decreasing the cost of search has no noticeable net effect on exports until 2003. The slow reaction of firms to shocks is a theme in all of our counterfactuals. The decrease in search costs appears to mainly encourage inexperienced firms to search harder. Since exporters start small, and this is reflected in a decrease in mean sales per client, the initial effect on aggregate exports is small. Over time, however, a successful exporter will ramp up her search behavior, so that aggregate exports ultimately grow relative to the baseline.

Exporters also react slowly to the fixed cost reduction in Figure 2.4, and different margins react with different speeds. While the number of active exporters does most of its jumping in 2002, the mean number of clients rises more gradually as it takes all exporters time to acquire the new equilibrium collection of clients.

Somewhat surprisingly, decreasing fixed cost does not cause mean sales per client to drop. Mean sales are affected by two margins. For a particular firm, mean sales per client will decrease as poor clients that would have been let go are allowed to stick around. On the other hand, lowering fixed costs also encourages highly productive firms to search harder. Since the typical match relationship at one of the best firms is highly lucrative, a new match can cause economy-wide mean sales per client to rise. That mean sales per client rise after decreasing fixed costs suggests that productive firms gain more new clients than unproductive firms.

Both a reduction in search costs and a reduction in fixed costs per shipment could be potentially interpreted as policy experiments. For instance, Proexport, the Colombian export promotion agency, has several programs aimed at helping firms find foreign clients. These range from publishing lists of potential buyers in their website to firm-specific studies and trips organized by Proexport (some of which the firm itself pays for). The introduction of this type of programs, or subsidized prices for them could lead to reduced search costs. As for the fixed cost per shipment, regulations may also affect these costs. The World Bank, for instance, estimates that in 2005 the fees associated with procedures to export goods amounted to $1,745 per one-container shipment.

Figure 2.4 shows the results of the experiment where the foreign market size suddenly increases by 20 percent. All matches become more lucrative. This mechanical rise in sales explains the sudden increase in exports and mean sales per client immediately after the shock. The gradual reaction of exports can be seen in the mean number of clients per exporter, which takes almost a decade to fully react to the shock.

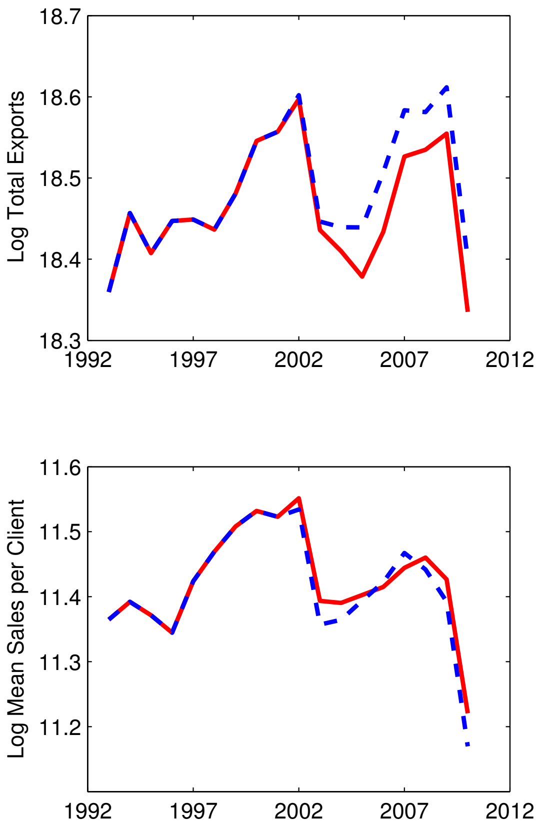


Figure 2.3: Time Series Effects of Search Cost Reduction

2.6 Summary

Customs records reveal tremendous turnover among Colombian manufacturers who export to the U.S.. In a typical year, 48 percent of these exporters are new to the U.S. market, and 81 percent of these new exporters will be gone two years hence. New exporters ship small quantities, so despite their numbers they account for only 12 percent of total Colombian exports in value terms. But each new cohort of Colombian exporters contains a small number of firms that survive and rapidly expand, growing many times faster than aggregate Colombian exports. They do so by adding U.S. clients to their customer base at a rapid rate.

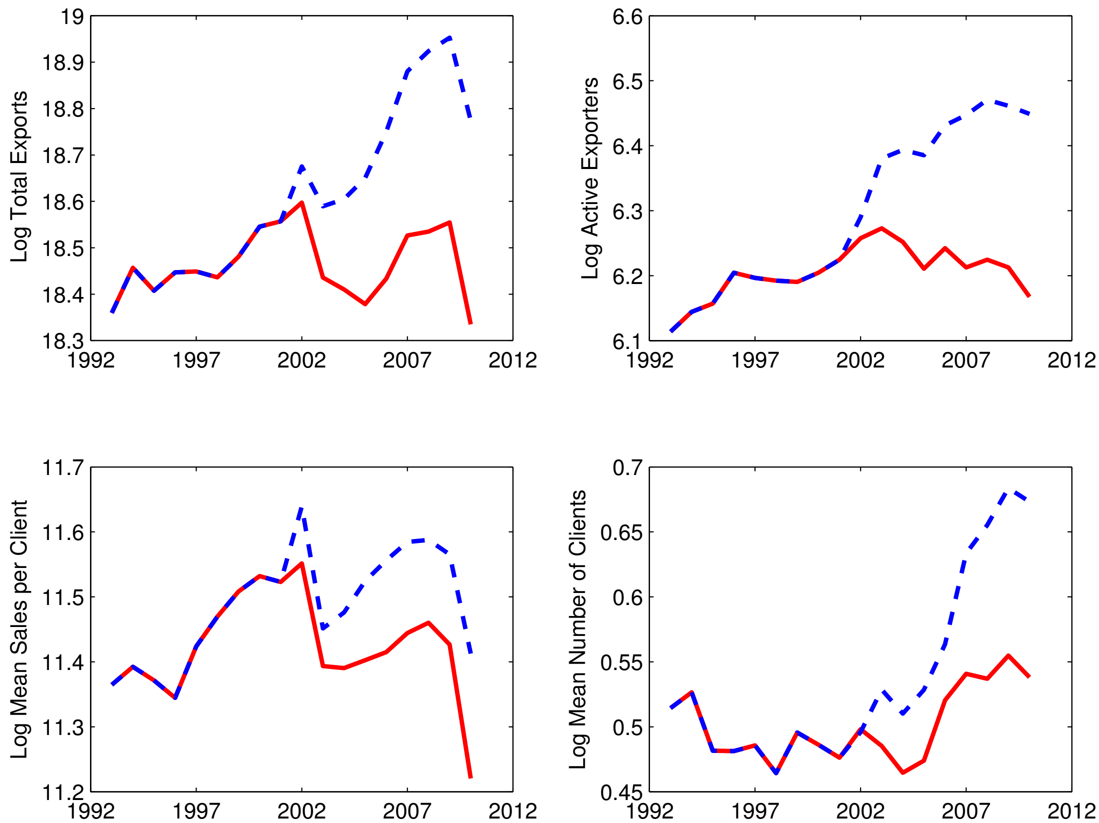


Figure 2.4: Time Series Effects of Fixed Cost Reduction

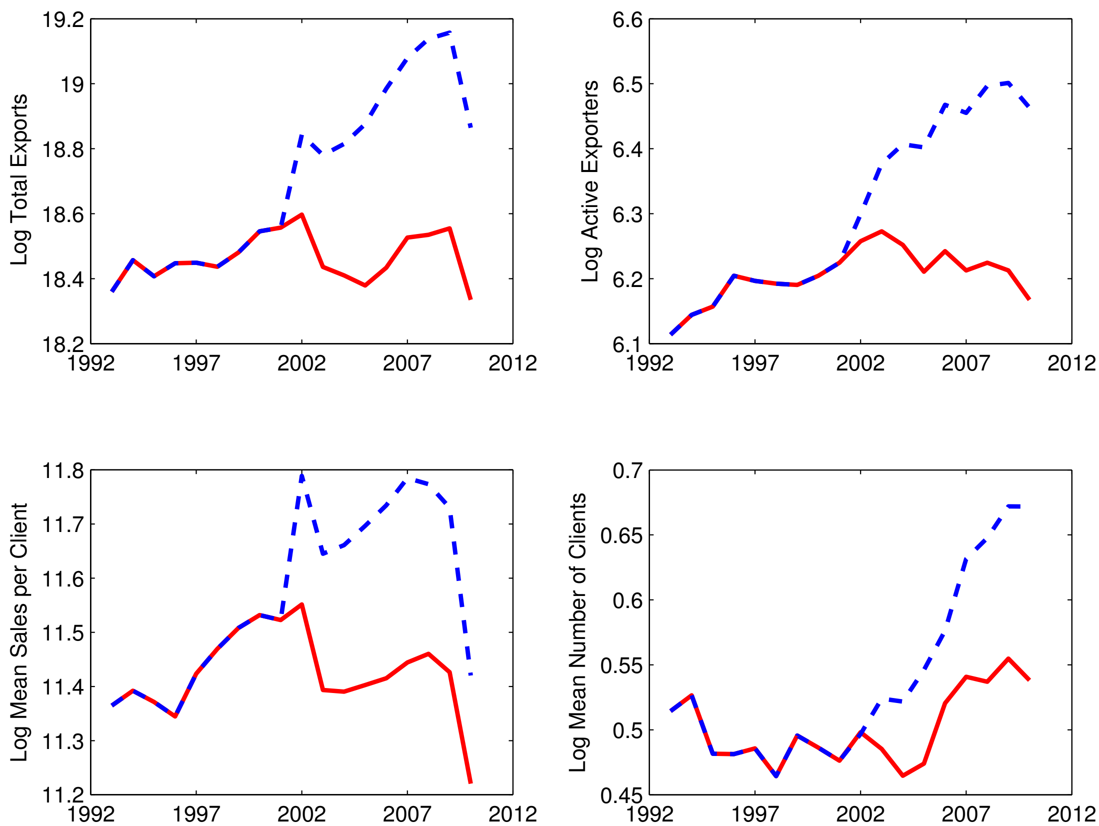


Figure 2.5: Time Series Effects of Positive Market-wide Shock

After documenting these patterns, we develop a continuous time model that explains them. Firms wishing to export must engage in costly search to identify potential buyers abroad. The buyers they encounter either reject their products or form finite-lived business relationships with them. Buyer who form business relationships with exporters send them favorable signals about the appeal of their products, and in doing so, encourage them to search more intensively for additional buyers. Successful business relationships also reduce search costs by improving sellers' visibility (network effects). Finally, sellers' search intensities depend upon their permanent idiosyncratic characteristics and market-wide conditions.

Fit using the method of simulated moments, the model replicates the patterns in customs records described above and allows us quantify several types of trade costs, including the search costs of identifying potential clients and the costs of maintaining business relationships with existing clients. It also allows us to estimate the network effect of previous exporting successes on the costs of meeting new clients, and to characterize the cumulative effects of learning on firms' search intensities. Both the learning effect and the network effect prove to be quantitatively important. Finally, our model provides a lens through which to view the seemingly unpredictable responses of export flows to exchange rate fluctuations.

If I walk with two others,

I must be able to learn something from one of them.

Analects of Confucious, Chapter VI

3

Peer Learning, Labor Mobility, and Knowledge Diffusion

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."[86]

This paper measures the role of 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 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 firms 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 current colleagues who already know about it, and fixed, potentially unobservable characteristics of a worker. The second part of the model is movement between firms. Moving is costly, payoffs also depend on permanent characteristics and unobservables, and the worker 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 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 [67] 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.* In order to measure peer spillover on test scores, Arcidiacono et al. [6] 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.† Sev-

*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.

†Jaffe et al. [56] is the classic citation, see Breschi and Lissoni [18] for a somewhat dated survey. Social networks are also important for knowledge diffusion. Conley and Udry [27] show that social networks in Ghana were important for the diffusion of technology related to pineapple growing.

eral reduced-form papers in this literature find evidence that workers take knowledge with them as they move between firms.[5,74,76‡] 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.[§] I also add to the literature by constructing a new data set tailored to addressing questions about inter-firm movement and knowledge diffusion. While patents have been used to infer the location of workers,[74] 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.[¶] 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.[22] That the spread of innovation is similar to the spread of infectious disease is not a new insight. Ken Arrow

[‡]Almeida and Kogut[5] estimate stronger geographical spillovers in locales with more movement between firms. Oettl and Agrawal[74] 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[76] 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.

[§]In Appendix C.1 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.

[¶]One paper using such a matched patent data set found an average of "a little over one" lifetime patents per inventor.[59] 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.

made such an observation in 1969, for instance.[‖] 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[45] or Rogers[79] similarly studied the empirical diffusion curves of narrowly defined technologies. More recent work by economists has focused on aggregate growth and diffusion models.[21,60,39,64,25]

The worker location choice model I develop builds on recent work by Kennan and Walker[58] 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[28] 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 macroeconomics literature on repeat and return migration.[34,37]

Figure 3.1 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.

[‖]'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

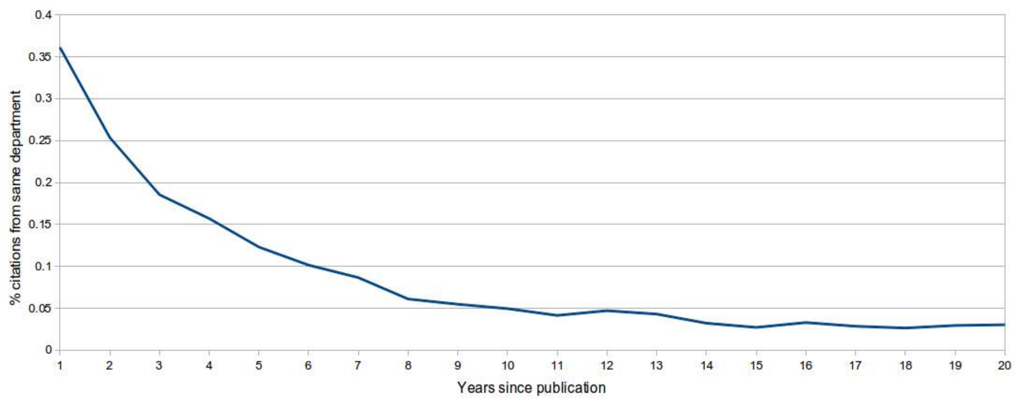


Figure 3.1: Citing paper location sharing over time.

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 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.

3.1 An Empirical Model of Knowledge Diffusion

3.1.1 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.[**]

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:[††]

$$\frac{e^{\alpha+\beta K(d,t-1)+h_i}}{1+e^{\alpha+\beta K(d,t-1)+h_i}} \tag{3.1}$$

$K(d, t-1)$ is the percentage of current colleagues who cited the paper by $t-1$. Equation (3.1) 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.

3.1.2 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 changing

[**] There is an extension in Section 3.6 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.

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

mobility among 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 3.1.1.‡‡

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[58] on interstate migration.* 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:

$$u_{i,t}(d) = W(d, \mathbf{X}_i) + \varepsilon_{i,d,t} \tag{3.2}$$

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:

$$W(d, \mathbf{X}_i) = w_v(\mathbf{X}_i) + w(d, \mathbf{X}_i) \tag{3.3}$$

The first part of the period payoff w_v depends only on personal characteristics like quality

‡‡An alternative reduced-form method to deal with bias is to use an instrumental variables approach. This is done in Section 3.7.

*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[28] 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.

and field. This component is the same at any department. The second part of the payoff w is department 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.[†]:

$$\ln w(d, \mathbf{X}_i) = \xi_0 + \xi_q q_i Q_d + \xi_f f_i F_d + \xi_h h_i F_d \tag{3.4}$$

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.[‡] 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.[§] The set of departments is $\mathcal{D}$. Suppress the permanent characteristic vectors $\mathbf{X}_i$ and write the recursive value function as:

[†]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 3.2.

[‡]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.

[§]In particular, if we add $\frac{\rho}{1-\rho} w_v$ to the left hand side of (3.5), 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.

$$V(d) = \rho \mathbb{E}_\varepsilon \big[\max\{V(d') + w(d') + \varepsilon_{i,d',t} - \mathbb{1}_{\{d' \neq d\}} \, \mathcal{C}\}_{d' \in \mathcal{D}} \big]$$
$$= \rho\gamma_e + \rho \ln \left(\sum_{d' \in \mathcal{D}} e^{V(d') + w(d') - \mathbb{1}_{\{d' \neq d\}} \, \mathcal{C}} \right) \tag{3.5}$$

The substitution of the expectation of the maximum of Type I Extreme Value errors in (3.5) follows Rust[82], and is derived in Appendix C.5. 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 C.4 that the natural operator on (3.5) is a contraction mapping. We can use the value function to get the probability of moving from department d to another department d':

$$Pr(d, d') = \frac{e^{V(d') + w(d') - \mathbb{1}_{\{d' \neq d\}} \mathcal{C}}}{\sum_{d'' \in \mathcal{D}} e^{V(d'') + w(d'') - \mathbb{1}_{\{d'' \neq d\}} \mathcal{C}}} \tag{3.6}$$

3.1.3 Summary

The model can be thought of as consisting of two parts. The learning and citing part is a stochastic process governed by (3.1). Only the observable department citation fraction $K(d, t-1)$ varies over time.¶ The part of the model governing movement between firms is a dynamic discrete choice problem, characterized by the value function (3.5) and the utility function (3.2). Solving the value function results in transition probabilities which depend on fixed observable and unobservable 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

¶See Section 3.6 for an extension in which there is a national knowledge spillover as well.

be discussed further in Section 3.3.1.

3.1.4 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:

$$h_i = \phi_Q Q_i^{(1)} + \phi_F F_i^{(1)} + h_i^*, \quad h_i^* \sim \mathcal{N}(0, \sigma^2) \tag{3.7}$$

Any level effect in (3.7) will be absorbed by α in the learning probability equation (3.1), and the size of parameter ξ_l in period payoff equation (3.4). 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 (3.4) depend on the current location of the academic which may change from year to year.

3.1.5 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 (3.7). 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:

$$
\begin{aligned}
Pr(C_i, \mathbf{M}_i | \mathbf{X}_{o,i}, \mathbf{Z}, \theta) &= \int Pr(C_i, \mathbf{M}_i | h_i, \mathbf{X}_{o,i}, \mathbf{Z}, \theta) dH \\
&= \int Pr(C_i, \mathbf{M}_i | \mathbf{X}_i, \mathbf{Z}, \theta) dH
\end{aligned} \tag{3.8}
$$

We can split the integrand in (3.8) into multiplicative terms:

$$
Pr(C_i, \mathbf{M}_i | \mathbf{X}_i, \mathbf{Z}, \theta) = Pr(C_i | \mathbf{X}_i, \mathbf{M}_i, \theta) Pr(\mathbf{M}_i | \mathbf{X}_i, \mathbf{Z}, \theta) \tag{3.9}
$$

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

$$
Pr(\mathbf{M}_i | \mathbf{X}_i, \mathbf{Z}, \theta) = \prod_{t=1}^{T-1} Pr(M_{i,t}, M_{i,t+1} | \mathbf{X}_i, \mathbf{Z}, \theta) \tag{3.10}
$$

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

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

$$Pr(C_i|\mathbf{X}_i, \mathbf{M}_i, \theta) = \left[(1-\gamma_{f_i}) + \gamma_{f_i} \prod_{t=1}^{T} \left(1 - 1_{\{M_{i,t}\neq\emptyset\}} \frac{e^{\alpha+\beta K(M_{i,t},t-1)+h_i}}{1+e^{\alpha+\beta K(M_{i,t},t-1)+h_i}} \right) \right]^{1_{\{C_i=\emptyset\}}}$$
$$\times \left[\gamma_{f_i} \prod_{t=0}^{C_i-1} \left(1 - 1_{\{M_{i,t}\neq\emptyset\}} \frac{e^{\alpha+\beta K(M_{i,t},t-1)+h_i}}{1+e^{\alpha+\beta K(M_{i,t},t-1)+h_i}} \right) \frac{e^{\alpha+\beta K(M_{i,C_i},C_i-1)+h_i}}{1+e^{\alpha+\beta K(M_{i,C_i},C_i-1)+h_i}} \right]^{1_{\{C_i\neq\emptyset\}}} \tag{3.11}$$

The two big multiplied terms in (3.11) 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:

$$Pr(\mathbf{C}, \mathbf{M}|\mathbf{X}_o, \mathbf{Z}, \theta) = \prod_i Pr(C_i, \mathbf{M}_i|\mathbf{X}_{o,i}, \mathbf{Z}, \theta) \tag{3.12}$$

3.2 Data Description

3.2.1 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 3.2.2 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. Recent independent research by Agrawal et al.[1] constructs a similar panel of academics also from Thomson-Reuters Web of Knowledge, but uses evolutionary biologists rather than economists.

I describe how the data was cleaned and filtered in Appendix C.3. 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.** This is a large, rich data set, containing thousands of economists, hundreds of departments from around the world, and more than one hundred thousand papers and citation records.

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

**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.

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 et al.[12]. 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.3.

3.2.2 Baseline Structural Model

In this section I describe the construction of variables specific to the estimation of the structural model described in Section 3.1. 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. If we were to include many papers in the model, we would need to let the potential interest parameters γ_0 and γ_1 vary across papers, leading to a large parameter space and imprecise estimates.[††]

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[50] and White[89]. 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 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.[‡‡] 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

[††] As a robustness check, I reestimate the entire model using Grossman and Hart[47], another influential paper published in 1986. In Appendix C.9 I present a comparison of the reestimated results to the baseline results, and find essentially no difference.

[‡‡] 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.

$Q_d \in [0,1]$.

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 3.1 and 3.2 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 3.1: 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 3.2: Department summary statistics

3.3 Identification and Estimation Routine

3.3.1 Identification and Causality

The main parameter of interest is β in (3.1), 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 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.[15,6,20*]

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 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.[†] The estimated

[*] 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[49] for a recent example). In this context, Rothstein[80] finds evidence that student fixed effects are not sufficient to control for endogenous sorting.

[†] 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.[57,31]

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 (3.1). 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 (3.1), 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 C is identified by the frequency of moves.

3.3.2 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.[88,72] In several of the hardest hit states, public universities implemented hiring freezes for various lengths of time.[61,71,30] I use data on state budget deficits in fiscal year 1991 to proxy for temporary, unanticipated hiring reductions at public universities in 1991.[42]

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 3.3 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.

	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 3.3: The net in-move effect of state budget shortfalls

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:

$$w_{1991}(d, \mathbf{X}_i) = e^{-\xi_{ex} b_d p_d} w(d, \mathbf{X}_i) \tag{3.13}$$

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 (3.10) are given by (3.14) rather than the original transition probabilities (3.6).[‡]

$$Pr_{1991}(d, d') = \frac{e^{V(d') + w_{1991}(d') - \mathbb{1}_{\{d' \neq d\}} \mathcal{C}}}{\sum_{d'' \in \mathcal{D}} e^{V(d'') + w_{1991}(d'') - \mathbb{1}_{\{d'' \neq d\}} \mathcal{C}}} \tag{3.14}$$

3.3.3 Implementation

I estimate the twelve parameters in the likelihood function (3.12) using Bayesian inference and Markov Chain Monte Carlo (MCMC). Description of the priors are contained in Table 3.4.

[‡]See Section 3.7 for an alternative reduced-form analysis using 1991 budget deficits as an instrument.

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.[§] 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.

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

[§]The exponential prior has been used for a similar purpose in the epidemiology literature.[22] 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 C.11 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 3.4: Priors

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 et al. [48]. 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. [48] 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 C.6 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.[41] 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.

3.4 Analysis of Results

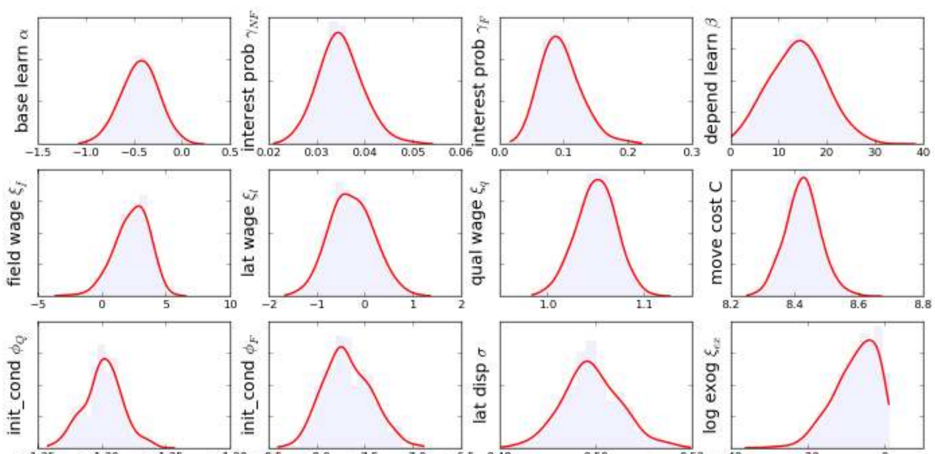


Figure 3.2: Posterior distributions

	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 3.5: Posterior moments

Posterior moments for the twelve estimated parameters are listed in Table 3.5, and parameter posterior distribution kernel densities and histograms are plotted in Figure 3.2. The first row of Figure 3.2 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 3.3 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 3.3 shows typical latent types in the data, which are below zero because of the initial condition equation described in Section 3.1.4. For the latent types in the data, an increase from 0% to 5% coworker knowledge raises annual learning probabilities by 35-60%.

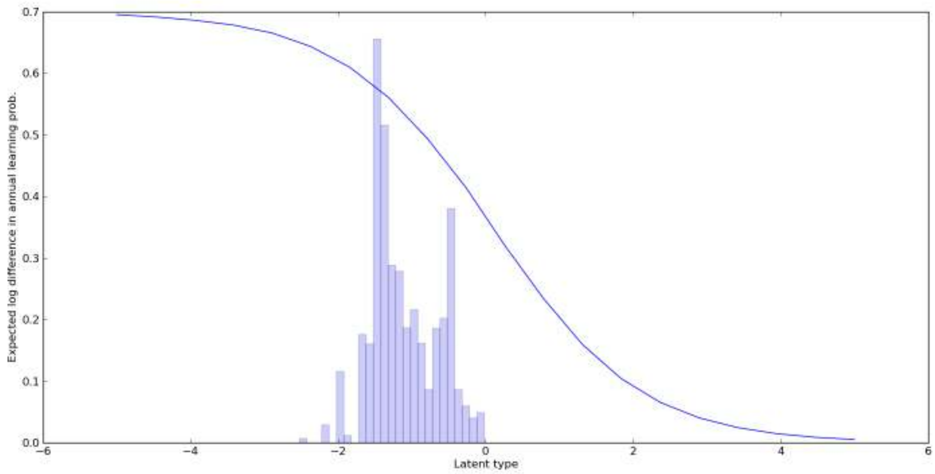


Figure 3.3: Annual learning probability percent increase, 0% to 5% coworker knowledge of new paper

The bottom two rows of Figure 3.2 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 $e^{-0.919\times 0.1} \approx 9\%$.[¶]

[¶]The entry and exit processes of academics is not modeled here, and the size of economics departments has been 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 C.8 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.

The latent type distribution parameter posteriors are located in the bottom row of Figure 3.2. 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.

3.4.1 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:

$$\lambda(t) = \alpha + \beta S(t) \tag{3.15}$$

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:

$$\frac{dS(t)}{dt} = (\alpha + \beta S(t))\,(1 - S(t)) \tag{3.16}$$

Solving for S gives:

$$S(t) = \frac{\alpha e^{(\alpha+\beta)t} - \alpha}{\alpha e^{(\alpha+\beta)t} + \beta} \tag{3.17}$$

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 3.4. 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.

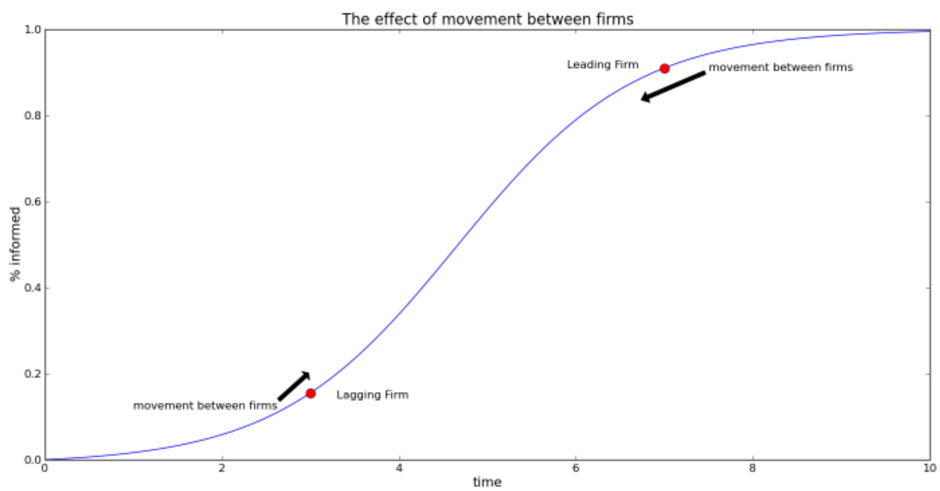


Figure 3.4: Movement between firms on a diffusion curve

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.[‖] This is not so. From (3.16), the effect of an additional worker becoming informed on the rate of diffusion is given by:

$$\frac{dS'(t)}{dS(t)} = -\left(\alpha + \beta S(t)\right) + \beta\left(1 - S(t)\right) \tag{3.18}$$

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:

$$-\left(\alpha + \beta s_l\right) + \left(\alpha + \beta s_h\right) + \beta\left(1 - s_l\right) - \beta\left(1 - s_h\right) = 2\beta(s_h - s_l) \tag{3.19}$$

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

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

[**] If $\beta = 0$ so that learning does not depend on coworkers, then (3.17) reduces to the exponential distribution. 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:

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.

3.4.2 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 aca-

---

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

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.

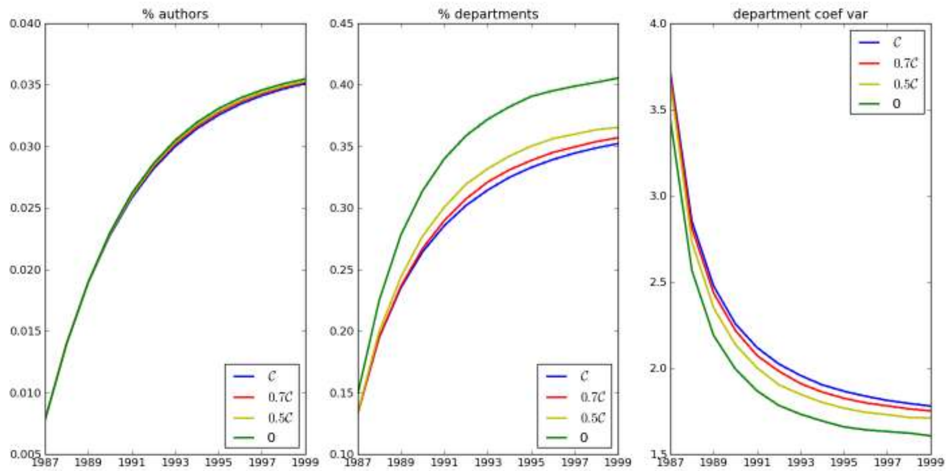


Figure 3.5: Counterfactual statistics, posterior expectations

demics 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 3.5 plots the expected evolution of those three statistics in the different scenarios. Figure 3.6 plots the posterior expectation of the log difference between counterfactual statistics and simulated data statistics. The bottom row of Figure 3.6 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. 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,

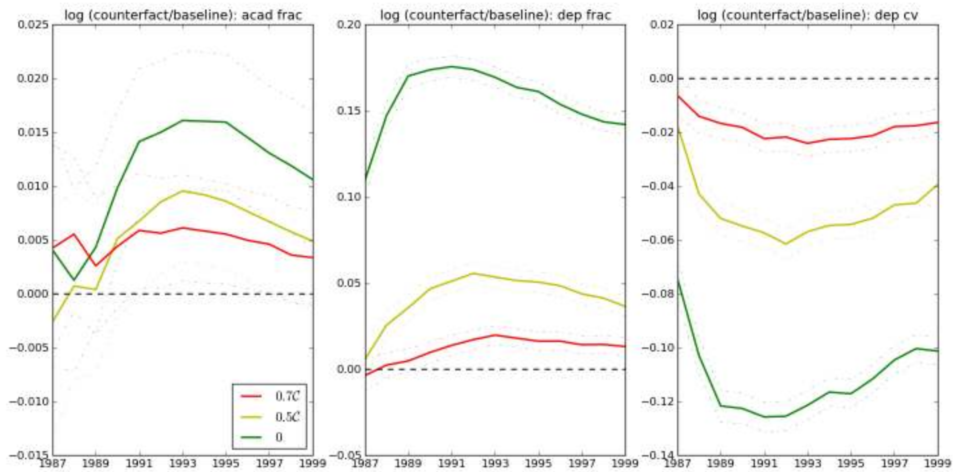


Figure 3.6: log change in posterior expectations

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.

3.5 Extension to 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.[††] 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 (3.1), but with different parameter values than in the United States:

$$\frac{e^{\alpha_c+\beta_c K(d,t-1)+h_i}}{1+e^{\alpha_c+\beta_c K(d,t-1)+h_i}} \tag{3.20}$$

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

[††] I am referring to the 221 Program. For more information on this program see Lixu[63]. I found the list of universities on Wikipedia.

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.[‡‡]

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 (3.1). 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 3.7 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

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

$$\frac{1}{9} = \alpha_c * \gamma_c * |A_c| \tag{3.21}$$

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

$$\frac{e^{\alpha+\beta 0.05+\bar{h}}}{1+e^{\alpha+\beta 0.05+\bar{h}}} - \frac{e^{\alpha+\bar{h}}}{1+e^{\alpha+\bar{h}}} = \frac{e^{\alpha_c+\beta_c 0.05+\bar{h}}}{1+e^{\alpha_c+\beta_c 0.05+\bar{h}}} - \frac{e^{\alpha_c+\bar{h}}}{1+e^{\alpha_c+\bar{h}}} \tag{3.22}$$

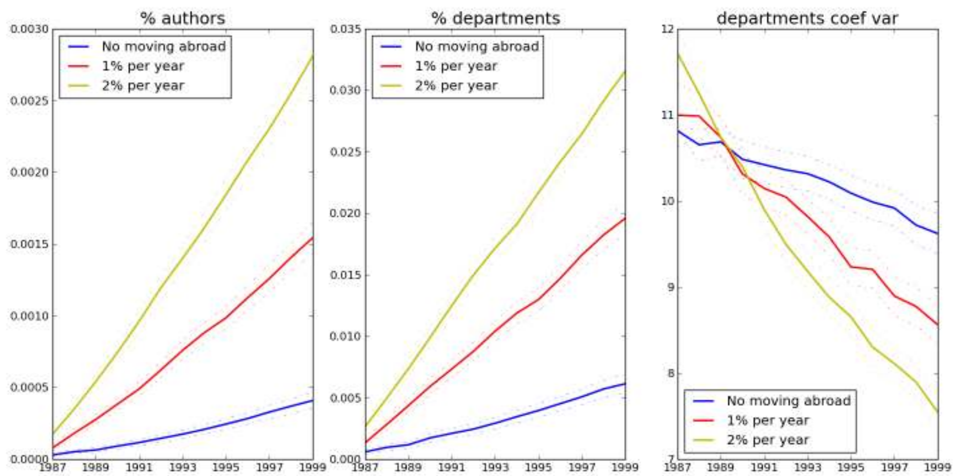


Figure 3.7: Expected Chinese knowledge diffusion

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 direction for future research. Past research has pointed to large welfare gains from reduction in barriers to migration.[26] 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.

3.6 Alternative Model Specifications

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 (3.1) 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 (3.23), $K(n, t-1)$ is the aggregate percentage of academics who cited the paper by $t-1$.

$$\frac{e^{\alpha+\beta K(d,t-1)+\beta_n K(n,t-1)+h_i}}{1+e^{\alpha+\beta K(d,t-1)+\beta_n K(n,t-1)+h_i}} \tag{3.23}$$

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 (3.1) all with f subscripts as in (3.24).

$$\frac{e^{\alpha_f+\beta_f K(d,t-1)+h_i}}{1+e^{\alpha_f+\beta_f K(d,t-1)+h_i}} \tag{3.24}$$

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 3.6

; 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 3.6: Posterior expectations, extensions versus baseline

compares the expectations of posteriors for the baseline model and extensions.* In the all extensions, the movement related parameters at the bottom of Table 3.6 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.

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

*Estimated posterior kernel densities for all extensions can be found in Appendix C.10.

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.

3.6.1 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 3.8 contains posterior means for the three model scenarios and data. When interpreting 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%

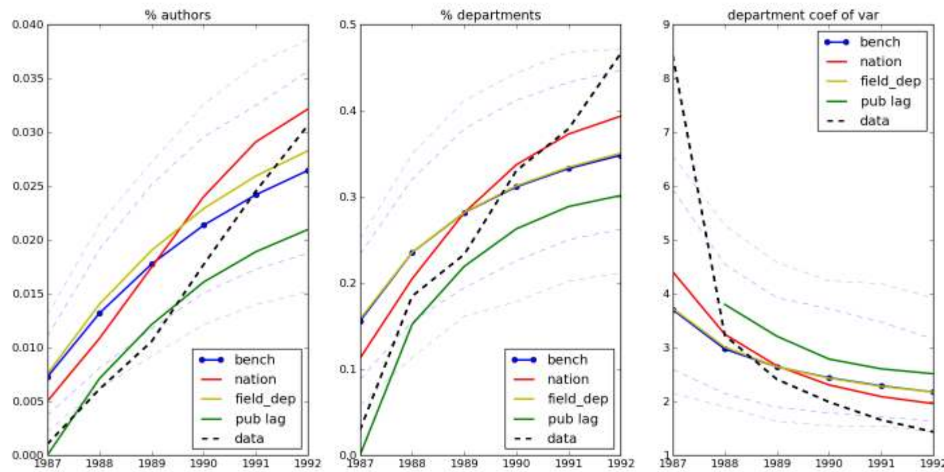


Figure 3.8: Model checking, simulations vs data

and 99% credible intervals of the baseline model are included as faint lines in Figure 3.8. 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.

3.7 Instrumented 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 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:

$$C_i^{1992} = 1_{\{\beta_X \mathbf{X}_{o,i} + \beta_K K_i + \varepsilon_i > 0\}} \tag{3.25}$$

And:

$$K_i = \Gamma_X \mathbf{X}_{o,i} + \Gamma_{b_d} b_{d,i} + v_i \tag{3.26}$$

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 3.7. 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[84] 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 C.1 contains two additional reduced-form exercises which provide evidence on the importance of location on knowledge diffusion.

3.8 Summary

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[7] 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.

	cit92	cit92	cit92	cit92_93	cit92_93	cit92_93
	b/se	b/se	b/se	b/se	b/se	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 3.7: Probit model

A

Chapter 1

A.1 Vindex Tables

Category	Vindex	SE
cigarettes	0.76	(0.014)
cars	0.72	(0.012)
clothing	0.70	(0.013)
furniture	0.68	(0.012)
jewelry	0.67	(0.015)
recreation 1	0.66	(0.012)
food out	0.61	(0.012)
alcohol home	0.60	(0.014)
barbers etc	0.60	(0.014)
alcohol out	0.59	(0.014)
recreation 2	0.57	(0.013)
books etc	0.57	(0.013)
education	0.56	(0.014)
food home	0.51	(0.014)
rent/home	0.49	(0.016)
cell phone	0.46	(0.016)
air travel	0.46	(0.014)
hotels etc	0.45	(0.013)
public trans	0.44	(0.015)
car repair	0.42	(0.014)
gasoline	0.39	(0.016)
health care	0.36	(0.014)
charities	0.34	(0.014)
laundry	0.33	(0.015)
home utilities	0.31	(0.015)
home phone	0.29	(0.015)
legal fees	0.26	(0.013)
car insur	0.22	(0.014)
home insur	0.16	(0.012)
life insur	0.16	(0.011)
underwear	0.12	(0.011)

Table A.1: Aggregate Vindex

	Interviewee age under 40				Interviewee age over 40			
	NEast	South	MWest	West	NEast	South	Mwest	West
Air	3.2	3.0	3.4	2.9	3.2	3.2	3.7	3.2
AlH	4.4	4.5	4.6	4.3	3.8	4.2	4.0	4.6
AlO	4.5	4.3	4.5	4.6	4.2	3.9	4.2	4.2
Bks	4.3	3.8	3.9	4.3	4.0	3.9	4.0	4.2
Brb	4.8	4.1	4.5	4.1	3.8	4.2	4.5	4.1
Bus	3.2	3.5	3.1	2.7	3.4	3.0	3.0	3.0
CIn	1.5	1.8	1.6	1.4	1.4	1.7	1.4	1.1
CMn	2.2	3.0	2.5	3.1	3.2	3.0	2.7	3.4
Car	4.8	5.2	4.9	4.9	5.1	5.2	5.5	5.0
Cha	2.5	2.5	2.7	3.0	2.4	2.3	2.3	2.0
Cig	5.3	5.0	5.3	5.5	5.4	5.4	5.6	5.7
Clo	5.3	5.1	5.3	5.8	4.9	4.7	4.9	4.8
Edu	4.0	3.9	3.8	3.7	4.1	4.0	4.2	3.8
FdH	3.4	3.8	3.3	3.8	3.9	3.6	3.3	3.7
FdO	4.7	4.3	4.6	4.8	4.1	4.1	4.5	4.2
Fee	1.7	1.8	1.7	1.3	2.0	1.9	2.0	1.9
Fur	4.2	4.9	5.0	4.9	5.0	4.9	4.7	4.8
Gas	2.4	2.4	2.4	2.4	2.9	3.0	2.8	2.9
HIn	1.3	1.2	1.1	0.6	1.3	1.4	1.0	1.0
Hom	3.7	3.8	3.3	3.7	3.7	3.4	3.3	3.4
Htl	3.6	3.2	3.5	2.9	3.6	3.2	3.0	3.0
Jwl	4.7	4.5	5.0	5.0	4.7	4.5	5.1	5.0
LIn	1.0	1.5	1.0	0.9	1.2	1.2	0.8	1.0
Lry	2.4	2.3	2.5	2.6	1.9	2.6	2.4	2.1
Med	1.7	2.4	2.9	2.3	2.7	2.8	2.4	2.8
Ot	4.8	4.7	4.8	5.0	4.6	4.3	5.0	4.8
Ot	4.1	4.2	3.8	4.3	4.3	4.1	3.9	4.1
Tel	2.1	1.8	2.0	2.2	1.9	2.4	2.2	1.7
Utl	2.5	1.9	2.0	1.6	2.0	2.4	2.1	2.7

Table A.2: Observation type probabilities by demographic category

A.2 Detailed Results

Good Cat	μ	std err	σ	std err	ψ	std err	z	std err
FdH	3.98	(0.011)	0.22	(0.002)	0.44	(0.003)	0.00	(0.000)
FdO	-0.48	(0.025)	0.82	(0.007)	-0.42	(0.006)	0.06	(0.001)
Cig	0.92	(0.020)	0.38	(0.003)	0.22	(0.005)	0.64	(0.001)
AlH	0.94	(0.016)	0.68	(0.006)	0.37	(0.005)	0.47	(0.002)
AlO	1.05	(0.026)	1.19	(0.007)	0.48	(0.008)	0.46	(0.002)
Clo	-0.81	(0.027)	1.01	(0.011)	-0.42	(0.006)	0.05	(0.000)
Lry	0.79	(0.031)	1.24	(0.010)	0.47	(0.009)	0.31	(0.002)
Jwl	0.61	(0.021)	0.90	(0.008)	0.32	(0.006)	0.57	(0.002)
Brb	0.07	(0.020)	0.64	(0.006)	0.11	(0.005)	0.09	(0.001)
Hom	4.17	(0.011)	0.19	(0.001)	0.23	(0.003)	0.00	(0.000)
Htl	0.09	(0.019)	0.60	(0.010)	0.06	(0.006)	0.52	(0.002)
Fur	-0.87	(0.032)	1.45	(0.015)	-0.29	(0.009)	0.17	(0.001)
Utl	2.50	(0.020)	0.31	(0.002)	0.27	(0.005)	0.04	(0.001)
Tel	2.12	(0.024)	0.45	(0.006)	0.37	(0.006)	0.01	(0.000)
HIn	-0.61	(0.032)	1.18	(0.008)	-0.22	(0.008)	0.19	(0.001)
Med	2.03	(0.030)	1.35	(0.014)	0.16	(0.008)	0.05	(0.001)
Fee	0.13	(0.027)	1.25	(0.012)	0.15	(0.007)	0.25	(0.002)
LIn	0.38	(0.023)	0.73	(0.006)	0.06	(0.006)	0.45	(0.001)
Car	-2.31	(0.028)	1.06	(0.008)	-0.86	(0.008)	0.76	(0.001)
CMn	-0.45	(0.023)	1.40	(0.012)	-0.23	(0.006)	0.13	(0.001)
Gas	0.92	(0.024)	0.53	(0.005)	-0.04	(0.006)	0.07	(0.001)
CIn	0.62	(0.018)	0.44	(0.005)	-0.02	(0.005)	0.22	(0.001)
Bus	0.78	(0.025)	0.99	(0.008)	0.33	(0.008)	0.63	(0.001)
Air	0.02	(0.014)	0.41	(0.008)	0.00	(0.004)	0.67	(0.002)
Bks	-0.75	(0.026)	0.89	(0.008)	-0.16	(0.007)	0.07	(0.000)
Ot1	-0.27	(0.027)	1.36	(0.012)	-0.04	(0.007)	0.29	(0.001)
Ot2	-0.72	(0.034)	0.89	(0.009)	-0.40	(0.009)	0.07	(0.001)
Edu	-0.21	(0.017)	0.86	(0.009)	-0.06	(0.005)	0.70	(0.002)
Cha	-0.06	(0.031)	1.35	(0.011)	-0.04	(0.009)	0.41	(0.001)
α	0.027	(0.000)						

Table A.3: US Parameter Estimates

Good Cat	μ	std err	σ	std err	ψ	std err	z	std err
Fdh/Fdo	3.79	(0.111)	0.13	(0.889)	0.01	(0.007)	0.00	(0.007)
Alh/Alo	0.70	(0.111)	1.08	(0.889)	0.22	(0.007)	0.47	(0.007)
Cig	0.08	(0.077)	3.72	(0.571)	0.42	(0.004)	0.10	(0.004)
Bks	2.02	(0.013)	0.72	(0.017)	-0.04	(0.001)	0.01	(0.001)
Edu	0.54	(0.022)	1.34	(0.050)	-0.22	(0.002)	0.03	(0.002)
Bus/Car	1.38	(0.020)	1.62	(0.040)	0.09	(0.002)	0.03	(0.002)
Utl	1.02	(0.071)	2.06	(0.488)	0.11	(0.003)	0.07	(0.003)
Tel	-0.50	(0.022)	1.44	(0.046)	-0.13	(0.005)	0.17	(0.005)
Clo/Jwl	1.27	(0.117)	1.79	(1.142)	0.37	(0.006)	0.30	(0.006)
Ot1/Ot2	0.98	(0.021)	1.32	(0.038)	-0.06	(0.006)	0.25	(0.006)
Fur/Lry/Bks	-0.72	(0.035)	0.79	(0.085)	-0.16	(0.007)	0.55	(0.007)
Med/Lin	-0.08	(0.062)	1.87	(0.267)	0.15	(0.007)	0.51	(0.007)
Hom/Htl	2.10	(0.011)	0.59	(0.012)	0.27	(0.001)	0.01	(0.001)
Fee/Cha	1.61	(0.018)	1.41	(0.032)	0.05	(0.001)	0.01	(0.001)
α	0.2618	(0.000)						

Table A.4: Chinese Parameter Estimates

B
Chapter 2

B.1 Data Checks

To investigate the quality of the exporter id (manuf_id) in the U.S. import records, we ran a series of robustness checks. The Colombian and U.S. data overlap for the years 2000-2008 and both contain measures of the value of exports as well as the number of exporting firms. If the manuf_id variable is error-prone and noisy, we would expect the U.S. data to over-report the number of Colombian firms exporting to the U.S. That is, each time a customs broker wrongly enters the data in the field, a new firm would be created. Table B.1 below summarizes the total value of exports to the U.S. and the number of Colombian firms, by year, for each data set.

The datasets align much more closely on value than they do on firm counts. The difference in value is never more than 10% while the firm count difference ranges from 18% to 74%. The

	Colombia		United States		% difference	
Year	# exporters	value	# exporters	value	# exporters	value
2000	1775	1038	2721	1140	53%	10%
2001	2026	995	2744	1019	35%	2%
2002	2230	870	2986	855	34%	-2%
2003	2800	1113	3579	1119	28%	1%
2004	3035	1379	4002	1415	32%	3%
2005	2861	1554	4288	1438	50%	-7%
2006	2689	1665	4361	1552	62%	-7%
2007	2420	1540	4175	1496	73%	-3%
2008	2161	1570	3758	1474	74%	-6%

Table B.1: Colombian versus U.S. Customs Records

differences are stable over time.

To look more closely at the cause of the difference in firm counts, we compared the number of firms across sources by HS2 categories. The counts in the LFTTD were higher than the Colombian data in only 28 of the 82 codes and by far the biggest differences are in HS codes 61 and 62: textiles. In these two product classes the U.S. data identifies 4025 more firms than the Colombian data. If we remove these two sectors from the list, the difference in firm counts flips and the Colombian data contain 1001 more firms than the LFTTD.

Interestingly, Title 19 of U.S. code specifically requires that the manuf_id variable for textile products represent the manufacturer of the textile products, not an intermediary. That is, for this sector in particular the manufacturer, not an intermediary must be reported on the CBP 7501 form. By contrast, prior work by several authors of this paper has shown (Marcela's 8/9/13 e-mail referenced this) that the Colombian data reports the exporter, which may or may not be the manufacturer. Given that revious research (Tybout, 2000 JEL) has shown that developing countries tend to have a disproportionately large share of small manufacturing firms, it is reasonable to assume that a large part of the reason why the U.S. data report so many more firms in the textile sector is that due to administrative reasons the U.S. data count many small manufacturers and the Colombian data are, in many cases, reporting aggregators and intermediaries.

As a final check of the integrity of the manuf_id variable - and the robustness of our main results - we experimented with a "fuzzy" version of the manuf_id variable that did not contain any street numbers in the string (a likely source of input errors). The effect of this is to reduce the number of Colombian firms in the data, an approximation of fixing any extraneous noise from data entry errors. Next we re-ran Table 2.7 with the fuzzy data and compared the results to the original version.

One of the key findings from Table 2.7 is the high match separation rates ranging from about 40% to 66%. Using the fuzzy version did not reduce the separation rates substantially and left the patterns intact. The fuzzy separation rates ranged from 26% to 62%, a drop of 6% on average. It does not appear that our results are sensitive to a modest reduction in data entry errors.

B.2 Moments for Restricted Models

Table B.2: Restricted versus Full Model Fit

	data $\widehat{M}$	*benchmark* $M_s(\Lambda)$	*no learning* $M_s(\Lambda^{NL})$	*no network* $M_s(\Lambda^{NN})$
Share of firms exporting				
$\widehat{E}(1_{X^f_{jt}>0})$	0.299	0.351	0.585	0.451
Log foreign sales on log domestic sales				
$\widehat{\beta}^{hf}_1$	0.727	0.515	0.923	0.575
$s\widehat{e}(\epsilon^{hf})$	2.167	1.424	0.843	1.146
log dom. sales autoreg.				
$\widehat{\beta}^h_1$	0.976	0.896	0.969	0.898
$s\widehat{e}(\epsilon^h)$	0.462	0.683	0.661	0.570
exporter exit hazards				
$\widehat{E}[1_{X^f_{jt}=0}\|A^c_{jt-1}=0]$	0.709	0.748	0.773	0.877
$\widehat{E}[1_{X^f_{jt}=0}\|A^c_{jt-1}=1]$	0.383	0.099	0.099	0.188
$\widehat{E}[1_{X^f_{jt}=0}\|A^c_{jt-1}=2]$	0.300	0.121	0.032	0.012
$\widehat{E}[1_{X^f_{jt}=0}\|A^c_{jt-1}=3]$	0.263	0.055	0.056	0.198
$\widehat{E}[1_{X^f_{jt}=0}\|A^c_{jt-1}=4]$	0.293	0.100	0.098	0.185
log sales per exporter by cohort age				
$\widehat{E}(\ln X^f_{jt}\|A^c_{jt}=0)$	8.960	9.306	9.608	8.541
$\widehat{E}(\ln X^f_{jt}\|A^c_{jt}=1)$	10.018	10.806	10.615	11.331
$\widehat{E}(\ln X^f_{jt}\|A^c_{jt}=2)$	10.231	10.755	10.431	11.037
$\widehat{E}(\ln X^f_{jt}\|A^c_{jt}=3)$	10.369	10.679	10.426	10.845
$\widehat{E}(\ln X^f_{jt}\|A^c_{jt}\geq 4)$	10.473	10.669	10.332	11.145
Log match sale autoregression				
$\widehat{\beta}^f_1$	0.811	0.613	0.105	0.268
$\beta^f_{\text{1st year}}$	0.233	0.370	0.056	0.087
$s\widehat{e}(\epsilon^f)$	1.124	0.503	0.287	0.425
Match death hazards				

$\widehat{E}[1_{X^f_{ijt}=0} \mid X^f_{ijt-1} > 0, A^m_{ijt-1} = 0]$	0.694	0.857	0.943	0.879
$\widehat{E}[1_{X^f_{ijt}=0} \mid X^f_{ijt-1} > 0, A^m_{ijt-1} = 1]$	0.515	0.329	0.452	0.337
$\widehat{E}[1_{X^f_{ijt}=0} \mid X^f_{ijt-1} > 0, A^m_{ijt-1} = 2]$	0.450	0.304	0.426	0.286
$\widehat{E}[1_{X^f_{ijt}=0} \mid X^f_{ijt-1} > 0, A^m_{ijt-1} = 3]$	0.424	0.281	0.434	0.332
$\widehat{E}[1_{X^f_{ijt}=0} \mid X^f_{ijt-1} > 0, A^m_{ijt-1} = 4]$	0.389	0.305	0.398	0.226
Match death prob regression				
$\widehat{\beta}^d_0$	1.174	1.640	1.843	2.087
$\widehat{\beta}^d_{\text{1st year}}$	0.166	0.203	0.031	0.055
$\widehat{\beta}^d_{\text{lsales}}$	-0.070	-0.100	-0.092	-0.140
$\widehat{se}(\epsilon^d)$	0.453	0.395	0.266	0.343
Match shipments per year				
$\widehat{E}\,(n^s)$	4.824	3.770	2.064	4.525
Transition probabilities, No. clients (n^c)				
$\widehat{P}[n^c_{jt+1} = 0 \mid n^c_{jt} = 1]$	0.618	0.534	0.677	0.643
$\widehat{P}[n^c_{jt+1} = 1 \mid n^c_{jt} = 1]$	0.321	0.358	0.255	0.307
$\widehat{P}[n^c_{jt+1} = 2 \mid n^c_{jt} = 1]$	0.048	0.082	0.056	0.045
$\widehat{P}[n^c_{jt+1} \geq 3 \mid n^c_{jt} = 1]$	0.013	0.024	0.010	0.004
$\widehat{P}[n^c_{jt+1} = 0 \mid n^c_{jt} = 2]$	0.271	0.260	0.456	0.165
$\widehat{P}[n^c_{jt+1} = 1 \mid n^c_{jt} = 2]$	0.375	0.321	0.291	0.306
$\widehat{P}[n^c_{jt+1} = 2 \mid n^c_{jt} = 2]$	0.241	0.281	0.166	0.427
$\widehat{P}[n^c_{jt+1} \geq 3 \mid n^c_{jt} = 2]$	0.113	0.135	0.086	0.100
Log sales per client on client no. regression				
$\widehat{\beta}^m_1$	2.677	0.842	0.944	3.887
$\widehat{\beta}^m_2$	-0.143	0.042	1.049	-1.451
$\widehat{se}(\epsilon^m)$	2.180	1.622	1.893	2.067
Client number inverse CDF regression				
$\widehat{\beta_1}^c$	-1.667	-1.587	-1.395	-1.655
$\widehat{\beta_2}^c$	-0.097	-0.280	-1.184	-1.420
$\widehat{se}(\epsilon^{n^c})$	0.066	0.128	0.062	0.069

C
Chapter 3

C.1 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.

C.1.1 Those Nearby are the First to Learn About New Ideas

The first exercise follows Jaffe et al.[56]. 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 C.1, 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 pub-

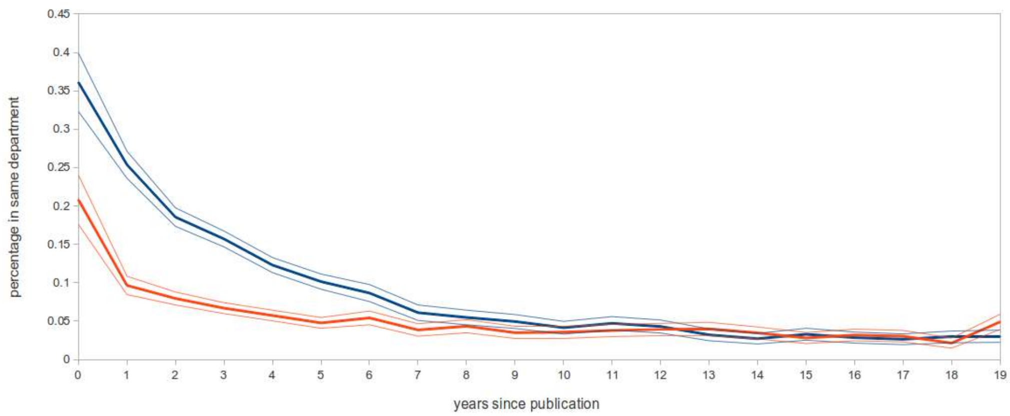


Figure C.1: Results of Jaffe exercise, Dark Blue = Citing Papers, Light Red = Reference Papers

lished those outside the originating department are just as likely as those in the department to cite the originating 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.[56].

C.1.2 Dealing with Endogenous Sorting

The Jaffe exercise above recently came under attack. Thompson and Fox-Kean[87] 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 C.2 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:

$$CY_i = \beta_0 + \beta_H Harv_i + \beta_J Jcits_i + \beta \mathbf{X}_i + \varepsilon_i \tag{C.1}$$

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

Table C.1 contains the estimation results from (C.1). 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.

	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 C.1: Effect of working at Harvard on year of Jensen citation

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.

C.2 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:

$$\lambda_{it} = \beta_0 + \beta_1 \text{dep\_frac}_{it} + \beta_2 \text{interest}_i + \varepsilon_{it} \tag{C.2}$$

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:

$$\lambda_{it} = \gamma_0 + \gamma_1 \text{dep\_frac}_{it} + \varepsilon_{it} \tag{C.3}$$

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

$$\mathbb{E}[\hat{\gamma_1}] = \beta_1 + \rho_{\{\text{dep\_frac,interest}\}} \sqrt{\frac{\sigma^2_{\text{interest}}}{\sigma^2_{\text{dep\_frac}}}} \tag{C.4}$$

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:

$$\lambda_{it} = \delta_0 + \delta_1 \text{dep\_frac}_{it} + \delta_2 \text{field} + \varepsilon_{it} \tag{C.5}$$

Now we can write the asymptotic expected value of the estimator $\hat{\delta}_1$ as (Hanushek and Jackson 1977):

$$\mathbb{E}[\hat{\delta}_1] = \beta_1 + \frac{\rho_{\{\text{dep\_frac,interest}\}} - \rho_{\{\text{dep\_frac,field}\}}\rho_{\{\text{field,interest}\}}}{1 - \rho^2_{\{\text{dep\_frac,field}\}}}\sqrt{\frac{\sigma^2_{\text{interest}}}{\sigma^2_{\text{dep\_frac}}}} \tag{C.6}$$

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.3 Data Construction

C.3.1 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.3.2 Constructing Fields

C.3.2.1 IDEAS Fields

In the main structural exercise we need fields for each academic. In some of the additional exercises reported in Section C.1, 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.3.2.2 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.[12]. 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.3.3 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 C.2 at the end of this file is a sample of included departments and fields.

DEPT	QUAL	JENS. FIELD
UNIV MICHIGAN	0.884	0.035
PURDUE UNIV	0.490	0.0036
PENN STATE UNIV	0.663	0.067
HARVARD UNIV	1.0	0.014
UNIV PENN	0.903	0.041
CORNELL UNIV	0.817	0.016
UNIV ROCHESTER	0.355	0.086
MIT	0.980	0.086
UNIV MARYLAND	0.798	0.060
STANFORD UNIV	0.932	0.028
UNIV DELAWARE	0.346	0.0
CARNEGIE MELLON UNIV	0.625	0.047
YALE UNIV	0.923	0.026
PRINCETON UNIV	0.971	0.051
UNIV ILLINOIS	0.557	0.031
BOSTON UNIV	0.913	0.012
UNIV N CAROLINA	0.403	0.0
...		
UNIV NEW MEXICO	0.038	0.140
CUNY HUNTER COLL	0.105	0.0
UNIV MIAMI	0.086	0.0
TUFTS UNIV	0.538	0.0
BRIGHAM YOUNG UNIV	0.336	0.0
UNIV CALIF IRVINE	0.644	0.0
UNIV HAWAII MANOA	0.144	0.0
EMORY UNIV	0.326	0.0
UNIV CALIF SAN DIEGO	0.855	0.056
WELLESLEY COLL	0.432	0.257
CUNY	0.375	0.0
DREXEL UNIV	0.067	0.0
MIDDLEBURY COLL	0.182	0.0
SANTA CLARA UNIV	0.115	0.0
UNIV CALIF SANTA CRUZ	0.615	0.0
SUNY ALBANY	0.394	0.0
TULANE UNIV	0.317	0.192
APPALACHIAN STATE UNIV	0.269	0.0
RENSSELAER POLYTECH INST	0.298	0.0
CHAPMAN UNIV	0.548	0.0

Table C.2: Selected departments.

NEP ABREV	FIELD NAME
NEP-ACC	Accounting & Auditing
NEP-AFR	Africa
NEP-AGE	Economics of Ageing
NEP-AGR	Agricultural Economics
NEP-ARA	Arab World
NEP-BAN	Banking
NEP-BEC	Business Economics
NEP-CBA	Central Banking
NEP-CBE	Cognitive & Behavioural Economics
NEP-CDM	Collective Decision-Making
NEP-CFN	Corporate Finance
NEP-CIS	Confederation of Independent States
NEP-CMP	Computational Economics
NEP-CNA	China
NEP-COM	Industrial Competition
NEP-CSE	Economics of Strategic Management
NEP-CTA	Contract Theory & Applications
NEP-ORE	Operations Research
NEP-PBE	Public Economics
NEP-PKE	Post Keynesian Economics
NEP-POL	Positive Political Economics
NEP-PPM	Project, Program & Portfolio M
NEP-PUB	Public Finance
NEP-REG	Regulation
NEP-RES	Resource Economics
NEP-RMG	Risk Management
NEP-SBM	Small Business Management
NEP-SEA	South East Asia
NEP-SOC	Social Norms & Social Capital
NEP-SOG	Sociology of Economics anagent
NEP-SPO	Sports & Economics
NEP-TID	Technology & Industrial Dynami
NEP-TRA	Transition Economics
NEP-TRE	Transport Economics
NEP-TUR	Tourism Economics
NEP-UPT	Utility Models & Prospect Theo
NEP-URE	Urban & Real Estate Economics

Table C.3: Selected fields.

C.4 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:

$$(Tf)(d) = \rho \ln \left(\sum_{d' \in \mathcal{D}} e^{f(d') + w(d') - \mathbb{1}_{\{d' \neq d\}} \, \mathcal{C}} \right) \tag{C.7}$$

We will verify that T is a contraction mapping using Blackwell's sufficient conditions. Recall Theorem 3.3 from Stokey[85]:

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 (C.7), monotonicity is immediate. Discounting is almost immediate as well. Take $a \geq 0$, and let $f(x) \in B(X)$:

$$
\begin{aligned}
[T(f+a)](d) &= \rho \ln \left(\sum_{d'} \left(e^{f(d') + w(d') - \mathbb{1}_{\{d' \neq d\}} \mathcal{C} + a} \right) \right) \\
&= \rho \ln \left(e^{a} \sum_{d'} \left(e^{f(d') + w(d') - \mathbb{1}_{\{d' \neq d\}} \mathcal{C}} \right) \right) \\
&= (Tf)(d) + \rho a
\end{aligned}
\tag{C.8}
$$

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

C.5 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: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[82], but it is not derived there either! I leave the derivation here for future puzzled American graduate students.

$$\mathbb{E}\left[\max\{X+\varepsilon_1, Y+\varepsilon_2\}\right] = \gamma_e + \ln\left(e^X + e^Y\right) \tag{C.9}$$

The CDF of the Gumbel distribution is $F(z) = e^{-e^{-z}}$, so we can write the distribution of the maximum in (C.9) as:

$$G(z) = e^{-\left(e^{X-z}+e^{Y-z}\right)} \tag{C.10}$$

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

$$g(z) = \left(e^{X-z}+e^{Y-z}\right)e^{-\left(e^{X-z}+e^{Y-z}\right)} \tag{C.11}$$

Now we can rewrite the LHS of (C.9) as:

$$\int_{-\infty}^{\infty} zg(z)dz \tag{C.12}$$

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

$$\int_0^{\infty} \left(\ln\left(e^X+e^Y\right) - \ln(t)\right)e^{-t}dt = \ln\left(e^X+e^Y\right) + \gamma_e \tag{C.13}$$

The substitution of γ_e is an identity.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[90] after a list of identities involving γ_e, but I was not able to find this particular identity there.

C.6 MCMC Diagnostics

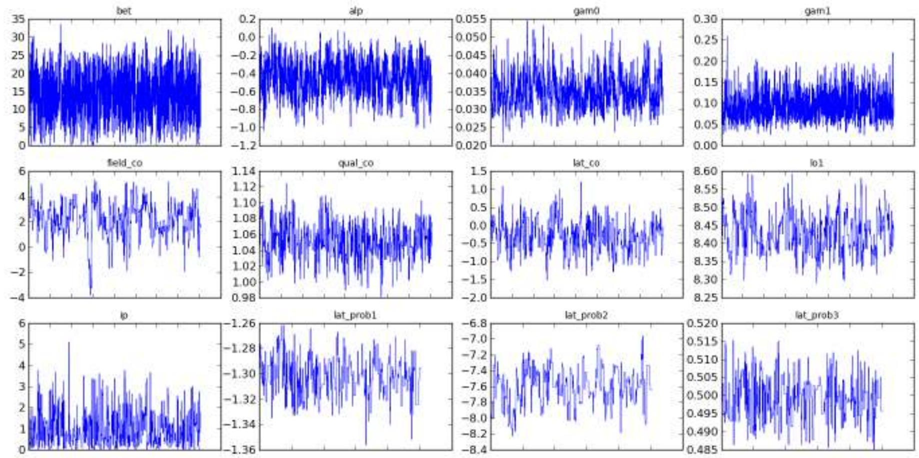


Figure C.2: Mixing plots

This appendix presents mixing plots and diagnostics from the parallel MCMC estimation. Eyeballing the plots in Figure C.2, everything looks good. Eyeballing, however, is not very reliable. Table C.4 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.

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 C.4: Gelman-Rubin Test

C.7 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[46]. Some researchers have been explicit about why they prefer patent data.[55] 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: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.

> [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[44].

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.

C.8 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.

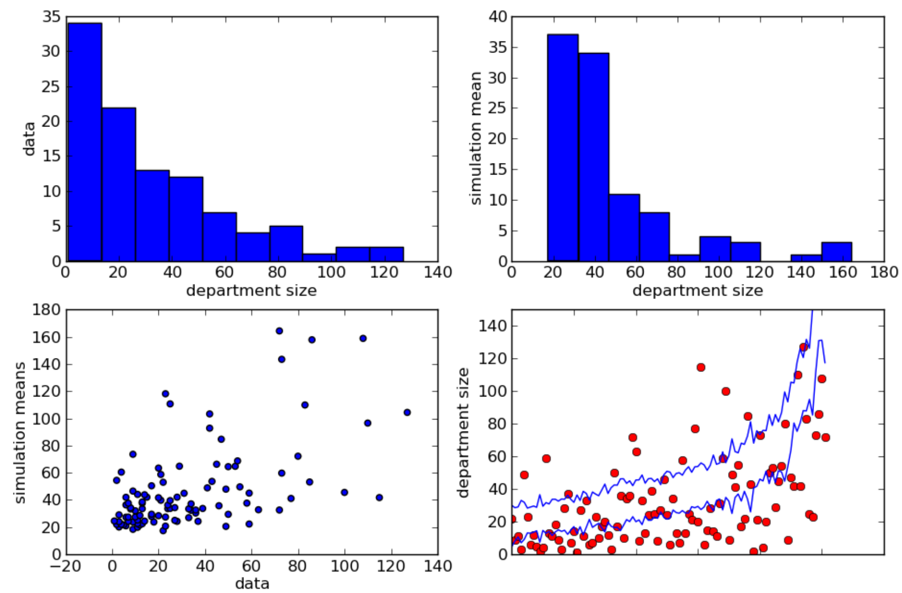


Figure C.3: TL: Histogram of data, TR: Histogram of simulation means, BL: Scatter plot data vs. simulation means, BR: Two standard deviations of simulation variation against data.

The top row of Figure C.3 contains a histogram of the data department sizes and a histogram of simulated mean department sizes. The histograms are qualitatively similar, although there are 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 C.3 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.

C.9 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 [47]. This is a slightly older version of the model in the baseline section. As can be seen in Table C.5, 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 C.5: Posterior moment comparison

C.10 Posterior Kernal Densities for Alternative Models

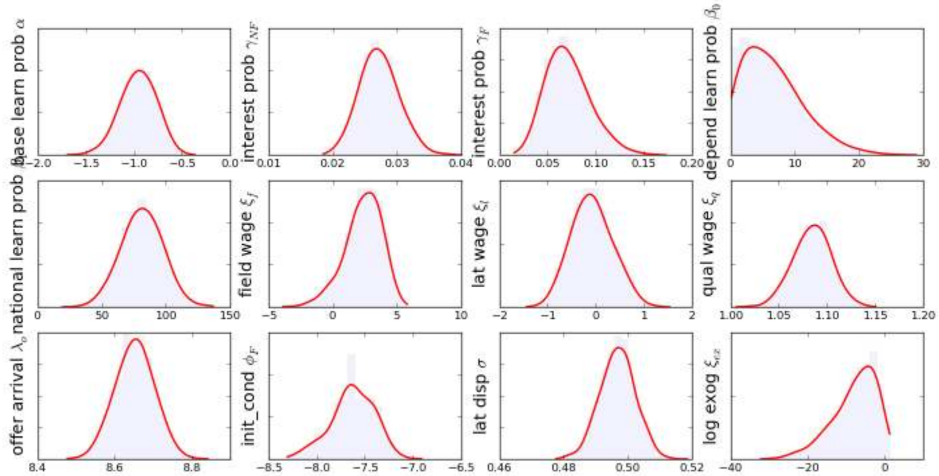


Figure C.4: Posteriors for national model

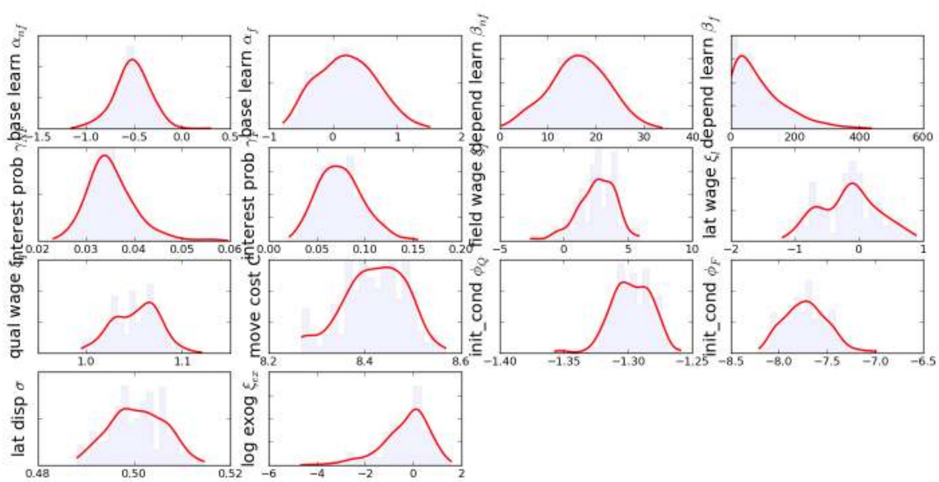


Figure C.5: Posteriors for field-specific model

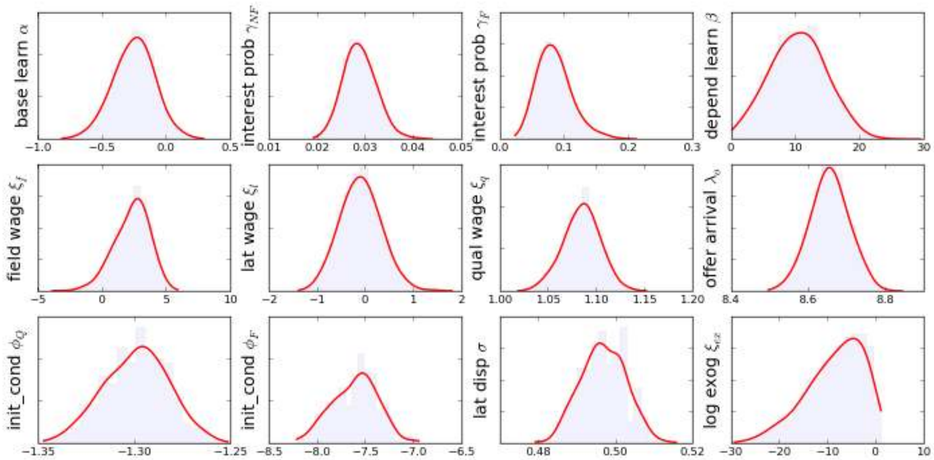


Figure C.6: Posteriors for publication lag model

C.11 Comparison of Baseline Priors

	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 C.6: Posterior moment comparison, alternate priors

References

[1] Agrawal, A., McHale, J., and Oettl, A. (2013). Why stars matter. *mimeo*.

[2] Albornoz, F., Calvo Pardo, H. F., Corcos, G., and Ornelas, E. (2012). Sequential exporting. *Journal of International Economics*, 88(1), 17–31.

[3] Alessandria, G. and Choi, H. (2007). Do sunk costs of exporting matter for net export dynamics? *The Quarterly Journal of Economics*, 122(1), 289–336.

[4] Alessandria, G., Kaboski, J. P., and Midrigan, V. (2010). Inventories, lumpy trade, and large devaluations. *American Economic Review*, 100(5), 2304–39.

[5] Almeida, P. and Kogut, B. (1999). Localization of knowledge and the mobility of engineers in regional networks. *Management science*, 45(7), 905–917.

[6] Arcidiacono, P., Foster, G., Goodpaster, N., and Kinsler, J. (2012). Estimating spillovers using panel data, with an application to the classroom. *Quantitative Economics*, 3(3), 421–470.

[7] Arcidiacono, P. and Miller, R. A. (2011). Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica*, 79(6), 1823–1867.

[8] Arkolakis, C. (2010). Market penetration costs and the new consumers margin in international trade. *Journal of political economy*, 118(6), 1151–1199.

[9] Arkolakis, C. (2011). *A unified theory of firm selection and growth*. Technical report, National Bureau of Economic Research.

[10] Baldwin, R. and Krugman, P. (1989). Persistent trade effects of large exchange rate shocks. *The Quarterly Journal of Economics*, 104(4), 635–654.

[11] Banerjee, A. V. and Duflo, E. (2007). The Economic Lives of the Poor. *The Journal of Economic Perspectives*, 21(1), 141–167.

[12] Barrett, C. B., Olia, A., and Bailey, D. V. (2000). Subdiscipline-specific journal rankings: whither applied economics? *Applied Economics*, 32(2), 239–252.

[13] Bernard, A., Eaton, J., Jensen, J., and Kortum, S. (2003). Plants and productivity in international trade. *The American Economic Review*, 93(4), 1268–1290.

[14] Bernard, A. B., Jensen, J. B., Redding, S. J., and Schott, P. K. (2007). Firms in international trade. *Journal of Economic Perspectives*, 21(3), 105–130.

[15] Betts, J. R. and Zau, A. (2004). Peer groups and academic achievement: Panel evidence from administrative data. *mimeo*.

[16] Bloch, F., Rao, V., and Desai, S. (2004). Wedding Celebrations as Conspicuous Consumption. *Journal of Human Resources*, 39(3), 675–695.

[17] Bolton, G. and Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *The American Economic Review*, (pp. 166–193).

[18] Breschi, S. and Lissoni, F. (2001). Knowledge spillovers and local innovation systems: a critical survey. *Industrial and corporate change*, 10(4), 975–1005.

[19] Brooks, E. L. (2006). Why don't firms export more? product quality and colombian plants. *Journal of Development Economics*, 80(1), 160–178.

[20] Burke, M. A. and Sass, T. R. (2013). Classroom peer effects and student achievement. *Journal of Labor Economics*, 31(1), 51–82.

[21] Caballero, R. J. and Jaffe, A. B. (1993). 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* (pp. 15–86). MIT press.

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

[23] Chaney, T. (2011). The Network Structure of International Trade. Available at Chaney's website.

[24] Charles, K. K., Hurst, E., and Roussanov, N. (2009). Conspicuous Consumption and Race. *The Quarterly Journal of Economics*, 124(2), 425–467.

[25] Chor, D. and Lai, E. L. (2013). Cumulative innovation, growth, and welfare improving patent policy. *mimeo*.

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

[27] Conley, T. G. and Udry, C. R. (2010). Learning about a new technology: Pineapple in ghana. *The American Economic Review*, (pp. 35–69).

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

[29] Das, S., Roberts, M. J., and Tybout, J. R. (2007). Market entry costs, producer heterogeneity, and export dynamics. *Econometrica*, 75(3), 837–873.

[30] Davis, G. E. (2011). *Presidential Installation*. University of Vermont.

[31] De Giorgi, G., Pellizzari, M., and Redaelli, S. (2010). Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics*, (pp. 241–275).

[32] Dixit, A. (1989). Hysteresis, import penetration, and exchange rate pass-through. *The Quarterly Journal of Economics*, 104(2), 205–228.

[33] Dominguez, Camilo, J., Eaton, J., Eslava, M., and Tybout, J. (2008). Search and learning in export markets: Case studies for colombia.

[34] Dos Santos, M. D. and Postel-Vinay, F. (2003). Migration as a source of growth: the perspective of a developing country. *Journal of Population Economics*, 16(1), 161–175.

[35] Drozd, L. A. and Nosal, J. B. (2012). Understanding international prices: Customers as capital. *The American Economic Review*, 102(1), 364–395.

[36] Duesenberry, J. S. (1967). *Income, Saving, and the Theory of Consumer Behavior*. Oxford University Press.

[37] Dustmann, C. (2003). Return migration, wage differentials, and the optimal migration duration. *European Economic Review*, 47(2), 353–369.

[38] Eaton, J., Eslava, M., Kugler, M., and Tybout, J. (2008). nthe margins of entry into export markets: Evidence from colombia, oin elhanan helpman, dalia marin and thierry verdier, eds., the organization of firms in a global economy.

[39] Eaton, J. and Kortum, S. (1999). International technology diffusion: Theory and measurement. *International Economic Review*, 40(3), 537–570.

[40] Fehr, E. and Schmidt, K. (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics*, 114(3), 817–868.

[41] Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, (pp. 457–472).

[42] Gold, S. D. (1995). *The fiscal crisis of the states: Lessons for the future*, volume 10. Georgetown University Press.

[43] Gourieroux, C. and Monfort, A. (1997). *Simulation-based econometric methods*. Oxford University Press.

[44] Greene, M. (2007). The demise of the lone author. *Nature*, 450(7173), 1165–1165.

[45] Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica, Journal of the Econometric Society*, (pp. 501–522).

[46] Griliches, Z. (1998). Patent statistics as economic indicators: a survey. In *R&D and productivity: the econometric evidence* (pp. 287–343). University of Chicago Press.

[47] Grossman, S. J. and Hart, O. D. (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *The Journal of Political Economy*, (pp. 691–719).

[48] Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive metropolis algorithm. *Bernoulli*, (pp. 223–242).

[49] Harris, D. N. and Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of public economics*, 95(7), 798–812.

[50] Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, (pp. 153–161).

[51] Heffetz, O. (2011). A Test of Conspicuous Consumption: Visibility and Income Elasticities. *The Review of Economics and Statistics*.

[52] Irarrazabal, A. A. and Opromolla, L. D. (2006). Hysteresis in export markets. *New York University, Working Paper*.

[53] Ireland, N. J. (1994). On Limiting the Market for Status Signals. *Journal of Public Economics*, 53, 91–110.

[54] Jackson, M. O. and Rogers, B. W. (2007). Meeting Strangers and Friends of Friends: How Random are Social Networks. *American Economic Review*, 97(3).

[55] Jaffe, A. B. and Trajtenberg, M. (2002). *Patents, citations, and innovations: A window on the knowledge economy*. MIT press.

[56] Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *the Quarterly journal of Economics*, 108(3), 577–598.

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

[58] Kennan, J. and Walker, J. R. (2011). The effect of expected income on individual migration decisions. *Econometrica*, 79(1), 211–251.

[59] Kim, J., Lee, S. J., and Marschke, G. (2009). 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* (pp. 321–348). University of Chicago Press.

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

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

[62] Li, S. (2002). Chinese Household Income Project. *Inter-university consortium for Political and Social Research*.

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

[64] Lucas, R. (2008). Ideas and growth. *Economica*, 76(301), 1–19.

[65] Luttmer, E. (2004). *Neighbors as Negatives: Relative Earnings and Well-being*. Technical report, National Bureau of Economic Research.

[66] Luttmer, E. G. (2007). Selection, growth, and the size distribution of firms. *The Quarterly Journal of Economics*, 122(3), 1103–1144.

[67] Mas, A. and Moretti, E. (2009). Peers at work. *The American Economic Review*, 99(1), 112–145.

[68] Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6), 1695–1725.

[69] Moav, O. and Neeman, Z. (2010). Status and Poverty. *Journal of the European Economic Association*, 8, 413–420.

[70] Moav, O. and Neeman, Z. (2012). Saving Rates and Poverty: The Role of Conspicuous Consumption and Human Capital. *The Economic Journal*, 122, 933–956.

[71] Money Magazine (1991). 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.

[72] Moore, S. (1991). State spending splurge: The real story behind the fiscal crisis in state government.

[73] National Bureau of Economic Research (2011). Consumer Expenditure Survey Family-Level Extracts. http://www.nber.org/data/ces_cbo.html.

[74] Oettl, A. and Agrawal, A. (2008). International labor mobility and knowledge flow externalities. *Journal of International Business Studies*, 39(8), 1242–1260.

[75] Perez-Truglia, R. (2013). Measuring the market value of non-market goods: The case of conspicuous consumption. *Journal of Socio-Economics*, 45, 146–154.

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

[77] Rauch, J. E. and Watson, J. (2003). Starting small in an unfamiliar environment. *International Journal of Industrial Organization*, 21(7), 1021–1042.

[78] Robson, A. and Samuelson, L. (2010). The Evoluationary Foundations of Preferences. *Handbook of Social Economics*, (pp. 221–310).

[79] Rogers, E. M. (1962). *Diffusion of innovations*. Simon and Schuster.

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

[81] Ruhl, K. J. and Willis, J. L. (2008). New exporter dynamics.

[82] Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica: Journal of the Econometric Society*, (pp. 999–1033).

[83] Shimer, R. (2005). The cyclical behavior of equilibrium unemployment and vacancies. *American economic review*, (pp. 25–49).

[84] Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65(3), 557–586.

[85] Stokey, N. L. (1989). *Recursive methods in economic dynamics*. Harvard University Press.

[86] Taiwan Panorama (2002). Don't taiwanese want to study abroad anymore?

[87] Thompson, P. and Fox-Kean, M. (2005). Patent citations and the geography of knowledge spillovers: A reassessment. *American Economic Review*, (pp. 450–460).

[88] Washington Post (1991). Public colleges in most states cut budgets this year; recession brings end to decade of growth, survey finds; further reductions expected.

[89] White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, (pp. 817–838).

[90] Whittaker, E. T. and Watson, G. N. (1996). *A course of modern analysis*. Cambridge university press.

David C. Jinkins

Penn State University
Department of Economics
Kern Building
State College, PA 16802

Phone: (814) 753-0181
Email: david.jinkins@gmail.com
Homepage: http://www.davidjinkins.com

Education

PHD Economics (expected), The Pennsylvania State University, 2014,
Fields: Trade, Development, Game Theory.

MA Economics, National Taiwan University, 2009,
Masters Thesis: "A Bridge Goes Both Ways: A Two-Way Migration Model with Incomplete Information and Social Capital".

BA Political Science, Carleton College, 2005.
Magna Cum Laude

Working Papers

Peer Learning, Labor Movement, and Knowledge Diffusion

A Structural Comparison of Conspicuous Consumption in China and the United States.

A Search and Learning Model of Export Dynamics (*with Jonathan Eaton, Marcela Eslava, C.J. Krizan, and Jim Tybout*).

Work in Progress

Trade on an Evolving Network (*with Jonathan Eaton, C.J. Krizan, Jim Tybout, and Daniel Xu*).

Market Shocks and Newspaper Ideology: Evidence from Taiwan (*with Chun-Fang Chiang and Long-Yu Chiou*)

Awards

Rosenberg Award, 2013

National Science Foundation East Asia and Pacific Summer Institute Fellowship, China 2012.

Penn State University, Graduate Research Assistant Fellowship 2009–Spring 2011.

Ministry of Education Taiwan Scholarship, Taiwan 2007–2009.

U.S. Fulbright Fellowship, Turkey 2006–2007.

Chang-Lan Fellowship, China 2003.

Personal

Citizenship: United States.

Languages: English, Mandarin Chinese, Turkish.

Technical: python, matlab, vim, LaTeX, Stata, html, css, bash, C++.