Title
Three Essays on Creative Destruction

Permalink
https://escholarship.org/uc/item/0qs699dv

Author
Igami, Mitsuru

Publication Date
2012

Peer reviewed|Thesis/dissertation
Three Essays on Creative Destruction

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Management

by

Mitsuru Igami

2012
This dissertation aims to advance our knowledge of long-run economic changes. It consists of three essays on strategic industry dynamics in retail services, agricultural commodities, and high-tech manufacturing, respectively. Although creative destruction is commonly understood as the replacement of old technologies by new ones, its true significance lies not in the transition of technologies per se but in either the reluctance or inability of old winners to innovate when faced with new challengers. Hence I emphasize the incumbent-entrant rivalry as the common theme across these essays.

Essay 1 assesses the impact of the entry of large supermarkets on incumbents of various sizes. Contrary to the conventional notion that big stores drive small rivals out of the market, data from Tokyo in the 1990s show that large supermarkets’ entry induces the exit of existing large and medium-size competitors, but improves the survival rate of small supermarkets. These findings highlight the role of store size as an important dimension of product differentiation and caution against size-based entry regulations.

Essay 2 studies the impact of international market structure on commodity prices, using a standard oligopoly model and exploiting historical variations in the structure of the international coffee bean market. The results suggest that, of the 75% drop in the real coffee price between 1988 and 2001, the end of a cartel treaty explains 49 points and the emergence of Vietnam as a major exporter explains another 9 points. I then discuss policy implications for competition, trade, and aid.
Essay 3 investigates why incumbent firms innovate more slowly than entrants. Theories predict cannibalization between existing and new products delays incumbents’ innovation, whereas preemptive motives accelerate it, and incumbents’ cost (dis)advantage would further reinforce these tendencies. To empirically quantify these three forces, I develop and estimate a dynamic oligopoly model using a unique panel dataset of hard disk drive (HDD) manufacturers (1981–98). The results suggest that despite strong preemptive motives and a substantial cost advantage over entrants, incumbents are reluctant to innovate because of cannibalization, which can explain at least 51% of the incumbent-entrant timing gap. I then discuss managerial and public-policy implications.
The dissertation of Mitsuru Igami is approved.

Raphael Thomadsen
Mariko Sakakibara
Hugo Hopenhayn
Daniel Ackerberg

Edward Leamer, Committee Chair

University of California, Los Angeles
2012
To Maki
All Fêtes
# Table of Contents

1 Does Big Drive Out Small?

Entry, Exit, and Differentiation in the Supermarket Industry ........................................ 1

1.1 Introduction .................................................................................................................. 1

1.2 Literature .................................................................................................................... 2

1.3 Industry and Data ......................................................................................................... 4

   1.3.1 Definition of Supermarket .................................................................................... 4

   1.3.2 Train Station-Centered Geographical Markets ................................................... 5

   1.3.3 The 1990 Deregulation and Its Historical Context ........................................... 7

   1.3.4 Descriptive Statistics of Towns and Supermarkets ........................................... 9

1.4 Empirical Analysis ....................................................................................................... 10

   1.4.1 Ordered Probit: Exit, Shrink, Stay Unchanged, or Expand ............................... 12

   1.4.2 Geographical Instruments for Entrants’ Town Choice ....................................... 14

   1.4.3 Tobit: Percentage Change in Incumbents’ Floor Size ....................................... 17

   1.4.4 Discussion ............................................................................................................ 20

1.5 Conclusion .................................................................................................................. 23

References ......................................................................................................................... 25

2 Oligopoly in International Commodity Markets:

the Case of Coffee Beans ................................................................................................. 27

2.1 Introduction .................................................................................................................. 27

   2.1.1 Related Literature ............................................................................................... 31

2.2 The International Coffee Bean Market ....................................................................... 33

   2.2.1 Importers, Exporters, and Product Characteristics ........................................... 34
2.2.2 The Cartel and the Cold War: Why the Rise and Fall of the International Coffee Agreement are Exogenous ........................................... 36
2.2.3 Geopolitics of Land, Labor, and Ethnicity: Why Vietnam’s Expansion is Exogenous ................................................................. 38
2.2.4 Summary of the Institutional Context .................................................. 41
2.3 Data ........................................................................................................ 42
  2.3.1 Variables ............................................................................................. 42
  2.3.2 Preliminary Data Analysis ................................................................... 45
2.4 Model ....................................................................................................... 46
  2.4.1 Demand and the Timing of the Game ................................................... 47
  2.4.2 The Cartel’s Coordination .................................................................. 48
2.5 Estimation and Results ............................................................................. 49
  2.5.1 Demand Estimates with Weather IV .................................................... 49
  2.5.2 Estimates for the Cartel’s Coordination .............................................. 53
2.6 Decomposing the “Coffee Crisis” ............................................................ 54
  2.6.1 Counterfactual Experiments ............................................................... 55
  2.6.2 Discussion and Policy Implications .................................................... 56
2.7 Conclusion ................................................................................................. 58

References ..................................................................................................... 72

3 Estimating the Innovator’s Dilemma:
Structural Analysis of Creative Destruction in the Hard Disk Drive Industry ................................................................. 77
  3.1 Introduction ............................................................................................. 77
    3.1.1 Related Literature ............................................................................. 83

vii
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2</td>
<td>Industry and Data</td>
<td>88</td>
</tr>
<tr>
<td>3.2.1</td>
<td>HDD Industry: Canonical Case of Innovation and Evolution</td>
<td>88</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Data Source</td>
<td>91</td>
</tr>
<tr>
<td>3.2.3</td>
<td>Focus: Transition from 5.25- to 3.5-inch Generations</td>
<td>93</td>
</tr>
<tr>
<td>3.3</td>
<td>Model</td>
<td>96</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Setup: Dynamic Discrete-Choice Game</td>
<td>96</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Demand</td>
<td>102</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Period Competition</td>
<td>103</td>
</tr>
<tr>
<td>3.3.4</td>
<td>Solution of Dynamic Game by Backward Induction</td>
<td>104</td>
</tr>
<tr>
<td>3.4</td>
<td>Estimation</td>
<td>105</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Estimation: Demand</td>
<td>105</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Estimation: Marginal Costs</td>
<td>108</td>
</tr>
<tr>
<td>3.4.3</td>
<td>Estimation: Sunk Costs of Innovation</td>
<td>108</td>
</tr>
<tr>
<td>3.5</td>
<td>Results</td>
<td>111</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Results: Demand</td>
<td>111</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Results: Marginal Costs</td>
<td>113</td>
</tr>
<tr>
<td>3.5.3</td>
<td>Results: Sunk Costs of Innovation</td>
<td>114</td>
</tr>
<tr>
<td>3.5.4</td>
<td>Results: Industry State, Policy, and Value</td>
<td>116</td>
</tr>
<tr>
<td>3.6</td>
<td>“Innovator’s Dilemma” Explained</td>
<td>118</td>
</tr>
<tr>
<td>3.6.1</td>
<td>Cannibalization (Counterfactual 1)</td>
<td>119</td>
</tr>
<tr>
<td>3.6.2</td>
<td>Preemption (Counterfactual 2)</td>
<td>121</td>
</tr>
<tr>
<td>3.6.3</td>
<td>Sunk Cost Gap (Counterfactual 3)</td>
<td>121</td>
</tr>
<tr>
<td>3.7</td>
<td>Policy Experiments</td>
<td>122</td>
</tr>
<tr>
<td>3.7.1</td>
<td>Broad Patent (Experiment 1)</td>
<td>125</td>
</tr>
<tr>
<td>3.7.2</td>
<td>R&amp;D Subsidy (Experiment 2)</td>
<td>128</td>
</tr>
</tbody>
</table>
3.7.3 Banning Non-compete Clauses (Experiment 3) ................... 129
3.7.4 No International Trade (Experiment 4) .......................... 130
3.8 Conclusion ............................................................. 132

References ................................................................. 135
LIST OF FIGURES

1.1  Towns and Railways in the Greater Tokyo Area .......................... 6
1.2  Number of Supermarket Openings ........................................... 8

2.1  Real Price of Coffee Beans .................................................. 28
2.2  Areas of Coffee Cultivation ................................................... 35
2.3  The Government-led Migration to the Central Highlands of Vietnam .... 40
2.4  The Geographical Expansion of Doi Moi and Coffee Production ........ 41
2.5  Price, Cost, and Markup ..................................................... 43
2.6  The Estimated Degree of Coordination among The ICA Exporters ($\hat{\mu}_t$) ... 54
2.7  Counterfactual Prices .......................................................... 56
2.8  Estimates and Simulation Based on Linear/Log-Linear Demand ........... 65

3.1  Shifting Generations of Technology ......................................... 81
3.2  CDF of Adoption Timing ...................................................... 89
3.3  Map of Computer Industry .................................................... 90
3.4  Evolution of the Industry Composition ..................................... 94
3.5  Quantity, Price, and Quality ................................................ 95
3.6  Illustration of the Dynamic Discrete Choice .............................. 96
3.7  Estimated Popularity (Unobserved Quality) and Marginal Costs ........ 113
3.8  Market Structure Dynamics .................................................. 116
3.9  Estimated Policy and Value Functions ...................................... 117
3.10 Incumbents-Entrant Timing Gap in Innovation ............................ 119
3.11 Explaining the Innovator’s Dilemma (Counterfactual Series I) .......... 120
3.12 Time Profile of Social Welfare (Baseline) .................................. 124
3.13 Policy Experiment 1 (Ex-ante and Ex-post Effects of Broad Patent) . . . . . . 127
3.14 Policy Experiments 2 and 3 . . . . . . . . . . . . . . . . . . . . . . . . . . . 129
3.15 Policy Experiment 4 (No International Trade) . . . . . . . . . . . . . . . . . 132
# List of Tables

1.1 Summary Statistics for Town Characteristics ............................................. 9  
1.2 Summary Statistics for Store-level Observations ................................. 10  
1.3 Summary Statistics for Store-level Observations (continued) .............. 11  
1.4 Ordered Probit Regressions of Decision to Exit < Shrink < Unchanged < Expand .......................................................... 13  
1.5 Binary IV-Probit Regressions (1 of 3) ............................................... 16  
1.6 Binary IV-Probit Regressions (2 of 3) ............................................... 17  
1.7 Binary IV-Probit Regressions (3 of 3) ............................................... 18  
1.8 Shrinkage: Two-Sided Tobit Regressions of Change in Floor Size (%) ... 19  
1.9 Expansion: One-Sided Tobit Regressions of Change in Floor Size (%) ... 20  
2.1 Summary Statistics ................................................................. 42  
2.2 Preliminary Regressions of Market Power on Shocks ......................... 45  
2.3 Demand Estimates ................................................................. 50  
2.4 Counterfactual Prices for Selected Years ....................................... 55  
2.5 Demand Estimates (Linear Specification) ....................................... 62  
2.6 Demand Estimates (Log-Linear Specification) ................................ 63  
2.7 Checks for Autocorrelation (Linear Specification) ........................... 66  
2.8 With Lagged Dependent Variable (Linear Specification) .................... 67  
2.9 Counterfactual Prices for Selected Years (II) ................................ 69  
3.1 Global PC Market Share by Units (%) ......................................... 91  
3.2 Preliminary Regressions of Innovation Timing on Firm Characteristics ... 93  
3.3 Demand Estimates for 5.25- and 3.5-inch HDDs ............................ 112
3.4 Demand Estimates for 5.25- and 3.5-inch HDDs (continued) . . . . . . . . . 112
3.5 Estimates of the Dynamic Parameters . . . . . . . . . . . . . . . . . . . . . . . 114
3.6 Comparison of Social Welfare across Policy Experiments . . . . . . . . . . . . 123
3.7 Brief History of Rodime’s 3.5-inch HDD Patent Affair . . . . . . . . . . . . . 125
I thank my advisors for all encouragement and educational experience. Specifically, I thank Edward Leamer for teaching me the right attitude in learning, Daniel Ackerberg for the excellent training in Empirical Industrial Organization, Hugo Hopenhayn for constantly setting high bars in research, Mariko Sakakibara for expanding the scope of my potential audience, Connan Snider for guiding me through the jungle of estimating dynamic games, and Raphael Thomadsen for his comprehensive, meta-level advisory throughout my UCLA Anderson years.

I thank the members of my first-year study group, Matthew Hill, Anthony Bruno Keats, and Wilko Ziggy Schultz-Mahlendorf, without whom I could not have survived the coursework. I also benefited from the Empirical I.O. reading group, whose members included Andraz Kavalar, Jinyoung Lee, Wilko Schultz, Jiyeon Seo, Yong Hyeon Yang, and Hisayuki Yoshimoto.

The following is the acknowledgement for each of the three papers constituting this dissertation. I list them separately because I wrote them in different years and places. I started working on Essay 1 in 2006, initially as my Master's thesis at the Graduate School of Economics, University of Tokyo. I drafted Essay 2 in 2009 to fulfill the second-year paper requirement in International Economics at UCLA. Later that year, I started data collection for Essay 3, which was my job-market paper in the academic year 2011–12. Financial support from the Ito Foundation (for Essays 1 and 2), the Nozawa Fellowship (for 1, 2, and 3), the UCLA CIBER (for 2 and 3), and the Dissertation Year Fellowship (for 3) are gratefully acknowledged.

**Essay 1:** For suggestions, I thank Jinyong Hahn, Hidehiko Ichimura, Toshiaki Iizuka, Edward Leamer, Rosa Matzkin, Hiroshi Ohashi, Connan Snider, Raphael Thomadsen, and seminar participants at the University of Tokyo, the *2007 Japanese Economic Association Annual Spring Meeting* at Osaka Gakuin University, and UCLA. I am grateful to the editor of *Review of Industrial Organization*, Lawrence White, and the two anonymous referees, for thoughtful comments. I thank Evan Gill, Matthew Khalil Hill, and Maki Komatsu for
editorial assistance, Saneaki Obata for industry expertise, and my local supermarkets CSN (small) and Life Extra (medium) for inspiration.

**Essay 2:** For suggestions, I thank Daniel Ackerberg, Andrew Ainslie, Bruce Carlin, Harold Demsetz, Jeffrey Dubin, Sebastian Edwards, Eric Fisher, Paola Giuliano, Jinyong Hahn, Hugo Hopenhayn, Ari Kang, Lawrence Karp, Edward Leamer, Rosa Matzkin, Ayako Obashi, Jeffrey Perloff, John Riley, Connan Snider, Raphael Thomadsen, Aaron Tornell, Sofia Berto Villas-Boas, Romain Wacziarg, and Mark Wright, as well as the participants at FREIT’s 4th *Empirical Investigations in Trade and Investment* Conference, the 10th Annual *Trans-Atlantic Doctoral Conference* at London Business School, UC Berkeley, and UCLA.

**Essay 3:** For suggestions, I thank Daniel Ackerberg, Ron Adner, Jinyong Hahn, Hugo Hopenhayn, Edward Leamer, Phillip Leslie, Rosa Matzkin, Matthew Mitchell, Ichiro Obara, Marc Rysman, Mariko Sakakibara, Connan Snider, Raphael Thomadsen, Kosuke Uetake, and Yong Hyeon Yang. I thank Minha Hwang for sharing engineering expertise and managerial insights into the manufacturing processes. I thank James Porter, the editor of *DISK/TREND Reports*, for sharing his encyclopedic industry knowledge and for making the reports available. I thank Clayton Christensen for inspiration and encouraging a new approach to the innovator’s dilemma. Previous versions of the paper were presented at IIOC, TADC, and REER. An earlier version of the paper received Best Student Paper Award at the 11th Annual REER at Georgia Tech.
VITA

2002 B.A. (Latin American Area Studies), University of Tokyo.


2007 M.A. (Economics), University of Tokyo.

2009 M.A. (Economics), UCLA.

PUBLICATIONS

CHAPTER 1

Does Big Drive Out Small?
Entry, Exit, and Differentiation in the Supermarket Industry

1.1 Introduction

This paper empirically assesses the impact of large supermarkets’ entry on existing supermarkets of various sizes. Contrary to the conventional notion that big stores drive small rivals out of the market, the results indicate that big entrants drive out big and medium-sized incumbents, while benefiting small stores. The outcome is consistent with economic theories of product differentiation; i.e., store size is providing an important dimension to differentiate among retailers.

The deregulation of the Tokyo supermarket industry in the early 1990s provides a suitable setting to evaluate the differential impacts of large stores’ openings on existing stores. First, I analyze the incumbent stores’ responses to the entry events by ordered probit regressions, where their four alternative actions are (1) exit, (2) shrink floor size, (3) stay unchanged, and (4) expand floor size. Second, since entry events are based on big retailers’ choices of towns to enter, this prompts a concern over potential selection biases. I conduct a series of IV probit regressions in which I instrument the entry events by the big entrants’ affiliations with particular geographical markets. Finally, I employ tobit regressions to analyze the magnitude of changes in floor size, using incumbent stores’ percentage change in floor size as the dependent variable.

The results suggest that large entrants displace large and medium-size incumbents, but
small supermarkets’ survival rate actually *improves*. Large and medium stores seem to compete directly with the new rivals, while small incumbents are insulated by product differentiation and even benefit from the positive demand externality (additional flow of shoppers).

These findings have direct public policy implications. Regulators around the globe often restrict the entry of large retail outlets.¹ Such (anti-)competition policies are often based on the premise that big stores drive out small ones. However, when store size is the source of differentiation across retail services, the unintended consequences of size-based entry regulations would appear to include: (i) softer competition among large retailers, hence limited, pricier choices for consumers on a daily basis; and (ii) forgone profit opportunities for small stores, who could have benefited from the positive externalities of new big entrants.

1.2 Literature

This research contributes to three strands of economic literature: First, the paper offers new empirical evidence to support economic theories of product differentiation, by comparing the effect of large supermarkets’ entry on competing stores of various sizes. D’Aspremont et al. (1979), Shaked and Sutton (1982), and Perloff and Salop (1985) showed that product differentiation could soften price competition.² However, as Borenstein and Netz (1999) noted, theoretical work on product differentiation has produced few corresponding empirical studies. I examine a particular type of product differentiation in the retail industry: store size. The results highlight the role of product differentiation in relaxing competition. More specifically, this study supports the theoretical prediction by Zhu et al. (2006) that the tradeoff between the business-stealing effect and the positive demand externality depends

---

¹Zoning laws in Britain, France, Germany, India, Japan, Korea, and Poland restrict the development of large stores (Lewis, 2004). Even in the United States, state and local authorities and state courts often make the final decisions on whether to allow the entry of a new Wal-Mart store (Sobel and Dean, 2006).

²Theoretical predictions vary with respect to the extent of differentiation. For example, D’Aspremont et al. (1979) show that, in order to soften price competition, two firms would maximally differentiate on a Hotelling product line. Neven and Thissé (1990) and Irmen and Thissé (1998) extend this framework to competition over multiple product attributes. In contrast, Anderson et al. (1992) show that firms that compete in two dimensions may locate together at the center of the market under certain conditions. Still, the basic insight remains the same: Differentiation opens the possibility for softening competition.
on the degree of differentiation between the entrant and the incumbents.

Second, this paper augments the empirical work on entry and exit, by introducing the product-differentiation aspect to the analysis in two ways: by examining the differential impacts of entry on the exit rates of the incumbents of different sizes; and by analyzing the incumbents’ responses in terms of store-size changes. It is only recently that the economics of product differentiation has been studied explicitly in conjunction with entry and exit. Mazzeo (2002) uses a static framework, while Ellickson (2007) uses a dynamic structure. This paper takes an alternative approach to capture the dynamics of the phenomena, by exploiting exogenous regulatory changes and studying incumbents’ responses to the entry of bigger stores.

Third, this paper sheds new light on the economic analysis of the “Wal-Mart effects;” i.e., inquiries into the competitive effects of big entrants on incumbents, by introducing the viewpoint of product differentiation. Shoppers world-wide have witnessed the rise of big retailers, such as Wal-Mart and Carrefour. The proliferation of these large stores has stirred the debate over the consequences of their entry into local markets. Some analysts credit them with lowering prices, raising productivity, and making wider product variety available (Hausman and Leibtag, 2007; Basker, 2005); others blame them for destroying jobs and local businesses (Wal-Mart Watch, 2005). Basker’s (2007) survey summarizes the discoveries to date and concludes that incumbents’ exit due to Wal-Mart range between two to five stores at a county level. Detailed panel data of supermarkets at a sub-county level allow me to address this issue while mitigating concerns over spurious correlations.\(^3\)

\(^3\)County-level observations may mask the rise and fall of towns; therefore using sub-county-level data increases the relevance of empirical analysis to actual shopping behavior and competition. Additionally, stores often change their size over time. If, for instance, a small store expands its floor area, a simple census might count it as a small store exit and, simultaneously, a medium store entry as if the latter drove out the former. Panel data are a convenient way to capture actual exit patterns.
1.3 Industry and Data

The supermarket industry in Tokyo in the early 1990s provides a suitable testing ground for evaluating the impact of large stores’ openings on incumbents’ exit rates for three reasons: First, stores of various sizes compete in this sector, offering a laboratory to analyze the differential impacts of large entrants on incumbent stores of different sizes. Second, it is relatively easy to identify local markets geographically. Unlike shopping for, say, fashion apparel or consumer electronics, which tends to cluster in urban centers, shoppers stay close to their home as far as the day-to-day purchase of food staples is concerned. Third, the (exogenous) relaxation of entry regulations in 1990 allows me to identify the impact of large stores’ entry by comparing the exit rates of incumbent supermarkets in similar towns with and without an entry event.

1.3.1 Definition of Supermarket

This study follows the industry standard in defining a supermarket as: (i) a self-service store, with (ii) floor space of at least 231m$^2$ (2,486ft$^2$) and/or minimum annual revenue of 100 million yen ($1 million), and with (iii) over 30% of revenue from food products including (but not limited to) fish, meats, and vegetables.\footnote{This (establishment level) definition of a supermarket is analogous to the one used in the U.S.: “a store selling a full line of food products and generating at least $2 million in yearly revenues” (Ellickson, 2007, p. 48).} My data on supermarkets – based on the 1990 and 1995 editions of the trade press Japan Supermarket Directory – contain retail establishments satisfying these conditions.\footnote{For each store, the directory lists its name, street address, year of opening, operating firm, location type, building structure, parking capacity, gross revenue, types of merchandise sold, floor area, rented area, and the number of employees.}

Practical meanings of the definition will be clearer in comparison with other retail formats. Conditions (i) and (ii) distinguish supermarkets from more traditional specialty shops that sell fish, meats, or vegetables, which typically offer customized services and are smaller in size. Additionally, condition (iii) ensures that a supermarket is an outlet primarily for the retailing of food, as distinguished from convenience stores, drug stores, tobacco shops,
or department stores. More recently, a larger supermarket with a wider lineup of household merchandise is often called a GMS (general merchandise store), hypermart, or superstore. The industry standard does not exclude these types of supermarkets from the definition; neither does this paper.\(^6\) Hence the likes of Aeon, Daiei, or Seiyu (the Japanese equivalent to Wal-Mart, Tesco, Carrefour, or Metro) are incorporated in the subsequent empirical analysis.

### 1.3.2 Train Station-Centered Geographical Markets

In the suburbs of the Greater Tokyo region (which spans Tokyo, Kanagawa, Saitama, and Chiba prefectures), relevant geographical markets can be identified by train stations along major railways (Figure 1.1). This is because daily shopping activities in suburban Tokyo are concentrated around train stations. Since trains are the predominant means of transport for commuters, stations provide focal points for both shoppers and retailers (Kawaguchi, 1996, p. 175).

Note that 91.7% of the supermarkets shown in Figure 1 are located within 1.5km from a station. Hence, the market database from Toyokeizai’s (a private business press and think tank) *Metropolitan Commercial Map 1995* compiles demographic and retail-related information by 240 suburban “towns” along major railway lines in the region.\(^7\) Each “town” contains a geographical area within a radius of approximately 1.5 kilometers (0.93 mile) from the train station.

This disaggregated (and non-administrative) unit of observation is expected to be the most relevant to the actual grocery shopping pattern, and hence competition, for three

---

\(^6\)This is also the convention in the U.S. supermarket industry (Ellickson, 2007).

\(^7\) *Metropolitan Commercial Map 1995* contains a number of localities that are better characterized as central business districts rather than suburban residential areas. Hence I drop 29 local markets in which the number of regular commuters exceeds that of residents. I also drop one locality in which the train station was established only after 1990. Finally, I drop four towns for which the data on train users are unavailable, and another four towns for which exit rates are not available. All of this leaves 202 towns.

The listed variables include population, the number of households and commuters, the principal means of transport between residence and the nearest train station, commercial land price, the number of retail shops, and their aggregate floor space and revenue as of 1991. In addition, the growth rate between 1984 and 1989 of the number of regular commuters is taken from the Institution for Transport Policy Studies' annual survey (1986 and 1991 issues), directed by the Ministry of Transport.
Figure 1.1: Towns and Railways in the Greater Tokyo Area

Note: Solid lines represent major railway lines. Subway lines in the central business districts are not shown. Markers represent the locations of large and medium supermarkets (hence of towns) as of 1994. This map is for illustration purposes only, so some towns analyzed in this paper are outside this map. However, note that, of the 242 large and medium supermarkets pictured here, only 20 (or 8.3%) are located far from stations, underlining this paper’s focus on station-centered “towns.”


reasons concerning physical structure of transportation, consumers’ shopping patterns, and supermarkets’ operations. First, for a typical railway line, the average distance between two adjacent suburban stations is 3.8km (2.4 miles). The 1.5km-radius towns comfortably split the terrain.

Second, approximately 1km is the radius for the standard trade area and geographic market for Japanese supermarkets. The exact distance may vary between 0.9km and 1.5km depending on town characteristics, but both the Japan Fair Trade Commission (2005) and

\[\text{Ideally, a formal check of the market definition, such as the SSNIP test, would be desirable for further assurance. Without detailed household-level information and price data, however, the task of defining local markets requires some simplifying assumptions (such as mine).}

To avoid using existing geographic boundaries (e.g., zip codes or counties), Ellickson and Misra (2008) resorted to cluster analysis. My approach is similar in spirit, although the need for cluster analysis is precluded by the focus on train stations. Exact physical features may differ across countries, but the clustering of stores in my data resonates with their findings in the U.S. data that “these store clusters are somewhat larger than a typical zip code, but significantly smaller than the average county.”
numerous industry experts agree on these numbers.

Third, consumer surveys also substantiate the previous point. Most activities of a typical suburban housewife/husband take place within 0.5km from her/his dwelling. So it is unsurprising that urban geographers decided to characterize a daily grocery shopping as: (i) conducted by a housewife/husband, (ii) either on foot or by bicycle, and (iii) within 1.5km from her/his home at maximum (Arai, 1996, pp. 58-62; Kawaguchi, 1996, p. 161).

1.3.3 The 1990 Deregulation and Its Historical Context

A priori, there are no natural classification criteria for the sizes of retail outlets. In the regulatory context of the Japanese retail sector, however, a suitable categorization arises from the laws that define large, medium, and small stores. The Large-scale Retail Law, introduced in March 1974, sought to cap the new openings of any retail store with floor space 1,500m² (16,146ft²) and up.\(^9\) Later, the 1979 revision of the law further added another target category: stores with 500-1,499m² (5,382-16,145ft²). This paper defines stores with floor space of 1,500m² and above, 500-1,499m², and less than 500m² as “large,” “medium,” and “small,” respectively, because these are the size categories that had defined the evolution of the sector.\(^10\)

The new entry of supermarkets was particularly hindered during the 1980s. The Ministry of International Trade and Industry (MITI) tightened the enforcement of the regulations in October 1981, publicly dissuading retailers from opening new large stores. Consequently, only small supermarkets could open during this period (Figure 1.2: the period marked by “1”).

In May 1990, however, some of the prohibitive conditions in the Large-scale Retail Law

---

\(^9\)This was not the first time that government regulations targeted larger stores. Since the inception of department stores a century ago, various forms of legal entry barriers existed in Japan. The first was the Department Store Law, introduced in the 1930s in response to conventional retailers’ political activism. The law targeted nascent department stores and restricted their entry and operation. See Minakata (2004) for the industry context.

\(^10\)Although the threshold for being “large” (1,500m²) is relatively low by international comparison, the actual size of the large entrants in my study (4,992m², or 53,716ft², on average) is comparable to those of major supermarkets in other countries including the U.K., where the average store size of supermarket chains range between 5,800ft² and 45,200ft² (Smith, 2004).
were relaxed, which prompted a boom of new large outlets. The regulators began to accept all of the entry requests, regardless of store sizes, and abolished so-called “entry control areas” that were previously untouchable for newcomers. The driving force behind this policy shift was the pressure from the U.S. government to “liberalize” Japan’s domestic markets during the trade talks in the late 1980s. One can therefore regard this entry deregulation as an exogenous change in the industry environment.\footnote{Even if the Japanese retailers had exercised considerable bargaining power over the timing and the extent of the entry deregulation, the analysis and conclusion of this paper would remain unchanged. In that case, the results would actually underestimate the true impact of new entry because the incumbents, who were in a position to influence the policy change, should have been better prepared than otherwise for the intensified competition from new entrants. This direction of bias would not favor my result.}

This study focuses on the years immediately after the deregulation (Figure 1.2: the period marked by “\(2\)”) in order to avoid confounding the effects from the subsequent waves of deregulation.

The exogenous change in regulatory policies allows me to address the timing of big retailers’ entry. To capture the changes following the 1990 deregulation, I employ the 1990 and 1995 editions of \textit{Japan Supermarket Directory}, which list the store information as of September 1989 and 1994. This sample period gives a sufficient time interval for observing new entries and incumbents’ responses, which typically take at least a year or two, while limiting the risk of confounding the effects of various policy changes in the late 1990s.\footnote{An “entry” is the opening for business in my data set. The majority of the entries occurred in the first half of the 1990-1994 interval. The latest entry events occurred in April 1994. I confirmed that dropping the two towns (out of 27) that experienced entries in 1994 does not alter the results materially.} Moreover,
five years would allow big retailers to open some outlets but not in all the promising towns, mainly because of the illiquid nature of markets for huge properties and credit constraints. The resulting variation across towns allows me to identify the impact.

1.3.4 Descriptive Statistics of Towns and Supermarkets

From these two sets of data, I reconstruct the market configuration for each of the 202 localities, by connecting stores’ street addresses to those corresponding to towns. Table 1.1 presents summary statistics.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Population growth rate (%)</td>
<td>202</td>
<td>16.4</td>
<td>20.6</td>
<td>-15.6</td>
<td>135.5</td>
</tr>
<tr>
<td>Population</td>
<td>202</td>
<td>88,957</td>
<td>33,956</td>
<td>13,766</td>
<td>182,590</td>
</tr>
<tr>
<td>Retail revenue per capita (mn yen)</td>
<td>202</td>
<td>1.35</td>
<td>0.96</td>
<td>0.22</td>
<td>9.56</td>
</tr>
<tr>
<td>Num. retail shops</td>
<td>202</td>
<td>1,162</td>
<td>758</td>
<td>197</td>
<td>4,704</td>
</tr>
<tr>
<td>Retail floor area (m²)</td>
<td>202</td>
<td>72,800</td>
<td>46,385</td>
<td>10,471</td>
<td>345,600</td>
</tr>
<tr>
<td>Num. incumbents: large</td>
<td>202</td>
<td>1.25</td>
<td>1.13</td>
<td>0.00</td>
<td>6.00</td>
</tr>
<tr>
<td>Num. incumbents: medium</td>
<td>202</td>
<td>1.77</td>
<td>1.35</td>
<td>0.00</td>
<td>6.00</td>
</tr>
<tr>
<td>Num. incumbents: small</td>
<td>202</td>
<td>1.50</td>
<td>1.48</td>
<td>0.00</td>
<td>6.00</td>
</tr>
</tbody>
</table>

Note: Following the regulatory thresholds, store size categories are defined based on floor area (Small: 1-499 square meters, medium: 500-1,499 square meters, and large: 1,500 square meters or larger.)

The average town counts 88,957 residents. Tokyo is comparable to other urban areas in terms of population density. With 4,430 people per km², it ranks as mere number 128 among the world’s 189 major urban areas. Prominent European cities such as Madrid (5,680/km²), Athens (5,500/km²), London (5,290/km²), and Barcelona (5,210/km²) surpass Tokyo. In short, Tokyo is not Hong Kong (25,740/km²).

The growth potential of demand is proxied by the growth rate between 1984 and 1989 of the number of “regular commuters,” the sample mean of which is 16.4%. There are on average 1,162 retail shops of all categories, with total floor space of 72,800m² (783,619ft²). I intend to measure the “depth” of demand by total retail revenue per capita (which averaged 1.35 million yen, or $13,500), which reflects the extent of shoppers from outside the town

---

14 Train users with fixed-route commutation tickets for one month or longer (teiki-ken).
and other household characteristics such as income and taste. In 1989, a typical local market had 1.50 small, 1.77 medium, and 1.25 large food supermarkets.

The outcome variable of interest is the incumbent supermarkets’ responses between 1989 and 1994: exit, shrink, stay unchanged, or expand. Tables 1.2 and 1.3 display the store-level descriptive statistics.

<table>
<thead>
<tr>
<th>Table 1.2: Summary Statistics for Store-level Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>1. All incumbent stores</td>
</tr>
<tr>
<td>Floor size in 1989 ($m^2$)</td>
</tr>
<tr>
<td>Floor size in 1994 ($m^2$)</td>
</tr>
<tr>
<td>Change in floor size (%)</td>
</tr>
<tr>
<td>Indicator: Exit</td>
</tr>
<tr>
<td>Indicator: Expansion</td>
</tr>
<tr>
<td>Indicator: Shrinkage</td>
</tr>
<tr>
<td>Indicator: treatment (large entrant)</td>
</tr>
<tr>
<td>2. Large incumbent stores</td>
</tr>
<tr>
<td>Floor size in 1989 ($m^2$)</td>
</tr>
<tr>
<td>Floor size in 1994 ($m^2$)</td>
</tr>
<tr>
<td>Change in floor size (%)</td>
</tr>
<tr>
<td>Indicator: Exit</td>
</tr>
<tr>
<td>Indicator: Expansion</td>
</tr>
<tr>
<td>Indicator: Shrinkage</td>
</tr>
<tr>
<td>Indicator: treatment (large entrant)</td>
</tr>
</tbody>
</table>

Note: Following the regulatory thresholds, store size categories are defined based on floor area (Small: 1-499 square meters, medium: 500-1,499 square meters, and large: 1,500 square meters or larger.)

Out of the 912 incumbent supermarkets in 1989, 94 exited, leaving 818 stores in 1994. Those who survived expanded their floor size by 2.9% on average. There were 252 large, 358 medium, and 302 small incumbents in 1989. Across all stores, 10% exited, 15% expanded, and 10% shrank their floor sizes (“stay unchanged” is the omitted category that accounts for the remaining 65%). These percentages do not vary much by size although small stores are slightly more likely to exit (16%).

1.4 Empirical Analysis

This section presents the findings from three sets of empirical analyses: (1) ordered probit regressions of incumbents’ decisions to exit, shrink, stay unchanged, or expand (Section
Table 1.3: Summary Statistics for Store-level Observations (continued)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3. Medium incumbent stores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Floor size in 1989 (m²)</td>
<td>358</td>
<td>948</td>
<td>306</td>
<td>500</td>
<td>1499</td>
</tr>
<tr>
<td>Floor size in 1994 (m²)</td>
<td>328</td>
<td>969</td>
<td>383</td>
<td>390</td>
<td>3800</td>
</tr>
<tr>
<td>Change in floor size (%)</td>
<td>328</td>
<td>4.0</td>
<td>35.0</td>
<td>-65.4</td>
<td>503.2</td>
</tr>
<tr>
<td>Indicator: Exit</td>
<td>358</td>
<td>.08</td>
<td>.28</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Indicator: Expansion</td>
<td>358</td>
<td>.16</td>
<td>.37</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Indicator: Shrinkage</td>
<td>358</td>
<td>.14</td>
<td>.35</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Indicator: treatment (large entrant)</td>
<td>358</td>
<td>.12</td>
<td>.32</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4. Small incumbent stores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Floor size in 1989 (m²)</td>
<td>302</td>
<td>300</td>
<td>135</td>
<td>57</td>
<td>499</td>
</tr>
<tr>
<td>Floor size in 1994 (m²)</td>
<td>254</td>
<td>313</td>
<td>141</td>
<td>62</td>
<td>385</td>
</tr>
<tr>
<td>Change in floor size (%)</td>
<td>254</td>
<td>4.3</td>
<td>30.9</td>
<td>-33.3</td>
<td>383.2</td>
</tr>
<tr>
<td>Indicator: Exit</td>
<td>302</td>
<td>.16</td>
<td>.37</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Indicator: Expansion</td>
<td>302</td>
<td>.10</td>
<td>.30</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Indicator: Shrinkage</td>
<td>302</td>
<td>.04</td>
<td>.19</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Indicator: treatment (large entrant)</td>
<td>302</td>
<td>.09</td>
<td>.28</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: Following the regulatory thresholds, store size categories are defined based on floor area (Small: 1-499 square meters, medium: 500-1,499 square meters, and large: 1,500 square meters or larger.)

1.4.1); (2) the set of similar (binary) probit regressions with geographical instruments for big entrants’ town choice (Section 1.4.2); and (3) tobit regressions of incumbents’ change in floor space (Section 1.4.3).

In all of the specifications, the identification of the entry effects on incumbents relies on the following two features of the study. First, the framework assumes independent local markets in which the following three events occur: (i) The incumbent supermarkets of various sizes, without knowledge of the entry deregulation, operate from before 1990;15 (ii) Upon the deregulation in 1990, the potential (big) entrants observe the existing store configuration in all local markets and choose towns to enter; and (iii) The incumbents observe the actual entrants and decide by 1994 whether to continue in business (and if so, whether to change own store size). These timing and informational assumptions are motivated by the historical/institutional background of the industry (see Section 1.3).

Second, the data set contains observations of similar towns (and individual stores within each of them) with and without entry events, both before (1989) and after (1994) the entry

---

15It is reasonable to assume a lack of anticipation because most incumbents had opened by the mid 1980s, long before the U.S.-Japan trade talks started discussing the retail deregulation.
deregulation. This variation in data, together with the institutional background, allows me to identify the impact of big entrants on existing supermarkets.

1.4.1 Ordered Probit: Exit, Shrink, Stay Unchanged, or Expand

The first set of results is based on ordered probit regressions (Table 1.4). The dependent variable is the four discrete alternatives for an incumbent: (i) exit; (ii) stay and shrink; (iii) stay unchanged; or (iv) stay and expand, ordered in this manner.

Formally, store $i$’s observed choice is

$$
y_i = \begin{cases} 
exit & \text{if } y_i^* \leq c_1 \\
shrink size & \text{if } c_1 < y_i^* \leq c_2 \\
stay unchanged & \text{if } c_2 < y_i^* \leq c_3 \\
expand size & \text{if } c_3 < y_i^*,
\end{cases}$$

(1.1)

where $c_1$, $c_2$, and $c_3$ are threshold parameters. I specify the latent variable $y_i^*$ representing incumbent supermarket $i$’s profit as

$$y_i^* = \alpha_{SIZE_i} + \beta_{SIZE_i} D_i + X_i \gamma + \varepsilon_i,$$

(1.2)

where $D_i$ is the dummy variable that indicates the entry of a new big supermarket (“treatment”) in the town in which store $i$ operates.

The vector $X_i$ includes the following town characteristics: Population growth rate between 1984 and 1989, Population, Retail revenue per capita, the Number of retail shops (of any kind), and Retail floor area, (i.e., the town’s total floor space that was dedicated to retail trade).\textsuperscript{16} It also incorporates the number of existing rival supermarkets in the town as of summer 1989, specified as a linear combination of the numbers of large, medium, and small

\textsuperscript{16}Except for Population growth rate, the values at the beginning of the deregulation are used. Each variable’s squared term is also included, to capture possible nonlinearities in the way that the town characteristics affect the outcome variables. More flexible specifications would be preferable in principle, but the relatively small sample size limits the extent of higher-order polynomials.
incumbents, and their squared terms. The error term ε_i is i.i.d. standard normal.

Large, medium, and small incumbents may have different intercepts, α_{SIZE_i}, where SIZE_i ∈ {large, medium, small}, with small as the omitted category. The coefficients of interest are the effects of new entry, β_{SIZE_i}, which are also allowed to vary by size classes of incumbents.

The estimation results in Table 1.4 suggest that the entry of new large supermarkets negatively affects large and medium incumbents, whereas the impact tends to be positive for small ones. Column (1) is the simplest specification with the entry dummy variable and incumbents’ size classes (large, medium, and small).

Table 1.4: Ordered Probit Regressions of Decision to Exit < Shrink < Unchanged < Expand

<table>
<thead>
<tr>
<th>Dep. var.:</th>
<th>Decision to Exit &lt; Shrink &lt; Unchanged &lt; Expand</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Treated: large</td>
<td>0.03 (.30)</td>
</tr>
<tr>
<td>Treated: medium</td>
<td>-0.56 (.20) ***</td>
</tr>
<tr>
<td>Treated: small</td>
<td>0.28 (.17) *</td>
</tr>
<tr>
<td>Treated * floor</td>
<td>1.69e-4 (0.81e-4) **</td>
</tr>
<tr>
<td>Floor</td>
<td>1.28e-5 (1.4e-5)</td>
</tr>
<tr>
<td>Large</td>
<td>0.29 (.10) ***</td>
</tr>
<tr>
<td>Medium</td>
<td>0.23 (.10) ***</td>
</tr>
<tr>
<td>Constant (=Small)</td>
<td>–</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
</tr>
<tr>
<td>Instruments</td>
<td>No</td>
</tr>
<tr>
<td>Observations</td>
<td>912</td>
</tr>
<tr>
<td>Pseudo R^2</td>
<td>.01</td>
</tr>
</tbody>
</table>

Note: Standard errors (clustered by 202 towns) in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%.

In addition to the three size classes, column (2) incorporates incumbents’ floor space (in m^2 as of 1989) together with its interaction term (Floor and Treated * Floor), controlling

---

17 An alternative specification that uses dummy variables was also tried (unreported), but does not materially alter the results.

18 Although there is no theoretical reason a priori for ordering shrink < unchanged < expand, I proceed with this particular ordering because it fits the data best.

19 I also run the same ordered probit regressions separately for each of the three size classes (unreported). The results are qualitatively similar but less precisely estimated due to smaller samples. An F-test of a null hypothesis β_{large} = β_{medium} = β_{small} is rejected at the 1% level, which indicates that the entry effects are indeed different across incumbents’ size classes.
for potential heterogeneity within each category. The positive coefficient estimates on this detailed size measure suggest that, within each size class, larger stores tend to fare better.

Finally, column (3) controls for the town characteristics ($X_i$ in equation (1.2)). In all of the specifications, the impact on medium stores is negative and statistically significant at 1% level. Negative impact on large stores and positive impact on small stores are less precisely estimated but consistently appear across all specifications (except for the case of large stores in column (1)).

1.4.2 Geographical Instruments for Entrants’ Town Choice

Potential selection biases are a cause of concern: It may be that the big-box stores were able to foresee the towns where incumbents were most likely to exit. In that case, the negative impacts on large and medium incumbents could be over-estimated (in magnitude), while the positive impacts on small incumbents might be under-estimated (again, in magnitude).

To address this issue of potential selection on unobservables, I construct instrumental variables (for the big entrants’ choices of town to enter) based on the following industry characteristics: Each chain retailer usually operates within predetermined geographic areas, which leaves only a subset of the total of the 202 towns in that firm’s choice set for entry. In particular, two different urban structures are relevant: (i) railway lines; and (ii) prefectures.

Regarding railways, many of the major retailers are closely affiliated with particular railway lines and their train operators. Some retailers are the grocery-store divisions of the conglomerates that also own and operate railway and property businesses.\textsuperscript{20} The institutional context is as follows: First, the railway transport enterprise involves massive property investment, in both land strips beneath railroads and train station structures. Second, over time, the train operators diversified into real estate brokerage/development activities and retail services due to strong complementarities that surround property deals. Third, since location is central to successful retail services, major retailers either are engaged in close

\textsuperscript{20}Examples include Odakyu, Keio, Tokyu, Keikyu, Seibu, and Tobu. See Masuda (2002) for the historical background on the railway networks and urban development in Tokyo.
It is therefore not surprising that the chain operators tend to open big stores along their affiliated railway lines, exploiting informational advantages in local real estate markets.\footnote{This advantage of an entrant is uncorrelated with incumbents’ decisions to exit or change floor size because it is a strictly private benefit. Let us also note that such an informational clout does not translate into an entrant’s ability to “kick out” incumbents (by means other than product-market competition). See Masuda (2002) for the close interaction between the railway transport, real estate, and retail businesses.}

With respect to prefectures, a chain retailer is often rooted in a certain prefecture as a matter of geographic origin. In addition to the informational advantages in local property markets, geographic familiarity brings two benefits. First, geographical proximity to its supplier/distribution networks facilitates the arrangement of logistics. Second, the familiarity with local rules (such as ordinances) ensures low probabilities of legal and/or political disruptions in opening new big stores. Therefore, the propensity of entry is higher when a town belongs to a major retailer’s prefecture of origin.\footnote{This instrumentation strategy is similar in spirit to the one employed in the Wal-Mart literature: A location’s physical distance from the company’s headquarter in Bentonville, Arkansas, predicts the likelihood and timing of new store openings in that town (see Basker, 2007).}

I identify each big retailer’s geographic specialization from the Appendices of the Metropolitan Commercial Map (Toyokeizai 1995) and various issues of Large-scale Retail Shops Directory (an annual publication also from Toyokeizai). Two sets of IVs are constructed: (IV-1) the number of big retail firms that operate along each railway line, and prefecture dummy variables; and (IV-2) the same railway-based IVs, and the number of big retail firms that operate from each prefecture.

Due to the institutional background in the above, a potential entrant enjoys significant informational and cost advantages when opening a new store in its familiar geographic areas (either along specific railway lines, in certain prefectures, or both).\footnote{See Section 3 for the details of the urban structure in the Greater Tokyo region. One rationale for such local specializations and informational advantages is the illiquid nature of markets for huge properties, which creates an environment characterized by imperfect information.} Thus, towns that happen to be located in the “backyards” of many big-box operators (i.e., towns that are located along railway lines and prefectures that are home to many big retailers) are more likely to experience entry events, for economic mechanisms that are unrelated to the likelihood of incumbents’ exit or expansion (i.e., the outcome variables of interest).
Tables 1.5, 1.6, and 1.7 show the results of binary IV-probit regressions. Ordered IV-probit regressions may achieve higher efficiency, in principle, but the actual estimation would become computationally expensive. As a logically consistent alternative method, I conduct three binary IV-probit regressions that divide an incumbent’s four choices differently: (Panel-A) “exit” or “shrink/unchanged/expand”; (Panel-B) “exit/shrink” or “unchanged/expand”; and (Panel-C) “exit/shrink/unchanged” or “expand.”

Table 1.5: Binary IV-Probit Regressions (1 of 3)

<table>
<thead>
<tr>
<th>Dep. var.: Decision Not to Exit</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. “Exit” or “Shrink/Unchanged/Expand”</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treated: large</td>
<td>-0.33</td>
<td>-1.36</td>
<td>-1.42</td>
<td>-2.04</td>
<td>-3.74</td>
</tr>
<tr>
<td></td>
<td>(.28)</td>
<td>(.58)***</td>
<td>(.66)**</td>
<td>(2.61)</td>
<td>(4.75)</td>
</tr>
<tr>
<td>Treated: medium</td>
<td>-0.49</td>
<td>-0.83</td>
<td>-0.90</td>
<td>-0.60</td>
<td>-0.42</td>
</tr>
<tr>
<td></td>
<td>(.31)</td>
<td>(.32)**</td>
<td>(.38)**</td>
<td>(1.22)</td>
<td>(4.01)</td>
</tr>
<tr>
<td>Treated: small</td>
<td>0.22</td>
<td>0.12</td>
<td>0.08</td>
<td>1.20</td>
<td>1.40</td>
</tr>
<tr>
<td></td>
<td>(.41)</td>
<td>(.41)</td>
<td>(.46)</td>
<td>(1.18)</td>
<td>(5.07)</td>
</tr>
<tr>
<td>Treated * floor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.17e-4</td>
<td>3.11e-4</td>
<td>3.18e-4</td>
<td>9.46e-4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.65e-4)*</td>
<td>(1.67e-4)*</td>
<td>(4.45e-4)</td>
<td>(7.84e-4)</td>
<td></td>
</tr>
<tr>
<td>Floor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.36e-5</td>
<td>8.62e-5</td>
<td>8.56e-5</td>
<td>7.03e-5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.24e-5)</td>
<td>(5.64e-5)</td>
<td>(6.18e-5)</td>
<td>(5.35e-5)</td>
<td></td>
</tr>
<tr>
<td>Large</td>
<td>0.59</td>
<td>0.29</td>
<td>0.25</td>
<td>0.42</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>(.17)***</td>
<td>(.25)</td>
<td>(.26)</td>
<td>(.33)</td>
<td>(.28)</td>
</tr>
<tr>
<td>Medium</td>
<td>0.47</td>
<td>0.43</td>
<td>0.45</td>
<td>0.50</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>(1.4)***</td>
<td>(1.4)***</td>
<td>(1.5)***</td>
<td>(1.8)***</td>
<td>(1.8)**</td>
</tr>
<tr>
<td>Constant (=Small)</td>
<td>0.98</td>
<td>0.96</td>
<td>1.18</td>
<td>0.96</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>(.10)***</td>
<td>(.10)***</td>
<td>(.56)**</td>
<td>(.70)</td>
<td>(1.90)</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Instruments</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>IV-1</td>
<td>IV-2</td>
</tr>
<tr>
<td>Observations</td>
<td>912</td>
<td>912</td>
<td>912</td>
<td>912</td>
<td>912</td>
</tr>
<tr>
<td>Pseudo R²</td>
<td>.03</td>
<td>.05</td>
<td>.09</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Note: Standard errors (clustered by 202 towns) in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%.

For each of Panels A, B, and C, columns (1), (2), and (3) use no IV, while (4) and (5) use IV-1 and IV-2, respectively. The results are qualitatively similar across columns and panels: An entry’s impact is negative on large and medium incumbents but positive on small ones. The order of magnitude is also similar. Thus, the potential issue of selection on unobservables is unlikely to be driving my baseline findings using ordered probit (Section 1.4.1).
### Table 1.6: Binary IV-Probit Regressions (2 of 3)

<table>
<thead>
<tr>
<th>B. “Exit/Shrink” or “Unchanged/Expand”</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treated: large</td>
<td>-0.13</td>
<td>-0.72</td>
<td>-0.46</td>
<td>-1.65</td>
<td>-3.25</td>
</tr>
<tr>
<td></td>
<td>(.35)</td>
<td>(.49)</td>
<td>(.54)</td>
<td>(1.42)</td>
<td>(2.50)</td>
</tr>
<tr>
<td>Treated: medium</td>
<td>-0.52</td>
<td>-0.65</td>
<td>-0.61</td>
<td>-0.10</td>
<td>-0.25</td>
</tr>
<tr>
<td></td>
<td>(.24)**</td>
<td>(.25)**</td>
<td>(.24)**</td>
<td>(.83)</td>
<td>(1.08)</td>
</tr>
<tr>
<td>Treated: small</td>
<td>0.37</td>
<td>0.33</td>
<td>0.35</td>
<td>0.91</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>(.41)</td>
<td>(.41)</td>
<td>(.46)</td>
<td>(1.03)</td>
<td>(1.10)</td>
</tr>
<tr>
<td>Treated * floor</td>
<td>–</td>
<td>1.30e-4</td>
<td>0.91e-4</td>
<td>4.45e-4</td>
<td>7.32e-4</td>
</tr>
<tr>
<td></td>
<td>(.81e-4)</td>
<td>(.83e-4)</td>
<td>(2.12e-4)**</td>
<td>(3.33e-4)**</td>
<td></td>
</tr>
<tr>
<td>Floor</td>
<td>–</td>
<td>0.53e-5</td>
<td>1.38e-5</td>
<td>0.60e-5</td>
<td>-0.19e-5</td>
</tr>
<tr>
<td></td>
<td>(.188e-5)</td>
<td>(.199e-5)</td>
<td>(2.63e-5)</td>
<td>(2.93e-5)</td>
<td></td>
</tr>
<tr>
<td>Large</td>
<td>0.09</td>
<td>0.06</td>
<td>0.05</td>
<td>0.12</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>(.14)</td>
<td>(.17)</td>
<td>(.18)</td>
<td>(.21)</td>
<td>(.22)</td>
</tr>
<tr>
<td>Medium</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.02</td>
<td>-0.02</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>(.12)</td>
<td>(.12)</td>
<td>(.12)</td>
<td>(.15)</td>
<td>(.14)</td>
</tr>
<tr>
<td>Constant (=Small)</td>
<td>0.83</td>
<td>0.83</td>
<td>0.55</td>
<td>0.25</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>(.09)**</td>
<td>(.09)**</td>
<td>(.49)</td>
<td>(.53)</td>
<td>(.59)</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Instruments</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>IV-1</td>
<td>IV-2</td>
</tr>
<tr>
<td>Observations</td>
<td>912</td>
<td>912</td>
<td>912</td>
<td>912</td>
<td>912</td>
</tr>
<tr>
<td>Pseudo $R^2$</td>
<td>.01</td>
<td>.01</td>
<td>.04</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

*Note:* Standard errors (clustered by 202 towns) in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%.

### 1.4.3 Tobit: Percentage Change in Incumbents’ Floor Size

In the preceding analyses, I characterize incumbents’ decisions as discrete choice problems. However, changes in floor size take continuous values. Some incumbents more than double their floor spaces by converting one-story buildings into two-story ones, while others increased their sizes only 10% by renting adjacent spaces in a commercial complex. Similarly, stores shrink floor sizes by different degrees.

In this section, I incorporate such heterogeneity in incumbents’ size changes by employing tobit regressions. The dependent variable is *the realized percentage change in floor size*. Recycling the notation from the preceding probit regressions, the observed outcome is now
Table 1.7: Binary IV-Probit Regressions (3 of 3)

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dep. var.: Decision Not to Exit, Shrink, or Stay Unchanged</strong></td>
<td><strong>(1)</strong></td>
<td><strong>(2)</strong></td>
<td><strong>(3)</strong></td>
<td><strong>(4)</strong></td>
</tr>
<tr>
<td>Treated: large</td>
<td>0.23</td>
<td>-0.66</td>
<td>-0.54</td>
<td>-3.56</td>
</tr>
<tr>
<td>(0.30)</td>
<td>(0.60)</td>
<td>(0.68)</td>
<td>(1.53)**</td>
<td>(1.85)</td>
</tr>
<tr>
<td>Treated: medium</td>
<td>-0.73</td>
<td>-0.89</td>
<td>-1.02</td>
<td>0.50</td>
</tr>
<tr>
<td>(0.33)**</td>
<td>(0.31)**</td>
<td>(0.30)**</td>
<td>(0.77)</td>
<td>(1.02)</td>
</tr>
<tr>
<td>Treated: small</td>
<td>0.27</td>
<td>0.22</td>
<td>0.31</td>
<td>2.22</td>
</tr>
<tr>
<td>(0.31)</td>
<td>(0.32)</td>
<td>(0.29)</td>
<td>(1.12)**</td>
<td>(1.25)*</td>
</tr>
<tr>
<td>Treated * floor</td>
<td>–</td>
<td>1.57e-4</td>
<td>1.33e-4</td>
<td>5.72e-4</td>
</tr>
<tr>
<td>(–)</td>
<td>(9.7e-4)</td>
<td>(1.03e-4)</td>
<td>(2.15e-4)**</td>
<td>(2.62e-4)</td>
</tr>
<tr>
<td>Floor</td>
<td>–</td>
<td>1.19e-5</td>
<td>0.46</td>
<td>-1.61e-5</td>
</tr>
<tr>
<td>(–)</td>
<td>(1.90e-5)</td>
<td>(2.00e-5)</td>
<td>(2.56e-5)</td>
<td>(2.85e-5)</td>
</tr>
<tr>
<td>Large</td>
<td>0.39</td>
<td>0.33</td>
<td>0.39</td>
<td>0.72</td>
</tr>
<tr>
<td>(0.13)**</td>
<td>(0.17)*</td>
<td>(0.18)**</td>
<td>(0.19)**</td>
<td>(0.21)**</td>
</tr>
<tr>
<td>Medium</td>
<td>0.36</td>
<td>0.35</td>
<td>0.31</td>
<td>0.23</td>
</tr>
<tr>
<td>(0.14)**</td>
<td>(0.14)**</td>
<td>(0.14)**</td>
<td>(0.17)</td>
<td>(0.17)</td>
</tr>
<tr>
<td>Constant (=Small)</td>
<td>-1.29</td>
<td>-1.30</td>
<td>-0.74</td>
<td>-1.12</td>
</tr>
<tr>
<td>(0.10)**</td>
<td>(0.10)**</td>
<td>(0.44)**</td>
<td>(0.58)*</td>
<td>(0.61)**</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Instruments</td>
<td>No</td>
<td>NO</td>
<td>No</td>
<td>IV-1</td>
</tr>
<tr>
<td>Observations</td>
<td>912</td>
<td>912</td>
<td>912</td>
<td>912</td>
</tr>
<tr>
<td>Pseudo R²</td>
<td>.02</td>
<td>.03</td>
<td>.06</td>
<td>–</td>
</tr>
</tbody>
</table>

*Note: Standard errors (clustered by 202 towns) in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%.*

The latent variable is the desired percentage change in floor space between 1989 and 1994. The underlying economic model assumes some fixed sunk costs that a store must incur when changing its floor size (i.e., when \( y_i \neq 0 \)).

Recall that almost two-thirds of the incumbents stay unchanged, hence the dependent variable has a mass of observations at \( y_i = 0 \). For this reason, I explicitly include \( y_i = 0 \) as a separate case in equation (1.3), which makes this censored regression different from the standard tobit.

I address the issue of lumpy observations by analyzing the exit/shrink and expansion
decisions separately. First, I concentrate on the floor reduction decision by running two-sided tobit where the focus is on the cases with $c_1 < y_i^* \leq c_2$. The percentage change in floor space is left-censored at $-100$ (i.e., exit) and right-censored at 0, the latter of which encompasses stores’ decisions to both “stay unchanged” and “expand.”

Second, I exclusively analyze the floor expansion decision by conducting one-sided tobit where I focus on the cases with $c_3 < y_i^*$. Here the percentage change in floor space is left-censored at 0 and not right-censored.

Table 1.8 presents two-sided tobit estimation results on incumbents’ decisions to shrink floor space. Here again, large and medium supermarkets respond “negatively” to the new entry of large rivals by shrinking their own size, or by exiting the town altogether. In contrast, small supermarkets are less inclined to shrink or exit.

<table>
<thead>
<tr>
<th>Dep. var.:</th>
<th>Change in Floor Size (%)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Treated: large</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>(54.79)</td>
<td>(111.22)</td>
<td>(113.45)</td>
<td>(73.30)</td>
<td>(704.96)</td>
<td></td>
</tr>
<tr>
<td>-33.13</td>
<td>-169.27</td>
<td>-132.86</td>
<td>-61.47</td>
<td>-663.08</td>
<td></td>
</tr>
<tr>
<td>(54.79)</td>
<td>(111.22)</td>
<td>(113.45)</td>
<td>(73.30)</td>
<td>(704.96)</td>
<td></td>
</tr>
<tr>
<td>Treated: medium</td>
<td>(65.51)</td>
<td>(63.21)</td>
<td>(64.38)</td>
<td>(48.97)</td>
<td>(306.81)</td>
</tr>
<tr>
<td>(40.95)**</td>
<td>(47.86)**</td>
<td>(94.75)**</td>
<td>(46.26)</td>
<td>(266.22)</td>
<td></td>
</tr>
<tr>
<td>-96.04</td>
<td>-128.44</td>
<td>-119.95</td>
<td>-12.49</td>
<td>-59.64</td>
<td></td>
</tr>
<tr>
<td>(63.15)</td>
<td>(63.21)</td>
<td>(64.38)</td>
<td>(48.97)</td>
<td>(306.81)</td>
<td></td>
</tr>
<tr>
<td>Treated: small</td>
<td>(65.51)</td>
<td>(63.21)</td>
<td>(64.38)</td>
<td>(48.97)</td>
<td>(306.81)</td>
</tr>
<tr>
<td>(40.95)**</td>
<td>(47.86)**</td>
<td>(94.75)**</td>
<td>(46.26)</td>
<td>(266.22)</td>
<td></td>
</tr>
<tr>
<td>65.51</td>
<td>55.71</td>
<td>56.01</td>
<td>30.33</td>
<td>224.67</td>
<td></td>
</tr>
<tr>
<td>(63.15)</td>
<td>(63.21)</td>
<td>(64.38)</td>
<td>(48.97)</td>
<td>(306.81)</td>
<td></td>
</tr>
<tr>
<td>Treated * floor</td>
<td>(65.51)</td>
<td>(63.21)</td>
<td>(64.38)</td>
<td>(48.97)</td>
<td>(306.81)</td>
</tr>
<tr>
<td>(23.49e-3)</td>
<td>(23.21e-3)</td>
<td>(8.60e-3)**</td>
<td>(102.75e-3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31.30e-3</td>
<td>25.46e-3</td>
<td>17.31e-3</td>
<td>153.36e-3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(24.39e-3)</td>
<td>(24.21e-3)</td>
<td>(8.60e-3)**</td>
<td>(102.75e-3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Floor</td>
<td>(65.51)</td>
<td>(63.21)</td>
<td>(64.38)</td>
<td>(48.97)</td>
<td>(306.81)</td>
</tr>
<tr>
<td>(46.35e-4)</td>
<td>(48.38e-4)</td>
<td>(12.34e-4)</td>
<td>(65.91e-4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31.75e-4</td>
<td>48.84e-4</td>
<td>4.22e-4</td>
<td>29.44e-4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(25.36)**</td>
<td>(33.29)</td>
<td>(33.78)</td>
<td>(9.76)</td>
<td>(47.31)</td>
<td></td>
</tr>
<tr>
<td>29.30</td>
<td>27.15</td>
<td>33.25</td>
<td>0.73</td>
<td>29.27</td>
<td></td>
</tr>
<tr>
<td>(22.58)</td>
<td>(22.64)</td>
<td>(22.55)</td>
<td>(5.29)</td>
<td>(29.39)</td>
<td></td>
</tr>
<tr>
<td>Constant (=Small)</td>
<td>(23.28)**</td>
<td>(23.13)**</td>
<td>(86.08)</td>
<td>(31.14)</td>
<td>(143.81)</td>
</tr>
<tr>
<td>146.36</td>
<td>144.41</td>
<td>111.59</td>
<td>12.11</td>
<td>53.10</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Standard errors (clustered by 202 towns) in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%.

24 More fundamentally, there is no theoretical reason to impose symmetry between shrinkage and expansion rates. Two separate censored regressions allow for the possibility of asymmetric effects of entry.
Table 1.9 tells a similar story regarding the floor expansion decision of incumbents. Medium stores tend not to expand their shopping space, followed by large incumbents to a lesser degree. Again, small supermarkets seem to respond more “aggressively” by expanding their size although coefficients are not very precisely estimated.

**Table 1.9: Expansion: One-Sided Tobit Regressions of Change in Floor Size (%)**

<table>
<thead>
<tr>
<th>Dep. var.:</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treated: large</td>
<td>11.98</td>
<td>-42.98</td>
<td>-24.45</td>
<td>-259.82</td>
<td>-133.97</td>
</tr>
<tr>
<td></td>
<td>(.23)</td>
<td>(48.91)</td>
<td>(50.15)</td>
<td>(150.37)*</td>
<td>(154.08)</td>
</tr>
<tr>
<td>Treated: medium</td>
<td>-66.47</td>
<td>-75.52</td>
<td>-79.45</td>
<td>66.90</td>
<td>82.17</td>
</tr>
<tr>
<td></td>
<td>(30.18)**</td>
<td>(30.93)**</td>
<td>(31.40)**</td>
<td>(76.88)</td>
<td>(74.11)</td>
</tr>
<tr>
<td>Treated: small</td>
<td>31.70</td>
<td>28.76</td>
<td>33.29</td>
<td>184.73</td>
<td>175.06</td>
</tr>
<tr>
<td></td>
<td>(25.68)</td>
<td>(25.77)</td>
<td>(28.35)</td>
<td>(114.36)*</td>
<td>(109.14)*</td>
</tr>
<tr>
<td>Treated * floor</td>
<td>–</td>
<td>9.13e-3</td>
<td>5.66e-3</td>
<td>43.65e-3</td>
<td>31.44e-3</td>
</tr>
<tr>
<td></td>
<td>(–)</td>
<td>(6.91e-3)</td>
<td>(7.05e-3)</td>
<td>(21.15e-3)**</td>
<td>(21.45e-3)</td>
</tr>
<tr>
<td>Floor</td>
<td>–</td>
<td>3.89e-4</td>
<td>-.16e-4</td>
<td>-16.22e-4</td>
<td>-10.43e-4</td>
</tr>
<tr>
<td></td>
<td>(–)</td>
<td>(18.98e-4)</td>
<td>(19.03e-4)</td>
<td>(25.03e-4)</td>
<td>(25.21e-4)</td>
</tr>
<tr>
<td>Large</td>
<td>21.92</td>
<td>19.97</td>
<td>23.88</td>
<td>53.88</td>
<td>44.66</td>
</tr>
<tr>
<td></td>
<td>(12.15)*</td>
<td>(15.45)</td>
<td>(15.42)</td>
<td>(21.23)**</td>
<td>(20.48)**</td>
</tr>
<tr>
<td>Medium</td>
<td>28.47</td>
<td>28.22</td>
<td>24.21</td>
<td>15.20</td>
<td>13.69</td>
</tr>
<tr>
<td></td>
<td>(11.22)**</td>
<td>(11.29)**</td>
<td>(11.24)**</td>
<td>(16.50)</td>
<td>(15.77)</td>
</tr>
<tr>
<td>Constant (=Small)</td>
<td>-111.75</td>
<td>-111.89</td>
<td>-47.51</td>
<td>-98.83</td>
<td>-103.68</td>
</tr>
<tr>
<td></td>
<td>(11.95)***</td>
<td>(11.97)***</td>
<td>(44.57)</td>
<td>(55.73)*</td>
<td>(54.88)*</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Instruments</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>IV-1</td>
<td>IV-2</td>
</tr>
<tr>
<td>Observations</td>
<td>912</td>
<td>912</td>
<td>912</td>
<td>912</td>
<td>912</td>
</tr>
<tr>
<td>Left-censored at 0%</td>
<td>775</td>
<td>775</td>
<td>775</td>
<td>775</td>
<td>775</td>
</tr>
<tr>
<td>Uncensored</td>
<td>137</td>
<td>137</td>
<td>137</td>
<td>137</td>
<td>137</td>
</tr>
<tr>
<td>Pseudo $R^2$</td>
<td>.01</td>
<td>.01</td>
<td>.02</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

*Note: Standard errors (clustered by 202 towns) in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%.*

### 1.4.4 Discussion

The basic finding from these empirical analyses is that the entry of new large rivals affect incumbents differently depending on the latter’s sizes. In other words, incumbents of different sizes face different incentives in responding to the entry shock.

This simple observation implies that large, medium, and small supermarkets offer differentiated services from the perspective of consumers. Hence the results highlight the role of product differentiation in relaxing competition, which is the common insight from numerous theoretical studies (see Section 1.2).
How do supermarkets’ retail services differ by store sizes? The most important differenti-
tation mechanism is that, while larger supermarkets offer a wider variety of merchandise for
weekend shopping trips, consumers use their nearest (often small) supermarkets for quick,
daily purchase of fresh foods (fish, vegetables, and meats). The latter is the principal mode
of shopping in Tokyo.

Three demand-side factors may explain the high frequency of fresh-food shopping. First,
houses (and therefore refrigerators) are small in Tokyo, leaving no storage space. Second,
many Japanese are obsessed with the freshness of meat, vegetables, and especially fish.
Third, the female labor-force participation rate is lower in Japan than in most developed
economies, so that there are still many “professional” housewives.

The most intriguing feature of the estimation results is that small incumbents are not
just insulated from the increased competition at the top end of the size spectrum. Small
supermarkets seem to benefit from the entry of new big rivals. An economic interpretation
of this finding would need to rely on some sort of positive externalities from the new entrant.
Since it is not very conceivable to imagine new entrants directly facilitating small incumbents
on the supply side, the positive externalities likely come from the demand side.

One example is the increased traffic of shoppers in town, attracted by new big retailers.
Such is the theoretical model of Zhu et al. (2006), which incorporates both product differ-
entiation and positive demand externality of new entry in the retail context. My findings
fit well with their prediction that the tradeoff between the business-stealing effect (i.e., in-
creased competition from new entrants) and the positive demand externality (i.e., increased
demand thanks to new entrants) hinges on the degree of differentiation between entrants
and incumbents.

Existing large and medium supermarkets suffer higher exit rates because they directly
compete with the new big rivals at the same end of the store-size spectrum. Small supermar-
kets, in contrast, stand at the other end of the product space, serving different demands, and
are thus insulated from the business-stealing effect of the new entry. On average, the small
incumbents benefit from the entrants probably because the latter attracts an additional flow
of shoppers to the neighborhood.

Aside from the decisions to exit, incumbents’ incentives in changing store size seem quite nuanced in some size categories. First, the incentive for medium supermarkets not to expand is understandable. They already suffer from the proximity (in size) to the new competitor. There is no reason to spend their shrinking profits to get even closer in size to their big rivals.

Second, the story could be more complicated for large incumbents. On the one hand, they, too, may want to follow medium stores’ strategy and distance themselves from the new entrants by shrinking floor space. On the other hand, however, large incumbents already belong to roughly the same size segment as new entrants, where their central appeal to consumers is to offer the widest variety of merchandises in town.

Hence, it would not be totally surprising if some of them choose to expand floor space, in an effort to regain their former position as the champion of one-stop shopping. They would race to the top end of the size spectrum to attract weekend shoppers back. I suspect that this mixture of incentives to shrink and expand might lie behind the estimation results on large incumbents, which are generally negative but not as precisely estimated as the response of medium stores.

Finally, why do the small supermarkets expand? Two mechanisms may possibly be at work. One is that the increased exits among medium stores leave some niche on the size spectrum. Even though it seems risky to become closer in size to the new big entrants, small stores’ initial positions are so distant from the high end of the product space that they might be able to capture the now under-served customer segment, without serious concerns over direct competition with the big entrant. The other possible reasoning is that the arrival of the new product (i.e., big-box retail service) somehow shifts upwards the entire product space effectively demanded by shoppers. The latter mechanism is reminiscent of Sutton’s (1991) endogenous sunk cost theory. These explanations are not mutually exclusive.

These potentially complicated incentives of floor shrinkage/expansion seem to suggest room for further investigation – both theoretical and empirical – on product differentiation
in conjunction with entry and exit.

1.5 Conclusion

Instead of driving out small rivals, large entrants seem to improve their survival prospects. This paper presents new evidence that supports the economic theory of product differentiation, and introduces the perspective of product differentiation to the empirical analysis of entry and exit.

I conduct ordered and IV probit regressions of each incumbent store’s responses (i.e., exit, stay and shrink, stay unchanged, or stay and expand). The results suggest that, even after accounting for both the (ordered) discrete nature of the decision problem and the potential issue of new entrants’ town selection based on the unobservable town characteristics, the main findings stand out: Large and medium incumbents are adversely affected by the new entries of big rivals, whereas small incumbents seem to benefit from them. The results from tobit regressions on the percentage change of floor space further confirms this contrast between large, medium, and small supermarkets.

These findings imply that store size functions as a key dimension of product differentiation among retailers. On the one hand, large and medium incumbents compete as closer substitutes to new large entrants. On the other hand, small supermarkets – thanks to a sufficient degree of differentiation – benefit from the increased traffic of shoppers that is generated by the entrants.

Consequently, this research critically examines the conventional notion that big drives out small, a notion that continues to motivate size-based entry regulations in many economies, both developed and emerging. Ironically, such policies appear to shield big retailers from competition and even preclude small stores from enjoying the increased customer flow that can be generated by large new entrants. The specifics of geographical and regulatory setting

\[\text{25}\text{Hence it is not surprising that the retail chains known for “hypermart” formats, such as Tesco and Carrefour, are introducing new store formats that are much smaller than their original sizes (Tesco Express and Carrefour Express). Even Wal-Mart, the synonym of big-box retailer, developed smaller store formats in Mexico and is now considering their transplantation to rural China (Financial Times, December 2, 2010).}\]
may differ by country and region, but these basic economic forces of product differentiation and entry/exit are likely to be at play in many markets.
REFERENCES


CHAPTER 2

Oligopoly in International Commodity Markets:  
the Case of Coffee Beans

2.1 Introduction

Despite its caffeine content, coffee can cause depression. Between 1988 and 2001, the price of coffee beans fell by 75 percent to an 89-year low of 50 cents per pound (see Figure 2.1),\footnote{Spot price (Brazilian and Other Arabica) at the New York Board of Trade, adjusted by US-CPI to the 2007 constant U.S. dollar. The 89-year low refers to the period since the beginning of price data in 1913.} with exporting countries and producers around the globe suffering the consequences.\footnote{The “coffee crisis” affected over 25 million farmers around the globe (Clark 2007, p.169). The 55 exporting countries I analyze in this paper constitute 42% of all developing countries and 59% of “highly indebted poor countries” (HIPCs). In Colombia, the plight of the coffee industry fueled the intensifying violence (Dube and Vargas 2009). Between 1988 and 1994, real GDP grew by only 0.96% annually in countries where coffee was the major export, whereas the figure was 3.19% for non-coffee exporters. The ensuing waves of sovereign debt cancellations cost taxpayers over $162 billion in richer countries (total debt relief for 51 coffee exporting countries, 1990–2007). In Rwanda, where coffee accounted for 57% of all exports, the commodity crisis exacerbated ethnic violence (Verwimp 2003).} Historically, there have been numerous crises caused by drops in commodity prices. Although the effects of these crises are well documented, few studies have analyzed their causes, which policymakers need to understand in order to alleviate the consequences. At the heart of this gap lies the common belief that commodity prices are volatile and hence unpredictable,\footnote{In the aftermath of the coffee crisis, some analysts noted, “Given the speculative nature of commodity markets it is difficult to attribute price changes to a sole source” (Giovannucci et al. 2004, p.xiv).} which is not necessarily true. Contrary to popular belief, cartels and other forms of imperfect competition are prevalent in the market structures for many commodities traded on international exchanges.\footnote{Cartel and imperfect competition are the norm rather than the exception in many commodity markets. Even when individual producers at the sub-national level are atomistic, exporting institutions and international agreements often create a virtual oligopoly of several nations at the international level. For example,}
prominent commodity crises in recent history: the “coffee crisis.” I have separately identified the price effects of (1) the breakdown of the international cartel agreement in 1989, (2) Vietnam’s emergence as a major exporter during the 1980s and 1990s, and (3) weather and other shocks to supply and demand.

To identify these effects, I exploit historical changes in the structure of the international coffee bean market. First, both the rise and fall of the international cartel agreement (the International Coffee Agreement: ICA, 1965–89) owe much to U.S. foreign policies motivated by Cold War geopolitics rather than internal market forces. Second, the rise of Vietnam as a major exporter since the late 1980s represents one of the rare successes of a developmental “big push” as distinct from the game-theory textbook case of a strategic new entrant to the

---

**Figure 2.1: Real Price of Coffee Beans**

![Graph showing the real price of coffee beans from 1960 to 2000.](image)

*Note:* The prices are in the 2007 constant US dollar. The Adjusted price plots the residual from regressing the original Unadjusted price on the weather shock variable, which is explained in Section 2.3.

*Source:* Commodity Research Bureau (CRB), Bureau of Labor Statistics (BLS), United States Department of Agriculture (USDA), and Author’s calculations.

---

market. Third, the major source of the price volatility—weather shocks disrupting supply, in particular frosts and droughts in Brazil—is observed and controllable. Finally, the market power of coffee bean exporters can be measured by markup, which is captured through measures of price and marginal costs.

The coffee bean market is interesting and important in itself, but focusing on this commodity has other merits. I have chosen to study the coffee bean market for several reasons. First, the coffee cartel was simple in that it was based on an export quota system. This feature allowed my analysis to focus on price, cost, and export quantities. Other international commodity agreements, in contrast, often employ more complex rules such as buffer stock inventories, which would materially complicate the analysis without adding significant insight. Second, of all price movements within primary commodities, that of the coffee bean is one of the most sensitive to the economic fundamentals of short-run supply and demand. By contrast, commodities such as crude oil and metals are storable and require a decade-long technological extraction process that entails much uncertainty. Third, with 55 exporting countries around the globe and trading volume second only to crude oil, the coffee bean could be called a “representative agent” of international commodity markets. It is hoped that some of the insights gained from this market can be carried over to other international commodity markets. Finally, some producing countries rely on the coffee bean for more than half of their entire export revenues (e.g., El Salvador and Uganda). Hence, the coffee price could be called their real exchange rate, fluctuations of which are tantamount to true macroeconomic shocks. These features render the study of the coffee bean industry particularly fruitful.

My empirical analysis proceeds as follows. As a preliminary data analysis, I regressed the measure of market power (markup, or the Lerner index) on the variables representing exogenous shocks to the market structure: (1) the cartel (breakdown) dummy, (2) the Vietnamese export quantity, and (3) weather shocks and other factors. The results conform to

---

5Coffee beans are storable to some extent, but storage is costly due to potential quality deterioration combined with low value relative to physical size. Hence only limited inventory exists, mostly for the purpose of smoothing logistics. Regarding production dynamics, a coffee tree takes approximately five years to reach its full production potential. However, drastic changes are rarely observed in a country’s production “capacities” (i.e., the number of coffee trees). Furthermore, shorter-run factors (trimming and fertilizers) alone may change the harvest quantity by as much as 50%. Thus issues of technological dynamics have limited implications.
expectation in that both (1) and (2) tended to have a negative impact on market power. However, economic interpretation of the coefficient estimates is not clear and is subject to a Lucas critique. That is, one cannot estimate for example what would have happened in 2001 in a “world with no cartel breakdown” or a “world with no Vietnam” without imposing some structure and assumptions.

Therefore I conducted a more formal empirical analysis by developing and estimating a structural model of demand and supply as follows. First, using weather shocks as the instrument for price, I estimated the demand for coffee beans. Then I characterized Vietnam as a “fringe” exporter outside the cartel whose output shifted the net demand curve inwards for all of the other countries. I could then measure the cartel’s market power, or the extent of cooperation among all of the other countries, in terms of the symmetric Cournot-equivalent number of competitors. Finally, using both the demand estimates and the collusion parameter estimates, I was able to infer the price impact of both the cartel’s breakdown and the emergence of Vietnam by contrasting the actual coffee price in 2001 with the prices under counterfactual scenarios of a continued cartel scheme, Vietnam’s absence, or both.

The results suggest that, of the 75 percent drop in the real coffee price between 1988 and 2001, the breakdown of the coffee cartel explains 49 points, the emergence of Vietnam as a major exporter explains another 9 points, and weather and other factors account for the remaining 17 points. These findings offer new insights into the intended and unintended consequences of international public policies on competition, trade, and development aid.

I have organized the rest of the paper as follows. The remainder of this section documents three distinct literatures that motivated this study. Section 2.2 summarizes the institutional and historical background of the international coffee bean market. Section 2.3 explains the dataset and shows the results of preliminary data analysis. Section 2.4 describes the model of the coffee export market. Section 2.5 reports the estimation results. Section 2.6 presents counterfactual experiments to identify the sources of the “coffee crisis.” Section 2.7 concludes with the discussion of policy implications.6

---

6The three appendices contain: (A) a sensitivity analysis of demand estimation comparing the linear and log-linear specifications, (B) checks for autocorrelation in demand estimation, and (C) an additional set of
2.1.1 Related Literature

The international commodity markets determine the prices and quantities of the most basic inputs to our economies, such as food and energy. But the Industrial Organization (IO) literature has not focused on these markets, nor have the trade and development literatures analyzed their market structures. This paper aims to fill this gap.

Three distinct bodies of literature motivated this study: IO, international trade, and miscellaneous studies on the international commodity markets. The first body is IO, where market power and cartels are central topics. Bresnahan (1982) pioneered the estimation of market power, followed by numerous applications. Genesove and Mullin (1998) cross-validated his estimated markups with directly observed measures. More recent contributions examined more complicated competitive environments. Röller and Steen (2006) studied the effectiveness of a legal cartel in the Norwegian cement market. Salvo (2010) considered a situation in which the threat of import competition limited the domestic oligopoly’s market power. He extended the framework by modeling latent imports as a competitive fringe. I have shared the spirits of these two papers, in structurally analyzing the market power of an explicit cartel, in the presence of a competitive fringe.

More fundamentally, I extended the scope of market power analysis to the international commodity markets, whose features pose three unique challenges: (1) geopolitical forces influence the market structure; (2) national governments are the decision makers; and (3) their coordination plays an important role in the form of international treaties and organizations. I incorporated geopolitical factors as the institutional context and in the data analysis (sections 2 and 3). I featured countries’ export decisions in my model (section 4). I estimated the cartel treaty’s effectiveness as the key structural parameter of my model (section 5).

The study of cartel and collusion forms another related strand within the IO literature. Seminal empirical works include Porter’s (1983) and then Ellison’s (1994) studies of the counterfactual analyses regarding hypothetical fundamentals-driven “commodity booms.”

railroad cartel. More recent papers (e.g., Fershtman and Pakes 2000, de Roos 2004) applied the dynamic oligopoly framework to tacit collusions, i.e., cooperation among rival firms without explicit contracts. Meanwhile, few studies focus on explicit collusions, i.e., cartel arrangements where agreements are explicitly stated and enforced. A rare exception is Röller and Steer (2006), who emphasized the importance of measuring the effectiveness of a particular cartel arrangement. My study contributes to this strand by measuring the actual market power of an explicit international cartel treaty, in a way that is structurally interpretable.

The second body of literature concerns oligopoly in international trade. Imperfect competition models are commonly used in this literature, but its emphasis has been on the welfare analysis of “strategic trade policies.” For example, Brander and Spencer (1985), Eaton and Grossman (1986), and Helpman and Krugman (1989) theoretically analyzed the strategic uses and welfare implications of tariffs, anti-dumping, quotas, and subsidies. Corresponding empirical works have measured and evaluated the actual effects of these trade policies (e.g., Berry, Levinsohn, and Pakes 1995, 1999; Feenstra and Levinsohn 1995; Goldberg 1995). My research follows their footsteps in applying an IO model to international trade, but with a different emphasis: countries, not companies, competing and cooperating in the product market. More recent papers on state trade enterprises (STEs) studied a similar situation in which exporting or importing decisions are made at the national government level. Whereas these studies focused on the rent-shifting effects in the strategic trade context, this paper aims to address a different issue: the measurement of market power and its price effects. Finally, this paper shares with that of Bagwell and Staiger (1990) the characterization of countries as strategic competitors. However, while they applied the tacit collusion model to explain “managed trade,” I focus on measuring the effectiveness of an explicit collusion based on an international treaty.

The third body of literature that motivated this paper is the extensive body of work across agricultural, development, and financial economics that studies the international commodity markets (see Deaton 1999 for an overview). Although this literature has documented the

---

8Hamilton and Stiegert (2002); Dong, Marsh, and Stiegert (2006); and McCorriston and MacLaren (2010).
welfare consequences of commodity price movements for national economies (e.g., Catão and Sutton 2002) as well as rural households (e.g., Varangis et al 2003), few studies have analyzed the determinants of commodity prices. This omission reflects the tendency of analysts to assume perfect competition, taking price fluctuations as random shocks. The IO approach remains rare, except perhaps for Karp and Perloff (1989, 1993). Since cartels and other forms of imperfect competition are common in these markets, however, the analysis of market power is crucial for understanding commodity prices. This paper fills the gap by using changes in market structure to explain one of the biggest commodity crises in history.

In summary, the international commodity markets represent important gaps both within and between these bodies of economic literature. This study is the first step in deepening our understanding of market power in these markets.

2.2 The International Coffee Bean Market

This section provides the institutional background of the international coffee bean market. First I describe the profiles of coffee importers and exporters, as well as product characteristics. Then I explain the historical context of the cartel treaty and fringe competitor, which is crucial for understanding the geopolitical nature of the changes in market structure. Specifically, I argue that both the rise and fall of the international cartel agreement, as well as the expansion of Vietnamese exports, were out-of-the-blue shocks to the export market, thus providing the econometrician with rich exogenous variations in data. Moreover, because the variations were direct reflections of U.S. foreign policies and Vietnam’s rural development projects, the subsequent empirical analysis is informative regarding public policies in competition, trade, and aid. Finally, weather shocks are another source of exogenous variations in data, which I explain in Section 3 and use as an instrument for demand estimation in Section 5.

---

9For example, see Akiyama and Varangis (1990), Mehta and Chavas (2008).

10Their main objective is to test particular dynamic oligopoly models. In contrast, this paper focuses on estimating market power and its price impacts.
2.2.1 Importers, Exporters, and Product Characteristics

Americans drink, on average, 1.76 cups of coffee a day, whereas Europeans down 2.04 cups. Even the green-tea-loving Japanese enjoy 1.42 cups of coffee.\textsuperscript{11} Together, the capitalist “first world” imported over 90 percent of the global coffee bean harvest in the sample period (1960–2006). Most of these countries (except for New Zealand and Israel) were “importing members” of the ICA. Their customs officials ensured that shipping carried the ICA’s stamp of approval, thereby monitoring and enforcing the exporters’ quota allocation. In essence, the ICA was a form of development assistance from the “first world” (see section 2.2 for details).

The socialist “second world,” or “non-member market,” imported approximately 5 percent of all coffee beans.\textsuperscript{12} This market was composed of fringe buyers and formed the unregulated “black” market. Transactions in this market were not systematically recorded, but evidence suggests that black-market prices were 30 percent to 40 percent lower than in the regulated market. This price level was roughly equal to the domestic farm-gate/wholesale prices in exporting countries. This finding implies marginal cost pricing by exporting countries, which is not surprising because they customarily dumped excess harvest (supply above quota) to the socialist bloc. Thus the “second world” played the role of the buyer of last resort, or a safety valve, for the “third-world” farmers.\textsuperscript{13}

To satisfy these peoples in the “North,” the “South” exports over $10 billion worth (or 80 million 60kg bags) of green coffee beans every year. Coffee cultivation requires tropical highlands, thereby determining the producing regions. No significant entry or exit has occurred since the 1960s, aside from the rise of Vietnam.\textsuperscript{14} Fifty-five countries across Africa,

\textsuperscript{11}Calculation based on the International Coffee Organization’s 2005 survey. I follow America’s National Coffee Association to equate an annual consumption of 1 kilogram with 0.42 cup per day (6.5 gram per cup).

\textsuperscript{12}They were outside the regime partly because they could not pay in hard currencies. See Bohman and Jarvis (1990) for a detailed account of the non-member market.

\textsuperscript{13}The overall import quantity in the socialist bloc was limited to about 5% of the world’s total, so parallel imports (so-called “migrating coffee”—cheaper black-market beans being resold to the “first world”) were not a major issue.

\textsuperscript{14}In this market, a country’s “entry” often meant national independence since most countries were former colonies, which is why, aside from data availability, I focus on the years since 1960.
Asia, and Latin America, with many small farms within each nation (see Figure 2.2), export coffee beans.

![Figure 2.2: Areas of Coffee Cultivation](image)

*Note:* r, m, or a indicates cultivation areas for robusta, mixture of robusta and arabica, or arabica species.

*Source:* Wikipedia.

Government bodies in these countries, such as national coffee marketing boards, regulated the industry, which worked as a domestic planning/coordination agency. Since the international cartel agreement (an export quota system) worked at the supra-national level, my analysis focuses on nation states as the export decision makers, and particularly on their collective behavior in the global market.\(^\text{15}\)

Green beans (coffee’s raw material form before roasting, grinding, and brewing) are commonly classified into the following four categories in the world’s commodity exchanges: Colombian Mild Arabica, Other Mild Arabica, Brazilian Natural Arabica, and Robusta. Like wine, coffee beans have subtle differences in aroma and taste by regions and even by farms.\(^\text{16}\)

Such subtleties notwithstanding, the price movements of all four types are closely related, with a correlation coefficient of over .93 (for monthly changes in prices). The cartel’s operation was based on the single indicator price (the weighted average of the four different coffee prices). Thus, to maintain consistency with the cartel’s practice and to focus on the measurement of its market power, I followed previous studies (Karp and Perloff 1993, Nakamura 1994).

\(^{15}\)Because this paper’s main objective is international competition/cooperation and market power, I abstract from political economy at the sub-national level. See Bohman, Jarvis, and Barichello (1996) for rich institutional details.

\(^{16}\)For example, Robusta beans are consistently priced lower than other beans because they taste like “burnt rubber,” so the U.S. roasters introduced a new process: steaming and spraying them with artificial flavors on them, and then selling these beans as “gourmet” coffee (Clark 2007, ch.6).
and Zerom 2010) in characterizing green coffee beans as homogeneous goods.

2.2.2 The Cartel and the Cold War: Why the Rise and Fall of the International Coffee Agreement are Exogenous

Since the 1930s, coffee exporting countries had made many attempts to form a cartel-like arrangement, but none of those attempts survived. A successful quota scheme takes monitoring and enforcement, both of which were lacking. However, the situation changed after 1959, when Fidel Castro and his fellow guerrillas took power in Cuba: the Cuban Revolution. After frosty interactions with the United States, the revolutionary government turned to the Eastern bloc.\(^{17}\) Cuba was a coffee-producing economy, and the revolutionary movement gained momentum from the countryside, just as Mao and Guévara had theorized and practiced. Given this development in the United States’ backyard, the Kennedy administration’s foreign policy prioritized the “fight against communism.”

Diplomatic alliance was one tool. Sending monetary aid was another. To maintain a good relationship with coffee-growing countries, the White House decided to provide monitoring/enforcement for its southern neighbors’ long-sought quota agreement. Thus the United States signed the ICA in 1962 and, after three years of negotiations in Washington, Congress approved the legislation for the customs inspection of “certificate of origin” (see below) in 1965.\(^{18}\) Other “Northern” governments (including Japan but not New Zealand or Israel) had already signed up.\(^{19}\)

The ICA was kept in place between 1965 and 1989 for the purpose of maintaining high coffee prices. Its primary mechanism was export quotas, which were allocated across countries in proportion to their historical market share.\(^{20}\) These quotas were then revised up or down

---

\(^{17}\)See Domínguez (1978) for the political economy of revolutionary Cuba.

\(^{18}\)Thus the early 1960s cannot be clearly characterized as either a non-cartel or cartel periods. For this reason, this paper does not intend to analyze the beginning of the ICA. See Bates (1997), ch.5. Other obstacles to the analysis of the coffee market’s earlier years include the changing identities of newly independent colonies and the lack of detailed data.

\(^{19}\)European governments had similar interests in stabilizing African economies.

\(^{20}\)See Bates (1997), ch. 6 for the details of the ICA’s functioning.
across the board in response to the movements of the indicator price. The decision-making at the International Coffee Organization (ICO), the ICA’s administrative body, relied on voting by both the exporting and importing member countries. All major decisions required a two-thirds majority. The votes were also allocated proportional to countries’ trade shares, so the largest exporters (notably Brazil and Colombia) as well as the United States held veto power.

The ICA used “certificate of origin” for the monitoring and enforcement of export quota. The importing members collected and returned these certificates to the ICO, so that all exports were publicly recorded. Penalties for excess shipments were imposed through the deduction of these shipments from the following year’s quota. Excess shipments in a second year were penalized by a doubling of the deductions in the next. A third violation would lead to the loss of voting rights and possible expulsion from the ICO.

In practice, the rules were leniently applied, and loopholes existed. For example, certificates of origin were not required for either imports from non-member countries (such as the socialist bloc, which comprised the perfectly competitive black market) or exports to the so-called “new markets” (i.e., countries in which coffee had not yet become a staple of consumption). The problem was that this “tourist” coffee could be re-exported to the “traditional” markets in the member countries. The ICA rules were gradually refined to prevent major deviations, but room for small-scale cheating remained. Political conflicts were also rampant. Since the details of data collection, monitoring, quota allocation, and the measures to close those loopholes affected countries differently, almost every aspect of the ICA’s operation was controversial, which complicated its internal political processes.

Despite these shortfalls, the ICA was widely considered as one of the most successful commodity agreements in history.\footnote{Exactly how successful is an open empirical question that this paper seeks to address.} Cooperation among producers remained problematic, but the crucial support came from the importing members in the form of the monitoring and enforcement of quotas. It also helped that the United States used its other aid programs to offer inducements and sanctions. As Bates (1997) documents: “The U.S. government was
intensely aware that the success of its broader development assistance programs in Latin America would be strongly affected by the success of the International Coffee Agreement.” (p.146)

But the honeymoon between the first and third worlds ended abruptly in 1989, when the second world imploded. The collapse of its Cold War enemy, and with it the rationale for hurting American consumers through high prices under the ICA, led the United States to withdraw its support. The ICA ended on July 4, 1989.\footnote{During the late 1980s, Secretary of State George Shultz demanded all international commodity treaties be reviewed by the State Department’s Division of Economic Affairs, staffed with economists from Chicago, who subsequently recommended the withdrawal from these treaties. See Bates (1997, ch.7) and Clark (2007, ch.6) for the details of this political process.}

Without American policing, none of the subsequent resuscitation efforts proved fruitful.

\subsection*{2.2.3 Geopolitics of Land, Labor, and Ethnicity: Why Vietnam’s Expansion is Exogenous}

Although in 1989 the developed world scrapped the subsidies the ICA embodied, the Vietnamese government was expanding coffee bean production on the back of foreign aid. After the U.S. exit from Saigon in 1975 and a fight against China in 1979, this war-torn country had received an investment from the Soviet bloc in the early 1980s to plant coffee trees. The communist world needed beans available for purchase without hard currency.\footnote{Even after the Berlin Wall fell and Soviet assistance was gone, Vietnam still had other donors, such as Japan, the World Bank, and the Asian Development Bank. American aid also resumed in 1991. See Library of Congress Federal Research Division (1987 and 2005), \textit{Vietnam Country Study}; and Mark E. Manyin, (2005), “U.S. Assistance to Vietnam,” Library of Congress.}

In a series of five-year plans, the central planners orchestrated widespread internal migration that prepared land and labor for coffee production, but their motives were more geopolitical than economic. From the perspective of the post-war strategists, Vietnam faced three challenges. First, the new government needed to establish control over land abandoned during wars and to secure border areas with China, Laos, and Cambodia. Second, the planners viewed ethnic minorities as internal political threats, most of whom lived in the forested upland areas and had sided with the United States during the war. Third, the peasantry...
backed revolutionary government suspected that their authority would be undermined by high and growing population densities in the urban areas of the Red and Mekong Deltas (Solem et al 2010).

One solution to these perceived problems was to encourage or coerce large numbers of Kinh (the majority ethnic population) to move from the north to the under-populated frontier zone of the Central Highlands, which turned out to be suitable for coffee cultivation (see map in Figure 2.3). The government designated these areas as New Economic Zones in which internal migrants produced coffee. The graph in Figure 2.3 shows that the population of the Central Highlands more than doubled between the mid 1980s and the turn of the century, from less than 2 million to over 4 million. A survey performed in 1998-1999 found that 89 percent of the population of the Central Highlands had not been born there (Duc 2002), which highlights the scale of the government-led migration.

In addition to these migration policies, the government adopted the Economic Renovation (Doi Moi) policy in 1986. This constituted economic reforms to create a “socialist-oriented market economy,” which allowed the establishment of an agricultural private sector. Part of Doi Moi was land reforms. The new Land Law of 1988 sought to grant land use rights to households. All land technically remained under the state ownership, but the second Land Law of 1993 further permitted individuals to trade their land use rights (Vuong 2001). The government promoted the expansion of coffee cultivation through subsidized land and loans, and it provided support for seedlings, fertilizer, irrigation, and agronomy (Giovannucci et al 2004). As a result, coffee production area grew five-fold in the 1990s (Figure 2.4), before it hit the geographical limit in the early 2000s.

Thus, the growth of Vietnam’s coffee bean production was primarily driven by the geo-

---

24 The region was previously populated by ethnic minorities such as Ede, Rong, Sedang, Tai, and Giay, most of whom relied on swidden (or slash-and-burn) agriculture. These groups were also incentivized to abandon nomadic behaviors (and join coffee farming) by the Fixed Cultivation and Sedentarization Program (Doutriaux et al 2008, Ha and Shively 2008).

25 The area under crop peaked in 2001 and slightly decreased thereafter. This drop reflects state farms’ retreat from unsuitable lands. In official statistics, several provinces on the periphery of coffee cultivation appear and then disappear around that time (Ministry of Agriculture and Rural Development). See also ICARD and Oxfam (2002).
The geographical expansion of *Doi Moi* policies, on the back of foreign aid and large-scale ethnic migration schemes. These policies, in turn, stemmed from the geopolitical concerns of the post-war central planners. Such historical developments do not necessarily preclude the importance of economic incentives. Certainly, the goal of *Doi Moi* was the gradual transition from central planning to a market-based economy. The planners promoted coffee because Vietnam could produce it cheaply for profitable export. For the purpose of subsequent econometric analyses, however, the crucial aspect of Vietnam’s modern history is that the growth of coffee production occurred for those idiosyncratic reasons in a series of five-year plans, and mostly in isolation from the year-on-year price fluctuations in the global market. By the late 1990s, Vietnam had dethroned Colombia as the world’s second-largest exporter (with Brazil still the largest), putting further downward pressure on the price of coffee beans. The Vietnamese government had not entered coffee trade negotiations before reaching its geographical limit of production in 2001. Vietnam finally became a member of the International Coffee Organization (ICO) on May 7, 2002, and of the World Trade Organization on January 11, 2007.
Figure 2.4: The Geographical Expansion of *Doi Moi* and Coffee Production


overall process of production growth would be best characterized as a (rare) success story of a developmental “big push.”

### 2.2.4 Summary of the Institutional Context

In summary, the international coffee bean market consists of the developed countries (as importers) and developing countries (as exporters). The latter enjoyed market power under the ICA (1965–89), the rise and fall of which were driven by the U.S. government’s Cold War concerns. The socialist bloc played the role of fringe importers, representing a small, unregulated market for exporters. The breakdown of the coffee cartel coincided with the emergence of Vietnam as a fringe exporter urged along by foreign aid, government-led migration, and market-oriented reforms. Thus both the cartel’s breakdown and Vietnam’s expanded exportation represent changes in market structure that are exogenous to the year-on-year price fluctuations in the global market. These changes provide the econometrician with rich variations in data, along with weather shocks to exporting countries, which I explain in the next section.
2.3 Data

This section first describes the data set. It then reports the preliminary data analysis, regressing the measure of market power on changes in the market structure (i.e., the cartel’s breakdown and the emergence of the new exporter; Vietnam). These regressions suggest the need for—and foreshadow the results of—subsequent structural analysis.

2.3.1 Variables

The data set spans the period 1960–2006 and contains the measures of prices, costs, exports, changes in market structure, and three other variables for instrumentation and control, at the annual frequency. Table 2.1 displays the summary statistics.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unit of Measurement</th>
<th>Obs.</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price ($P_t$)</td>
<td>US cents</td>
<td>47</td>
<td>243</td>
<td>145</td>
<td>50</td>
<td>818</td>
</tr>
<tr>
<td>Cost ($mc_t$)</td>
<td>US cents</td>
<td>30</td>
<td>127</td>
<td>61</td>
<td>41</td>
<td>283</td>
</tr>
<tr>
<td>Markup ($m_t \equiv \frac{P_t}{mc_t}$)</td>
<td>-</td>
<td>30</td>
<td>.31</td>
<td>.16</td>
<td>.02</td>
<td>.55</td>
</tr>
<tr>
<td>World Export ($Q_t$)</td>
<td>Million 60kg bags</td>
<td>47</td>
<td>65.2</td>
<td>14.0</td>
<td>41.8</td>
<td>94.3</td>
</tr>
<tr>
<td>of which Cartel ($Q^C_t$)</td>
<td>Million 60kg bags</td>
<td>47</td>
<td>62.0</td>
<td>10.2</td>
<td>41.8</td>
<td>79.1</td>
</tr>
<tr>
<td>of which Vietnam ($q^V_t$)</td>
<td>Million 60kg bags</td>
<td>47</td>
<td>3.2</td>
<td>5.3</td>
<td>0.0</td>
<td>18.7</td>
</tr>
<tr>
<td>Cartel Breakdown ($I_t$)</td>
<td>0/1 indicator</td>
<td>47</td>
<td>.38</td>
<td>.49</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Weather Shocks ($W_t$)</td>
<td>Month (fraction)</td>
<td>47</td>
<td>.51</td>
<td>.44</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Buyers’ GDP ($X_t$)</td>
<td>Trillion US dollars</td>
<td>47</td>
<td>15.7</td>
<td>6.3</td>
<td>5.9</td>
<td>27.2</td>
</tr>
<tr>
<td>Tea Price ($Z_t$)</td>
<td>US cents</td>
<td>47</td>
<td>478</td>
<td>233</td>
<td>204</td>
<td>983</td>
</tr>
</tbody>
</table>

*Note:* Prices and GDP are expressed in the 2007 constant U.S. dollars. Cost data (hence markup, too) are available only for 30 years.

*Source:* Commodity Research Bureau (CRB) for the prices of coffee and tea, United States Department of Agriculture (USDA) for the export quantities, International Coffee Organization (ICO) for the farm-gate prices (measures of Cost), and International Monetary Fund (IMF) for the importing countries’ GDP. The Markup, Cartel Breakdown, and Weather Shocks are constructed by the author.

First, I observed both the *international coffee bean price* (the benchmark price index calculated by the ICO based on the spot prices in NYBOT) and the exporting countries’ *domestic farm-gate prices* (the prices the growers receive in these countries). Because the

---

27I use the market-share-weighted average of the farm-gate prices across countries, but all of the subsequent results are robust to other aggregation rules such as simple averaging or using only the top producers like Brazil and Colombia.
latter reflect what exporters/governments pay the coffee growers, it represents the marginal cost of exporting. Studies on the international “black” market (see section 2) suggest the prevailing price in that unregulated, perfectly competitive market equals those farm-gate prices in exporting countries. This is yet another way in which the farm-gate price (≈ the black-market price) reflects the opportunity cost (marginal cost) for a country to export to the ICA-member (i.e., regulated) market. Given these data, I was able to measure exporters’ market power by markup (price-cost margin, or the Lerner index): 

\[ m_t \equiv \frac{P_t - mc_t}{P_t} \]

(see Figure 2.5).

Figure 2.5: Price, Cost, and Markup

Second, I examined the exports of coffee by country. Three quantities are particularly important for the subsequent analyses: world exports, cartel members’ collective exports, Vietnam’s export (denoted by \( Q_t \), \( Q^C_t \), and \( q^V_t \), respectively). This \( q^V_t \) reflects the sudden acceleration of the Vietnamese production, a government-led “big push” (see section 2.2.3 for the institutional background). Together with \( q^V_t \), the cartel breakdown dummy (which

---

This interpretation can be supported by at least three models of an exporting country’s domestic economy: (1) the domestic wholesale coffee bean market is competitive; (2) the government/exporter acts as a benevolent social planner for the entire domestic economy; and/or (3) the government/exporter acts as the representative of the coffee producers by first paying them the marginal cost of production and then transferring the export surplus to the producers as well. See Bohman, Jarvis, and Barichello (1996) for the various models of exporting countries’ domestic market structure that are consistent with this interpretation.
equals zero before 1989 and one since then) represents changes in the market structure due to the end of the Cold War (see section 2.2.2 for details).

Another important variable is weather shocks, $W_t$. In the spirit of Roll’s (1984) analysis of orange juice and weather, I use weather shocks as an instrument for the price in demand estimation (see section 2.5). Weather is the most important shock for coffee production. The trade publication *The CRB Commodity Yearbook* contains news on events that influence commodity prices, from which I collected the record of major frosts and droughts in producing countries, most importantly Brazil, as well as Central America, East Africa, and South East Asia. Since damaged coffee trees usually take two years to recover, I code one for 24 months starting from the month of a weather shock, and zero otherwise.\(^{29}\) Because $P_t$ and $W_t$ are monthly series in the original data source, I annualize them by taking the averages (according to the international coffee year, which starts in October and ends in September) in order to match the data frequency of the other variables.

Other shocks on the demand side will serve as control variables. Developed economies’ aggregate real GDP, $X_t$, embodies both population and per-capita income of the importing countries. Since the world coffee demand grew roughly proportional to the size of population, this variable effectively detrends the demand.\(^{30}\) Finally, the real price of tea, $Z_t$, would control for a potential substitution of tea for coffee.\(^{31}\) The inclusion of these variables is informed by Karp and Perloff’s (1993) and Nakamura and Zerom’s (2010) studies of coffee demand.

\(^{29}\) The estimation results are robust to changes in the way I code the weather variable, including different treatments of the instances of multiple frost/drought shocks.

\(^{30}\) See Appendix B for the benefits of this feature in addressing autocorrelation issues.

\(^{31}\) The tea price is suitable as a control variable as it evolves independently from the developments in the coffee market for the following reasons. First, the climate zone for tea growing is different from that for coffee and subject to different weather shocks. Second, over half of the tea production is for domestic consumption in places such as China, India, Sri Lanka, Bangladesh, and Japan, rather than for exports. Third, unlike the coffee trade, the international trade of tea is competitive except for the brief period during the Great Depression (i.e., outside my sample period). See Gupta (2001) for details.
2.3.2 Preliminary Data Analysis

Those changes in the market structure reflect the geopolitical developments that are exogenous to the year-on-year price fluctuations in the export market, so simple regressions could be informative of the relationship between the markup and the market structure. Such regressions will also clarify the basic sources of identification for the subsequent structural analysis. Thus, as a preliminary data analysis, I regress the measure of market power (the Lerner index) in year $t$, $m_t$, on those “shocks” to the market structure and controls:

$$m_t = b_0 + b_1 I_t \{ \text{Cartel Breakdown} \} + b_2 q^V_t + b_3 W_t + b_4 X_t + b_5 Z_t,$$

where $I_t \{ \text{Cartel Breakdown} \}$ is the indicator function that equals one from 1989, $q^V_t$ is the Vietnamese exports, $W_t$ is weather shocks (the number of months with frost/drought damage), $X_t$ is the importing countries’ GDP (population times per-capita income), and $Z_t$ is the price of tea (arguably an important substitute for coffee).

Table 2.2: Preliminary Regressions of Market Power on Shocks

<table>
<thead>
<tr>
<th>Dep. var.: $m_t$ (Markup)</th>
<th>OLS</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_t$ (Cartel Breakdown)</td>
<td>-.27***</td>
<td>-.27***</td>
<td>-.22***</td>
<td>-.21**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.03)</td>
<td>(.04)</td>
<td>(.07)</td>
<td>(.09)</td>
<td></td>
</tr>
<tr>
<td>$q^V_t$ (Vietnam’s Export)</td>
<td>-.15***</td>
<td>.03</td>
<td>.14</td>
<td>.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.04)</td>
<td>(.03)</td>
<td>(0.08)</td>
<td>(0.91)</td>
<td></td>
</tr>
<tr>
<td>$W_t$ (Weather Shocks)</td>
<td>.05</td>
<td>.08*</td>
<td>.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.04)</td>
<td>(.04)</td>
<td>(.04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_t$ (Buyers’ GDP)</td>
<td>.02</td>
<td>-.84**</td>
<td>.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.18)</td>
<td>(.31)</td>
<td>(.44)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Z_t$ (Tea price)</td>
<td>.02</td>
<td>-.01</td>
<td>.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.02)</td>
<td>(.03)</td>
<td>(.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>.47***</td>
<td>.38***</td>
<td>.47***</td>
<td>.27</td>
<td>2.71**</td>
</tr>
<tr>
<td></td>
<td>(.03)</td>
<td>(.03)</td>
<td>(.03)</td>
<td>(0.55)</td>
<td>(.97)</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>.70</td>
<td>.27</td>
<td>.69</td>
<td>.71</td>
<td>.63</td>
</tr>
<tr>
<td>Number of observations</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
</tbody>
</table>

Note: ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. Heteroskedasticity- and autocorrelation-consistent (up to 5 lags) standard errors are in parentheses. I chose 5-year lags as maximum because a new coffee tree needs three to five years to reach full production potential, but different lags (bandwidths) do not materially alter the HAC standard errors.

Table 2.2 shows the results of these preliminary regressions. First, the cartel’s breakdown
in 1989 seems likely to be the primary cause of the coffee crisis, with big and measurable impacts across all regressions ([1], [3], [4], and [6]). Second, the Vietnamese export shows some correlation in [2], but the results are mixed. Third, the control variables mostly carry expected signs (+), although column [5] exhibits a counterintuitive outcome for $X_t$ (negative and statistically significant). Overall, the results seem reasonable and foreshadow the conclusions of the subsequent analysis.

However, these preliminary regressions fall short of separately identifying the effects of the cartel’s breakdown and the expansion of Vietnam. To further disaggregate these different forces at play, in the next section, let us consider each variable’s economic meaning in explaining the “coffee crisis.”

### 2.4 Model

The regressions in the previous section suggest the likely price impacts of the cartel breakdown, the emergence of Vietnam, and weather shocks, but the results are not very clear. This is attributed to multicollinearity and potentially complicated interactions. One cannot reliably decompose the coffee crisis into these factors without some modeling. More fundamentally, the extent to which the cartel optimally restrained its own exports (thereby affecting prices) would depend on the output from Vietnam as well as demand-side conditions, in a highly nonlinear manner. Hence, one cannot predict what the coffee price would have been in 2001 (say) in a “world with a continued cartel treaty” or a “world without Vietnam.” Some structure and assumptions are needed to simulate such counterfactual scenarios and to derive welfare implications of the “coffee crisis.”

For these purposes, I developed an estimable model of demand and supply for the coffee bean export market, which is parameterized by the cartel’s degree of coordination. The model also features Vietnam as a fringe competitor, reflecting the unique institutional context of its coffee-sector development (see Section 2.2.3).
2.4.1 Demand and the Timing of the Game

World demand \( Q(P_t) \) determines the relationship between world exports, \( Q_t \), and price \( P_t \). Let \( P(Q_t) \) represent the inverse demand function. Vietnam is a new, fringe exporter. To consistently measure the coordination among all of the other exporters (the cartel treaty members), one should subtract Vietnamese exports from global demand to calculate the net demand curve that cartel exporters face. More formally, the setup is as follows:

1. Every year \( t \), the world demands coffee beans. The (inverse) demand schedule, \( P(Q_t) \), is common knowledge. The aggregate supply, \( Q_t \), consists of the Vietnamese export \( q_t^V \) and those from the other 54 countries collectively called “the cartel” \( Q_t^C \): \( Q_t = q_t^V + Q_t^C \).

2. The new, fringe producer, Vietnam, exports its entire harvest for the year, \( q_t^V \). This \( q_t^V \) is the result of government-led ethnic migration, the Doi Moi policy, and foreign aid, and was therefore determined independently (see section 2.2.3 for details).\(^{32}\)

3. Given \( P(Q_t) \) and \( q_t^V \), that is, given the demand curve net of fringe supply, the 54 cartel members decide their collective exports, \( Q_t^C \).

Thus, the framework is a combination of Stackelberg and Cournot games. Its “Stackelberg” feature is that there are first and second movers although the first mover (Vietnam) represents a policy shock rather than an active, strategic player as in the standard case. The setup also has a Cournot component in the sense that the second mover is not a single decision maker but a collection of export quantity-setting players, whose coordination may not be perfect.

\(^{32}\)In principle, Vietnam can be alternatively modeled as a strategic first mover (i.e., \( q_t^V \) depends on \( P(Q_t) \) and \( Q_t^C \)), although this assumption would be inconsistent with the historical facts. The final results do not change significantly even under this alternative assumption. I have chosen the baseline assumption of nonstrategic Vietnam because of the country’s unique institutional context and because I can then maintain the time-consistent interpretation of the coordination parameter over time.
2.4.2 The Cartel’s Coordination

The collective exports, $Q^C_t$, reflects the cartel members’ coordination (or lack thereof). Consider a Cournot model with $N$ symmetric countries shipping coffee beans. Country $i$ in year $t$ chooses export, $q_{it}$, to maximize its profit:

$$\pi_{it} = P (Q^C_t, q^V_t) q_{it} - C (q_{it}, W_t),$$

where $P (\cdot)$ is the inverse demand function, $Q^C_t$ is the sum of exports from all countries excluding Vietnam (i.e., $\sum_{i \neq V} q_{it}$), and $C (\cdot)$ is the cost function. The cost shifter $W_t$ affects the marginal cost $mc_{it} = \partial C (q_{it}, W_t) / \partial q_{it}$. The first order condition of country $i$’s profit maximization is

$$P_t + \frac{\partial P}{\partial Q} q_{it} = mc_{it};$$

hence

$$- (P_t - mc_{it}) \left( \frac{\partial P}{\partial Q} \right)^{-1} = q_{it},$$

or equivalently,

$$\mu_t \equiv (P_t - mc_t) \frac{- \left( \frac{\partial P}{\partial Q} \right)^{-1}}{Q^C_t} = \frac{1}{N}.$$  

I defined the expression on the left hand side as the collusion (coordination, or conjectural variation) parameter, $\mu_t$. If the cartel exercises some market power above the Nash equilibrium (Cournot competition) price level, this parameter will be greater than the actual $1/N$ in data and approach 1 as the cartel’s coordination becomes perfect (i.e., as their collective action more closely resembles a hypothetical monopolist’s behavior). Thus, the collusion parameter $\mu_t$ maps the markup (and hence the price) into an effective number of symmetric-Cournot competitors. For these reasons, I interpreted $\mu_t$ as the structural parameter that represents the cartel countries’ coordination level. The goal of the next section is to estimate its empirical counterpart to quantify what the cartel actually achieved.

Before moving on to the estimation of $\mu_t$, let us consider why this static measure of market power is a sufficient statistic for the purpose of this study. As an alternative modeling
approach, it would be interesting to consider a dynamic game within the ICA cartel, in the spirit of *tacit* collusion models. The ICA, however, is an *explicit* collusion regime that is *externally* enforced by the importing countries.\(^{33}\) Occasional episodes of cheating notwithstanding, an exporting country cannot deviate much from its allocated quota because the U.S. government and other importing countries monitored and enforced the quota agreement. Moreover, quota is annual and stable over years. Hence the actual functioning of the ICA leaves little room for truly dynamic strategic interactions, either in the form of repeated game or in terms of production capacity dynamics. Thus, in the institutional context under study, the static notion of market power (embedded in \(\mu_t\)) is adequate for measuring the performance of the cartel treaty.\(^{34}\) I estimate \(\mu_t\) in the next section.

### 2.5 Estimation and Results

The goal of this section is to estimate the structural model of the previous section. This section first presents the demand estimates for coffee beans, which I then used to adjust the observed markup, \(m_t\). This normalized measure of market power, \(\mu_t\), is directly interpretable as the *structural parameter of the ICA’s coordination*, and will be useful for simulations in the next section.

#### 2.5.1 Demand Estimates with Weather IV

Demand estimation provides the basis for the subsequent analyses. Table 2.3 shows the estimates based on the following linear form:

\[Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \epsilon_t\]

\(^{33}\)Therefore the level of \(\mu_t\) during the ICA period is fixed by the developed countries’ foreign policies. This observation motivates my assumption that \(\mu_t\) is constant within each of the two subsample periods (before and after 1989).

\(^{34}\)Even if the cartel’s market power had complicated dynamics, as suggested in Cortes’s (1999) critique of the estimation approach by Bresnahan (1982), one could still interpret the static parameter of market power, \(\mu_t\), as a logically consistent “reduced-form” representation of all of the possible dynamic interactions, under the assumption of a constant level of collusion. That is sufficient for the purpose of this paper, which is to quantify how successful the ICA was and to identify its price impact, in the geopolitical context that effectively sets \(\mu_t\) at a constant level.
\[ Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 X_t + \alpha_3 Z_t + \alpha_4 Z_t P_t + \varepsilon_t, \]  

(2.6)

where \( Q_t \) is the quantity demanded, \( P_t \) is the world price, \( X_t \) is the demand shifter (the importing countries’ GDP, i.e., buyers’ population times income per capita), \( Z_t \) is the demand rotator (the price of tea leaves), and \( \varepsilon_t \) is the unobserved shock.\(^{35}\) This specification follows Karp and Perloff’s (1993) earlier study on the coffee demand, and the results also are comparable to their work.

### Table 2.3: Demand Estimates

<table>
<thead>
<tr>
<th>Dependent var.: ( Q_t )</th>
<th>OLS</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_t ) (Coffee price)</td>
<td>-58.43*** (-21.07)</td>
<td>-45.62* (24.11)</td>
</tr>
<tr>
<td>( X_t ) (Buyers’ GDP)</td>
<td>28599*** (1131)</td>
<td>28784*** (912)</td>
</tr>
<tr>
<td>( Z_t ) (Tea price)</td>
<td>16.88 (11.63)</td>
<td>25.83* (14.73)</td>
</tr>
<tr>
<td>( Z_t P_t ) (Interaction)</td>
<td>.0341*** (.0093)</td>
<td>.0354*** (.0099)</td>
</tr>
<tr>
<td>Constant</td>
<td>79371*** (6253)</td>
<td>38695* (20126)</td>
</tr>
</tbody>
</table>

| Adjusted \( R^2 \) | .37 | .92 | .92 | .35 | .92 | .92 |
| Num. obs. | 47 | 47 | 47 | 47 | 47 | 47 |

| Partial \( R^2 \) for \( P_t \) | - | - | - | .52 | .73 | .40 |
| F-stat for \( P_t \) | - | - | - | 6.93 | 11.01 | 15.69 |
| Partial \( R^2 \) for \( Z_t P_t \) | - | - | - | - | .71 | .30 |
| F-stat for \( Z_t P_t \) | - | - | - | - | 3.64 | 10.92 |

Note: ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Heteroskedasticity- and autocorrelation-consistent (up to 5 lags) standard errors are in parentheses. I chose 5-year lags as maximum because a new coffee tree needs three to five years to reach full production potential, but different lags (bandwidths) do not materially alter the HAC standard errors. \( P_t \) and \( Z_t P_t \) are instrumented by the coffee supply shifters (weather shocks) \( W_t \) and their interaction with the tea price \( Z_t W_t \). See Appendices A and B for the full results.

In Table 2.3, columns [1], [2], and [3] present the OLS estimates. In columns [4], [5], and [6], I used weather, \( W_t \), as an instrumental variable (IV) for the price, \( P_t \), to address the potential simultaneity bias (i.e., the price could be high partially due to some positive demand shocks). As the first-stage (partial) \( R^2 \)’s indicate, weather shock (\( W_t \)) is a strong

\(^{35}\)The error term \( \varepsilon_t \) is allowed to be serially correlated in my estimation of the standard errors. The potential concern over autocorrelation in the residuals receives full consideration in Appendix 2.B. Appendix 2.A presents demand estimates and all of the empirical results based on the log-linear specifications.
instrument for \( P_t \). The results look reasonable with negative price coefficients (ranging between -29.68 and -58.43) and positive coefficients on both the buyers’ GDP and the price of the substitute good (except for [6], where the interaction term is negative and statistically insignificant). In what follows, I will use [5] as the baseline result because the use of IV is preferable to OLS in the presence of potential simultaneity bias, and because the key price coefficients (on \( P_t \) and \( Z_t P_t \)) are more precisely estimated in [5] than in [4] and [6], with more intuitive sign and magnitude.\(^{36}\)

Before proceeding to the estimation of the supply-side parameter, I discuss potential dynamics on the demand side. In particular, I considered a possibility that changes in the demand side might cause the markup to shrink, thereby offering an alternative explanation for the coffee crisis. A rise in the price elasticity of demand would lower the equilibrium price. After estimating the price coefficients for different periods, however, I found no statistically significant change across decades. Hence I ruled out the demand-side explanation for the price decline.

Why do changes in importing countries fail to alter the price coefficient systematically? First, the evolution of consumers’ tastes is not uniform or monotonic across the globe. According to America’s National Coffee Association, per-capita consumption in the United States peaked in 1962 at 3.12 cups per day, but its decline since then has happened only gradually over five decades.\(^{37}\) This development is likely more than offset by the population growth during the same period, along with the increasing per-capita consumption in other populous countries such as Japan and Germany.

\(^{36}\)The price coefficient in [5] (-49.05) implies the price elasticity of 0.174 at the mean price and quantity. The existing research also reported rather low price elasticities around 0.2. See Bates (1997, Appendix). Compared to these figures, the price coefficient in [6] (-29.68) is too small in magnitude, which is another reason to prefer [5] over [6]. The J test after the estimation of [6] does not reject the estimates in [5] at the 1%, 5%, and 10% levels.

The low price elasticity is likely due to the fact that green-bean costs account for only a small fraction of the retail prices for coffee drinks. In other words, although physically essential, green beans are a cost-wise negligible input for the final goods.

\(^{37}\)Trade press suggests several underlying developments behind such long-run changes in demand: (1) the introduction of substitutes such as bottled water and caffeinated energy drinks; (2) the price wars between Coca Cola and Pepsi that since the 1990s have lowered cola prices (arguably a close substitute for coffee); and (3) an increased “health-consciousness” that somehow deterred many consumers from drinking coffee.
Second, corporate buyers of coffee beans have not accumulated monopsony (oligopsony) power to match the monopoly power of exporters. These buyers are manufacturing firms that process beans into roasted or ground coffee products. Despite occasional mergers and acquisitions, the roasting industry was still comprised of 219 companies in 1992, 215 in 1997, 284 in 2003, and 253 in 2008 (IBISWorld 2008). Explaining the declining bean price by an increase in the roasters’ oligopsony power is therefore difficult. Furthermore, since the governments of the importing countries policed the cartel agreement, the companies in those countries were unlikely to have influenced the exporting countries’ market power on their own.

Finally, there is Starbucks Coffee, the largest coffee shop chain in the world. Could Starbucks’ meteoric rise in the late 1990s and early 2000s somehow increase the price-sensitivity of global coffee demand? Odds are against this explanation given the following facts. Even as late as 2007, Starbucks purchased less than 5 percent of the total global green-bean exports. This number alone would rule out the Starbucks hypothesis. In a broader context, coffee shops are a new channel of coffee retailing, but the overall per-capita consumption of coffee in the United States remained roughly constant between the late 1980s and 2000s. In other words, consumers merely switched from supermarkets to coffee shops as a point of purchase, and only partially. Moreover, although new to America, coffee houses have been around for centuries in Europe and Japan. London’s coffee houses opened in the seventeenth century—at around the same time modern corporations were invented—and the first coffee shop opened in Japan in 1878. Hence Starbucks Coffee and the emergence of coffee shops in the United States are unlikely to have changed the global pattern of coffee bean demand.

Therefore I believe it is reasonable to focus on the supply-side explanations such as the cartel’s collapse and extra coffee shipments from Vietnam. The estimated demand function

---

38I abstract from the multi-layered buyers. See Nakamura and Zerom (2010) for the analysis of vertical relationships and incomplete pass-through at the wholesale and retail levels in the United States.

39This calculation is based on Starbucks Corporation’s Form 10-K for the fiscal year ending September 30, 2007. As of September 30, 2007, the company had $324 million in fixed-price purchase commitments, $49 million in price-to-be-fixed contracts, and $339 million in unroasted bean inventories. Evaluated at the prevailing spot price of $1.35 per pound, the total amounts to 1.816 million 60-kg bags, which the company deemed enough stock of green coffee through the fiscal year 2008. This represents a mere 4.5% of the world export in 2007 (88.148 million 60-kg bags).
allows us to isolate the price impact of Vietnam’s export. The remaining task is to now quantify the market power (the extent of coordination) of all the other exporters both before and after 1989.

2.5.2 Estimates for the Cartel’s Coordination

I measured the extent of cooperation among the cartel countries by observing the markup and the net demand curve the countries face. To facilitate the interpretation of the measure, one can normalize the price-cost margin by the price coefficients of the demand estimate. The linear demand in equation (2.6) implies the following inverse demand function:

\[ P(Q_t) = \frac{1}{\alpha_1 + \alpha_4 Z_t} [Q_t - (\alpha_0 + \alpha_2 X_t + \alpha_3 Z_t + \varepsilon_t)], \]  

(2.7)

hence

\[ \left( \frac{\partial P}{\partial Q} \right)^{-1} = \alpha_1 + \alpha_4 Z_t. \]

Combine this equation with equation (2.5) to obtain the empirical counterpart to the collusion parameter:

\[ \hat{\mu}_t \equiv (P_t - mc_t) \frac{(\alpha_1 + \hat{\alpha}_4 Z_t)}{Q_t^C} \in (0, 1), \]  

(2.8)

which reflects the extent of cooperation. This parameter \( \mu_t \) takes values between zero and one, and equals \( 1/N \), where \( N \) is the effective number of (symmetric Cournot) competitors.

Figure 2.6 displays the evolution of \( \hat{\mu}_t \) and the implied effective number of competitors. The pre-1989 average of \( \hat{\mu}_t \) is 0.067, which is equivalent to a 15-firm Cournot competition, far from the monopoly level. Given that 54 countries are in the cartel, however, this level of market power is economically significant. For the post-1989 period, the normalized markup is 0.015 on average, or equivalent to a 66-firm Cournot competition, which is in line with the actual number of players. The mean difference between 0.067 and 0.015 is statistically significant at the 1-percent level. This gap reflects the breakdown of the ICA, under the assumption of constant \( \mu_t \) during and after the ICA, respectively.\(^{40}\) To assess its price

\(^{40}\) The 1989 lapse of the ICA roughly coincides with the abolition of national coffee boards in some exporting
impact, one can predict the coffee price in a hypothetical “world with a continued cartel treaty” by simulating the post-1989 exports and prices based on the pre-1989 parameter value (0.067). This is the main idea of the counterfactual analysis in the next section.

2.6 Decomposing the “Coffee Crisis”

Based on the estimated model, this section presents counterfactual simulations to identify the sources of the coffee crisis and discuss policy implications.
2.6.1 Counterfactual Experiments

One can use the estimated models of supply and demand to decompose the coffee crisis (the 75 percent drop in the real price between 1988 and 2001) into three different economic forces by simulating counterfactual cases (with/without the cartel and with/without Vietnam). The simulation results are displayed in Table 2.4 and Figure 2.7.

<table>
<thead>
<tr>
<th></th>
<th>Scenario 1</th>
<th>Scenario 2</th>
<th>Scenario 3</th>
<th>Actual Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>1988</td>
<td>199</td>
<td>198</td>
<td>199</td>
<td>198</td>
</tr>
<tr>
<td>1991</td>
<td>200</td>
<td>197</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>2001</td>
<td>165</td>
<td>147</td>
<td>52</td>
<td>50</td>
</tr>
</tbody>
</table>

Note: The counterfactual 1988 prices in Scenarios 1 and 3 are already slightly different from the Actual Price because these are the cases in which Vietnam’s exports is hypothetically set at zero. In contrast, the Actual Price and Scenario 2 reflect some positive exports from Vietnam, which had already started to grow.

Each of the four columns in Table 2.4 represents an alternative market structure. First, the “best case” for the coffee exporters (and the “worst case” for the coffee importers) entails the continuation of the cartel scheme (even after 1989) and the absence of Vietnam (Scenario 1). The price in 2001 would be $1.65 per pound. The gap between the actual 1988 price ($1.98) and $1.65 reflects weather and other transient shocks. Second, the presence of Vietnam (Scenario 2) pushes the 2001 price down to $1.47. This change represents the Vietnam effect. Third, the reverse case (Scenario 3: with the breakdown of cartel and without Vietnam) sees a large drop in price to $0.52. This decrease is due to the impact of the cartel’s breakdown. Finally, one would label the gap between $0.52 and the actual 2001 price ($0.50) the “interaction” effect of the two factors.

The comparison of these three cases against the actual prices in 1988 ($1.98) and 2001 ($0.50) allows one to decompose the coffee crisis into

\[ 1.48 = 0.33 \text{ (others)} + 0.18 \text{ (Vietnam)} + 0.95 \text{ (cartel)} + 0.02 \text{ (interaction)} , \]
Figure 2.7: Counterfactual Prices

Note: Scenario 1 is the most profitable case for the coffee exporters based on (1) the traditional exporters’ maintaining the degree of collusiveness at 0.066 (1981-88 average) even after 1988, and (2) zero exports leaving Vietnam in all years. Scenario 2 assumes only (1). Scenario 3 assumes only (2). The Actual case witnessed both the cartel breakdown and Vietnam’s export expansion.

or, in percentage terms

$$75\% = 17\% + 9\% + 49\%,$$

where the interaction term is dropped due to rounding. Thus the results suggest that, of the 75 percent drop in the real coffee price between 1988 and 2001, the breakdown of the coffee cartel explains 49 points and the emergence of Vietnam as a major exporter contributed further 9 points. Other factors, such as weather shocks, account for the remaining 17 points.

2.6.2 Discussion and Policy Implications

How big are these numbers? In this subsection, I discuss the results from three aspects: (1) the economic significance of the cartel’s market power; (2) the cartel’s price impact and its broader policy implications; and (3) Vietnam’s impact and its lessons for aid policies.

First, the international cartel of coffee exporters achieved a moderate but significant level of collusive pricing—moderate because the effective level of cooperation (market power) was
far from the monopoly level, and significant in the sense that over 50 countries managed to achieve an outcome as collusive as in a 15-player quantity competition. This is a remarkable level of coordination given the large number of producers and the associated difficulty of aligning their interests in an international political process.

Second, the counterfactual analysis reveals that the quota agreement’s lapse in 1989 explains a major part of the coffee crisis. This result implies that the policies affecting international market structure should be taken seriously. Although international commodity markets are often perceived and modeled as perfectly competitive, trade policy coordination could bring about major price consequences on the basis of changing market power. More specifically, international treaties (and STEs) may lead to an oligopolistic export market, even when millions of farms exist at the sub-national level.

Third, this market structure creates an environment in which development aid aimed at enhancing productive capacity in certain countries may harm producers in other (competing) countries by an effect akin to “business-stealing” in game-theoretic models. This implication becomes clearer in a simple accounting of global social welfare. Vietnam earned $0.8 billion in export revenue in 2001 compared to $1.7 billion in losses by all other exporters, for a net loss of $0.9 billion in global producer surplus. On the demand side, coffee drinkers in richer economies benefited from the lower price. The consumer surplus increased by $1.9 billion. Accounting for lost producer surplus, we may conclude that the world’s social welfare increased by $1 billion. Whether foreign aid should decrease the exporter surplus is another question.

These observations allow us to predict the likely impact of future policies. At a recent ICO meetings, Dr. Dang Kim Son, the general director of Vietnam’s Institute of Policy and

---

41Therefore, although the total official development assistance to Vietnam (not necessarily earmarked for coffee) reached over $1 billion each year, other developing economies, ironically, “lost” almost double that amount. I based this calculation on the preceding analysis under the actual post-ICA trade regime.

42Thus this study finds new lessons of the “unintended consequences” (Hirschman 1981) of foreign aid, outside the traditional debates over “Dutch disease” regarding the effects of natural resource discoveries or foreign transfers on the real exchange rate. For “Dutch disease,” see Rajan and Subramanian (2011), van Wijnbergen (1985), and Corden and Neary (1982). Edwards (1984) investigates this issue in the context of coffee exports in Colombia.
Strategy for Agriculture and Rural Development, recommended several measures for “sustainable development” including: (1) establishing a national coffee board that formulates sector strategies; (2) public investment in the monitoring system of supply and demand; and (3) international cooperation for monitoring and information sharing. These policies echo the ICA, except for the lack of enforcing agents such as the United States. Since monitoring and enforcement by importers was crucial to the ICA’s success, any efforts to resuscitate the ICA must consider some mechanisms to replace these functions. One possibility is that current information and communication technology might enable the real-time monitoring of quantities and prices across the globe. Conditional on the successful alignment of major exporting countries’ interests and the development of credible enforcement/punishment schemes, such multilateral arrangements might achieve significant price increases, as the simulations demonstrate in the previous subsection.

In contrast, little price impact can be expected from any single country’s unilateral plan to cut coffee cultivation area. Policy recommendations on rural development often include such plans (e.g., ICARD and Oxfam 2002), but the current international market structure would likely preclude any measurable price impact.

2.7 Conclusion

This paper exploits historical variations in data and an oligopoly model to separately identify the effects of changes in market structure (the breakdown of the cartel agreement and the emergence of a new exporter) on market power and price. Two findings emerge. First, the export quota agreement maintained moderate but economically significant market power. Second, counterfactual experiments suggest that, of the 75 percent drop in the price of coffee between 1988 and 2001, the disappearance of market power (the breakdown of the cartel in 1989) explains 49 points, the emergence of Vietnam as a major exporter explains 9 points, and other shocks such as weather explain the remaining 17 points. Thus the changing

---

43 Dang Kim Son, “Highlights of Vietnam Coffee Sector,” presentation at the third World Coffee Conference in Guatemala City, Guatemala (February 27, 2010).
market structure explains most of this commodity crisis.

These results reveal that, contrary to the conventional perception of commodity markets as perfectly competitive, economic forces of oligopoly are fully at play. In addition to the coffee market, the markets for crude oil, diamonds, base metals, and tropical agriculturals, to name but a few, also have a reputation for cartels and imperfect competition. Economic analysis of internationally traded commodities would gain much from explicitly incorporating market power.

These findings point to the room for cross-fertilization in public policies between competition, international trade, and development aid. Regarding the cartel treaty, my estimate of the ICA’s price impact suggests that market power can be used as a device for development assistance.\(^{44}\) That was indeed the intention of the U.S. foreign policy during the Cold War era. Likewise, the coffee price declined when the United States withdrew its support for the ICA in 1989. The falling coffee price was presumably another intended consequence from the perspective of the U.S. foreign policy makers, but the magnitude of spillover effects (including the exporting countries’ sovereign debt cancellations and the intensifying violence in Colombia and Rwanda) suggests that a more gradual withdrawal might have reduced the total cost of handling all such contingencies abroad.

In contrast, Vietnam would probably represent an unintended consequence of foreign aid. To the extent that Western and Asian donors helped fund the country’s coffee expansion, their aid policy effectively cut the price of coffee beans. In an oligopolistic market, assistance to a certain exporting country may hurt the other exporting countries. Thus, rather than adding new wealth to the developing world as a whole, coffee development in Vietnam mostly transferred the other producers’ surplus to that country. Such a policy is reasonable as far as Vietnamese policymakers are concerned, but not necessarily so from the donors’ viewpoint.

Hence the study calls for better coordination among different divisions of public policies. Specifically, the assessment of product-market impacts should be an integral part of designing

\(^{44}\)Somewhat differently, “fair trade” schemes also try to enhance the market power by artificially creating room for product differentiation.
trade and aid policies. In other words, when giving to the poor, policymakers should always let their left hands know what their right hands are doing.

45 Afghanistan could easily become another example of unintended consequences. On one hand, analysts often discuss economic development of the country in the context of the U.S. defense strategy to pacify the region. On the other hand, the country produces over 80% of the global heroine supply. Hence crop-substitution programs may seem like a solution on both fronts. However, if Afghanistan were to start exporting cotton, for example, the decision could potentially destabilize the country’s neighbors, Pakistan and India, who are major exporters of this commodity.
Appendix 2.A: Sensitivity Analysis 1 (Linear vs Log-linear Demand)

I show the full results of the demand estimation below. Column [11] of Table 2.5 is the baseline in the main text. The price coefficient of -49.05 implies a price elasticity of demand (at mean price and quantity) of 0.174, which is broadly consistent with the estimates under the log-linear specification in Table 2.6, as well as those in the previous literature. For example, the log-linear estimates of columns [1], [5], and [7] in Table 2.6 imply elasticities of 0.25, 0.123, and 0.131, respectively.46

To assess how the main findings vary under the log-linear specification, in Figure 2.8, I compare the empirical results based on the linear and log-linear demand specifications in three aspects: (1) the market power parameter estimates \{\hat{\mu}_t\}, (2) the simulated cartel exports \{\tilde{Q}_t\}, and (3) the counterfactual prices \{\tilde{P}_t\}.

First, the market power parameter estimates \{\hat{\mu}_t\} retain the main qualitative feature: the market power dropped in 1989. However, the drop is smaller in the log-linear case. The immediate cause of this discrepancy is as follows: \[ \hat{\mu}_t^L = -\frac{dQ}{dP} \frac{P_t-m_a}{Q_t} \] in the linear case, whereas \[ \hat{\mu}_t^{LL} = -\frac{d\log Q}{d\log P} \frac{P_t-m_a}{P_t} \] in the log-linear case. Hence, as \(P_t\) drops, \(\hat{\mu}_t^L\) will drop a lot (because the numerator decreases while the denominator increases), explaining a larger fraction of the variation in \(P_t\). Meanwhile, \(\hat{\mu}_t^{LL}\) will not drop much (because both the numerator and the denominator decrease), accounting for only a small fraction of the price fluctuation. This difference between \(\hat{\mu}_t^L\) and \(\hat{\mu}_t^{LL}\) also foreshadows the diverging results in the price simulations discussed below.

Second, the cartel’s simulated export quantities (under different scenarios) are almost identical between the linear and log-linear cases. Third, even though the simulated prices are derived from those similar export quantities, the price sequences based on the log-linear demand are much more volatile than those based on the linear demand.

The fact that, under the log-linear demand specification, the price elasticity of demand remains constant across different quantity levels drives this outcome. It follows that, in a

---

46 As Genesove and Mullin (1998) summarize, the two functional forms can be derived from the same general demand curve \(Q(P) = \beta (\alpha - P)^\gamma\). The linear case is when \(\gamma = 1\), whereas the log-linear form corresponds to \(\alpha = 0\) and \(\gamma < 0\). So they are not just two ad-hoc choices by the analyst.
### Table 2.5: Demand Estimates (Linear Specification)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_t$ (Coffee price)</td>
<td>-58.43***</td>
<td>-88.85***</td>
<td>-22.02***</td>
<td>-29.97***</td>
<td>-47.29***</td>
<td>-35.44***</td>
</tr>
<tr>
<td></td>
<td>(21.07)</td>
<td>(13.35)</td>
<td>(3.60)</td>
<td>(4.07)</td>
<td>(7.75)</td>
<td>(7.84)</td>
</tr>
<tr>
<td>$X_t$ (Buyers' GDP)</td>
<td></td>
<td>26421***</td>
<td>35681***</td>
<td>28599***</td>
<td>34373***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1654)</td>
<td>(3285)</td>
<td>(1131)</td>
<td>(4403)</td>
<td></td>
</tr>
<tr>
<td>$Z_t$ (Tea price)</td>
<td>-70.95***</td>
<td></td>
<td>21.30***</td>
<td>16.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6.26)</td>
<td>(7.43)</td>
<td>(11.63)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Z_tP_t$ (Interaction)</td>
<td>.1199***</td>
<td></td>
<td>.0341***</td>
<td>.0096</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.0183)</td>
<td>(0.0093)</td>
<td>(0.0132)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>79371***</td>
<td>104380***</td>
<td>173</td>
<td>-4137</td>
<td>-27137</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6252)</td>
<td>(3939)</td>
<td>(4447)</td>
<td>(2625)</td>
<td>(16708)</td>
<td></td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>.37</td>
<td>.84</td>
<td>.91</td>
<td>.92</td>
<td>.92</td>
<td>.92</td>
</tr>
<tr>
<td>Num. obs.</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>47</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dep. var.: $Q_t$</th>
<th>[7]</th>
<th>[8]</th>
<th>[9]</th>
<th>[10]</th>
<th>[11]</th>
<th>[12]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_t$ (Coffee price)</td>
<td>-45.62*</td>
<td>-96.94***</td>
<td>-30.11***</td>
<td>-31.92***</td>
<td>-49.05***</td>
<td>-29.68**</td>
</tr>
<tr>
<td></td>
<td>(24.11)</td>
<td>(15.91)</td>
<td>(6.00)</td>
<td>(4.00)</td>
<td>(8.20)</td>
<td>(14.06)</td>
</tr>
<tr>
<td>$X_t$ (Buyers' GDP)</td>
<td></td>
<td>24716***</td>
<td>36135***</td>
<td>28785***</td>
<td>37396***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1493)</td>
<td>(2789)</td>
<td>(912)</td>
<td>(5187)</td>
<td></td>
</tr>
<tr>
<td>$Z_t$ (Tea price)</td>
<td>-72.86***</td>
<td></td>
<td>22.84***</td>
<td>25.83*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6.25)</td>
<td>(5.54)</td>
<td>(14.73)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Z_tP_t$ (Interaction)</td>
<td>.1399***</td>
<td></td>
<td>.0354***</td>
<td>-.0055</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.0253)</td>
<td>(0.0099)</td>
<td>(0.0270)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>76542***</td>
<td>104468***</td>
<td>6353</td>
<td>-4423*</td>
<td>-38695*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(7697)</td>
<td>(3211)</td>
<td>(4773)</td>
<td>(2595)</td>
<td>(20126)</td>
<td></td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>.35</td>
<td>.83</td>
<td>.90</td>
<td>.92</td>
<td>.92</td>
<td>.92</td>
</tr>
<tr>
<td>Num. obs.</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>47</td>
</tr>
</tbody>
</table>

Partial $R^2$ for $P_t$: .52
Partial $R^2$ for $Z_tP_t$: - .71

**Note:** ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. Heteroskedasticity- and autocorrelation-consistent (up to 5 lags) standard errors are in parentheses.

I chose 5-year lags as maximum because a new coffee tree needs three to five years to reach full production potential, but different lags (bandwidths) do not materially alter the HAC standard errors. $P_t$ and $Z_tP_t$ are instrumented by the coffee supply shifters (weather shocks) $W_t$ and their interaction with the tea price $Z_tW_t$.

Consequently, the 75-percent drop in the price between 1988 and 2001 is decomposed differently under the log-linear specification: 10 points are due to the cartel’s breakdown and 5 points are due to Vietnam’s expansion. Miscellaneous (or ”residual”) factors explain the high-volume region (as in the years after the mid-1990s), even a wide variation in quantity will correspond to only a small change in price. Hence, even though export quantity varies widely across different counterfactual scenarios, the resultant prices do not differ much from the actual price series.
Table 2.6: Demand Estimates (Log-Linear Specification)

<table>
<thead>
<tr>
<th>Dep. var.: ( \ln Q_t )</th>
<th>OLS</th>
<th>[1]</th>
<th>[2]</th>
<th>[3]</th>
<th>[4]</th>
<th>[5]</th>
<th>[6]</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \ln P_t ) (Coffee price)</td>
<td>-.2500***</td>
<td>-.5881**</td>
<td>-.0767***</td>
<td>-.0919***</td>
<td>-.1231*</td>
<td>.3778***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.0407)</td>
<td>(.2623)</td>
<td>(.0132)</td>
<td>(.0163)</td>
<td>(.0717)</td>
<td>(.1129)</td>
<td></td>
</tr>
<tr>
<td>( \ln X_t ) (Buyers GDP)</td>
<td>.4004***</td>
<td>.4613***</td>
<td>.4297***</td>
<td>.5351***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.0163)</td>
<td>(.0632)</td>
<td>(.0523)</td>
<td>(.0541)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \ln Z_t ) (Tea price)</td>
<td>-.8825***</td>
<td>.0714</td>
<td>.5871***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.2318)</td>
<td>(.0685)</td>
<td>(.1598)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \ln Z_t \ln P_t ) (Interaction)</td>
<td>.0948**</td>
<td>.0066</td>
<td>-.0821***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.0437)</td>
<td>(.0109)</td>
<td>(.0205)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>12.39***</td>
<td>16.46***</td>
<td>10.40***</td>
<td>9.891***</td>
<td>10.36***</td>
<td>6.735***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.1960)</td>
<td>(.1355)</td>
<td>(.1025)</td>
<td>(.5516)</td>
<td>(.1573)</td>
<td>(1.012)</td>
<td></td>
</tr>
<tr>
<td>Adjusted ( R^2 )</td>
<td>.51</td>
<td>.83</td>
<td>.92</td>
<td>.93</td>
<td>.92</td>
<td>.93</td>
<td></td>
</tr>
<tr>
<td>Num. obs.</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dep. var.: ( \ln Q_t \ )</th>
<th>IV</th>
<th>[7]</th>
<th>[8]</th>
<th>[9]</th>
<th>[10]</th>
<th>[11]</th>
<th>[12]</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \ln P_t ) (Coffee price)</td>
<td>-.1311**</td>
<td>-1.0479***</td>
<td>-.0782***</td>
<td>-.0907***</td>
<td>-.0526</td>
<td>.5080*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.0650)</td>
<td>(.3419)</td>
<td>(.0142)</td>
<td>(.0190)</td>
<td>(.0585)</td>
<td>(.2784)</td>
<td></td>
</tr>
<tr>
<td>( \ln X_t ) (Buyers GDP)</td>
<td>.3990***</td>
<td>.4625***</td>
<td>.3844***</td>
<td>.5570***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.0156)</td>
<td>(.0628)</td>
<td>(.0407)</td>
<td>(.0697)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \ln Z_t ) (Tea price)</td>
<td>-1.3058***</td>
<td>.0718</td>
<td>.7157**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.3121)</td>
<td>(.0708)</td>
<td>(.3369)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \ln Z_t \ln P_t ) (Interaction)</td>
<td>.1736***</td>
<td>-.0037</td>
<td>-.1038**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.0586)</td>
<td>(.0083)</td>
<td>(.0501)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>11.76***</td>
<td>18.91***</td>
<td>10.42***</td>
<td>9.879***</td>
<td>10.44***</td>
<td>5.91***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.3704)</td>
<td>(1.807)</td>
<td>(.1090)</td>
<td>(.5372)</td>
<td>(.1275)</td>
<td>(2.06)</td>
<td></td>
</tr>
<tr>
<td>Adjusted ( R^2 )</td>
<td>.39</td>
<td>.83</td>
<td>.92</td>
<td>.93</td>
<td>.92</td>
<td>.93</td>
<td></td>
</tr>
<tr>
<td>Num. obs.</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td></td>
</tr>
</tbody>
</table>

| Partial \( R^2 \) for \( P_t \) | .61 | .53 | .68 | .63 | .67 | .35 |
| F stat for \( P_t \)             | 9.14 | 32.37 | 25.14 | 43.31 | 17.96 | 36.27 |
| Partial \( R^2 \) for \( Z_t P_t \) | – | .53 | – | – | .66 | .34 |
| F stat for \( Z_t P_t \)         | – | 37.23 | – | – | 11.64 | 43.62 |

Note: ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. Heteroskedasticity- and autocorrelation-consistent (up to 5 lags) standard errors are in parentheses.

I chose 5-year lags as maximum because a new coffee tree needs three to five years to reach full production potential, but different lags (bandwidths) do not materially alter the HAC standard errors. \( \ln P_t \) and \( \ln Z_t \ln P_t \) are instrumented by the coffee supply shifters (weather shocks) \( W_t \) and their interaction with the tea price \( \ln(Z_t)W_t \).

I view the results based on the log-linear demand as a mathematical artifact rather than an intrinsic economic insight because the log-linear demand curve is not very realistic regarding the nature of coffee demand in the high-volume (or, equivalently, low-price) region. Although the constant-elasticity assumption (inherent in this specification) dictates that coffee consumption goes up as the price of coffee goes down, there is a physical limit to how much one can drink. This observation casts doubt on the price simulations for the period the remaining 60 points, which is reminiscent of the low explanatory power given to \( \hat{\mu}_t^{LL} \).
after the mid-1990s, when the market was experiencing high volumes and low prices. The linear demand curve is the preferred model for this region.

Having cast doubts on the log-linear results, however, I find the fact that all of the discrepancies between the linear and log-linear cases are clearly attributable to the latter’s mathematical properties, leaving no mystery, reassuring. Finding that the relative importance of the two effects under investigation—the cartel’s market power and the emergence of the new producer—is not significantly altered between the two specifications is also encouraging; that is, the breakdown of the cartel had a somewhat bigger effect on the coffee price, and Vietnam’s rise was also a non-negligible factor.
Figure 2.8: Estimates and Simulation Based on Linear/Log-Linear Demand
Appendix 2.B: Sensitivity Analysis 2 (Autocorrelation)

The baseline demand estimation attempts to correct for the residual’s autocorrelation only with respect to standard errors. Here I display more comprehensive checks and corrections for the autocorrelation.

Table 2.7: Checks for Autocorrelation (Linear Specification)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>P_t (Coffee price)</td>
<td>-58.43***</td>
<td>-88.85***</td>
<td>-22.02***</td>
<td>-29.97***</td>
<td>-47.29***</td>
<td>-35.44***</td>
<td>60.52***</td>
<td>-99.55***</td>
<td>-29.96***</td>
<td>-31.91***</td>
<td>-46.97***</td>
<td>-27.62***</td>
</tr>
<tr>
<td>(15.18)</td>
<td>(14.93)</td>
<td>(5.52)</td>
<td>(5.39)</td>
<td>(11.33)</td>
<td>(13.19)</td>
<td></td>
<td>(15.25)</td>
<td>(18.66)</td>
<td>(5.75)</td>
<td>(5.53)</td>
<td>(11.46)</td>
<td>(20.05)</td>
</tr>
<tr>
<td>X_t ( Buyers’ GDP)</td>
<td>26421***</td>
<td>35681***</td>
<td>28599***</td>
<td>34373***</td>
<td></td>
<td></td>
<td>25008***</td>
<td>36122***</td>
<td>28414***</td>
<td>37417***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1518)</td>
<td>(3548)</td>
<td>(1354)</td>
<td>(4405)</td>
<td></td>
<td></td>
<td>(1474)</td>
<td>(3521)</td>
<td>(1341)</td>
<td>(6239)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z_t (Tea price)</td>
<td>-70.95***</td>
<td>21.30***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4.69)</td>
<td>(7.60)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z_t P_t (Interaction)</td>
<td>0.1199***</td>
<td>0.0341***</td>
<td>0.0096</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.0182)</td>
<td>(0.0124)</td>
<td>(0.0187)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>79371***</td>
<td>104380***</td>
<td>173</td>
<td>-32751***</td>
<td>-4137</td>
<td>-27137</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3789)</td>
<td>(3046)</td>
<td>(4584)</td>
<td>(12424)</td>
<td>(3664)</td>
<td>(16631)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Durbin-Watson d-stat</td>
<td>0.44</td>
<td>1.33</td>
<td>1.84</td>
<td>2.22</td>
<td>2.10</td>
<td>2.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Box-Ljung Q-stat</td>
<td>102.14</td>
<td>30.03</td>
<td>24.50</td>
<td>33.25</td>
<td>29.79</td>
<td>33.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>.00</td>
<td>.09</td>
<td>.27</td>
<td>.04</td>
<td>.10</td>
<td>.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dep. var.: Q_t | [7]   | [8]   | [9]   | [10]  | [11]  | [12]  |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>P_t (Coffee price)</td>
<td>-60.52***</td>
<td>-99.55***</td>
<td>-29.96***</td>
<td>-31.91***</td>
<td>-46.97***</td>
<td>-27.62***</td>
</tr>
<tr>
<td>(15.25)</td>
<td>(18.66)</td>
<td>(5.75)</td>
<td>(5.53)</td>
<td>(11.46)</td>
<td>(20.05)</td>
<td></td>
</tr>
<tr>
<td>X_t ( Buyers’ GDP)</td>
<td>25008***</td>
<td>36122***</td>
<td>28414***</td>
<td>37417***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1474)</td>
<td>(3521)</td>
<td>(1341)</td>
<td>(6239)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z_t (Tea price)</td>
<td>-71.84***</td>
<td>22.80***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(6.52)</td>
<td>(7.33)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z_t P_t (Interaction)</td>
<td>0.1350***</td>
<td>0.0319**</td>
<td>-0.0081</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.0303)</td>
<td>(0.0148)</td>
<td>(0.0364)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>79515***</td>
<td>105010***</td>
<td>5701</td>
<td>-34177***</td>
<td>-3488</td>
<td>-39263</td>
</tr>
<tr>
<td>(4321)</td>
<td>(3503)</td>
<td>(4580)</td>
<td>(12308)</td>
<td>(3546)</td>
<td>(24637)</td>
<td></td>
</tr>
<tr>
<td>Durbin-Watson d-stat</td>
<td>0.46</td>
<td>1.34</td>
<td>1.89</td>
<td>2.25</td>
<td>2.11</td>
<td>2.25</td>
</tr>
<tr>
<td>Box-Ljung Q-stat</td>
<td>98.92</td>
<td>30.03</td>
<td>24.50</td>
<td>33.25</td>
<td>29.79</td>
<td>33.49</td>
</tr>
<tr>
<td>p-value</td>
<td>0.00</td>
<td>0.19</td>
<td>0.16</td>
<td>0.03</td>
<td>0.10</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Note: *** *, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. Heteroskedasticity-robust standard errors are in parentheses. P_t and Z_t P_t are instrumented by the coffee supply shifters (weather shocks) W_t and their interaction with the tea price Z_t W_t.

To see whether concerns for serial correlation exist in the first place, Table 2.7 shows the two tests for the autocorrelation in the residual. Durbin-Watson’s d-statistics value should fall in the range of [0, 4], with values “far” from 2 indicating the presence of autocorrelation. The p-value is for the chi-squared test based on Box-Ljung’s Q-statistics, with values less than .05 commonly taken as evidence of autocorrelation.
The results suggest the specifications in columns [3], [5], [9], and [11] do not suffer from autocorrelation in the residuals. The inclusion of $X_t$ (log of the consumer countries’ GDP) seems to ameliorate the serial correlation issues by eliminating a time trend from the $Q_t$ (export quantity) series.

Still, by explicitly including the lagged dependent variable ($Q_{t-1}$) in the right-hand side, I can try correcting for what autocorrelation might exist. The results in Table 2.8 further confirm that my baseline demand estimates (hence all of the subsequent empirical results) are robust to the handling of autocorrelation.

Table 2.8: With Lagged Dependent Variable (Linear Specification)

| Dep. var.: $Q_t$ | OLS | | | | | |
|---|---|---|---|---|---|
| $Q_{t-1}$ (Lag) | .8717*** | .5717*** | .1781 | .0612 | .0352 | .0474 |
| | (.0590) | (.1285) | (.1534) | (.1448) | (.1743) | (.1645) |
| $P_t$ (Coffee price) | -10.45** | -34.02** | -18.57*** | -28.87*** | -45.79*** | -33.24* |
| $X_t$ ( Buyers’ GDP) | 21847*** | 33673*** | 27535*** | 3178*** |
| | (4695) | (6351) | (5586) | (6281) |
| $Z_t$ (Tea price) | -27.93*** | 20.80** | 17.77 |
| | (10.34) | (8.28) | (11.51) |
| $Z_tP_t$ (Interaction) | .0412** | .0329** | .0071 |
| | (.0195) | (.0152) | (.0217) |
| Constant | 11936*** | 44568*** | 72 | -31325** | -3746 | -2757 |
| | (4276) | (13096) | (4956) | (4448) | (16899) |
| Box-Ljung Q-stat | 37.47 | 24.88 | 26.94 | 33.73 | 29.57 | 33.51 |
| p-value | .01 | .25 | .17 | .04 | .10 | .04 |

| Dep. var.: $Q_t$ | IV | | | | | |
|---|---|---|---|---|---|
| $Q_{t-1}$ (Lag) | .8763*** | .5479*** | .0706 | .0187 | .0245 | .0316 |
| | (.0645) | (.1894) | (.1965) | (.1925) | (.2063) | (.1952) |
| $P_t$ (Coffee price) | -9.76 | -42.40 | -27.68*** | -32.29** | -45.32*** | -22.30 |
| | (7.96) | (26.41) | (8.36) | (9.33) | (17.08) | (23.99) |
| $X_t$ ( Buyers’ GDP) | 23364*** | 35665*** | 27613*** | 37966*** |
| | (5340) | (8076) | (6474) | (7798) |
| $Z_t$ (Tea price) | -28.80* | 23.48** | 31.82* |
| | (16.88) | (9.30) | (17.82) |
| $Z_tP_t$ (Interaction) | .0506 | .0302 | -.0182 |
| | (.0398) | (.0182) | (.0380) |
| Constant | 11202** | 47063** | 4997 | -34359* | -3103 | -45218* |
| | (5577) | (21009) | (5328) | (14570) | (4252) | (24270) |
| Box-Ljung Q-stat | 37.83 | 24.88 | 26.94 | 33.73 | 29.57 | 33.51 |
| p-value | .01 | .32 | .17 | .04 | .11 | .05 |

Note: ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. Heteroskedasticity-robust standard errors are in parentheses. $P_t$ and $Z_tP_t$ are instrumented by the coffee supply shifters (weather shocks) $W_t$ and their interaction with the tea price $Z_tW_t$.

Specifically, the coefficient estimates for $Q_{t-1}$ (lag) are not significantly different from...
zero in models [3] through [6] and [9] through [12]. The p-values for the Box-Ljung test are also comfortably above .05 in columns [2], [3], [5], [8], [9], and [11]. Moreover, the price coefficient estimates are almost identical to the original results in Table 2.5 for at least [3], [5], [9], and [11]. Therefore we can conclude the baseline demand estimates are not sensitive to the way I address (or do not address) the autocorrelation issues.

Finally, considering potential reasons for the persistence in $Q_t$ over time is interesting, even though my baseline case does not exhibit such issues. One factor may be the addictive nature of coffee consumption, which would give rise to a sticky demand over time through habit-formation dynamics. Historically speaking, however, per-capita consumption in the largest coffee importing country, the United States, more or less plateaued in the 1960s. Although Americans “discovered” coffee houses in the 1990s, with chain-store operations proliferating into the 2000s, the phenomena merely changed the composition and location of coffee consumption in subtle ways.

In contrast, less-developed countries are more likely to have formed a caffeine addiction during the sample period. However, simply controlling for GDP—that is, income and population growth combined—seems enough to eliminate any habit-formation effect. Perhaps the pure “addiction effect” would be difficult to separate from the “income effect” when both factors work in the same direction during the same period of time.

On the supply side, potential “stickiness” factors include the gradual evolution of production capacity (including planting and growing of coffee trees and construction of transportation infrastructure) and the inventory/storage adjustment at various stages of the supply chain in both exporting and importing countries (although the extent of storage seems limited). However, these types of “investment” dynamics would not play major roles unless major unexpected shocks were to arise, because as long as the rigid quota regime is in place—which it was in this case—each producing country has little incentive to engage in wild investment activities. Hence the only occasions where time-to-build and inventory stickiness would matter are when frosts and droughts of once-in-a-century magnitude hit Brazil’s coffee producing regions. Such events are, by definition, rare and therefore unlikely to have driven the dynamics in my dataset over the entire sample period.

68
Appendix 2.C: Fundamentals-driven Commodity Booms

In this section, I complement the main analysis of the commodity crisis (bust) with another set of counterfactual experiments that explores the possibility of commodity booms. Although the popular press often attributes rising prices to “speculators,” these exercises demonstrate that economic fundamentals alone, such as growing coffee demand from China, cross-border land deals, and multilateral export agreements, can cause even tenfold price increases.

To what extent can economic fundamental explain commodity booms such as those we have witnessed in recent years? This question is directly relevant to regulatory policies since, at the height of commodity booms in 2007 and 2008, many observers almost mechanically blamed “speculators” and called for curbing trade in futures contracts. To shed light on this issue, this section extends the scope of counterfactual experiments to (1) global demand growth, (2) increased market power, and (3) the combination of both. Table 2.9 summarizes the results.

Table 2.9: Counterfactual Prices for Selected Years (II)

<table>
<thead>
<tr>
<th>(cents/lb)</th>
<th>Actual</th>
<th>Scenario 4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>China drinks coffee</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Cross-border M&amp;A</td>
<td>No</td>
<td>No</td>
<td>Regional</td>
<td>Global</td>
<td>Regional</td>
<td>Global</td>
</tr>
<tr>
<td>1998</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
</tr>
<tr>
<td>2001</td>
<td>50</td>
<td>64</td>
<td>460</td>
<td>879</td>
<td>1,069</td>
<td>2,096</td>
</tr>
<tr>
<td>2006</td>
<td>109</td>
<td>131</td>
<td>556</td>
<td>1,016</td>
<td>1,262</td>
<td>2,429</td>
</tr>
</tbody>
</table>

Source: CRB; Author’s calculation.

Demand-driven Boom: When China Drinks Coffee

The analysis of almost any commodity market frequently discusses the growing demand from emerging economies, notably China. This factor is particularly relevant for the coffee market. Although the current Chinese consumption level is equivalent to about one cup per person per year, the figures for the UK, Japan, the United States, Germany, and Finland are 1.01, 1.42, 1.75, 2.55, and 5.33 cups, respectively, per capita per day. A mere catch-up

---

48According to the 2005 surveys conducted by the International Coffee Organization (available online).
with the Japanese standard implies the doubling of the global demand for coffee.49

Scenario 4 simulates the price effect of this doubled demand by hypothetically cutting the price coefficient of demand by half for 1999 and afterwards. The prices in 2001 and 2006 are 28 percent and 20 percent higher than the actual prices, respectively.50

Supply-driven Boom: When Brazil Grabs Africa

To push the market-power viewpoint further, one can consider “merger simulations” as another set of counterfactual experiments. One possible “merger simulation” is the continuation of recent trends in cross-border land deals.51 If such transactions lead to concentrated ownership (or managerial decision making) of farms, that would imply a higher market power. Another possibility is the increased coordination in export policies across producing countries.52

Suppose coffee export policies are harmonized at regional levels: Latin America and the Caribbean (20 countries), Asia (10 countries), and Africa (25 countries). If the output decisions are made at this level, market power will rise to $\mu = 0.33$ (equivalent to a symmetric 3-firm Cournot competition). This assumption underlies Scenario 5. The effect is spectacular, with the price in 2006 surging fivefold to $5.56/lb, even in the presence of Vietnam as an outsider producer.

49Such a change is not unrealistic. In Shanghai, for example, Starbucks Coffee had 21 stores in 2003. It now serves coffee at 117 stores (as of February 2010): a 457% growth.

50Still, the doubled demand would not nearly compensate for the lost market power, as embodied in Scenario 1 and Scenario 2, in terms of price. The reason is that, with negligible market power at $\mu = 0.015$, coffee exporting countries are behaving quite competitively, to the extent that a lower price elasticity of demand is left largely unexploited.

In reality, doubling the supply instantaneously would be impossible, mainly because of the lack of availability of land in suitable climate zones and because new coffee trees require 4 to 6 years to reach full production capacity. However, one can still interpret the results as an indication of likely outcomes in the long run, given certain levels of market power.

51Foreign land acquisitions have increased in recent years—for example, South Korean wheat projects in Sudan, India’s investment in Ethiopia, or Kuwaiti rice deals with Cambodia.

The International Food Policy Research Institute (Economist, May 23, 2009) calculates that foreigners have negotiated between 15 million and 20 million hectares of farmland in deals in developing countries since 2006, equivalent to about one-fifth of all of the farmland of the European Union, worth up to $30 billion.

52The euro and the European Central Bank, for instance, present a successful case of regionally integrated (albeit monetary) policies.
What if $\mu = 1$, equivalent to the perfect collusion (monopoly) level? This scenario is the case of globally integrated coffee-supply decision making (Scenario 6). The coffee price in this scenario approaches $10.00 per pound. Buying out coffee farms in poor countries and then reducing the global output to a monopoly level would be a profitable use of Brazil’s sovereign wealth funds. Alternatively, foreign aid can buy influence in other producers’ export policies.

**Demand- and Supply-driven Boom: With Both China and Brazil**

Finally, Scenario 7 and Scenario 8 combine the supply- and demand-side counterfactuals by supposing the doubled demand and regionally or globally integrated supply decisions. Because the increased market power in these scenarios exploits the lower demand elasticity, the prices in 2006 are twelve- and twenty-twofold higher, respectively.

Thus, the combination of growing demand and some supply-side realignment can easily raise the price by more than tenfold. All of the counterfactual scenarios in this section reflect radical thought experiments, but the results suggest the possibility of wild fluctuations in commodity prices when those factors are at play. In the face of growing demand, public policies concerning agricultural land-related FDI can potentially impact the commodity market significantly.
REFERENCES


74


CHAPTER 3

Estimating the Innovator’s Dilemma: Structural Analysis of Creative Destruction in the Hard Disk Drive Industry

3.1 Introduction

“In the long run we are all dead,”¹ and firms and technologies are no exception. Netflix’s movie download service has grown fast, whereas Blockbuster, a once-dominant DVD rental chain, filed for bankruptcy protection in 2010 after a reluctant pursuit of an online distribution service. Amazon is now selling everything from electronic books to disposable diapers, whereas Borders, America’s number-two book retailer, liquidated its shops in 2011 after belated efforts to introduce its own electronic reader. These examples may seem extreme, but even when introducing a new technology is not too difficult, “the old winners tend not to adapt well; the new entrants face lower cost of entry, and quickly become successful,” as a former CEO of Intel warned based on the experience of the world’s biggest chip maker (Grove 1996). Some incumbents never introduce a new technology/product despite shrinking demand for their existing products, a puzzling phenomenon called “the innovator’s dilemma” (Christensen 1997). This paper asks why incumbent firms are slower than entrants to innovate, and empirically tests three theoretical determinants of incumbents’ innovation.²

¹John Maynard Keynes, A Tract on Monetary Reform (1923), Ch. 3.

²I use the words innovation, diffusion, and technology adoption/choice interchangeably in this paper because I am studying a case of technological transition that exhibits all of these features. Alternatively, innovation can be more narrowly defined as invention and its first commercialization, as distinct from its subsequent spread, but such a distinction does not seem adequate for the situation in the HDD industry, where a technological roadmap is widely shared among engineers and managers throughout the industry.
So why do incumbents delay innovation? Viewed from a microeconomic perspective, the
determinants of innovation timing include (1) cannibalization, (2) different costs, (3) pre-
emption, and (4) institutional environment (Hall 2004, Stoneman and Battisti 2010). First,
the benefits of introducing a new product are smaller for incumbents than for entrants be-
cause of cannibalization, to the extent that the old and new goods substitute for each other.
By introducing new goods, incumbents are merely replacing their old source of profits, so
Arrow (1962) calls this mechanism the “replacement effect.” Second, incumbents may face
higher costs of innovation because of organizational inertia. Economic theory, as well as case
studies, suggest that as firms grow larger and older, their R&D efficiency diminishes (e.g.,
Schumpeter 1934);\(^3\) although, \textit{a priori}, hypothesizing that incumbency confers some advan-
tages due to accumulated R&D capital is equally plausible (e.g., Schumpeter 1942). Hence,
whether incumbents have a cost advantage or disadvantage is an open empirical question.
Third, market structure dynamics play an important, countervailing role, as theories pre-
dict incumbents should innovate more aggressively than entrants to preempt potential rivals
(e.g., Gilbert and Newbury 1982) under various oligopolistic settings. Finally, the impact of
these three determinants will change under different institutional contexts, such as the rules
governing patents, non-compete clauses,\(^4\) R&D subsidies, or international trade. In total,
these three competing forces (plus institutional contexts) determine innovation timing. Can-
nibalization delays incumbents’ innovation, whereas preemptive motives accelerate it, and
incumbents’ cost (dis)advantage would further reinforce these tendencies. Given this “tug
of war” between the three theoretical forces, I propose to explicitly incorporate them into a
unified model and estimate how much each of them matters empirically.

The goal of this paper is to empirically quantify these competing forces behind “the
innovator’s dilemma” in the hard disk drive industry, which is a highly relevant setting.
This industry is a canonical case of “creative destruction” (Schumpeter 1942) or “disruptive
\(^3\)The existing literature suggests various reasons for incumbents’ inertia, such as bureaucratization
(Schumpeter 1934), hierarchy (Sah and Stiglitz 1986), the loss of managerial control (Scherer and Ross
1990), and informational, cognitive, or relationship reasons (Grove 1996, Christensen 1997).

\(^4\)A non-compete clause is a type of employment contract that restricts employees from competing with
their former employer firms. Such contracts work as entry barriers when the employees of existing firms
leave their employers to start new businesses (called “spin-outs”).
innovation” (Christensen 1997), where cohorts of firms come and go with the generational transitions of technologies. I construct a unique dataset from the industry publications, DISK/TREND Reports (1977–99), which record a comprehensive set of firms (both incumbents and potential entrants) for more than two decades. First, I build and estimate a dynamic oligopoly model that explicitly incorporates cannibalization, heterogeneous sunk costs, and preemption (dynamic strategic interactions), all of which endogenously determine the timing of innovation and the evolution of market structure. Then I measure the effects of the three forces (i.e., estimate the innovator’s dilemma) by contrasting the outcome of the estimated model with those of three counterfactual simulations in which firms ignore each of these incentives, respectively. Finally, to study broader implications of the phenomena, I simulate evolutions of the industry under four alternative institutional settings: (1) a broad patent regime, (2) a ban on non-compete clauses, (3) R&D subsidies for incumbents, and (4) a ban on foreign goods.

The estimation results suggest that despite strong preemptive motives and a substantial cost advantage over entrants, incumbents are reluctant to innovate early because of cannibalization, which can explain at least 51% of the timing gap. The results from the four policy simulations highlight the pro-innovation effect of competition, even though the overall effectiveness of public policies seems somewhat limited. These findings represent a contribution to the innovation literature in three respects (Hall 2004, Stoneman and Battisti 2010, Cohen 2010). First, through the modeling of strategic “creative destruction,” I provide an empirically tractable microeconomic foundation of the phenomena that are central to both innovation and industry evolution. Second, by quantifying the three determinants of innovation timing (and its heterogeneity between incumbents and entrants), I separately measure the importance of each of these theoretical incentives in an actual industry setting. The most interesting finding is that incumbents may lag behind entrants, despite their advantage in innovation costs, which suggests a substantial part of what researchers have previously

\[5\]In concluding her literature survey on innovation and diffusion, Hall (2004) suggests “there is room for an approach that is more structural and grounded in the choice problem actually faced by the adopter.” This paper implements her recommendations in this respect. The following two contributions are the fruits of this approach.
understood as organizational inertia could potentially be reinterpreted as an effect of cannibalization. Third, by simulating alternative competitive environments, I derive implications for both managerial and public policies. For example, I find a ban on international trade discourages innovation and hurt consumers. However, social welfare sometimes improves under anti-competitive policies, such as broad patents. Ironically, welfare improves not through promotion of innovation but through cost savings from preventing “excess” entry/innovation.

The timing of innovation in general and the incumbent-entrant timing gap in particular are important for both businesses and policymakers. Who innovates and survives better (and why) is a central question for individual firms. The timing gap also has broader implications for public policies because it is a symptom of the fundamental heterogeneity between incumbents and entrants. Discussing pro-innovation competition policies, Bresnahan (2003) stresses the importance of innovation by industry outsiders and new entrants that often results in Schumpeterian changes: “For society to have a rapid rate of technical progress, we need innovative competition from outsiders as well as innovation incentives for incumbents.” Depending on how incumbents’ and entrants’ incentives differ, competition and innovation policies will have different consequences. Understanding the determinants of the timing gap is the first step toward designing a pro-innovation competition policy. For these purposes, I have chosen to study the HDD industry.

An HDD is a component of a personal computer (PC) that stores information. Desktop PCs used 5.25-inch HDDs during the 1980s, but 3.5-inch HDDs became popular during the 1990s, so those firms that exclusively manufactured 5.25-inch HDDs disappeared by the turn of the century (see Figure 3.1). For studying the long-run dynamics of firms and technologies, the HDD industry is ideal for three reasons. First, both technologies and firms are empirically tractable (i.e., the old and new technologies/products are different enough to represent distinct investment opportunities for firms, yet similar enough to compete within the same market as the “secondary information storage device” in desktop PCs).6 That many firms competed, both incumbents and entrants, is also helpful for econometric purposes. Second,
an unusually long panel dataset is available in the form of an annual industry publication series, the DISK/TREND Reports (1977–99). I obtained hard copies of the 23 volumes, interviewed the editor, and manually digitized the quantitative as well as qualitative contents. This 23-year period is long enough to cover the rises and falls of multiple generations of technologies and firms. Third, this industry is suitable for assessing the welfare impacts of public policies in competition and innovation, because contentious policy issues arose during its history, such as (frivolous) claims of patent infringement and restrictions on spin-out activities based on non-compete clauses.

Figure 3.1: Shifting Generations of Technology

Note: Shipment-based recognition of firms. Major firms only.

The task of quantifying the three forces calls for a structural approach, because incentives to innovate are sensitive to the technological and institutional context of an industry. Moreover, these incentives all interact with each other in a complex manner, are not directly observable, and create a situation in which both innovative activities and market structure evolve endogenously. Hence, absenting natural experiment-like episodes, some modeling is needed to identify these factors. In addition, policy evaluation at an industry level must address the Lucas critique (i.e., one cannot predict how innovation and competition would evolve under hypothetical policies, such as an alternative patent system, without estimating structural parameters). For these reasons, this paper takes a structural approach. I build, estimate, and simulate a model that incorporates the technological and institutional features of the HDD industry, as well as the three theoretical forces.
My empirical analysis proceeds in three steps as follows. First, I estimate demand using a standard discrete choice model for differentiated goods (the old- and new-generation HDDs). That is, I let the data tell the degree of substitution between the old and new goods and hence of cannibalization. Second, I recover marginal costs of production, implied from the first-order conditions of static competition (multi-product Cournot). From these demand and cost estimates, I calculate the static equilibrium profits in every state of the industry, that is, given any number of active firms in the market. These profit estimates embody the relationship between market structure and profitability and hence give a preliminary indication of the extent to which preemption motivates firms’ introduction of a new technology/product. Third, I feed these static (period) profits into a dynamic model to estimate the sunk costs of innovation. The model features two types of firms, incumbents and entrants, so I estimate the sunk costs separately for each type. I explicitly incorporate firms’ dynamic discrete choice between entering, exiting, continuing operation with the old product, or introducing the new product. I fully incorporate preemptive motives, because firms interact strategically and are forward-looking with rational expectations over the endogenous evolution of market structure. To reflect the computer industry’s ever-changing nature, I make my model non-stationary, allowing demand, costs, and hence value and policy functions to change over years.

Conceptually, this third step is simple. I employ maximum likelihood estimation (MLE) to find the sunk cost parameter values that would maximize the likelihood of observing the actual innovation and entry/exit behaviors in the data. Intuitively, I invoke a “revealed preference” argument for every firm-year observation, comparing the benefits and costs of different alternatives and then inferring the sunk-cost size that is consistent with the observed action. Computationally, however, this procedure poses two technical challenges. One is the possibility of multiple equilibria. I address this issue through parsimonious modeling (small choice sets, a small state space, and period-by-period solutions) to guarantee the uniqueness of equilibrium under certain configurations, along with numerical and analytical randomization. The other problem is the computational burden of calculating the equilibrium strategies (choice probabilities) and expected values. That is, for each set of
candidate parameter values in the MLE procedure, I need to solve the dynamic game for its equilibrium play. I address this issue by coding the most computationally intensive routines (the calculation of expected values) in the C language within the MATLAB platform.

I have organized the rest of the paper as follows. The remainder of this section explains how this research contributes to the literature on innovation and industry dynamics. Section 3.2 summarizes the technological and institutional background of the HDD industry. Section 3 describes the model. Sections 3.4 and 3.5 explain the estimation procedure and results. In section 3.6, I quantify the three economic forces behind “the innovator’s dilemma.” In section 3.7, I evaluate welfare consequences of four different policies in innovation and competition. Section 3.8 concludes with a discussion of the implications for managerial practices and public policies.

3.1.1 Related Literature

This paper studies innovation and industry dynamics using a structural approach. As such, three bodies of literature motivate this study: competition and innovation, market structure dynamics, and the structural estimation of dynamic games. I aim to enhance the first two bodies by providing a microeconomic foundation of the “disruptive innovation” phenomena (i.e., the generational transitions of technologies and firms) and by quantifying the three incentives to innovate that have been prominent in theory but have few empirical counterparts. In doing so, I build on and extend the frameworks developed in the third body of literature by featuring a radical product innovation and incumbent-entrant heterogeneity (hence entry/exit dynamics as well) in a dynamic oligopoly model. I summarize the findings and remaining tasks of the literature in the following section.

Innovation

Many papers, both theoretical and empirical, have studied the relationship between competition and innovation, with mixed predictions and inconclusive evidence (see Gilbert [2006] and Cohen [2010] for detailed surveys). Arrow (1962) predicted an incumbent monopolist
has less incentive to innovate than perfect competitors because of the “replacement effect” (i.e., the substitution between the old and new technologies), against which others theorized the preemptive motive for an incumbent monopolist to innovate more aggressively than an entrant (e.g., Gilbert and Newbery 1982, Reinganum 1983, Fudenberg and Tirole 1986). Empirical works have simplified and recast these predictions as two competing hypotheses regarding the effect of market structure on innovation, typically regressing R&D spending (or other measures of innovative activities) on the market share concentration (or measures of market power such as markups) in a cross-sectional dataset of industry-/firm-level observations. The findings are mixed. Horowitz (1962), Hamberg (1964), Scherer (1967), and Mansfield (1968) were the first among many to find a positive relationship, whereas Williamson (1965), Bozeman and Link (1983), Mukhopadhyay (1985), and Blundell et al. (1999) found a negative effect of concentration. Moreover, Scherer (1967) found a non-linear, “inverted-U” relationship, later replicated by Scott (1984), Levin et al. (1985), and Aghion et al. (2005). Gilbert (2006) attributed the inconclusiveness of evidence to (1) the failure to control for contingencies highlighted by theorists, (2) the presence of fundamental heterogeneities across industries, and (3) the shortcomings of data and methods.

Another limitation was that whereas theories emphasized the heterogeneity of incentives to innovate between incumbents and entrants, the earlier data analyses considered the behavior of incumbents alone. This omission was problematic also from an empirical perspective because “there is abundant evidence from case studies (...) new entrants contribute a disproportionately high share of all really revolutionary new industrial products and processes” (Scherer 1980). Geroski (1989, 1991, 1994), Acs and Audretsch (1991), and Gans et al. (2002) partially filled this gap by analyzing the relationship among entry, competition, and innovation. However, rather than showing simple effects of one on the others, Geroski’s studies suggested other industry characteristics such as technological opportunity and appropriability simultaneously determined these three variables. Thus, to the list of concerns Gilbert (2006) raised, Cohen (2010) added a fundamental critique that market structure was a function of innovation itself, and entry, competition, and innovation were simultaneously determined by more structural factors such as demand and technological opportunities.
Some papers employed instrumental variables to address this simultaneity problem (Howe and McFetridge 1976, Levin et al. 1985, Blundell et al. 1999, Aghion et al. 2005), but Cohen (2010) concluded “cross-sectional analyses (...) have offered little insight into the actual role of these industry-level factors.” The consideration of the underlying industry dynamics remains a major challenge as well. This paper addresses these issues by focusing on a specific high-tech industry, explicitly incorporating the technological context of the industry into a dynamic model, and estimating the structural parameters of the model.

Industry Evolution (Market Structure Dynamics)

The co-evolution of technology and competition has played a central role in the studies of industry evolution, or market structure dynamics, which is the second body of literature that motivates this study. Theoretical models and qualitative case studies constitute the bulk of this literature, whereas data analyses are scarce due to data limitations and the simultaneity issue. Prominent models include those of Nelson and Winter (1978, 1982), Jovanovic (1982), Hopenhayn (1992), Ericson and Pakes (1995), Klepper (1996), and Sutton (1998). Researchers developed these theories alongside the documentation of empirical regularities (e.g., Mueller and Tilton 1969, Abernathy 1978, Abernathy and Utterback 1978, Utterback 1979, Gort and Klepper 1982, Klepper and Graddy 1990, and Klepper and Simons 2005).

Technology and market structure evolve particularly closely in a moment of “disruptive innovation,” when an industry experiences the generational transitions of firms and technologies. Numerous case studies record such instances: Tushman and Anderson (1986), Mitchell (1989), Henderson and Clark (1990), Henderson (1993), Ehrnberg and Sjöberg (1995), Christensen (1997), and Tripsas (1997). More recent papers formally model these generational transitions, such as Adner and Zemsky (2005) and Klepper and Thompson (2006). However, a quantitative empirical work has not yet been released, probably because the drastic nature of the phenomena poses challenges to both data collection and empirical methods. Given the context of this literature, I propose a formal empirical analysis by developing a structural model of the Ericson-Pakes (1995) class and applying it to data from the HDD industry, the canonical case of “disruptive innovation” studied by Christensen (1993, 1997), Chesbrough
Structural Estimation of Industry Dynamics

Structural analysis of industry dynamics is the third body of literature on which this study builds (for an overview, see Ackerberg et al. [2007] and Doraszelski and Pakes [2007]). The purpose of developing and estimating a structural model is to consider both theory and data explicitly, so invisible, theoretical effects and their welfare implications could be quantified through counterfactual simulations. Two approaches exist to estimate Markov-perfect equilibrium (MPE) models of the Ericson-Pakes (1995) class. The “full-solution” approach started with Rust’s (1987) analysis of investment under uncertainty (bus engine replacement), whereas Hotz and Miller (1993) pioneered the “two-step” approach. The first approach requires intensive computation because a full solution of a dynamic programming problem is needed for each set of candidate parameter values; the second approach demands a large dataset to alleviate such a computational burden by a first-stage non-parametric estimation procedure. The first approach has been pursued by Su and Judd (2008), who proposed a re-formulation of the problem as Mathematical Programming with Equilibrium Constraints (MPEC) and the use of state-of-the-art solvers, by Besanko et al. (2010), who used a homotopy or path-following algorithm to trace the entire set of equilibria, and by Weintraub et al. (2008a, 2008b), who proposed oblivious equilibrium (OE) as a more tractable approximation concept for MPE. The second approach has been further developed by Aguirregabiria and Mira (2007), Bajari et al. (2007), Pakes et al. (2007), and Pesendorfer and Schmidt-Dengler (2008), who proposed various two-step estimators for dynamic games. I have chosen to follow the first approach by alleviating the computational burden through parsimonious modeling (symmetry among the same type of firms, relatively small choice sets and state space, and a finite-horizon environment). Since I study a geographically globalized industry, the dataset does not contain a large number of independent markets, which is an empirical setting where non-parametric estimation in the second approach might not function properly.

These dynamic structural frameworks have been applied to study both innovation and entry/exit (i.e., evolution of market structure), but only independently. I propose to incorpo-
rate both simultaneously, in order to study generational transitions of technologies and firms. Four recent papers studied various forms of innovation. Schmidt-Dengler (2006) took a full-solution approach to analyze U.S. hospitals’ adoption timing of magnetic resonance imaging (MRI) technology. Goettler and Gordon (2011) also took the same approach to study the introduction of faster microprocessors by Intel and AMD. Xu (2008) applied an OE framework to the cost reduction among Korean electric motor makers. Finally, Eizenberg (2009) studied the introduction of PCs with faster chips. I follow Schmidt-Dengler (2006) and Goettler and Gordon (2011) in taking a full-solution approach in estimating a dynamic oligopoly; to investigate incumbent-entrant heterogeneity as well as market structure dynamics, I extend the scope of analysis to include entry/exit.

Structural analysis of entry/exit started from the estimation of static models (e.g., Berry 1992, Mazzeo 2002, Seim 2006, Jia 2008). More recent papers employed dynamic models: Ryan (2011) studied the cost of environmental regulation in the cement industry, Collard-Wexler (2010) studied the role of demand fluctuations in the concrete industry, Dunne et al. (2009) studied several retail industries, and Kalouptsidi (2010) focused on time to build in the bulk shipping industry. This paper shares the focus on industry evolution with the second strand of literature, but my empirical approach also builds on the first strand, namely, Seim’s (2006) characterization of an entry/exit game. Specifically, I employ a finite-horizon setup to reflect the HDD industry’s non-stationary environment, which allows me to solve for an industry equilibrium by backward induction, one (static) subgame at a time. More substantively, I extend the framework for analyzing entry/exit to incorporate incumbents’ technology adoption, so I can study their technology choice in relation to entry/exit behavior and incumbent-entrant heterogeneity.

In short, this paper presents the first structural analysis of “disruptive innovation” (to my knowledge), bridging the frameworks to analyze innovation and entry/exit in a simple model. This model is empirically tractable and motivated by Christensen’s (1993, 1997) case study as well as my own reading of the original data source. The next section summarizes the institutional background of the HDD industry.
3.2 Industry and Data

This section describes the key features of the HDD industry and explains why it is particularly suitable for the study of innovation and industry evolution.

3.2.1 HDD Industry: Canonical Case of Innovation and Evolution

The HDD industry provides a particularly fruitful example for the study of technological change and industry dynamics.

First, the HDD industry is the canonical example of “disruptive innovation.” Multiple generations of technologies were born, matured, and died within a decade or two. A generation was defined by the diameter of disks used: 14-, 8-, 5.25-, 3.5-, and 2.5-inch (see Figure 3.1). The introduction of a new HDD of smaller diameters required a significant technological investment because each firm had to go through a process of trial and error in determining the adequate configuration of components, then build new assembly lines, and finally establish a reliable process for volume production. Along with each generation, a cohort of firms came and went, many of which delayed the adoption of a newer technology. Pooling the observations across five generations (4 transitions), Figure 3.2 plots the timing of innovation separately for incumbents (i.e., firms already active in the previous generation) and entrants (i.e., firms that appeared for the first time as the producer of new-generation HDDs). Only about a half of all incumbents ever innovated into a new generation. Even among those that did, their timing was approximately two years later than entrants. Those that never adapted gradually disappeared along with the shrinking demand for the old products. Changes in technology and market structure are pervasive in many industries, but the HDD market has witnessed one of the fastest, most unrelenting, and most easily measurable turnovers of products and firms.

Second, a detailed industry data book series, the DISK/TREND Reports (1977–99), is available for this industry. From the original reports, I construct a comprehensive panel of the world’s HDD manufacturers by digitizing each firm-year observation. The sample period is long enough to capture five generations of technologies, two of which I analyze in detail.
Third, a high-tech manufacturing sector with rapid growth and innovation is precisely the type of industry that is most relevant to the discussion of pro-innovation public policies. Moreover, the HDD market’s fairly competitive structure (a total of 178 unique firms over 23 years) and geographical outreach (firms from the Americas, Asia, and Europe compete in a global market) underline the potential generalizability of the findings.

Figure 3.3 depicts the position of HDD manufacturing within the broader context of the computer industry. The market structure of the HDD industry was less concentrated than in other sub-sectors such as memory (which Samsung Electronics dominated), microprocessors (Intel and AMD), display (Samsung), operating systems (Microsoft), or various internet services (e.g., Google in search, or Facebook and Twitter in social media). Monopolies and duopolies are interesting subjects, but the performances of these “outlier” firms were so remarkable that a researcher would need to focus on their idiosyncrasies in great detail. By contrast, the HDD industry saw such a massive wave of new entries that the number of firms (pooled across all generations) reached over 100 in the late 1980s. Many of them later exited in a “shakeout” phase, which is a typical development in many sectors, including manufacturing and services, as documented by Carroll and Hannan (2000). Still, the top maker’s market share in 1998, the final year of my dataset, was below 20%.
Another attractive feature of the HDD industry is that these firms originate from all over the world, not just Silicon Valley. The dataset allows comparison of firms from different regions. Moreover, perhaps because HDDs are not as directly exposed to consumers as are cars, computers, or other household electronics products, the sector largely avoided government interventions. Except in Brazil and France, national governments did not intervene as a matter of trade or industrial policies. Thus the dataset is reasonably free from political complications.

My analysis of the HDD industry takes the developments in the global PC market (i.e., HDD’s “downstream” industry) as given for the following reasons. First, the growth in PC demand was primarily driven by hobbyists during the 1980s and then by office automation and growing popularity among households during the 1990s. Second, the price and performance of the central processing unit (CPU) and operating system (OS) determined most of the cost and perceived quality of PCs, and hence the overall demand for PCs and their replacement purchase cycles. Although the quality improvement of HDDs contributed to the enhanced performance of PCs in terms of storage capacity, Intel and Microsoft (“Wintel”) were perceived to be the leaders of the PC industry. Third, the market structure of PC makers is rather competitive, with more than 100 firms across the globe. As Table 3.1
shows, even the combined market share of the top five makers was less than 50%. Moreover, the vendors (brands) and manufacturers of PCs were often different; that is, many less well-known manufacturers made products for famous brands such as Compaq and Hewlett-Packard. Hence the market structure of actual manufacturers is less concentrated than what vendors’ market share suggests.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sinclair 23.9</td>
<td>IBM 12.3</td>
<td>IBM 11.4</td>
<td>Compaq 10.0</td>
</tr>
<tr>
<td>2</td>
<td>Apple 13.6</td>
<td>Commodore 11.4</td>
<td>Apple 9.1</td>
<td>IBM 8.6</td>
</tr>
<tr>
<td>3</td>
<td>Commodore 13.0</td>
<td>Apple 7.8</td>
<td>Commodore 8.3</td>
<td>P-B NEC 6.0</td>
</tr>
<tr>
<td>4</td>
<td>Tandy 12.8</td>
<td>Amstrad 5.9</td>
<td>NEC 5.8</td>
<td>Apple 5.9</td>
</tr>
<tr>
<td>5</td>
<td>Atari 4.5</td>
<td>NEC 5.3</td>
<td>Compaq 4.0</td>
<td>HP N/A</td>
</tr>
<tr>
<td>Others</td>
<td>32.3</td>
<td>57.2</td>
<td>61.5</td>
<td>69.5</td>
</tr>
</tbody>
</table>

Note: Market share based on worldwide unit shipments.
Source: Gartner Dataquest, Wikipedia.

Likewise, I do not explicitly model the developments in the “upstream” industry: HDD components, such as read-write heads, platters, and motors. Some HDD manufacturers make these components in-house, whereas others procure them from electronics parts makers. According to James Porter, the editor of DISK/TREND Reports, there is no clear advantage/disadvantage concerning whether to make or to buy.

### 3.2.2 Data Source

I manually construct the panel data of 1,378 firm-year observations from DISK/TREND Reports (1977–99), an annual publication series edited by the HDD experts in Silicon Valley. I digitize each firm-year observation, which is accompanied by half a page of qualitative descriptions (on the characteristics of the firm, managers, funding, products, production locations, as well as major actions taken in that year with their reasons) in the original publication. Not all information is amenable to quantitative analysis, but some of the firms’ characteristics are suitable for regressions. For example, firms’ age and size (in terms of revenues and profits, either company-wide or specifically for the HDD business) are readily
codifiable. Firms’ organizational forms, regions of origin, and the initial HDD generations in which they started manufacturing, are also digitized as categorical variables.

As a preliminary, descriptive analysis, I regress the timing of innovation (i.e., the shipment of new-generation HDDs) on these firm characteristics. Table 3.2 reports the estimation results based on a standard duration model (Cox proportional hazard estimates). The estimates suggest incumbents are 50% less likely than entrants to innovate in a given year, even after controlling for all of the observed characteristics.

I employ two different definitions of entrants in these regressions. The first definition is data-driven and narrower: a potential entrant is recognized when a new firm announces the product specifications without actually manufacturing or shipping them. The second definition is more conceptually motivated and broader: for all of the potential entrants that announced or shipped HDDs, I count as “time at risk (of innovation)” all years since the industry-wide establishment of new-generation standards. That is, regardless of whether these new firms were already incorporated or not, I interpret that their founders were considering the innovation and entry, from the moment the industry consensus emerged on the physical size of new-generation HDDs.

An auxiliary dataset, also from DISK/TREND Reports, containing the prices and shipment quantities of HDDs, accompanies this panel data of firms. For each year, the reports record the average transaction price and total quantity for each of the generation-quality categories (5 generations and 14 quality levels in total).

Researchers in both economics and management repeatedly confirm the accuracy, relev-

---

8 However, age is not necessarily comparable across firms that had roots in different industries (e.g., manufacturers of card punchers, typewriters, automobile components, or coin laundries). Not all firms disclosed division-level revenue/profit information. For these reasons, I omit these variables in the following regressions.

9 Econometrically, regressions based on the first definition would over-estimate the entrants’ propensity to innovate because this definition recognizes entrants only when they were most serious about innovation. In contrast, regressions based on the second definition would probably under-estimate entrants’ propensity to innovate because all years until actual shipments are interpreted as evidence against their innovativeness, regardless of whether they existed as firms or not. Table 3.2 reports only the results based on the second definition, in order to show that entrants were twice more likely to innovate than incumbents, even when I use the definition that would bias against such findings.
Table 3.2: Preliminary Regressions of Innovation Timing on Firm Characteristics

<table>
<thead>
<tr>
<th>Duration model (Cox proportional hazard estimates)</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent variable: Decision to Innovate</td>
<td>.41***</td>
<td>.53***</td>
<td>.50***</td>
</tr>
<tr>
<td>Incumbent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial generation of entry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.25-inch</td>
<td></td>
<td>.84</td>
<td>.71</td>
</tr>
<tr>
<td>3.5-inch</td>
<td></td>
<td>.52***</td>
<td>.46***</td>
</tr>
<tr>
<td>2.5-inch</td>
<td></td>
<td>.34***</td>
<td>.91</td>
</tr>
<tr>
<td>Organizational Form</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specialized HDD start-up</td>
<td></td>
<td>1.04</td>
<td>1.01</td>
</tr>
<tr>
<td>Computer maker</td>
<td></td>
<td>1.39</td>
<td>1.32</td>
</tr>
<tr>
<td>HDD component maker</td>
<td></td>
<td>.55</td>
<td>.54*</td>
</tr>
<tr>
<td>Region of Origin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U.S.</td>
<td>1.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asia</td>
<td>1.91**</td>
<td>1.90**</td>
<td></td>
</tr>
<tr>
<td>Europe (West)</td>
<td>1.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Europe (East)</td>
<td>.14*</td>
<td>.16*</td>
<td></td>
</tr>
<tr>
<td>Industry state</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of firms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of firms squared</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of subjects (firms)</td>
<td>437</td>
<td>437</td>
<td>437</td>
</tr>
<tr>
<td>Number of innovations</td>
<td>190</td>
<td>190</td>
<td>190</td>
</tr>
<tr>
<td>Time at risk of innovation</td>
<td>2,591</td>
<td>2,591</td>
<td>2,591</td>
</tr>
<tr>
<td>Number of observations</td>
<td>1,842</td>
<td>1,842</td>
<td>1,842</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>−1,018</td>
<td>−997</td>
<td>−990</td>
</tr>
</tbody>
</table>

Note: Coefficients greater (less) than 1 indicate higher (lower) propensities to innovate. Omitted categories are “Potential entrant,” “8-inch,” “Other electronics maker (horizontal diversification),” and “Brazil.” ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. Standard errors in parentheses.

3.2.3 Focus: Transition from 5.25- to 3.5-inch Generations

I analyze the technological transition from the 5.25- to 3.5-inch generations, which I will call the “old” and “new” generations henceforth. This subsample of the dataset spans 18 years (1981–98) and 259 firm-years. I concentrate on these generations because they competed directly with each other in the desktop PC market. Although transitions between the other generations showed similar developments, 14-, 8-, and 2.5-inch HDDs were used in different segments of the computer industry, that is, 14-inch for mainframe computers, Asia.

8-inch for minicomputers, and 2.5-inch for notebook PCs. By focusing on 5.25- and 3.5-inch generations, I avoid confounding factors that might originate from diverging trends in different segments downstream. These two generations were also historically the most important of all generations in terms of volume and revenue.

Figure 3.4 shows the numbers of firms in four different states: (1) “old-only,” (2) “both,” (3) “new-only,” and (4) “potential entrant.” Incumbents start as (1) and become (2) upon the adoption of new technology. Entrants start as (4) and become (3) upon adoption (entry).

![Figure 3.4: Evolution of the Industry Composition](image)

*Note:* “Old-only” and “new-only” firms produce 5.25- and 3.5-inch HDDs, respectively. “Both” represents incumbents that adopted the new technology, hence producing both of the two generations. “Potential entrant” is identified by the announcement of product specifications (without actual shipment).

The two generations of HDD experienced a fast growth in volume and a steady decline in price (Figure 3.5, top). Over the years, the average quality (information storage capacity) of HDDs improved at an approximately constant rate (Figure 3.5, bottom). These developments were typical of those in many computer-related industries.
Note: Both 5.25- and 3.5-inch HDDs serve the same market, namely, desktop personal computers. Quality is measured by average capacity per unit for each generation.
3.3 Model

This section presents my industry equilibrium model. The first subsection outlines the dynamic game of technology adoption and entry/exit. The second subsection explains the demand side. The third subsection explains the spot-market competition. The fourth subsection shows how I solve the dynamic part of the model by backward induction.

3.3.1 Setup: Dynamic Discrete-Choice Game

Time is discrete with finite horizon \( t = 0, 1, 2, \ldots, T \). Two goods, old and new, are imperfect substitutes from the buyers’ viewpoint. Each of these goods requires a specific generation of technology for production, old and new.

Two Types, Four States

There are two types of firms (“incumbents” and “entrants”) and four individual states (“old-only,” “both,” “new-only,” and “potential entrant”), as illustrated in Figure 3.6.

Figure 3.6: Illustration of the Dynamic Discrete Choice

<table>
<thead>
<tr>
<th>t</th>
<th>t+1</th>
<th>t+2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incumbent(s):</td>
<td>Adopt</td>
<td>Stay</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entrant(s):</td>
<td>Adopt</td>
<td>Stay</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Note: “Old-only” and “New-only” firms produce 5.25- and 3.5-inch HDDs exclusively. “Both” represents incumbents that adopted the new technology, hence producing both of the two generations. “Potential entrant” is recognized (in data) by the announcement of product specifications without actual shipment.

“Incumbents” produce old goods from time 0, using the old-generation technology. In any time period, an incumbent may choose to adopt the new-generation technology by paying a
sunk cost and starting to produce both old and new goods from the next period. Hence an incumbent starts in the “old-only” state (a production technology status in which a firm can produce only old goods) and may elect to transition to the “both” state (a status in which a firm can produce both old and new goods).

“Entrants” are the other type of firms. They are not active in the market at time 0. Each of them appears in a predetermined year (observed in data), at which moment they may choose to adopt new technology and enter the market, or quit the prospect of entry. That is, by paying a sunk cost, a potential entrant becomes an actual entrant in the subsequent period in the “new-only” state, a production technology status in which a firm can produce only new goods.\footnote{I do not consider entrants’ adoption of the old technology because it rarely happens in the data once the new technology becomes available. For the same reason, I rule out the alternative to “wait.”}

Hence a firm belongs to any one of the four states, $s_{it} \in \{\text{old, both, new, pe}\}$. The industry state summarizes all firms’ states, $s_t = (N_t^{\text{old}}, N_t^{\text{both}}, N_t^{\text{new}}, N_t^{\text{pe}})$, as the numbers of firms in each of four states. Let $s_{-i,t}$ denote the numbers of competitors for firm $i$, which is simply $s_t$ minus 1 in firm $i$’s own category.

**Period Profit**

Firm $i$’s single-period profit, $\pi(s_{it}, s_{-i,t}, W_t)$, depends on its own state $s_{it}$, its competitors’ state $s_{-i,t}$, and the characteristics of demand and cost $W_t$. $s_{it}$ and $s_{-i,t}$ change endogenously as a result of each firm’s dynamic decision, whereas $W_t$ evolves exogenously. Old and new goods are imperfect substitutes; that is, products are differentiated across generations. The logit demand system characterizes their substitution pattern (section 3.2). Products are homogeneous within each generation, with the spot-market competition characterized by a symmetric Cournot game (section 3.3).

**Dynamic Discrete-choice Problem**

Each firm aims to maximize its expected present value. The interest rate is assumed to be positive and constant over time, resulting in a constant discount factor $\beta \in (0, 1)$ per period. In each period, events occur in the following order:
1. Each active firm observes its private cost shocks $\varepsilon^0_{it}$, $\varepsilon^1_{it}$, and $\varepsilon^2_{it}$, associated with exiting, staying, and innovating (if a firm’s state is old), respectively. A potential entrant draws $\varepsilon^0_{it}$ (for quitting) and $\varepsilon^2_{it}$ (for entry/innovation) but not $\varepsilon^1_{it}$ because it does not sell anything and hence cannot wait on the sidelines.

2. Firms make dynamic decisions, namely exit, stay, or innovate.

3. Active firms compete in the spot market and earn profits, $\pi(s_{it}, s_{-it}, W_t)$.

4. Dynamic decisions are implemented. Specifically, exiting firms exit and receive their sell-off values $\phi + \varepsilon^0_{it}$. Staying firms receive $\varepsilon^1_{it}$. Adopting incumbents pay $\delta \kappa^{inc} + \varepsilon^2_{it}$. Potential entrants receive $\varepsilon^0_{it}$ (if they quit) or pay $\delta \kappa^{ent} - \varepsilon^2_{it}$ (if they adopt/enter) to become active.

5. The industry takes on a new state, $s_{t+1} = (N^{old}_{t+1}, N^{both}_{t+1}, N^{new}_{t+1}, N^{pe}_{t+1})$.

Hence the decision problems for active firms in each of the three states are\(^\text{12}\)

\[
V^{old}_{t}(s_t) = \pi^{old}_t(s_t) + \max \left\{ \phi + \varepsilon^0_{it}, \beta E \left[ V^{old}_{t+1}(s_{t+1}) \right] + \varepsilon^1_{it}, \beta E \left[ V^{both}_{t+1}(s_{t+1}) \right] - \delta \kappa^{inc} + \varepsilon^2_{it} \right\},
\]

\[
V^{both}_{t}(s_t) = \pi^{both}_t(s_t) + \max \left\{ \phi + \varepsilon^0_{it}, \beta E \left[ V^{both}_{t+1}(s_{t+1}) \right] + \varepsilon^1_{it} \right\}, \text{ and}
\]

\[
V^{new}_{t}(s_t) = \pi^{new}_t(s_t) + \max \left\{ \phi + \varepsilon^0_{it}, \beta E \left[ V^{new}_{t+1}(s_{t+1}) \right] + \varepsilon^1_{it} \right\},
\]

subject to the perceived law of motion governing the industry state, $s_t$ (see below). For a potential entrant, the problem is simply

\[
\max \{ \varepsilon^0_{it}, \beta E \left[ V^{new}_{t+1}(s_{t+1}) \right] - \delta \kappa^{ent} + \varepsilon^2_{it} \}.
\]

\(^{12}\)For notational simplicity, I suppress $\varepsilon^0_{it}$, $\varepsilon^1_{it}$, and $\varepsilon^2_{it}$ from $V_i(s_t)$, where they should also be included as payoff-relevant individual state variables.
Non-stationary Environment and Solution Concept

I assume the HDD industry reaches its terminal state in 1998, when the state stops evolving. I solve the model backward over 18 years in the spirit of Subgame-perfect equilibrium (SPE). In reality, the industry keeps evolving after the sample period, but the point is that the 5.25-inch HDDs all but disappeared by 1998. Since my purpose is to analyze the economic incentives surrounding the transition from the 5.25- to 3.5-inch HDDs and firms’ turnover and the environment is non-stationary, I believe the finite-horizon setup is a reasonable representation of history during the sample period.\textsuperscript{13}

Strictly speaking, I am using not exactly SPE but Perfect Bayesian Equilibrium (PBE) as a solution concept. The game involves private information, in the form of private cost shocks associated with dynamic actions (entry/exit and innovation). A firm never observes the realizations of these shocks for its rivals, which precludes the existence of subgames. However, the past realizations of private cost shocks do not matter \textit{per se} because only the current market structure (along with time and a firm’s own draws of current private shocks) affects the firm’s payoff. My model inherits this convenient property from the Markov-perfect equilibrium (MPE) models of industry dynamics, although I refrain from assuming a stationary environment as in the latter models. Hence what we assume about off-path beliefs is irrelevant.\textsuperscript{14}

Multiple Equilibria

The multiplicity of MPE is the common cause of concern in the studies of dynamic oligopoly, which motivated the development of the two-step estimation methodologies that bypass the issue. However, the two-step approach is too data-intensive for my industry/data

\textsuperscript{13}Recent applications of dynamic game and its estimation techniques often use MPE, which is adequate for a stationary environment. In contrast, I am studying a topic and industry whose chief characteristics are non-stationary. That is, the market size grows explosively, the production costs drop steadily over time, and the demand for the old-generation products eventually disappears as the new technology/product becomes mainstream. The research question is how much each of the three theoretical forces determines whether and when incumbents decide to adopt new technology. Given the motivation and dataset, I choose to model the industry’s non-stationary evolution as is rather than try to stationarize the environment and apply techniques suitable for MPE.

\textsuperscript{14}For this reason, I may also proceed with Sequential Equilibrium instead of PBE.
setting (i.e., a single global market). My alternative approach is three-fold: (1) compartmentalization, (2) computational randomization, and (3) analytical randomization.

First, my dynamic game is rather simple, with small state space and choice sets, symmetry within each firm type, and iid private cost shocks. This setup alone can provide the uniqueness of equilibrium under certain market structure (i.e., only few types of firms) and payoff profiles (i.e., clear advantage/disadvantage of particular types), in the spirit of Seim’s (2006) one-shot entry game with private cost shocks. The non-stationary, finite-horizon environment allows for a solution by backward induction, which proceeds with period-by-period and state-by-state solutions of such (one-shot) simultaneous-move games of entry/exit and innovation. In this sense, I “compartmentalize” the multi-period game and its multiple equilibria issue into more manageable games.

Second, several equilibria are possible in the remaining subset of market configurations. In such cases, my computational algorithm (which starts from an arbitrary initial strategy profile) picks up an arbitrary equilibrium. This kind of equilibrium selection is akin to “tilting the playing field” one way or the other, favoring particular types of firms, and would in principle bias the estimation results. However, since my application involves several thousand industry states (per period) over 18 years, this numerical “randomization” over (at most several) multiple equilibria results in a rather “level playing field” on average, from the perspective of the entire game.\(^\text{15}\)

Third, I am currently investigating a more analytical approach to randomization, at the model level. Specifically, I consider replacing the simultaneous-move assumption in the current model with a sequential-move assumption akin to Stackelberg games. Since I maintain the assumption of symmetry within each of the four firm types, this alternative structure would achieve “more uniqueness.” Furthermore, to avoid “tilting the playing field” due to particular sequencing, I aim to incorporate random ordering of sequential moves. My preliminary attempts indicate the results are quantitatively similar to the current ones (based on numerical randomization).

\(^{15}\)As a sensitivity analysis, I experimented with various starting points (i.e., different assortment of randomly picked equilibria across states and years), which barely changed my estimates (in Section 5.3).
Beliefs (Perceived Law of Motion)

For rules governing firms’ expectations, alternatives include rational expectations and perfect foresight. Regarding firms’ beliefs about rivals’ moves, \( s_{-i,t+1} \), I assume rational expectations. That is, a firm correctly perceives how its rivals make dynamic decisions up to private cost shocks, \((\varepsilon_{it}^0, \varepsilon_{it}^1, \varepsilon_{it}^2)\) iid extreme value. This setup allows for dynamic strategic interactions, which are a prerequisite for incorporating cannibalization and preemptive motives into the model. I “knock out” this feature in one of the counterfactuals (no-preemption case, in section 6.4), where firms instead perceive the industry state as evolving exogenously, in the spirit of non-stationary Oblivious Equilibrium (Weintraub, Benkard, Jeziorski, and Van Roy 2008).

With respect to the evolution of demand and production costs, I assume firms’ perfect foresight. From the theoretical perspective, this choice reflects my analytical focus on strategic interactions and adoption costs rather than informational factors related to demand uncertainty. I do not intend to disregard the role of information in studying investment under uncertainty. Rather, the issue is the empirical tractability of informational factors given the current dataset. In my view, this assumption is simplistic but not distortionary because firms’ beliefs are homogeneous regardless of their types or individual state in any given period. Hence it is unlikely to affect the incumbent-entrant asymmetry this paper tries to explain.

Model Primitives

There are dynamic and static components of model primitives. Dynamic primitives are the discount factor \( \beta \), the mean sell-off value \( \phi \), the base sunk costs of technology adoption \( \kappa^{inc} \) and \( \kappa^{ent} \), the annual rate of sunk cost change \( \delta \), a dynamic equilibrium concept, and the informational assumptions made on firms’ perceived law of motion for the industry state.

---

16Adaptive expectations might be another interesting modeling choice.

17Although they are beyond the scope of this paper, information-related topics present fascinating directions for future research, including (1) beliefs about new HDDs’ profitability that are potentially heterogeneous across firms, and (2) the updating of these beliefs through own experimentation and learning from rivals.
Static primitives determine the period profit function: demand parameters, cost parameters, and a static equilibrium concept. The next two subsections explain the details of these static model components.

### 3.3.2 Demand

I capture the substitution pattern across generations of HDDs using the logit model of differentiated products. The dynamic discrete-choice model (section 3.1) captures HDDs’ differentiation across generations and assumes homogeneity within each generation.

A buyer \( k \) purchasing an HDD of generation \( g \) enjoys utility,

\[
  u_{kg} = \alpha_0 + \alpha_1 p_g + \alpha_2 I(g = \text{new}) + \alpha_3 x_g + \xi_g + \epsilon_{kg},
\]

where \( p_g \) is the price, \( I(g = \text{new}) \) is the indicator of new generation, \( \xi_g \) is the unobserved characteristics (most importantly, “design popularity” among buyers, as well as other unobserved attributes such as “reliability”), and \( \epsilon_{kg} \) is the idiosyncratic taste shock over generations. The outside goods offer the normalized utility \( u_{k0} \equiv 0 \), which represent removable HDDs (as opposed to fixed HDDs) and other storage devices.

Let \( \bar{u}_g \equiv \alpha_0 + \alpha_1 p_g + \alpha_2 I(g = \text{new}) + \alpha_3 x_g + \xi_g \) represent the mean utility from a generation-\( g \) HDD whose market share is

\[
  m_s_g = \frac{\exp (\bar{u}_g)}{\sum_t \exp (\bar{u}_t)}.
\]

The shipment quantity of generation-\( g \) HDDs is

\[
  Q_g = m_s_g M,
\]

\footnote{I suppress the time subscript \( t \) for the simplicity of notation. The demand side is static in the sense that buyers make new purchasing decisions every year. This assumption is not restrictive because multi-year contracting is not common and there are hundreds of buyers (computer makers) during the sample period.}

\footnote{Tape recorders, optical disk drives, and flash memory.}
where \( M \) is the size of the HDD market including the outside goods (removable HDDs and other storage devices). Practically, \( M \) reflects all desktop PCs to be manufactured globally in a given year.

Berry’s (1994) inversion provides the linear relationship,

\[
\ln \left( \frac{m_{sg}}{m_{s0}} \right) = \alpha_1 p_g + \alpha_2 I (g = new) + \alpha_3 x_g + \xi_g, \tag{3.4}
\]

where \( s_g \) represents the market share of HDDs of generation \( g \), and \( m_{s0} \) is the market share of outside goods (removable HDDs and other devices). The inverse demand is

\[
p_g = \frac{1}{-\alpha_1} \left[ -\ln \left( \frac{m_{sg}}{m_{s0}} \right) + \alpha_2 I (g = new) + \alpha_3 x_g + \xi_g \right]. \tag{3.5}
\]

### 3.3.3 Period Competition

The spot-market competition is characterized by multi-product (i.e., old and new goods) Cournot competition. Marginal costs of producing old and new goods, \( mc^{old} \) and \( mc^{new} \), are assumed to be common across firms and constant with respect to quantity. Firm \( i \) maximizes profits

\[
\pi_i = \sum_{g \in A_i} \pi_{ig} = \sum_{g \in A_i} (p_g - mc_g) q_{ig}
\]

with respect to shipping quantity \( q_{ig} \) \( \forall g \in A_i \), where \( \pi_{ig} \) is the profit of firm \( i \) in generation \( g \), and \( A_i \) is the set of generations produced by firm \( i \). Firm \( i \)'s first-order condition with respect to its output \( q_{ig} \) is

\[
p_g + \frac{\partial p_g}{\partial Q_g} q_{ig} + \frac{\partial p_h}{\partial Q_g} q_{ih} = mc_g, \tag{3.6}
\]

with \( g, h \in \{old, new\}, g \neq h \), if firm \( i \) produces both old and new HDDs. The third term on the left-hand side is dropped if a firm makes only one generation.
3.3.4 Solution of Dynamic Game by Backward Induction

I assume the state stops evolving after year $T$. Hence the terminal values associated with a firm’s states, $s_{iT} \in \{old, both, new\}$, are

$$
(V_{T}^{old}, V_{T}^{both}, V_{T}^{new}) = \left( \sum_{\tau=T}^{\infty} \beta^{\tau} \pi_{T}^{old} (s_{T}) , \sum_{\tau=T}^{\infty} \beta^{\tau} \pi_{T}^{both} (s_{T}) , \sum_{\tau=T}^{\infty} \beta^{\tau} \pi_{T}^{new} (s_{T}) \right),^{20}
$$

In year $T - 1$, an old-only firm’s problem (aside from maximizing its period profit) is

$$
\max \left\{ \phi + \varepsilon_{i,T-1}^{0}, \beta E \left[ V_{T}^{old} (s_{T}) | s_{T-1} \right] + \varepsilon_{i,T-1}^{1}, \beta E \left[ V_{T}^{both} (s_{T}) | s_{T-1} \right] - \delta^{T-1} \kappa_{inc} + \varepsilon_{i,T-1}^{2} \right\}.
$$

I follow Rust (1987) to exploit the property of the logit errors, $\varepsilon_{it} = (\varepsilon_{it}^{0}, \varepsilon_{it}^{1}, \varepsilon_{it}^{2})$, and their (conditional) independence over time, to obtain a closed-form expression for the expected value before observing $\varepsilon_{it}$,

$$
E_{\varepsilon_{i,T-1}} \left[ V_{T-1}^{old} (s_{T-1}, \varepsilon_{i,T-1}) | s_{T-1} \right] = \pi_{T-1}^{old} (s_{T-1}) + \gamma \\
+ \ln \left[ \exp (\phi) + \exp \left( \beta E \left[ V_{T}^{old} (s_{T}) | s_{T-1} \right] \right) \\
+ \exp \left( \beta E \left[ V_{T}^{both} (s_{T}) | s_{T-1} \right] - \delta^{T-1} \kappa_{inc} \right) \right],
$$

where $\gamma$ is the Euler constant (0.5772). Similar expressions hold for the other two types:

$$
E_{\varepsilon_{i,T-1}} \left[ V_{T-1}^{both} (s_{T-1}, \varepsilon_{i,T-1}) | s_{T-1} \right] = \pi_{T-1}^{both} (s_{T-1}) + \gamma \\
+ \ln \left[ \exp (\phi) + \exp \left( \beta E \left[ V_{T}^{both} (s_{T}) | s_{T-1} \right] \right) \right],
$$

$^{20}$Alternatively, I might consider anchoring the terminal values to some auxiliary data that would cover the periods after 1998, the final year of my data set.
In this manner, I can write the expected value functions from year $T$ all the way back to year 0. The associated choice probabilities (policy functions) will provide a basis for the maximum likelihood estimation (in section 4.3).

### 3.4 Estimation

My empirical approach takes three steps. First, I estimate the system of demand for differentiated products. Second, I recover the marginal costs of production implied by the demand estimates and the first-order conditions of the firms’ period-profit maximization. These two steps generate the measure of period profit in each year and state, which forms the basis for estimating the dynamic parameters. Third, I estimate the dynamic parameters (the sunk costs of technology adoption/entry and the sell-off value upon exit) using the solution to my dynamic model.

#### 3.4.1 Estimation: Demand

The empirical demand analysis proceeds at the level that is more detailed than generation $g$. The purpose is to facilitate identification and control of HDD product attributes other than $g$, such as observed quality $x$ and unobserved “popularity” $\xi$.

In data, the unit of observation is the combination of generation, quality, buyer category (PC makers and distributors/end-users), geographical regions (U.S. and non-U.S.), and year $t$. For notational simplicity, I denote the generation-quality pair by “product category” $j$ and suppress subscripts for the latter three dimensions. I estimate the simple logit model as well as its nested version (with nests on generations) so as to empirically confirm the
economic significance of substitution across generations.

These empirical considerations lead to the following recasting of the demand model. A buyer \( k \) purchasing an HDD of product category \( j \), that is, a combination of generation \( g \) (diameter) and quality \( x \) (storage capacity in megabytes), enjoys utility

\[
u_{kj} = \alpha_0 + \alpha_1 p_j + \alpha_2 g_j + \alpha_3 x_j + \xi_j + \sigma \zeta_{kgj} + (1 - \sigma) \epsilon_{kj}, \tag{3.8}\]

where \( p_j \) is the price, \( \xi_j \) is the unobserved characteristics, \( \zeta_{kgj} \) is the idiosyncratic taste shock over generations of HDDs, and \( \epsilon_{kj} \) is the idiosyncratic taste shock over generation-quality bins.

The coefficients, \( \alpha_0 \) through \( \alpha_3 \) and \( \sigma \), are the taste parameters to be estimated. The nest parameter \( \sigma \in [0, 1] \) measures the importance of substitution \textit{within} generation \( g \) relative to that \textit{across} generations. For example, \( \sigma = 1 \) indicates substitution is only within \( g \), whereas \( \sigma = 0 \) indicates a plain logit model without nests. \( \epsilon_{kj} \) is iid extreme value (over buyers and bins); that is, its cumulative distribution function is \( F(\epsilon_{kj}) = \exp \left( - \exp \left( - \epsilon_{kj} \right) \right) \). \( \zeta_{kgj} \) is distributed such that the marginal distribution of the composite error term, \( \sigma \zeta_{kgj} + (1 - \sigma) \epsilon_{kj} \), is also iid extreme value.

I estimate the logit and nested logit models of demand using OLS and instrumental variables (IVs). Berry’s (1994) inversion allows the econometrician to run a linear regression,

\[
\ln \left( \frac{m_{sj}}{m_{s0}} \right) = \alpha_1 p_j + \alpha_2 g_j + \alpha_3 x_j + \sigma \ln \left( m_{sj|gj} \right) + \xi_j, \tag{3.9}\]

for the estimation of the nested logit model, where \( m_{sj} \) represents the market share of HDDs of category \( j \), \( m_{s0} \) is the market share of outside goods (removable HDDs), and \( m_{sj|gj} \) is the market share of category-\( j \) HDDs \textit{within} its generation \( g_j \).

**Identification**

The demand parameters are identified by the time-series and cross-sectional variations in data (subscripts omitted for notational simplicity) as well as the (nested) logit functional form. The sample period is the 18 years between 1981 and 1998. There are three sources of
cross-sectional variation. First, an HDD’s product category (denoted by $j$) is a pair of generation (diameter, or form-factor) and quality (information storage capacity in megabytes). There are two generations and 14 discrete quality levels according to the industry convention. Second, data are recorded by buyer category, PC makers and distributors/end-users. Third, data are recorded by geographical category, U.S., and non-U.S.

The OLS estimation relies on the assumption that $E[\xi_j|p_j, g_j, x_j] = 0$; that is, the price of a category-$j$ HDD is uncorrelated with that particular category’s unobserved attractiveness to the buyers. However, one might suspect a positive correlation between them because an attractive product category would command both higher willingness to pay and higher cost of production.

In the IV estimation, I use the following variables as instruments for $p_j$ (and also $ms_{j|g_j}$ in nested logit): (1) the prices in the other region and user category, (2) the number of product models and firms, and (3) the number of years since first introduction. The first IV is used by Hausman (1996) and then by Nevo (2001). The identifying assumption is that production cost shocks are correlated across markets, whereas taste shocks are not. This assumption is consistent with the industry context that HDD makers from across the globe compete in both the United States and elsewhere, whereas end users of HDDs (and hence of PCs) are more isolated geographically.

The second IV was used by Bresnahan (1981) and Berry et al. (1995) and exploits the proximity of rival products (in product space), that is, the negative correlation between the number of models/firms, markup, and price in oligopolies. The identifying assumption is that taste shocks in any given period are not correlated with the number of models/firms in a particular product category $j$. Firms need to make product-introduction decisions in prior years, without observing taste shocks in particular regions/user types in the following years. More importantly, such dynamic decisions are driven by the sum of discounted present values of future profits, which is affected only negligibly by taste shocks in any single period and for particular regions/user types. Hence this identifying assumption is plausible as long as particular regions’/user types’ taste shocks are not extremely serially correlated.
The third IV relies on steady declines in the marginal costs of production over years. In the HDD industry, costs dropped because of design improvements, reduced costs of key components, and offshore production in Singapore, Malaysia, Thailand, and the Philippines. This overall tendency holds at the product category level as well. The identifying assumption is that taste shocks are not correlated with such time patterns on the cost side, which may or may not be the case.

Thus the first IV is the most preferable and so is used in most IV estimations, whereas the second and third IVs are used in the case where the first IV is not available, namely, the robustness check with an alternative market definition.

3.4.2 Estimation: Marginal Costs

For each year, we can infer the marginal costs of production, \( mc_{old} \) and \( mc_{new} \), from equation (3.6), namely, the first-order conditions for the firms’ static profit maximization problems. Because the unit observation in the HDD sales data is product category level—and not firm or brand level—I maintain, as identifying assumptions, symmetry across firms (up to individual state) and constant marginal cost with respect to quantity.

3.4.3 Estimation: Sunk Costs of Innovation

I set the discount factor \( \beta \) at values between .82 and .94.\(^2\) I do not intend to estimate it because its identification is known to be impractical in most cases (c.f., Rust 1987). Likewise, the rate of drop in sunk costs, \( \delta \), is difficult to estimate directly from the following procedure, so instead I will assume \( \delta \) equals the average rate of decline in \( mc_{new} \) over time.

The contribution of an old firm \( i \) in year \( t \) to the likelihood is

\[
f^{old}(d_{it}|s_t; \phi, \kappa^{inc}, \delta) = \begin{cases} 
pr^{old}(d_{it} = \text{exit}) I(d_{it} = \text{exit}) & \text{if } d_{it} = \text{exit} \\
pr^{old}(d_{it} = \text{stay}) I(d_{it} = \text{stay}) & \text{if } d_{it} = \text{stay} \\
pr^{old}(d_{it} = \text{adopt}) I(d_{it} = \text{adopt}) & \text{if } d_{it} = \text{adopt} 
\end{cases}
\]

---

\(^2\)Values of \( \beta \) outside this range result in either computational errors or unintuitive parameter estimates (e.g., negative \( \phi \)).
where \( pr^{\text{old}}(\cdot) \) is the probability that an old-only firm takes a particular action \( d_{it} \):

\[
pr^{\text{old}}(d_{it} = \text{exit}) = \frac{\exp(\phi)}{Q},
\]

\[
pr^{\text{old}}(d_{it} = \text{stay}) = \frac{\exp\left( E_sV^\text{old}_{t+1}(s_{t+1}) \right)}{Q},
\]

\[
pr^{\text{old}}(d_{it} = \text{adopt}) = \frac{\exp\left( E_sV^\text{both}_{t+1}(s_{t+1}) - \delta^t\kappa^{\text{inc}} \right)}{Q}.
\]

where \( Q = \exp(\phi) + \exp\left( E_sV^\text{old}_{t+1}(s_{t+1}) \right) + \exp\left( E_sV^\text{both}_{t+1}(s_{t+1}) - \delta^t\kappa^{\text{inc}} \right) \). Similarly, the contributions of the other three types of firms are

\[
f^{\text{both}}(d_{it}|s_t; \phi) = pr^{\text{both}}(d_{it} = \text{exit})^{I(d_{it}=\text{exit})} pr^{\text{both}}(d_{it} = \text{stay})^{I(d_{it}=\text{stay})},
\]

\[
f^{\text{new}}(d_{it}|s_t; \phi) = pr^{\text{new}}(d_{it} = \text{exit})^{I(d_{it}=\text{exit})} pr^{\text{new}}(d_{it} = \text{stay})^{I(d_{it}=\text{stay})},
\]

\[
f^{\text{pe}}(d_{it}|s_t; \kappa^{\text{ent}}, \delta) = pr^{\text{pe}}(d_{it} = \text{quit})^{I(d_{it}=\text{quit})} pr^{\text{pe}}(d_{it} = \text{adopt})^{I(d_{it}=\text{adopt})},
\]

where

\[
pr^{\text{both}}(d_{it} = \text{exit}) = \frac{\exp(\phi)}{\exp(\phi) + \exp\left( E_sV^\text{both}_{t+1}(s_{t+1}) \right)},
\]

\[
pr^{\text{both}}(d_{it} = \text{stay}) = \frac{\exp\left( E_sV^\text{both}_{t+1}(s_{t+1}) \right)}{\exp(\phi) + \exp\left( E_sV^\text{both}_{t+1}(s_{t+1}) \right)},
\]

\[
pr^{\text{new}}(d_{it} = \text{exit}) = \frac{\exp(\phi)}{\exp(\phi) + \exp\left( E_sV^\text{new}_{t+1}(s_{t+1}) \right)},
\]

\[
pr^{\text{new}}(d_{it} = \text{stay}) = \frac{\exp\left( E_sV^\text{new}_{t+1}(s_{t+1}) \right)}{\exp(\phi) + \exp\left( E_sV^\text{new}_{t+1}(s_{t+1}) \right)},
\]

\[
pr^{\text{pe}}(d_{it} = \text{quit}) = \frac{\exp(0)}{\exp(0) + \exp\left( E_sV^\text{new}_{t+1}(s_{t+1}) - \delta^t\kappa^{\text{ent}} \right)}, \text{ and}
\]

\[
pr^{\text{pe}}(d_{it} = \text{adopt}) = \frac{\exp\left( E_sV^\text{new}_{t+1}(s_{t+1}) - \delta^t\kappa^{\text{ent}} \right)}{\exp(0) + \exp\left( E_sV^\text{new}_{t+1}(s_{t+1}) - \delta^t\kappa^{\text{ent}} \right)}.
\]

Year \( t \) has \( N_t \equiv (N_t^{\text{old}}, N_t^{\text{both}}, N_t^{\text{new}}, N_t^{\text{pe}}) \) active firms in each state, of which \( X_t \equiv (X_t^{\text{old}}, X_t^{\text{both}}, X_t^{\text{new}}) \) decide to exit. Also, \( E_t \equiv (E_t^{\text{old}}, E_t^{\text{pe}}) \) firms (incumbents and potential entrants) decide to adopt the new technology. The joint likelihood for year \( t \) of observing
data \((N_t, X_t, E_t)\) is

\[
P(X_t, E_t, N_t) = \left(\frac{N_{\text{old}}^t}{X_{\text{old}}^t}\right) \left(\frac{N_{\text{old}}^t - X_{\text{old}}^t}{E_{\text{old}}^t}\right) p_{\text{old}}(d_{it} = \text{exit})^{X_{\text{old}}^t} \times p_{\text{old}}(d_{it} = \text{stay})^{N_{\text{old}}^t - X_{\text{old}}^t - E_{\text{old}}^t} \times p_{\text{old}}(d_{it} = \text{adopt})^{E_{\text{old}}^t}
\]

\[
\times \left(\frac{N_{\text{both}}^t}{X_{\text{both}}^t}\right) p_{\text{both}}(d_{it} = \text{exit})^{X_{\text{both}}^t} \times p_{\text{both}}(d_{it} = \text{stay})^{N_{\text{both}}^t - X_{\text{both}}^t}
\]

\[
\times \left(\frac{N_{\text{new}}^t}{X_{\text{new}}^t}\right) p_{\text{new}}(d_{it} = \text{exit})^{X_{\text{new}}^t} \times p_{\text{new}}(d_{it} = \text{stay})^{N_{\text{new}}^t - X_{\text{new}}^t}
\]

\[
\times \left(\frac{N_{\text{pe}}^t}{E_{\text{pe}}^t}\right) p_{\text{pe}}(d_{it} = \text{adopt})^{E_{\text{pe}}^t} \times p_{\text{pe}}(d_{it} = \text{quit})^{N_{\text{pe}}^t - E_{\text{pe}}^t}
\]

(3.10)

The overall joint likelihood for \(t = 0, 1, 2, \ldots, T - 1\) is

\[
P(X, E, N) = \prod_{t=0}^{T-1} P(X_t, E_t, N_t).
\]

Thus, the maximum likelihood estimators for the mean sell-off value \(\phi\) and the base sunk costs of technology adoption \(\kappa^{\text{inc}}\) and \(\kappa^{\text{ent}}\) are

\[
\arg \max_{\phi, \kappa^{\text{inc}}, \kappa^{\text{ent}}} \ln [P(X, E, N)].
\]

(3.11)

**Identification**

Intuitively, I rely on a revealed-preference argument to identify the sell-off value and the sunk costs. For each firm \(i\) in year \(t\), I compare the benefits and costs of the three dynamic alternatives (exit, stay, and adopt), each of which is associated with the parameters \((\phi, \kappa^{\text{inc}}, \text{and } \kappa^{\text{ent}})\) and the value of being in a particular state in a given year \((E_t V_{\text{old}}^{\text{old}}(s_{t+1}), E_t V_{\text{both}}^{\text{both}}(s_{t+1}), \text{and } E_t V_{\text{new}}^{\text{new}}(s_{t+1}))\). These values of dynamic alternatives are, in turn, based on the model of a dynamic discrete-choice game as well as the period profits earned by “old-only,” “both,” and “new-only” firms across years and across different industry states (see sections 4.1 and 4.2). Thus, in principle, these dynamic parameters are identified by both time-series and cross-sectional variations.
3.5 Results

This section reports the estimation results.

3.5.1 Results: Demand

Tables 3.3 and 3.4 display demand estimates. I employ two market definitions, broad (1 and 2) and narrow (3 and 4). The former definition aggregates observations across both regions (U.S. and non-U.S.) and user types (computer makers and distributors/end users), in a manner consistent with the industry’s context of a single, global market. However, the dataset contains richer variations across regions and user types, which we can exploit for improved precision of estimates. Moreover, the most plausible IVs (Hausman-Nevo IVs) become available under the narrower market definition (i.e., by region/user type). For these reasons, I present results under both market definitions.

The IV estimates in columns (2) and (4) are generally more intuitive and highly statistically significant than the OLS estimates in columns (1) and (3). Specifically, the price coefficient is negative ($\hat{\alpha}_1 < 0$), whereas both smaller size (3.5-inch diameter = new generation) and quality (the log of storage capacity) confer higher benefits ($\hat{\alpha}_2 > 0, \hat{\alpha}_3 > 0$) to the buyers.

In columns (5) through (8), I report results for the nested logit model, which nests HDDs by diameter (generation). The nest parameter estimate ($\hat{\sigma} = .49$) in column (8) suggests within- and cross-generation substitutions are equally important, implying the presence of cannibalization effects when an incumbent considers the introduction of new-generation products. Results (5), (6), and (7) are difficult to interpret, with positive (or at least not significantly negative) price coefficients. Moreover, the nest parameter estimates are close to or above 1, the theoretical upper bound. I attribute these unintuitive results to the lack of adequate instrumentation. Although column (6) uses some IVs, the broad market definition precludes the use of Hausman-Nevo IVs.

I use column (4), the logit IV estimates under the narrow market definition, as my baseline
Table 3.3: Demand Estimates for 5.25- and 3.5-inch HDDs

<table>
<thead>
<tr>
<th>Model: Logit</th>
<th>Market definition:</th>
<th>Broad</th>
<th>Narrow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation method:</td>
<td>OLS IV</td>
<td>OLS IV</td>
<td></td>
</tr>
<tr>
<td>Price ($000)</td>
<td>-1.66*** -2.99***</td>
<td>-0.93** -3.28***</td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>(.36) (.64)</td>
<td>(.38) (.58)</td>
<td></td>
</tr>
<tr>
<td>Nests of Diameters</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>(-) (-)</td>
<td>(-) (-)</td>
<td></td>
</tr>
<tr>
<td>Diameter = 3.5-inch</td>
<td>.84** .75</td>
<td>1.75*** .91**</td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>(.39) (.50)</td>
<td>(.27) (.36)</td>
<td></td>
</tr>
<tr>
<td>Log Capacity (MB)</td>
<td>.18 .87***</td>
<td>.04 1.20***</td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>(.25) (.31)</td>
<td>(.22) (.27)</td>
<td></td>
</tr>
<tr>
<td>Year dummies</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Region/user dummies</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>.43</td>
<td>.29</td>
<td></td>
</tr>
<tr>
<td>Number of obs.</td>
<td>176</td>
<td>176</td>
<td></td>
</tr>
<tr>
<td>Partial $R^2$ for Price</td>
<td>.08</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Partial $R^2$ for Nest</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Number of obs.</td>
<td>405</td>
<td>405</td>
<td></td>
</tr>
</tbody>
</table>

Note: Dependent variable is $\ln(ms_j/ms_0)$. Standard errors in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 3.4: Demand Estimates for 5.25- and 3.5-inch HDDs (continued)

<table>
<thead>
<tr>
<th>Model: Nested Logit</th>
<th>Market definition:</th>
<th>Broad</th>
<th>Narrow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation method:</td>
<td>OLS IV</td>
<td>OLS IV</td>
<td></td>
</tr>
<tr>
<td>Price ($000)</td>
<td>.08 4.22***</td>
<td>-.05 1.63***</td>
<td></td>
</tr>
<tr>
<td>(5)</td>
<td>(.19) (1.53)</td>
<td>(.10) (1.06)</td>
<td></td>
</tr>
<tr>
<td>Nests of Diameters</td>
<td>1.01*** 2.29***</td>
<td>.98*** .49***</td>
<td></td>
</tr>
<tr>
<td>(6)</td>
<td>(.05) (.39)</td>
<td>(.04) (.15)</td>
<td></td>
</tr>
<tr>
<td>Diameter = 3.5-inch</td>
<td>1.96*** 4.27***</td>
<td>2.24*** 1.70***</td>
<td></td>
</tr>
<tr>
<td>(7)</td>
<td>(.20) (.76)</td>
<td>(.16) (.31)</td>
<td></td>
</tr>
<tr>
<td>Log Capacity (MB)</td>
<td>.06 1.19**</td>
<td>.08 .65***</td>
<td></td>
</tr>
<tr>
<td>(8)</td>
<td>(.09) (.53)</td>
<td>(.07) (.24)</td>
<td></td>
</tr>
<tr>
<td>Year dummies</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Region/user dummies</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>.85</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>Number of obs.</td>
<td>176</td>
<td>176</td>
<td></td>
</tr>
<tr>
<td>Partial $R^2$ for Price</td>
<td>-</td>
<td>.32</td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>-</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>Partial $R^2$ for Nest</td>
<td>-</td>
<td>.37</td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>-</td>
<td>.25</td>
<td></td>
</tr>
<tr>
<td>Number of obs.</td>
<td>405</td>
<td>405</td>
<td></td>
</tr>
</tbody>
</table>

Note: Dependent variable is $\ln(ms_j/ms_0)$. Standard errors in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

result for the subsequent analyses. Although the nested logit result (8) is also reasonable, the dynamic model does not incorporate product differentiation within diameters, thus leaving
no rationale for preferring (8) over (4). I refrain from using the results based on the broader market definition. Specifically, result (2) is similar to (4) and highly intuitive, but I am concerned about the limited availability of IVs and the reduced variation in data, which sometimes leads to unintuitive results like (6).

All eight estimates incorporate year dummies and also allow for the time-varying unobserved product quality by diameter ($\xi_{jt}$ in equations [3.8] and [3.9]; note I suppress time-subscripts in these formulae for notational simplicity). I use equation (3.9) to recover $\hat{\xi}_{jt}$ as residuals. Figure 3.7 (left panel) shows the evolution of $\hat{\xi}_{jt}$ for both old and new HDDs. Because $\hat{\xi}_{jt}$ reflect old and new HDDs’ relative appeal to the buyers (but unobserved to the econometrician), I interpret and refer to them as “popularity” henceforth. These unobserved popularities of the old and new products switched in 1992, suggesting the 3.5-inch replaced 5.25-inch as the mainstream HDD type.

Figure 3.7: Estimated Popularity (Unobserved Quality) and Marginal Costs

![Graph depicting estimated popularity and marginal costs over time.](image)

*Note: Results based on the IV estimates of logit demand system.*

### 3.5.2 Results: Marginal Costs

From the demand estimates and firms’ first-order conditions, I infer marginal costs of production (Figure 3.7, right). The continual drop in the marginal costs reflects two tendencies in the industry. First, HDDs required increasingly fewer parts due to design improvements, probably reflecting to some extent learning by doing. Second, offshore production in Singapore and other South East Asian locations became prevalent, reducing primarily the cost
of hiring engineers. Together these developments represent important channels of "process innovation."  

22. The new HDDs’ marginal cost declines at the average annual rate of 6.12%, which I assume equals the rate of drop in the sunk costs of adoption; that is, \( \delta = .9388 \) because the adoption cost of new technology directly relates to the production of new HDDs.  

23. Alternatively, one may assume time-invariant sunk costs (i.e., \( \delta = 1 \)). However, I believe sunk costs dropping in line with production costs is more natural since both costs are concerned with the manufacturing of the same goods. In principle, one can try to estimate \( \delta \) directly as a part of the dynamic model. In practice, however, such estimates tend to be unreliable, probably due to the same issue that plagues the estimation of \( \beta \), the discount factor. Because changes in \( \delta \) or \( \beta \) move almost everything in the model in the same direction, their identification seems impractical.

### 3.5.3 Results: Sunk Costs of Innovation

Table 3.5 shows the maximum likelihood estimates of the mean sell-off value, \( \phi \), and the base sunk costs of adopting new technology, \( \kappa^{inc} \) and \( \kappa^{ent} \). Each column represents a set of coefficient estimates given a particular value of the discount factor, \( \beta \).

<table>
<thead>
<tr>
<th>(Billion $)</th>
<th>( \beta = .94 )</th>
<th>( .92 )</th>
<th>( .90 )</th>
<th>( .88 )</th>
<th>( .86 )</th>
<th>( .84 )</th>
<th>( .82 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sell-off value (( \phi ))</td>
<td>12.51</td>
<td>8.87</td>
<td>5.69</td>
<td><strong>4.00</strong></td>
<td>2.74</td>
<td>2.04</td>
<td>1.41</td>
</tr>
<tr>
<td>Incumbents’ cost (( \kappa^{inc} ))</td>
<td>2.62</td>
<td>2.50</td>
<td>2.38</td>
<td><strong>1.99</strong></td>
<td>1.89</td>
<td>1.77</td>
<td>1.72</td>
</tr>
<tr>
<td>Entrants’ cost (( \kappa^{ent} ))</td>
<td>8.69</td>
<td>8.25</td>
<td>7.08</td>
<td><strong>4.60</strong></td>
<td>2.95</td>
<td>2.06</td>
<td>1.42</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-129.3</td>
<td>-117.1</td>
<td>-113.3</td>
<td>-110.2</td>
<td>-107.3</td>
<td>-104.2</td>
<td>-101.6</td>
</tr>
</tbody>
</table>

*Note: Standard errors are not available because of a step function-like shape of likelihood, which is typical of dynamic discrete games and leads to either zero or very large standard errors.*

Estimates tend to decrease with \( \beta \) because a lower discount factor implies a lesser value of doing business in the HDD market, hence lower values associated with entry/exit, too. The log likelihood also increases with \( \beta \), suggesting a slightly better fit (in terms of the choice probabilities, upon which the likelihood is based) for estimates based on lower \( \beta \). As \( \beta \) approaches .80, however, the sell-off value estimate \( \hat{\phi} \) drops close to zero and then turns negative (not reported). The adoption cost estimates \( \hat{\kappa}^{inc} \) and \( \hat{\kappa}^{ent} \) also turn negative with lower \( \beta \)s. Such results are economically implausible because, taken literally, they would
suggest firms are somehow “penalized” upon exit (instead of earning sell-off value) and “rewarded” upon innovation (instead of paying sunk cost).

Consequently, I choose $\beta = .88$ estimates as my baseline result: the choice that best reflects the tradeoff between data fit and economic sensibility. On the one hand, the $\beta = .88$ estimates fit data considerably better than $\beta = .94, .92,$ or $.90$. On the other hand, the coefficient estimates with $\beta = .88$ are sufficiently higher than zero, allowing straightforward interpretations. Furthermore, the simulation of market structure based on $\beta = .88$ performs better than simulations based on lower $\beta$ in terms of the peak number of firms in “both” and “new-only” states, that is, active players that have adopted new technology, because a lower $\beta$ leads to lower values of doing business. This lowering of value induces too many exits and too few adoptions to match the number of firms observed in the data. Hence a higher $\beta$ is desirable in this aspect of data fit (in state evolution, in contrast to the fit in choice probabilities as measured by log likelihood in Table 3.5). Finally, the existing empirical studies of dynamic oligopolies conventionally used $\beta = .95 \sim .90$, a range that is close to my preferred model.\(^{24}\)

The baseline ($\beta = .88$) estimates show two important features. One is that the entrants’ base sunk cost is higher than the sell-off value upon exit ($\hat{\phi} < \hat{\kappa}_{\text{ent}}$), which implies “no free lunch;” that is, entering this market for the sole purpose of exiting and running away with sell-off value does not pay off. The other important feature is that the adoption cost is lower for incumbents than for entrants ($\hat{\kappa}_{\text{inc}} < \hat{\kappa}_{\text{ent}}$); therefore, the seeming “inertia” of incumbents does not stem from their innate cost disadvantage. The explanation lies in other incentives, as I explore in detail with counterfactual analyses in the next section. The result that $\hat{\kappa}_{\text{inc}} < \hat{\kappa}_{\text{ent}}$ does not necessarily mean incumbents are entirely free from organizational, informational, or other commonly attributed disadvantages. Rather, my estimates simply suggest incumbents enjoy a certain innovation-cost advantage over entrants in net terms.

\(^{24}\)Collard-Wexler (2010), Schmidt-Dengler (2006), Ryan (2011), and Goettler and Gordon (2011) chose $\beta = .95, .94, .90,$ and $.90$, respectively, in their studies of the markets of concrete, hospitals, cement, and microprocessors. I believe a lower discount factor would be more adequate for the HDD industry because of its fast pace of technological changes.
capabilities over the years, which outweigh other potential disadvantages associated with being larger and older. Determining the exact contents of $k^{inc}$ and $k^{ent}$ is beyond the scope of this paper, but incumbents’ (net) cost advantage will have important welfare implications (see section 7).

3.5.4 Results: Industry State, Policy, and Value

Figure 3.8 compares the evolutions of the industry state in the data with the estimated model. “Model” (right panel) displays the evolution of the simulated industry state, which is the average equilibrium path based on the estimated choice probabilities (policy functions) with $\beta = .88$.

Figure 3.8: Market Structure Dynamics

Note: “Model” (right panel) displays the mean evolution of simulated industry state, based on the estimated choice probabilities (policy functions) with $\beta = .88$.

Overall, the estimated model replicates three key features of the data, albeit in a slightly smoother manner. First, the number of adopting incumbents (firms in “both” state) peaks at a level lower than entrants (“new-only”). Specifically, the peak numbers of “both” and “new-only” are 7 and 10 in the data, compared with approximately 5 and 10 in the model. Second, during the second half of the sample period, the survival rate is higher for adopting incumbents than for entrants, resulting in similar numbers of survivors as of 1998 (i.e., 4 “both” and 3 “new-only” in the data, compared with approximately 5 “both” and 5 “new-only” in the model). Third, the number of non-adopting incumbents (“old-only”) declines precipitously during the 1980s and then more slowly during the 1990s before reaching zero.
in 1998. Thus the estimated model captures the key data features of innovation and market structure dynamics.

Figure 3.9: Estimated Policy and Value Functions

Note: Policy and value functions along the mean equilibrium path, based on $\beta = .88$.

The simulation in Figure 3.8 (right) is based on the solution of the dynamic discrete game, that is, the optimal strategies and payoffs. Figure 3.9 shows the estimated policy functions (i.e., the optimal choice probabilities) and value functions (i.e., the attractiveness of each alternative). Three features are important. First, incumbents become increasingly more eager to innovate in years approaching 1988, with a peak adoption rate of 37% (Figure 3.9, left, “Old-only: Adopt”). After 1988, the adoption rate plummets to 1% and recovers only slightly toward 1997, the final year of dynamic decision-making. The estimated equilibrium values reflect this policy trajectory. The value of being an “old-only” firm starts relatively high (Figure 3.9, right, “Old-only”), but the values of being “both” and “new-only” gradually catch up and surpass that of “old-only” by 1988, as the new-generation HDDs gain in popularity. At this point, innovation becomes most attractive to incumbents: hence the increasing adoption rate toward 1988. However, after 1988, the value of being “old-only” is so low incurring sunk costs and joining the herd of (already numerous) new-HDD producers no longer pays off.

The second important feature is the high adoption rate among entrants. Except for 1981, potential entrants’ equilibrium probability of adoption (entry) is consistently above 80% (Figure 3.9, left, “Potential entrant: Adopt”). The estimated policy matches the data well, with most potential entrants deciding to adopt as well. Third, an increasing number
of firms exit toward the end. All classes of firms show this tendency because the value of staying in the industry declines as the game approaches the terminal year, 1998. However, the firms in “both” state, that is, the adopting incumbents, temporarily back this downward trend between 1989 and 1993 because the number of rivals in the new HDD category starts to decrease while its profitability finally begins to take off. In general, the time profile of value function, including whether and how much $V_t^{both}$ trends upward in the middle, depends on the discount factor, with a lower $\beta$ leading to “bumpier” time-series.

### 3.6 “Innovator’s Dilemma” Explained

This section answers the first question of the paper, namely, why incumbents are slower than entrants in innovation. I quantify the effects of the three theoretical forces that determine the incumbent-entrant timing gap in technology adoption: cannibalization, sunk-cost gap, and preemption. To measure each effect, I compare the timing gaps in the estimated baseline model with a counterfactual simulation in which that particular incentive mechanism is absent.

Figure 3.10 summarizes the results of the counterfactual analyses. The timing gap between incumbents and entrants is measured by the percentage-point differences between incumbents’ and entrants’ CDF of adoption timing (c.f., Figure 3.2), averaged over years. “Baseline” is the estimated model’s outcome (10.32 percentage points). The other three values (5.27, 16.45, and 51.73 percentage points) represent the simulated counterfactuals in which I “shut down” particular economic incentives.

The comparison of the counterfactuals against the baseline suggests the following: (1) cannibalization can explain 51% of the timing gap; (2) without preemptive motives (and other dynamic strategic incentives), incumbents would have further delayed innovation by as much as 59%; and (3) contrary to the prior impression of “organizational inertia,” incumbents enjoy a cost advantage over entrants (and hence the elimination of this cost advantage would have led to incumbents’ much longer delay). These timing-gap outcomes are derived from the simulations shown in Figure 3.11, which compares the evolutions of the industry state
in the baseline model with three counterfactuals.

In the following three subsections, I explain the setup, result, and interpretation of each counterfactual.

### 3.6.1 Cannibalization (Counterfactual 1)

I eliminate the cannibalization factor from incumbents’ adoption behavior by isolating the adoption decision (production of new HDDs) from the profit maximization regarding old HDDs. In other words, I effectively split each incumbent firm into two separate entities: a “legacy” division that takes care of the manufacturing of old HDDs and a “corporate venture” in charge of developing new HDDs. The former division acts as an independent “old-only” firm that decides whether to stay or exit in each year, but without the third alternative to adopt new technology and become “both.” The latter division acts like a “potential entrant” with staying power, which can choose to adopt (and become “new-only”), wait, or exit. Thus each incumbent in this no-cannibalization counterfactual is two separate firms dedicated to
old and new HDDs in isolation.\textsuperscript{25}

Incumbents (their “corporate venture” divisions, to be precise) are much more eager to adopt new HDDs than in the baseline case. Consequently, an approximately equal numbers of incumbents and entrants produced the new HDDs during the 1990s.

Free of the cannibalization concerns regarding their own old-HDD business, more incumbents (their “corporate venture” divisions) start producing new HDDs earlier. Cannibalization can explain half of the actual timing gap between incumbents and entrants. On a

\textsuperscript{25}An alternative approach to isolate the cannibalization factor would be to directly alter the HDD demand system in such a way that old and new HDDs no longer substitute for each other. Computationally, this alternative approach is easier than the approach I chose, because the latter substantially increases the effective number of firms. A drawback of the alternative approach is that the counterfactual demand system needs to be specified in a rather arbitrary manner. There must be two different markets, of size $M_{\text{old}}^t$ and $M_{\text{new}}^t$, when in fact only one market existed (with size $M_t$), but I see no obvious way to split $M_t$ into $M_{\text{old}}^t$ and $M_{\text{new}}^t$, and the outcomes depended heavily on the way I split the market (not reported).
separate note, the sudden exit of “old-only” divisions of incumbents is interesting (Figure 3.11, top right, “Incumbent (legacy division”) ). It reflects the limited value of staying in the old-HDD business with no prospect of starting new HDD production.

### 3.6.2 Preemption (Counterfactual 2)

Preemption is a dynamic strategic motive. In an oligopolistic environment, some firms’ early adoption would reduce the incremental profits available to late adopters. An incumbent has incentives to preempt other incumbents as well as potential entrants. Thus, the silencing of preemption requires that firms do not perceive the evolution of industry state (the numbers of firms in “old-only,” “both,” and “new-only” states) as something they can influence by their own actions. In the no-preemption counterfactual, firms take the evolution of industry as exogenous to their dynamic decisions.\(^{26}\)

In the absence of preemptive motives, incumbents’ delay increases substantially, to 16.45 percentage points from 10.32 percentage points in the baseline model. The number of “both” firms (adopting incumbents) grows more slowly.

Each firm ignores its rivals’ decisions, so the nature of the dynamic game changes fundamentally from that of strategic entry/exit to a single-agent optimal stopping problem. An incumbent does not need to act aggressively to deter the rivals, so the innovation rate becomes lower and the incumbent-entrant gap wider.

### 3.6.3 Sunk Cost Gap (Counterfactual 3)

An important finding from estimating the baseline model is the sunk cost advantage of incumbents relative to entrants, the estimates of which were 1.99 and 4.60, respectively. What if incumbents no longer enjoyed this cost advantage? To eliminate the cost difference, this counterfactual sets the sunk costs at 3.30 for both incumbents and entrants

\(^{26}\)One might alternatively label this counterfactual as a “no dynamic strategic interaction” or “dynamic monopolistic competition” scenario. I choose to call it a “no-preemption” case to highlight the economic incentives I believe are at the heart of firms’ technology adoption decisions.
\( (\kappa^{inc} = \kappa^{ent} = \bar{\kappa} = 3.30) \).\(^{27}\)

Incumbents’ innovation is discouraged. At most, only two “both” firms are active in the market. By contrast, more “new-only” firms thrive even toward the end of the 1990s.

A 66% increase in sunk cost is sufficient to suppress most incumbents’ adoption. Since technology adoption is essentially a dynamic discrete choice problem, a material change in the cost of choosing a particular alternative is bound to have large repercussions on the outcome. Another interesting feature is the higher survival rate among entrants. Seven “new-only” firms survive until 1998, whereas only five survive in the baseline case. With few incumbents adopting and directly competing in the new HDD category, entrants can enjoy higher profits and hence improved survival prospects.

### 3.7 Policy Experiments

In this section, I evaluate public policies concerning innovation and competition. I conduct counterfactual simulations and compare measures of social welfare. Specifically, I experiment with four policies: (1) broad patent on new HDDs, (2) R&D subsidies to incumbents, (3) ban on non-compete clauses, and (4) ban on international trade. The purpose of these experiments is to inform policy design as well as to deepen our understanding of the interactions between innovation, competition, and welfare.

Table 3.6 summarizes the welfare analysis. Rows represent different policy simulations, including the benchmark cases. Columns list the components of social welfare: (A) consumer surplus, (B) producer surplus, (C) sell-off value upon exit, and (D) sunk costs of technology adoption. Social welfare is their sum. Given the finite-horizon setup, I display social welfare figures separately for the sample period (1981 through 1998) and for the years since 1999 (omitted due to space constraints). The latter consists of the terminal values of (A) and (B)

---

\(^{27}\)Alternatively, \( \bar{\kappa} = 1.99 \) (i.e., entrants enjoy the low cost of incumbents) is an equally plausible setup. However, even in the baseline model, most potential entrants decide to adopt anyway, which limits the upside for entrants’ adoption rate. Another possible configuration is \( \bar{\kappa} = 4.60 \) (i.e., incumbents are as “handicapped” as entrants). This setup results in few incumbents’ adoption, only toward the end of the sample period (not reported). Although interesting, this result is too extreme to be compared with the baseline case. Therefore, I chose to show the results for the “mean” counterfactual \( (\bar{\kappa} = 3.30) \) for more meaningful comparison.
Table 3.6: Comparison of Social Welfare across Policy Experiments

<table>
<thead>
<tr>
<th>(Billion $)</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>(A+B+C+D)</th>
<th>Change from Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Benchmarks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Baseline</td>
<td>127.7</td>
<td>9.4</td>
<td>20.0</td>
<td>-39.5</td>
<td>117.6</td>
<td>0%</td>
</tr>
<tr>
<td>• U.S.</td>
<td>102.5</td>
<td>4.5</td>
<td>9.6</td>
<td>-18.0</td>
<td>98.6</td>
<td>0%</td>
</tr>
<tr>
<td>• Non U.S.</td>
<td>25.2</td>
<td>4.9</td>
<td>10.4</td>
<td>-21.6</td>
<td>18.9</td>
<td>0%</td>
</tr>
<tr>
<td>2. Planner</td>
<td>410.7</td>
<td>0.0</td>
<td>0.0</td>
<td>-0.8</td>
<td>409.9</td>
<td>248.7%</td>
</tr>
<tr>
<td>3. Monopolist</td>
<td>60.3</td>
<td>16.6</td>
<td>0.0</td>
<td>-0.8</td>
<td>76.0</td>
<td>-35.3%</td>
</tr>
<tr>
<td><strong>Experiments</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Broad Patent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Pre-announced</td>
<td>109.6</td>
<td>16.4</td>
<td>6.2</td>
<td>-13.9</td>
<td>118.2</td>
<td>0.5%</td>
</tr>
<tr>
<td>• Surprise in 1988</td>
<td>67.9</td>
<td>15.1</td>
<td>17.0</td>
<td>-39.0</td>
<td>61.0</td>
<td>-48.2%</td>
</tr>
<tr>
<td>2. R&amp;D Subsidy</td>
<td>127.2</td>
<td>9.4</td>
<td>20.7</td>
<td>-45.4*</td>
<td>111.9</td>
<td>-4.8%</td>
</tr>
<tr>
<td>3. Ban NCC</td>
<td>129.6</td>
<td>7.6</td>
<td>18.5</td>
<td>-47.5</td>
<td>108.2</td>
<td>-8.0%</td>
</tr>
<tr>
<td>4. No Trade</td>
<td>107.4</td>
<td>16.8</td>
<td>6.7</td>
<td>-31.1</td>
<td>99.7</td>
<td>-15.2%</td>
</tr>
<tr>
<td>• U.S.</td>
<td>88.5</td>
<td>12.7</td>
<td>0.0</td>
<td>-12.4</td>
<td>88.8</td>
<td>-10.0%</td>
</tr>
<tr>
<td>• Non U.S.</td>
<td>18.9</td>
<td>4.1</td>
<td>6.7</td>
<td>-18.7</td>
<td>11.0</td>
<td>-88.9%</td>
</tr>
</tbody>
</table>

*Note: Each number is the sum of discounted present values as of 1981. Includes government subsidies.*

The largest social welfare component is consumer surplus (A). However, the second largest item, adoption costs (D), tends to vary most across different scenarios because the number and timing of technology adoption will drastically alter the total sunk costs. In contrast, consumer surplus stays within a narrower range as long as several firms are supplying old and new HDDs. Therefore, the final welfare outcome (A+B+C+D) mainly reflects the tradeoff between the benefits (A) and costs (D) of innovation.

This tradeoff is clear in Figure 3.12, which graphs the time profile of social welfare in the baseline case. To illustrate the relative magnitude of surpluses and costs in different years, each welfare component is displayed without time discounting. Three features are noteworthy. First, the size of the consumer surplus grows over time with the growth of market size ($M_t$). Second, as the number of firms declines during the 1990s, the increased market power leads to higher producer surplus. The sell-off value is also visible because of continual exits. Third, adoption costs are huge and accrue in the early years, so the net social welfare is negative during most of the 1980s, when demand for HDDs is limited. Once the sunk costs are paid, however, all future surpluses count toward a net increase in social
welfare. Consequently, the world enjoys a net gain of $117.6 billion between 1981 and 1998, and $229.9 billion if we include all years after 1998. If we consider the evolution of the HDD industry an investment project, the social “internal rate of return” would be 33.5%.\textsuperscript{28}

Figure 3.12: Time Profile of Social Welfare (Baseline)

Note: Numbers are not time discounted. Excludes terminal values (i.e., surpluses accruing after 1998).

To find theoretical benchmarks, I calculate the welfare profiles under the hypothetical scenarios in which a social planner and a monopolist, respectively, make production decisions. In both cases, only one incumbent adopts new technology because that is the cheapest way to produce new HDDs. The planner implements marginal-cost pricing ($p_g = mc_g$), whereas the monopolist charges profit-maximizing prices. Interestingly, both the planner and the monopolist will choose to innovate in the same year, 1986, when the overall demand for HDD just begins to take off. Their optimal innovation timing coincides because the time profiles of their costs are identical, as they both operate a single incumbent firm and consider its optimal adoption year. Their benefits calculations also have similar time profiles based on the overall size of potential demand, although the ways in which the planner and the

\textsuperscript{28}The internal rate of return (IRR) is the discount rate at which the project’s net present value becomes zero. The IRR of 33.5\% implies the social return from the entire historical development in the HDD is positive as long as the social discount factor is above .665.
monopolist evaluate (appropriate) surpluses are different, namely, consumer surplus versus producer surplus.

### 3.7.1 Broad Patent (Experiment 1)

The question of whether broad patents encourage innovation is particularly relevant to the HDD industry, in which Rodime, a Scottish firm that was among the first adopters of the 3.5-inch technology, claimed patent on the whole concept of 3.5-inch HDDs in 1986 (see Table 3.7). After years of lawsuits between Rodime and its rivals, the U.S. Court of Appeals for the Federal Circuit (CAFC) eventually rejected the claim in 1995 and 1996, but several HDD makers gave up the court battles before the rulings and agreed to pay license fees to Rodime.

Table 3.7: Brief History of Rodime’s 3.5-inch HDD Patent Affair

<table>
<thead>
<tr>
<th>Year</th>
<th>Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>Rodime became independent from Burroughs’ 5.25” HDD plant (Glenrothes, Scotland).</td>
</tr>
<tr>
<td>1983</td>
<td>Rodime became the first maker to achieve volume production of 3.5-inch HDDs.</td>
</tr>
<tr>
<td>1986</td>
<td>Rodime surprised the industry by obtaining a patent on the concept of a 3.5” drive. Rodime sued Miniscribe and Conner Peripheral for patent infringement. IBM sued Rodime, which countersued IBM.</td>
</tr>
<tr>
<td>1988</td>
<td>The 3.5” patent affair headed for a long tour of the U.S. federal court system. Miniscribe opted out by taking a license from Rodime.</td>
</tr>
<tr>
<td>1989</td>
<td>Rodime moved to Singapore for production efficiency, but neared bankruptcy and got some financing. Top management was completely overhauled in early 1989.</td>
</tr>
<tr>
<td>1991</td>
<td>Patent affair ended when IBM and Conner Peripheral took licenses, as well as Fujitsu and Alps Electric. Several other firms were in negotiation. Rodime pursued joint ventures with Japan’s JVC and firms in Taiwan and Korea, but in mid-1991 announced it would file for bankruptcy and cease manufacturing operations. It planned to remain active in pursuing licensing revenues from 3.5” HDD patents.</td>
</tr>
<tr>
<td>1994</td>
<td>High legal expenses and falling license revenues put financial pressure on Rodime.</td>
</tr>
<tr>
<td>1995</td>
<td>In September 1995, a U.S. appeals court ruled some of Rodime’s patent claims invalid, a ruling in favor of Quantum. Rodime still argued that other patent claims were valid, in a separate legal action against Seagate.</td>
</tr>
<tr>
<td>1996</td>
<td>Appeals court rulings in 1995 and 1996 appear to have weakened Rodime’s negotiating position, but it continues to argue that other patent claims are still valid.</td>
</tr>
</tbody>
</table>

*Source: DISK/TREND Reports, various years.*

Although Rodime’s claims were considered outrageous in the industry at the time, and ended in a legal grey zone, studying what the welfare consequences would have been had the patent system and CAFC’s rulings been different is worthwhile. I propose two separate experiments, one designed to study the *ex-ante* impact of a pre-announced broad patent
regime and the other to study the impact of *ex-post* “surprise” court rulings.

In the first counterfactual, only the first adopter(s) is/are allowed to manufacture new HDDs, and this legal arrangement is pre-announced by the patent authority before 1981, before the game begins. The setup allows for multiple adopters as long as they belong to the first cohort. The motivation for this permissive regime is to reflect the custom of defensive patenting in the computer-related industry, where rival firms tend to hold competing intellectual-property claims and engage in lawsuits and countersuits. Firms typically use patents to enhance their bargaining power in negotiating favorable terms in cross-licensing agreements.29

In the second counterfactual scenario, Rodime’s rivals ignore the company’s patent claims until 1988, when the CAFC, a centralized appellate court for patent cases established in 1982, announces its surprise ruling to honor Rodime’s patent infringement claims, paving the way for a legal monopoly of the 3.5-inch technology.30

Figure 3.13 compares the evolutions of the industry state in the two counterfactual simulations. The pre-announced broad patent regime (left panel) induces a relatively simple market structure. Seven of the 11 incumbents decide to adopt new HDDs in 1981, after which further adoption becomes illegal. No entrants can adopt. Social welfare increases by 0.5% between 1981 and 1998, but, ironically, this gain does not come from increased innovation. Rather, social welfare increases because patents prevent innovation by many firms, thereby saving adoption costs.

In the second, “surprise court ruling” experiment (right panel), the industry evolves just as in the baseline case until 1988, the year when the CAFC hypothetically entrenched Rodime’s legal monopoly. All “new-only” firms immediately go out of business, and all “both” firms except for Rodime itself are forced back to “old-only” status by 1989, a position that is so unattractive the firms all subsequently elect to exit the industry.

---

29 Alternatively, one might consider the patent regime in which only and exactly one firm can make 3.5-inch HDDs. I omit this scenario because the outcome is almost equivalent to the benchmark “monopolist” case.

30 Patent lawsuits usually end with the payment of damages. For simplicity, this experiment assumes that the payments are large enough to make Rodime’s rivals indifferent with respect to staying versus exiting.
Broad patents function as a hard entry barrier. Because adoption cost is an important component of social welfare, limited entry translates into a huge saving in societal cost. Limited competition reduces the 1981–98 consumer surplus by 14.2% ($18.1 billion), but the adoption cost reduction of 64.8% ($25.6 billion) compensates for this loss. Seven adopters seem sufficient to ensure relatively competitive supply and pricing. The innovation timing is front-loaded to 1981, which is a favorable development from the viewpoint of pro-innovation policy.

In stark contrast with the previous experiment, social welfare drops by 48.2% in the ex-post patent scenario because most firms had already paid the sunk costs and started production of 3.5-inch HDDs by 1988, so that there is no cost saving as in the "pre-announced" patent regime. Instead, the only major change is that the industry becomes a monopoly from 1990 and consumers suffer.

In the literature, Lerner (1994) found a positive correlation between broader patents and increased innovative activities, whereas Sakakibara and Branstetter (2001) found no measurable relationship.\(^{31}\) In contrast, the results of my experiments point to the possibility that allowing for broad patents might decrease innovative activities. These diverging results are not necessarily at odds with each other for two reasons. First, even when the number of total innovators decreases in the long run, if the timing is earlier then the number of

\(^{31}\) Other theoretical and empirical investigations on this topic include Gilbert and Shapiro (1990), Klemperer (1990), Jaffe and Lerner (2004), Chaudhuri, Goldberg, and Jia (2006), and Qian (2007).
innovators will be higher in the short run. In other words, a tradeoff seems to be present between short-run and long-run levels of innovative activities, and patents may change this balance. These changes in the timing and number of innovations embody the ex-ante effect of patents. Second, because patents are legal monopoly rights by definition, if one firm’s patent on a particular product/technology is strictly honored and protected, rival firms must be excluded from the same product/technology. This ex-post effect of broad patents is highlighted in the second experiment. In principle, however, we might still think those excluded firms might seek technological opportunities elsewhere and end up conducting innovative activities outside that particular market.

### 3.7.2 R&D Subsidy (Experiment 2)

Is subsidizing the adoption of new HDDs socially worthwhile? R&D subsidies are rampant, as arguing against the virtues of innovation is difficult. Whether subsidies are really desirable is another issue, which I explore in this experiment.

I set $\kappa^{inc} = 0$ in this counterfactual, letting the government subsidize 100% of incumbents’ innovation cost. This setup does not mean innovations are socially costless; rather, the cost simply accrues to the public sector, which is reflected in Table 3.6 as adoption cost (D). This R&D subsidy specifically targets incumbents because their costs are lower. More fundamentally, for the government to identify and contact potential entrants is not always feasible.32

Subsidies accelerate incumbents’ adoption. The peak number of innovating incumbents (“both”) reaches eight in 1987, compared with the baseline result of five in 1989. However, the total social welfare is 4.8% lower than in the baseline case because more innovators means higher total adoption costs.

Subsidies successfully encourage more incumbents to adopt, and in earlier years. However, this aggressive adoption does not translate into a significant increase in consumer surplus for two reasons. First, a higher number of exits offsets a higher number of adopters. Second,

---

32Raising enough capital to lobby the government is also difficult for potential entrants.
potential entrants are less likely to enter due to preemption, and even if they do, they are more likely to exit.

This result echoes some of the findings in the existing research. The effect of R&D subsidy has been an active area of innovation studies, where a crucial policy question is whether R&D subsidies increase or decrease innovative activities in general, and private R&D spending in particular.\textsuperscript{33} My experiment suggests R&D subsidy would increase the number of adopters, but public R&D spending would substitute for private R&D spending without inducing a complementary increase in private R&D spending, because of the once-and-for-all nature of innovation in the technological context of the HDD industry. Although the nature of R&D efforts varies by industry, this experiment suggests a potential explanation for why studies tend to find small effects of public R&D spending in inducing private R&D.

\subsection*{3.7.3 Banning Non-compete Clauses (Experiment 3)}

As Scherer (1980) and Bresnahan (2003) emphasized as the (then) chief economists of the Federal Trade Commission, entrants are important sources of innovation.\textsuperscript{34} Therefore, entry facilitation is a potential channel through which public policy may promote innovation. One


\footnote{Managers seem to agree, too. See Grove (1996) for the former Intel CEO’s account.}

\textit{Note:} See sections 7.2 and 7.3 for the details of counterfactuals.
possible policy instrument concerns the legality of the private contracts governing the ease of starting new businesses. Specifically, the non-compete clause (NCC, or the covenants not to compete [CNC]) is the type of employment contract that limits the actions of personnel working at an existing firm and considering starting their own businesses (or working for rival firms). Many HDD makers are headquartered in the American state of California, where NCCs are banned. Nevertheless, exceptions to this ban exist, such as the cases involving mergers, and the practice was common during the sample period. Of the 14 potential entrants in my data, five Californian firms are founded by former managers and engineers from existing HDD makers, a typical Silicon Valley phenomenon. Thus a strict ban on NCC may have important welfare implications through innovation and competition.

In this counterfactual, the five potential entrants founded by industry veterans appear two years earlier than in the actual data.

The two-year front-loading of the five entrants results in their earlier adoption. This front-loading does not materially alter the trajectory of incumbents. Social welfare decreases by 8% because of increased adoption costs.

As in the case of more aggressive adoption by incumbents (section 7.2), earlier adoption tends to generate limited extra consumer surplus while raising the adoption costs substantially. Consumer surplus does not increase substantially, because the market size (the number of potential buyers) is still small during the early years. Adoption costs increase because they are both initially high and discounted less.

3.7.4 No International Trade (Experiment 4)

Disputes over intellectual property rights often become international trade disputes in the computer-related industries. For example, Apple sued HTC, a leading manufacturer of the Android-based smart phone in Taiwan, in July 2011 for two patent infringement claims related to touch-screen technologies in iPhones. HTC countersued Apple, but, meanwhile,

---

35 For instance, Finis Conner had to wait for two years before leaving Seagate Technology to found Conner Technology in the late 1990s, because he had been a party to the merger between Seagate Technology and Conner Peripheral in years prior to his departure.
the U.S. Commerce Department started contemplating banning U.S. imports of HTC’s smart phones. A similar dispute between Apple and Samsung led to a temporary ban on Samsung’s Galaxy tablet computer in Germany and Australia. Given the global nature of competition in these high-tech industries, intellectual property and competition policy issues tend to become international trade issues at the same time. Although the international trade of HDDs avoided political interventions during the sample period, other sectors of the computer industry were full of trade disputes around the same time.\textsuperscript{36} Because the HDD industry data are relatively free from political complications, they provide a clean laboratory in which to experiment with the welfare impact of trade barriers.

This counterfactual examines an environment in which no HDD trade occurs between the U.S. and the rest of the world. On the demand side, the world market (its pool of potential buyers) is split into the U.S. and non-U.S. On the supply side, American firms can only serve the U.S. market, whereas non-American firms serve the rest of the world. Thus this experiment is one of alternative market sizes and market structures.

In both the United States and elsewhere, consumer surpluses as well as the total adoption costs decrease. The net result is a 15.2\% decrease in worldwide social welfare. Interestingly, the effect is asymmetric between the two regions. The U.S. social welfare drops by only 10\% while the rest of the world suffers an 88.9\% reduction.

On the supply side, trade restrictions are anti-competitive in limiting the numbers of effective competitors in both the United States and outside. Moreover, the smaller market size on the demand side implies smaller spoils of innovation for potential adopters, discouraging the production of new HDDs. These two effects lead to more concentrated market structures. The latter effect translates into a lower adoption rate and hence cost savings. Since these cost savings accrue to producers, the effect of trade barriers is positive for the U.S. producers. However, non-U.S. producers’ surplus declines because the number of firms originating from outside the United States is disproportionately high compared to the size

\textsuperscript{36}For example, Japan’s Ministry of International Trade and Commerce restricted the imports and manufacturing of IBM’s computers during the 1960s. The U.S. Commerce Department imposed “anti-dumping” duties on Samsung’s memory chips (DRAM) in the early 1990s.
of the market. Thus, whereas the concentration effect (i.e., a smaller number of firms compete) dominates in the United States, to the advantage of American HDD makers, outside the United States, the market size effect (i.e., fewer buyers of HDDs exist) dominates. In other words, consumers across the globe would suffer from the lack of trade, and so would non-American producers. American firms would be the only parties to benefit from the ban on trade.

3.8 Conclusion

This paper presents an industry equilibrium model of innovation and entry/exit with dynamic strategic interactions. I estimate the model using an 18-year panel of HDD market data. Contrary to Schumpeter’s (1934) earlier conjecture, incumbents actually enjoy an advantage over entrants in terms of innovation sunk costs: a result that echoes his later observation (Schumpeter 1942). This finding implies incumbents are not slower than entrants because of an efficiency handicap.

Then what explains “the innovator’s dilemma,” or the incumbent-entrant timing gap? The first set of counterfactual simulations quantifies the three theoretical forces that determine incumbents’ delayed adoption timing compared to entrants. The results suggest cannibalization between old and new technologies/products can explain at least a half of the dilemma. Because old and new HDDs substitute for each other, the introduction of
new HDDs would dampen profits from the old HDDs. Taking this “replacement effect” into consideration, rational incumbents can expect relatively small benefits from innovation compared to new entrants. Thus creative destruction (i.e., the process of transitions from old to new technologies along with firms’ turnover) takes place not necessarily due to old firms’ lack of creativity, but because of their rational unwillingness to destroy old sources of profits. This finding explains why new entrants, despite their cost disadvantage, sometimes manage to creatively destroy old winners in the industry: the rational innovator’s dilemma.

Another finding from the first set of simulations (estimating the innovator’s dilemma) is that dynamic strategic incentives, such as preemption, are important determinants of innovation and industry evolution. Without strategic interactions, the incumbent-entrant timing gap would have been wider by 59%. Thus certain aspects of industry dynamics call for explicit—although often computationally burdensome—modeling of game-theoretic interactions among firms.

In the second set of simulations, I experiment with various public policies related to innovation and competition. The main finding is that competition is generally pro-innovation, as highlighted in the counterfactual simulation of a “no international trade” scenario. However, the welfare implication of competition is more nuanced. When many firms adopt new technology, the duplication of effort often turns out to be socially wasteful. Three factors explain this outcome. First, the estimated sunk cost of innovation is economically significant. Second, once more than four or five firms compete in the same (new) product category, a further increase in the number of adopters adds little to consumer surplus. Third, although buyers generally do benefit from early production of new goods, the market size is small during the early years. Thus the earlier adoption by many firms would result in negligible gains at sizeable costs. This finding explains why the simulated R&D subsidies lead to more innovation but reduced social welfare, and also why a (pre-announced) broad-based patent, which represents a fairly anti-competitive intellectual property regime, results in increased social welfare.

Although some of these results would be specific to the HDD market, the economic incentives studied here are quite general and could be expected to operate in many oligopolistic
markets. Similar trajectories of creative destruction are also widely observed across industries, especially in the computer-related sectors. These considerations lead me to believe this paper’s analytical framework, as well as its empirical findings, could be applicable elsewhere: old winners can survive the “gale of creative destruction” only through creative self-destruction.
References


