INDUSTRIES IN RAPID CHANGE: ESSAYS ON HEALTH CARE, BANKING AND SOFTWARE.

by

John Kim

A.B. Economics
Harvard University, 1995

SUBMITTED TO THE DEPARTMENT OF ECONOMICS IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1998

© 1998 John Kim. All rights reserved

The author hereby grants permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part.

Signature of Author .................................................................

Department of Economics
May 15, 1998

Certified by .................................................................
Richard Schmalensee
Gordon Y Billard Professor, Industrial Economics
Thesis Advisor

Certified by .................................................................
Jerry Hausman
John & Jennie S. MacDonald Professor, Econometrics and Public Finance
Thesis Advisor

Accepted by .................................................................
Peter Temin
Elisha Gray II Professor of Economics
Chairman, Departmental Committee on Graduate Studies
INDUSTRIES IN RAPID CHANGE: ESSAYS ON HEALTH CARE, BANKING AND SOFTWARE.

by

John Kim

Submitted to Department of Economics on May 15, 1998 in partial fulfillment of the requirement for the degree of Doctor of Philosophy.

ABSTRACT

This dissertation explores three different sectors all undergoing rapid organizational change: health care, retail banking and software technology. In the health sector, the continued penetration of managed care organizations has brought with it a fundamental change in the financial incentives faced by physicians. This has brought to the forefront the issue first raised in Arrow’s seminal 1963 paper: Do physicians operate in financially neutral environments? Chapter one represents a contribution to this ongoing debate by examining the case of cesarean versus natural child delivery. By using the subsample of Medicaid births, I find that a $100 increase in the differential between cesarean and natural reimbursement leads to a .7 percent increase in the likelihood of cesarean delivery. I interpret this to be striking evidence that ob/gyns do indeed respond significantly to financial incentives.

In the banking sector, increased consolidation among the large regional automated-teller-machine (ATM) networks has brought to the forefront the upstream access prices the network charges to retail banks as well as the downstream price charged to consumers for using ATMs. Chapter two attempts to model the unique nature of competition in this industry: Banks who compete at the retail level have backward integrated a portion of their assets (their ATM machines) to form the shared ATM network which sells access to foreign ATMs to the various banks in the network. The member banks themselves are also residual claimants to the profits of the network entity. The analysis shows that upstream prices charged to member banks are lowest when the shareholders in the network entity are network neutral banks, i.e. their depositors rely on other banks’ ATMs as much as other banks’ depositors rely on their ATMs. There also appears to be a tradeoff between keeping access prices low at the upstream level to retail banks versus keeping access prices low to the end-consumer.

Few industries in recent memory have undergone the kind of transformation we see in high-tech. The pace of innovation has made both hardware and software technology a revolutionary force in the economy. Chapter three is the first attempt to put structure on the serendipitous process of how VCs price these new and unproven high-tech software ventures. The results indicate that how far along a company is to a completed product is the primary motivator of how pricing is done. In addition, investor sentiment towards the high-tech sector in general represented by the previous year’s return on the Nasdaq
Composite Index appears to have a significant effect on pricing of early stage deals. Finally, prior experience on the management side as well as the technical team leads to large premiums over first-time entrepreneurs.

This thesis concludes with an essay on the disaggregation of the software value chain. In particular, it focuses on how the rapid pace of innovation in this sector has led to more firms outsourcing their entire information systems to systems integrators, and these same systems integrators looking upstream to the startup community for the technical innovation.
Acknowledgements

What a privilege it's been to be a graduate student in economics at MIT. I learned from the best professors, studied with the best students, all in the best possible environment for learning. I owe much to this institute as a whole.

I give my thanks first to my family. My debt to them cannot be done justice here in a paragraph, and so I will just say thank you.

I owe much to the entire industrial organization and econometrics faculty at MIT. In particular Glenn Ellison, Paul Joskow, Whitney Newey and my two advisors: Richard Schmalensee and Jerry Hausman. Additional gratitude goes to my two classmates Alan Sorensen and Mark Rainey. Their insight, wisdom and willingness to help someone who was always a step or so behind will not be forgotten. Finally, I would like to acknowledge the help of Christopher Velluto of National Economic Research Associates. He has been a terrific mentor, and my work with him has demonstrated to me the relevance of economic thinking in real world issues.

With that said, my greatest debt belongs in all circumstances to my Lord and Savior Jesus Christ. He makes all things possible, and the completion of this mundane thesis is a trivial demonstration of His power, but a testimony nonetheless. From the day I stepped onto this campus, I have been watched and cared for by the Holy Spirit, and it gives me great joy to know that I need not verbalize anything for Him to understand my gratitude. Blessed by our God.

"The Lord is my shepherd; I shall not want." Psalm 23:1.
# Table of Contents

<table>
<thead>
<tr>
<th>Chapter 1</th>
<th>Physician Fees and Procedure Intensity: The Case of Cesarean Delivery</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Introduction</td>
<td>10</td>
</tr>
<tr>
<td>1.2 Theory</td>
<td>13</td>
</tr>
<tr>
<td>1.3 Previous Literature</td>
<td>15</td>
</tr>
<tr>
<td>1.4 Data and Empirical Strategy</td>
<td>21</td>
</tr>
<tr>
<td>1.5 Basic Results</td>
<td>25</td>
</tr>
<tr>
<td>1.6 Conclusions</td>
<td>28</td>
</tr>
<tr>
<td>Table 1A: Means of Dataset</td>
<td>31</td>
</tr>
<tr>
<td>Table 1B: Medicaid Fee Difference</td>
<td>32</td>
</tr>
<tr>
<td>Table 1C: Medicaid Fee Difference Normalized by Private Vaginal Fee</td>
<td>32</td>
</tr>
<tr>
<td>Table 2: Logit Regression Results</td>
<td>33</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 2</th>
<th>Impact of Proprietary Positions and Equity Interest in the Pricing of Network ATM Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Introduction</td>
<td>35</td>
</tr>
<tr>
<td>2.2 Background on ATM Networks</td>
<td>37</td>
</tr>
<tr>
<td>2.3 Previous Literature</td>
<td>41</td>
</tr>
<tr>
<td>2.4 The Basic Model</td>
<td>44</td>
</tr>
<tr>
<td>2.5 Discussion of Results</td>
<td>54</td>
</tr>
<tr>
<td>2.6 Policy Implications</td>
<td>61</td>
</tr>
<tr>
<td>2.7 Conclusions</td>
<td>63</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 3</th>
<th>Pricing of High-Tech Software Startups: Art or Science?</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Introduction</td>
<td>66</td>
</tr>
<tr>
<td>3.2 Background on High-Tech Startups</td>
<td>68</td>
</tr>
<tr>
<td>3.3 Data and Empirical Strategy</td>
<td>71</td>
</tr>
<tr>
<td>3.4 Results</td>
<td>78</td>
</tr>
<tr>
<td>3.5 Conclusions</td>
<td>81</td>
</tr>
<tr>
<td>Table 3: OLS Regression Results</td>
<td>83</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 4</th>
<th>Disaggregation of the Software Value Chain</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Introduction</td>
<td>84</td>
</tr>
<tr>
<td>4.2 Downstream Disaggregation: Long-Term Contracts</td>
<td>87</td>
</tr>
<tr>
<td>4.3 Upstream Disaggregation: Spot Market for Technologies</td>
<td>91</td>
</tr>
<tr>
<td>4.4 Conclusions</td>
<td>93</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bibliography</th>
<th></th>
</tr>
</thead>
</table>
Introduction

This thesis examines three separate industries—health care, retail banking, and software—that share a common characteristic: they are all undergoing a tremendous amount of organizational change. The health sector in the United States has been redefining itself since 1993 with the advent of managed care organizations. There has been a discernible shift in the balance of power away from physicians towards the case managers who run these new health maintenance organizations more like traditional businesses. In light of these changes, the question of how physicians respond to financial incentives remains paramount. The retail banking industry has just completed a wave of consolidation among the regional ATM networks—the upstream providers for access to non-proprietary ATM machines. Over the last few months, the issue of “double charging” (or surcharging) has taken center stage where both the bank which a consumer has an account with and the bank who’s ATM the consumer is using for a transaction separately charge the consumer for using the ATM. In the software technology realm, the pace of innovation has skyrocketed with the arrival of Internet and Object-Oriented technologies. Information technology has become primary in the agendas of both large and medium sized companies due to its potential to impact both the supply and demand side of a firm’s business. As a result of this pace of innovation, we are witnessing an industry-wide organizational response best described as disaggregation of the software value chain.

Chapter one represents a contribution to a debate health economists and policymakers have engaged in over the last thirty-five years: to what extent to physicians respond to financial incentives? This issue has taken on additional significance over the
last five years with the changes in financial incentives faced by more physicians. The traditional incentive structure under fee-for-service insurance where physicians receive payment for all services rendered has been replaced by schemes like capitation. Under this new payment structure implemented by health maintenance organizations (HMOs) and other managed care organizations, physicians are rewarded monetarily for limiting the intensity of the interventions used on their patients. To shed some light on this issue, I examine the particular case of cesarean versus natural child delivery faced by an ob/gyn. By taking advantage of the significant variation across time and states in the differential level of reimbursement determined by Medicaid, I am able to use changes in fee differentials to identify treatment intensity responses. The conclusion reached here is that ob/gyn's do respond significantly to changes in the fee environments. From a welfare perspective, this finding suggests that at the least, physicians may have the propensity to compromise their clinical integrity for financial gain. A more pessimistic view would question more fundamentally the wisdom behind the recent changes made to the standard fee-for-service incentives.

Chapter two looks at the pricing of network ATM access from a game-theoretic perspective. To consumers, ATM access has become a necessary part of the service offering rendered by their retail bank. In response to this, in most areas of the country retail banks have come together to form an entity called shared ATM networks. These networks sell access to non-proprietary ATMs to the various member banks in the network who in turn typically resell it as a component of the n-dimensional vector of services provided to end-consumers. The interesting element of this industry is the fact that in many of these dominant regional ATM networks, the member banks are themselves the residual claimant
to the profits of the network. Given the payment scheme instituted each time a consumer from bank \( x \) performs an ATM transaction at bank \( y \), an interesting incentive problem arises. The shareholding banks in the network set the upstream terms of transaction between bank \( x \), bank \( y \), and the network itself while also competing downstream as a bank reselling ATM access with its traditional service offering. The equilibrium in this model indicates that when the shareholders of the network are composed of banks with large proprietary networks (as opposed to banks with small proprietary networks of ATM machines), the upstream access price charged to member banks is relatively higher while the downstream price charged to consumers relatively lower. From a static equilibrium perspective, there appears to be a divergence in the policy one would pursue to intervene on behalf of consumers versus intervention on behalf of the smaller community bank. I deal with this and other issues in greater depth in chapter two.

Chapters three and four explore issues in the software industry: the financing of software technology ventures and the industry's organization response to the rapid pace of innovation. Information technology has taken on a tremendous amount of significance in all sectors of our economy because of its ability to provide fundamental product-side and cost-side advantages over the competition. However, early adopters are often times purchasing unproven technologies which may not necessarily work well and perhaps even hinder the overall productivity of a firm. The driving force behind this rapid innovation is centered not on the established industry players but more so in the startup environments in Silicon Valley, California and Rte. 128, Massachusetts. The financing for these ventures is executed predominantly by the venture capitalist community, and I examine the factors a venture capitalist take into account when determining the price of these pre-revenue stage
companies. Using a somewhat limited data-set, I find that prior experience of management, investor bullishness on the technology sector in general, and the product stage of a company go a long way towards determining the valuations of these early-stage deals.

These three industries present some of the most interesting and policy-relevant issues for economists today. The welfare of the masses is very much affected by physician response to financial incentives, the information technology revolution, and access to shared ATM networks. A casual glance on any given day of the leading newspapers and magazines signals the profound impact these industries will likely have well into the next decade. These industries should continue to provide an arena not only for academic study by economists, but a setting for us to improve the welfare of society by using the tools of analysis at our disposal.
Chapter 1
Physician Fees and Procedure Intensity: The Case of Cesarean Delivery

1.1 Introduction

A major topic of public policy concern in the U.S. in recent years has been the rapid growth in the expenditures of the major public insurance programs, Medicaid and Medicare. Medicaid, which provides health insurance to low income women and children, disabled, and elderly, has grown by almost 400% over the past decade, with total state and federal expenditures of $156 billion in 1995. Medicare, which provides universal coverage to the elderly, has grown by 250%, with expenditures of $180 billion in 1995. These two programs now amount to 18% of the federal budget, and have accounted for 30% of the growth in the budget over the past decade\(^1\). As a result, there is enormous pressure to reign in program costs. But doing so through channels such as restricted eligibility or reduced beneficiary benefits has run into considerable political difficulty. Instead, much of the attention to cost control in these programs has focused on provider reimbursement, for example through reducing physician fees.

In theory, cutting fees to physicians is a simple mechanism for controlling costs. In practice, behavioral responses by physicians renders this a much more difficult proposition than simple static analysis would imply. On the one hand, there is concern that lowering fees might lead to deterioration in quality of care for the publicly insured. On the other hand, there is also concern about physician "induced demand", whereby providers may

---

\(^1\) U.S. Congress, Committee on Ways and Means (1996).
make up a large share of fee reductions through increased volume, limiting any cost
savings from reductions in fees.

In fact, there are enormous health care literatures on the consequences of fee changes
in these two programs. But these literatures have had quite different focuses. The
literature on Medicare has focused on the effect of fee changes on the quality of care, and it
has largely concluded that lowering fees will raise treatment intensity; as a result, estimates
of the savings to the government from fee reductions include an offset for induced demand
responses. (Christensen (1992) and Rice (1989)). The literature on Medicaid, on the other
hand, has focused on patient access, assessing whether reductions in fees may lead
providers to avoid potential Medicaid patients. But there has been remarkably little work
on the effect of reduced Medicaid fees on the quality of care of Medicaid patients,
conditional on gaining access to the system.

It is therefore tempting to apply the induced demand findings of the Medicare
literature to estimate the effects of fee changes in the Medicaid program. In fact, however,
there is reason to believe that the evidence from the Medicare program may not apply to
the case of Medicaid. The reason emerges from the basic economics of physician
behavior. In a standard model such as that of McGuire and Pauly (1987), fee changes will
have both income and substitution effects on physicians. As McGuire and Pauly point out,
the induced demand hypothesis is simply the implication of a model where income effects
dominate substitution effects. This may be a relevant concern in the Medicare program,
since surgery on Medicare patients is primarily performed by physicians who specialize in
the Medicare population. But for physicians treating Medicaid patients, those patients will
generally be a smaller share of their patient pool, so that income effects will be much less
important. In this context, it is therefore possible that substitution effects will dominate, so that lowering fees will lower treatment intensity.

We investigate this issue within the context of a particular example, cesarean section delivery of childbirth. Cesarean delivery is a particularly useful example because the underlying costs of the procedure in terms of physician time and intensity are considered to be similar to the alternative, vaginal birth (Hsiao, 1988), yet reimbursement has traditionally been higher. Moreover, there is tremendous variation across state Medicaid programs in the differential reimbursement of cesarean delivery. In this paper we consider the effect of variations in this differential on the likelihood of cesarean delivery for the Medicaid population.

In particular, we use discharge abstract data from 11 states over the 1988-1992 period to model the effect of the cesarean reimbursement differential on substitution towards cesarean delivery. We match to this data information on Medicaid fee differentials, and we estimate the effect of these differentials on the mode of delivery for childbirth. We find that, on net, there is a strong positive effect of fee differentials on use of cesarean delivery, in contrast to the induced demand findings of the Medicare literature. We estimate that a $100 increase in the absolute differential between cesarean and vaginal delivery reimbursement would increase the cesarean delivery rate by .7 percentage points. This finding suggests that cutting reimbursement for cesarean delivery under the Medicaid lowers the intensity of treatment of childbirth.

Our paper proceeds as follows. Section 1.2 sets up a simple model of physician behavior which provides the framework for integrating the previous views. Section 1.3
reviews the previous literature on physician behavior. Section 1.4 presents our data and estimation strategy, and our results are presented in section 1.5. Section 1.6 concludes.

1.2 Theory

We begin our analysis with a brief theoretical formulation of the physician intensity decision. Our analysis closely follows the models of McGuire and Pauly (1991) and Gruber and Owings (1996). Our goal here is to highlight the ambiguous prediction for the effect of fee changes on treatment intensity, and to focus on the relative role of income and substitution effects.

The physician obtains utility from income, but disutility from inducement. His separable utility function has the form:

\[ U(Y, I) = U(Y) + U(I) \]  

(i)

where \( Y \) is income and \( I \) is total inducement. We make the standard assumptions on the physician's utility function: \( U_y > 0, U_I < 0, U_{yy}, U_{II} < 0 \). \( Y \) is defined as

\[ Y = B\{1 - a(i)\}Y_n + Ba(i)Y_c \]  

(ii)

\[ Y = BY_n + Ba(i)m \]

where \( B \) is the total number of births, \( a(i) \) is the share of total deliveries which are by cesarean as a function of inducement per birth \( i \), and \( m \) is equal to the difference between the income from a cesarean versus the income from a natural delivery, \( m = Y_c - Y_n \). \( m \) is a non-negative number throughout our sample period.

We posit \( a'(i) > 0 \) and \( a''(i) = 0 \). Since inducement has no natural units, we define the \( a(i) \) as linear in \( i \). We can think of \( a(0) \) as equal to some \( p \) which has the interpretation
as cesarean rate if there were no inducement at all, i.e. the clinically necessary cesarean rate. \( I \), defined as total inducement, is equal to

\[
I = Bi,
\]

(total number of births times inducement per birth).

If we maximize equation (i), the first-order-conditions with respect to inducement per birth, \( i \), gives us the following:

\[
\frac{dU}{di} = U_y a'(i^*) m + U_i = 0
\]  

(iv)

\( i^* \) being the optimally chosen level of inducement per birth. If we fully differentiate the first-order condition with respect to \( m \), the difference in cesarean and natural delivery reimbursement, we get the following result:

\[
\frac{di^*}{dm} = \frac{-U_{yy}aa'(i^*)m - U_y \frac{a'(i^*)}{B}}{U_{yy}(a'(i^*)m)^2 + U_m}
\]  

(v)

Since the second derivatives with respect to both income and inducement are negative, the denominator is strictly less than zero. The first term on the numerator above is positive, and the second term is negative. Hence, the sign of the equation above is ambiguous. However, upon taking a closer look, equation (v) clearly separates the income and substitution effects associated with the fee change. The \( U_y \) term in the numerator represents the substitution effect while the \( U_{yy} \) term represents the income effect. When substitution effects dominate, then \( \frac{di^*}{dm} > 0 \), and an increase in the differential leads to greater inducement, and hence more cesareans. When income effects dominate \( \frac{di^*}{dm} < 0 \), and...
an increase in the fee differential leads to fewer cesareans. This basic result provides the theoretical underpinning for much of what follows.

1.3 Previous Literature

**Physician Fees and Treatment Intensity**

There is a sizeable health economics literature on the effects of fee changes on procedure intensity\(^2\). Much of this work has centered on a "natural experiment" from Colorado in 1977. Prior to that year, prevailing Medicare charges were calculated separately in each of the ten areas of Colorado, and were substantially higher in urban areas of the state. In 1977, the payment system was changed so that charges for calculated for the state as a whole, resulting in a large increase in fees for nonurban physicians, with a freeze for urban physicians. In a series of articles, Rice (1983, 1984) found a strong negative relationship between fee levels and treatment intensity, for surgical, medical, and ancillary services. This finding was confirmed by Christensen (1992), who re-analyzed the Colorado data and estimated an overall elasticity of -0.5 for volume with respect to fees. Thus, the evidence from the Colorado experience is strongly supportive of a backward bending supply curve for physician services.

Other evidence provides a more mixed picture. Escarce (1993) and Yip (1994) studied the effect of reduced reimbursement for "overpriced" medical procedures under Medicare through the 1987 OBRA legislation. Escarce finds no clear pattern across surgical specialties in the relationship between fees and treatment intensity. Yip performs

---

\(^2\) This is only one strand of the general "induced demand" literature. There are a number of other studies which consider the effect of direct changes in demand for physician services, for example through changes in surgeon density. See, for example, Fuchs (1978) or Cromwell and Mitchell (1986); for criticisms of this approach, see Phelps (1986) or Dranove and Wehner (1996).
a detailed study across physician subgroups and finds a mixture of results depending on the particular specialty group. For example, in the market for coronary artery bypass grafting (CABG), the large reduction in Medicare fees for thoracic surgeons led not only to a large increase in CABGs performed in the Medicare population, but there were significant spillover effects into the private sector leading to a higher CABG rate among the privately insured. However, in markets involving other specialty groups—general surgeons, orthopedic surgeons, and urologists—she does not find this same strong within sector income effect or additional spillover effect as she finds in thoracic surgeon population. This mixed evidence is also reflected in Hurley, Labelle, and Rice’s (1990) study of physician fees and treatment intensity in Ontario. They examine physician responses to fee changes for 28 procedures, and find significant physician responses in 13 of them.

Another source of evidence is fee controls, in the U.S. and abroad. Wedig, Mitchell and Cromwell (1989) examine the quantity response of physicians in Alabama, Connecticut, Washington and Wisconsin to the Medicare fee freeze from July 1984 to May 1986. During this time interval, Medicare payments as well as the amount physicians were permitted to balance-bill patients were also frozen at July 1984 levels. They find that the dollar value of services per Medicare beneficiary increased about $24 from Q3 1984 to Q4 1986 (a 21% increase) while physicians per capita increased only 1.6%. Taken together, these figures imply that quantity per physician rose nearly 20% during the fee freeze. Such rapid output growth per physician under the fee control is consistent with the target income hypothesis. Rice and Labelle (1989) perform a similar analysis on fee controls for nine Canadian provinces over the period from 1975 to 1985. In four of the nine provinces, real unit fees decreased. However, activity per physician increased substantially in those four
provinces, such that only two of the four experienced a decline in total billings. The other two experienced quantity increases which more than offset the reduction in fees. Rice and Labelle interpret this finding that physicians can, and do, generate demand in response to real fee reductions in order to maintain their pre-reduction level of income. Feldman and Sloan (1988) also look at Canadian data (but from an earlier period) and find contrasting evidence. They arrive at what they consider conclusive evidence that fee controls worked. Fees were controlled under the universal health insurance plans introduced during the late 1960s and early 1970s. National Physician expenditures as a percentage of GNP was 1.21 percent in 1970, 1.16 percent in 1975 and 1.10 percent in 1980. Overall, Feldman and Sloan argue that there is little evidence for demand inducement in primary care physician services, and if demand inducement does indeed exist in markets for surgical services, it is less severe than previously thought.

This literature has focused primarily on the Medicare program, or on national fee controls. There has been little work on the effect of Medicaid fees on treatment intensity. Most of the work on Medicaid has focused on the effects of fees on access to care, through inducing provider participation in the program. Beginning with the work of Sloan, Mitchell, and Cromwell (1978) and Hadley (1979), and continuing through Held and Holahan (1985), Mitchell (1991), and Adams (1994), many researchers documented a strong correlation between higher (relative to Medicare or private payers) Medicaid fees and access to care. But the only paper of which we are aware that directly focuses on the relationship between Medicaid fees and treatment intensity is Decker (1994), who finds that physicians spend more time with their Medicaid patients when fees increase.

---

3 Although recent work has found a significantly weaker relationship; see Decker (1994) or Baker and Royalty (1996).
Can the results from the Medicare program, or from national fee controls, necessarily be applied to the case of Medicaid? We would argue that they may not be applicable, for a simple reason: physicians are much less specialized in Medicaid than they are in Medicare (and certainly less than in the case of national fee controls, which affect all treatment).

According to our data, the median physician who performs a surgical procedure on a Medicare patient derives over 60% of his caseload from Medicare. In contrast, for the median physicians who delivers Medicaid babies, only 32% of his caseload comes from the Medicaid program.\(^4\) Thus while a single-payer model may be a reasonable way to model the effects of Medicare fee changes (since physician’s who perform Medicare procedures work primarily in the Medicare sector), modeling Medicaid fee changes requires a multiple-payer framework. MacGuire and Pauly introduce an additional payer into their single-payer model, and the effect of a fee reduction from payer \(x\) on the intensity of treatment in sector \(x\) is largely similar to the single-payer model. However, the distinction is the following: To the extent that income from sector \(x\) represents smaller and smaller fraction of total income, the income effect becomes less potent. Thus in general the likelihood of physicians responding to fee reductions from a given payer \(x\) with increased quantity of services to that sector \(x\) is mitigated as the share of total income coming from payer \(x\) declines. Instead, the physician is more likely to look at other income sources to recoup his lost profits. Given these facts, it may be that while income effects dominate in the case of Medicare, substitution effects dominate in the case of Medicaid. As a result, even as Medicare fee reductions lead to increased intensity of treatment, Medicaid fee

\(^4\) These calculations are based on the roughly 20% subsample of observations in our data for whom physician identifier data are available. This may not be a random subsample, and we are unable in some cases to distinguish individual physicians from physician groups, so these figures should be taken as suggestive only.
reductions may lead to reduced intensity, since income effects are small on average for physicians delivering Medicaid infants.

*Financial Incentives and Cesarean Delivery*

There is a separate, and sizeable, literature that focuses specifically on the cesarean section delivery decision. The first type of study focuses on differences in cesarean utilization across payer types. Historically, the financial incentive to perform a cesarean has been greatest among the privately insured, less so among the publicly insured, and least prevalent among the uninsured. The 1989 average differential in fees for was $560 for the privately insured and $127 for Medicaid population. As for the uninsured, the likelihood that the physician bill will not be paid at all whether a cesarean is performed or not renders an almost negligible financial bias towards cesareans. Stafford (1990) exploits this variation and examines cesarean rates by payment source in a sample of California hospitals. He finds results that are consistent with the financial incentives outlined above; however, the fact that there may be omitted characteristics by payer class which are correlated with the cesarean decision casts some doubt on a causal interpretation (i.e. privately insured women tend to be older). Other studies have looked to declining fertility rates and increased ob/gyn density as possible explanation for higher incidence of cesarean delivery. Tussing and Wojdowycz (1992), examining New York data, finds no correlation between ob/gyn density and cesarean delivery rates. Again, such a study suffers from the possibility that omitted regional differences are correlated both with higher ob/gyn density and cesarean utilization. Gruber and Owings (1996) study the effect of declining fertility in the United States. Fertility fell by 13.5 percent during the 1970-1982 period and cesarean
utilization increased by over 240% in the same period. They posit that this decline increased the income pressure on physicians leading them to substitute the more highly reimbursed cesarean delivery. Their primary result is that a within state 10% decline in the fertility rate is associate with a within state .97 percentage point increase in the cesarean rate. While their results are significant, only 16-32% of the total increase in cesarean usage can be explained by this income shock.

Recent work by Keeler and Fok (1996) is closely related to our work. They study the impact of an insurance reform under California Blue Cross that equalized fees for vaginal and cesarean delivery, a relative decline in cesarean fees of 21%. Using data from before and after the reform, the authors find only a modest 0.7% reduction in cesarean delivery rates, which appears to indicate little intensity response to fee changes. The authors report that this reflects similar experiences with fee changes under Blue Cross plans elsewhere.

There are a number of reasons why these results do not necessarily have predictive power for the effects of Medicaid fee changes. Our reasoning is analogous to the argument set forth earlier on why physician responses to Medicaid fee changes may go in the opposite direction from Medicare fee changes. The primary difference in the two samples is that the average physician in Keeler and Fok's sample may be getting a substantial proportion of his income from the Health Plan which implemented the fee changes. However, the average ob/gyn in our sample receives a relatively small proportion of his income from Medicaid births. Quantity responses to fee changes in sector $x$ attributable to income effects are attenuated (and substitution effects subsequently augmented) in sector $x$ if the proportion of income coming from sector $x$ is small. In addition, the equalization of fees in their study corresponds to a $1000 decline in the differential for cesarean versus
natural delivery. Hence, the modest .7% decline discovered by Keeler and Fok could simply be a "canceling out" of income and substitution effects (see equation 5). In contrast we would expect the substitution effect to be the dominant term for a given change in Medicaid fee incentives.

1.4 Data and Empirical Strategy

Our primary source of data for this analysis is the Healthcare Access and Utilization Project (HCUP) data, collected by the Agency for Health Care Policy and Research (AHCPR). This survey collects information from a random sample of discharges across eleven states, over the 1988-1992 period.\(^5\) Hospitals are sampled randomly in these states, along five strata, to represent the universe of community-based hospitals in the U.S. The strata are: geographic region; ownership; location; teaching status; and bed size. For each discharge, the survey collects information on source of payment, patient demographics, hospital characteristics, diagnoses, and procedures. Our sample is all diagnoses with a diagnosis of childbirth.

We match to these data information on Medicaid reimbursement of vaginal and cesarean childbirth. These data come from several sources. For 1988 and 1992, we use data provided by the American College of Obstetricians and Gynecologists (ACOG). For

---

\(^5\) The states are: Arizona, California, Colorado, Florida, Illinois, Iowa, Massachusetts, New Jersey, Pennsylvania, Washington, and Wisconsin. In 1988, data are only available for eight of the states (Arizona, Pennsylvania, and Wisconsin are not included). Beginning in 1993, the sample was expanded to ??? states. We do not include Arizona in our analysis, since they do not have a traditional Medicaid program.
1989, we use data from PPRC (1991); we also use these data to supplement missing values in 1988. For 1990, we use data from Holahan (1993), supplemented by additional information from ACOG. For 1991, we use data from Singh et al. (1993).

We use these data to run logit regressions of the form:

$$CSEC_{it} = f(a + \beta_jFEEDIFF_{it} + \beta_2X_{it} + \beta_3Z_{it} + \beta_d d + \beta_t t + \epsilon_{it})$$  \hspace{1cm} (vi)

where $i$ indexes individuals, $j$ indexes hospitals, $s$ indexes states, and $t$ indexes years

$CSEC$ is a dummy for cesarean section delivery

$FEEDIFF$ is the differential for cesarean delivery (relative to vaginal)

$X$ is a set of individual characteristics

$Z$ is a set of hospital characteristics

$d_i$ is a complete set of state dummies

$t_i$ is a complete set of year dummies

In this model, we estimate the effect of fee differentials on the odds of cesarean delivery. The fee differential is measured in two ways. First, we simply use the dollar differential between the cesarean and vaginal reimbursement rates ($m$ in our model above). While this follows naturally from the theory presented in Part I, there is a drawback in that it may not be properly capturing the financial incentive to physicians. The reason is that the theory as presented assumes that physicians operate in a single-payer domain.

However, as aforementioned the average ob/gyn receives the majority of his income from non-Medicaid patients. Hence the extent to which physicians will be willing to trade off the disutility of inducement for the utility of income will depend on their alternative

---

* We are grateful to Susheela Singh for providing us with some unpublished data from that study.
income earning prospects. Following the logic of McGuire and Pauly’s two payer model, the incentive to respond to fee changes in sector $x$ with quantity inducement in sector $x$ is dependent on the prospects of recouping that income from the alternative sector. As a result, a given dollar increment may provide less incentive for Medicaid inducement in areas with very high reimbursement for privately insured vaginal delivery, since physicians can substitute his time out of Medicaid births into privately insured births. We therefore re-estimate the model using the fee differential divided by the reimbursement for a privately insured vaginal delivery as our index for the financial incentive.

We also control in the estimation for a variety of factors which potentially determine the cesarean decision. We include a vector of characteristics of the woman: dummies for age groups (20-24; 25-29; 30-34; 35-39; 40-44; 45+); dummies for white, black, or neither; and dummies for median income in the zip code of residence. There are eight possible categories for median income: 0-$15,000, $15,000-$20,000, $20,000-$25,000,... $40,000-$45,000, $45,000+. We also control for characteristics of the hospital in which the woman gave birth: number of discharges; type of ownership (public or for-profit, with non-profit being the excluded group); and teaching status (whether the hospital is a member of the Council of Teaching Hospitals).

In addition, we include a full set of year dummies, to control for underlying trends in the cesarean delivery rate. And we include a full set of state dummies as well, to control for secular differences in the rate of cesarean delivery across the states in our sample.

With state dummies included in the model, our estimates are identified by within-state changes in fees over time. There are large within state changes in cesarean fee differentials

---

7 For both race and income, the excluded group is those with missing data: 46.1% of our sample does not have recorded information on race, and 8% does not have recorded information on income.
during our sample period: California’s absolute differential, for example, fell from $306 in 1988 to 0 by 1990; Colorado’s differential rose from $72 in 1988 to $413 in 1990; and Illinois’ differential rose from $75 in 1989 to $207 in 1992. These large changes allow us to identify the effect of fees on cesarean delivery rates, holding constant underlying differences across the states.

The means of our data set are presented in Table 1A. The cesarean delivery rate was 17.8% for our sample period in contrast to 22.5% in the private sector. The mean age for a patient was a 24.3 years in contrast to 29.3 for a private sector patient. Roughly ten percent of the women had a previous cesarean, and the various measures of “complicated birth” are each present less than 10% of the time. Most (62%) deliveries occurred in private, non-profit hospitals with very few coming in private, for-profit. Our data is reasonably evenly distributed across the time period, but over 60% is concentrated in two states: California and Florida. Since our model is identified off within state variation in fees across time, it is good that California is well represented since it experienced a large reduction in fee differential- from $306 in 1988 down to $0 in 1990. Unfortunately, Florida had zero fee differential across the whole period.

Table 1B shows the Medicaid fee differential between cesarean and natural delivery across states from 1988-1992. There is a substantial amount of variation both across states and time- it reaches as high as $440 in Colorado for 1991, to as little as $0 for Florida throughout our sample period (other state-year pairs also exhibit zero differential). All states with the exception of Florida and Pennsylvania exhibited non-trivial variation across time. In particular, California and New Jersey all went from $0 difference in some years to several hundred-dollar difference in others. For the nation as a whole, the 1989 private fee
differential was $561 on average versus $127 for the Medicaid program—this $434
difference amounts to 29% of the average private vaginal fee in 1989.

Beneath it Table 1C shows the fee differential divided by the private vaginal
delivery reimbursement. As aforementioned, the purpose here is to show that the
"magnitude" of the financial incentive goes beyond simple dollar differences for Medicaid
deliveries and takes into account alternative means an ob/gyn has to recoup income lost
from Medicaid patients. The distinction between these two measures of financial incentives
is apparent when looking at California and Iowa data for 1988. The absolute fee difference
is $50 higher in California, which might lead one to believe the financial incentive is
stronger in California for cesarean. However, when divided by the private fee, the Iowa
measure turns out to be more than twice as large (.35 versus .15). Hence the incentive to
induce may be stronger in Iowa because of less promise in the alternative sector.

1.5 Basic Results

Our basic findings are shown in Table 2, which presents logit estimation of
equation (vi). In the second row of each column, we interpret the logit coefficient by
showing the effect of a $100 increase in fees, or a 10% rise relative to private fees for
vaginal delivery.

In contrast to the Medicare literature, we find a highly significant positive coefficient
on the fee differential, either specified as absolute dollars, or normalized by the private fee.
Specified as absolute dollars, we find that each $100 dollar increase in fees leads to a 0.7
percentage point, or 3.9%, rise in cesarean delivery rates. Normalize by private fees, we
find that each 10% rise relative to the private vaginal delivery fee leads to a 1.5 percentage point, or 8.4%, rise in cesarean delivery rates.

As noted earlier, in 1989, the national average private fee differential for cesarean delivery was $561 on average, as compared to $127 for the Medicaid program; this $434 differential amounts to 29% of the private vaginal fee in 1989. Our results therefore imply that if the Medicaid program raised its fee differential to the private level, there would be a rise in cesarean delivery rates of 3.04 percentage points (17.1%) to 4.35 percentage points (24.4%). As noted earlier, the cesarean delivery rate for private patients in our sample is 5.7 percentage points higher than the Medicaid rate. Thus, our findings suggest that raising the Medicaid fee differential to the private sector level would reduce the gap in cesarean delivery rates by between 53% and 76%. That is, the majority of the difference in cesarean delivery rates between private sector and Medicaid patients can be explained by fee differentials.

The control variables are generally significant as well. We find a strong pattern of increasing cesarean delivery rates with age. C-section rates are lower for all those with reported race relative to those with unreported race (the omitted race category), but they are highest for whites, lower for blacks, and even lower for other non-whites. Cesarean delivery rates are lower at teaching hospitals; this is consistent with evidence from New York State in Tussing and Wojtowycz (1986). They are also higher at for-profit hospitals, relative to non-profits, which is consistent with the findings of a number of papers, including Gruber and Owings (1996). Surprisingly, however, they are lower at large hospitals and higher at government hospitals; both of these findings are at odds with the previous literature. One reason may be that previous estimates have either used the
privately insured, or all patients from an era when Medicaid represented a smaller share of births. Factors which cause high cesarean rates among the privately insured population may not apply to the Medicaid population.

We find that the odds of cesarean delivery rise with median income of the patient’s zip code as well. This may be driven by the fact that physicians who operate in wealthier areas have an increased predilection to conducting cesareans. There are generally more privately insured patients in the more affluent zip codes, and the higher cesarean rate among the privately insured may spill over to greater cesarean incidence among Medicaid mothers. Finally, there is a very strong correlation between cesarean delivery and previous cesarean birth, as in previous work. It is common medical practice to have subsequent deliveries to a cesarean birth also be cesarean.

Alternative Explanation: Endogenous Fees

We find these results to provide striking evidence that higher fee differentials under Medicaid lead to higher cesarean delivery rates. An alternative explanation for this finding, however, is that higher fee differentials are chosen by states when there is high demand for cesarean delivery. This argument loses much of its force once we include state fixed effects in the regression, since it would imply that Medicaid programs react to changes in cesarean demand. Moreover, we have included a rich set of individual characteristics which would capture any demographic or economic shifts correlated with demand for cesareans, for example through the aging of the Medicaid population or changes in the income distribution.

But one natural channel through which this response could occur is fetal or maternal health. There is a clear correlation between cesarean delivery and underlying fetal and
maternal health (Gruber and Owings, 1996). It is also possible, although admittedly not likely, that a deterioration in fetal or maternal health which gave rise to more cesarean deliveries could motivate state policymakers to raise the cesarean differential as well.

In order to control for this possibility, we include in the regression three factors which further control the underlying riskiness of the birth: breech presentation, fetal distress and maternal distress. In the United States, more than one in seven women experience complications during labor and delivery arising from either conditions existing prior to pregnancy (i.e. diabetes, hypertension, or infectious diseases) or pathological conditions which develop during pregnancy (i.e. eclampsia or placenta praevia). These problems can be life threatening to the mother and baby, thus cesarean delivery is clinically the safest solution. As column 3 and 4 in table 2 show; however, the results on our measures of financial incentives are further strengthened when these additional covariates are inserted.

The same $100 increase in the absolute fee differential leads to a full 1% rise in cesarean delivery rates (as opposed to the earlier .7% rise for a $100 increase). Likewise, a 10% rise relative to the private vaginal fee leads to a 1.9% increase in cesarean delivery rates (as opposed to the earlier 1.5% rise for a $100 increase). The coefficients themselves on these regressors are also large in magnitude and highly significant. The logit interpretation of the coefficients indicate that the presence of breech presentation increases the likelihood of cesarean by 54%, maternal distress increases likelihood by 15% and fetal distress by 27%.

We feel this provides sufficient evidence against the notion that endogenous fees are driving our results.

1.6 Conclusions
Provider fee policy remains the tool of choice for policy-makers in trying to reign in program costs. As a result, the effect of fees on treatment intensity has received enormous attention in the context of the Medicare literature; but there has been scant attention paid to this question in the context of Medicaid. This is unfortunate, because as we have argued the results for Medicare cannot be naively applied to the case of Medicaid. If the induced demand type results that have been found for Medicare are due to physician income effects, they may apply much less strongly in the Medicaid environment, where a smaller share of a physician's caseload comes from the public insurance program.

Indeed, in contrast to much of the Medicare literature on fees and treatment intensity, we find a positive relationship between the fee differential for cesarean delivery and the rate of cesarean delivery. Our findings suggest that this fee effect is sufficiently large to explain over half, and up to three-quarters, of the differential cesarean delivery rate between the (more highly reimbursed) private sector and Medicaid. One explanation for our finding, relative to the earlier work on fee changes under Medicare in Colorado, is physician income effects. But the Medicare literature has been mixed, and there are a host of differences between our example and the cases studied for Medicare. A resolution of the source of difference between these papers awaits further study.

The welfare implications of our findings are unclear. We have demonstrated that lowering Medicaid fees will lower the rate of cesarean delivery, but we have not provided evidence on whether the marginal cesarean deliveries that are foregone are beneficial for mothers and infants. In fact, the findings of Currie, Gruber, and Fischer (1995) indicate that raising physician fees under Medicaid causes a significant reduction in infant mortality rates. It is difficult, however, to separate the mechanism through which fees are affecting
infant outcomes. Future work on this topic could usefully explore whether the changes in treatment intensity uncovered here have real effects on patient outcomes.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cesarean Delivery Rate</td>
<td>0.178</td>
</tr>
<tr>
<td># of discharges</td>
<td>22520.810</td>
</tr>
<tr>
<td># of beds</td>
<td>458.310</td>
</tr>
<tr>
<td>Public Hospital</td>
<td>0.327</td>
</tr>
<tr>
<td>Private Non-profit Hospital</td>
<td>0.620</td>
</tr>
<tr>
<td>Private For-profit Hospital</td>
<td>0.053</td>
</tr>
<tr>
<td>Teaching Hospital</td>
<td>0.343</td>
</tr>
<tr>
<td>White</td>
<td>0.201</td>
</tr>
<tr>
<td>Black</td>
<td>0.085</td>
</tr>
<tr>
<td>Race Missing</td>
<td>0.461</td>
</tr>
<tr>
<td>Not White or Black</td>
<td>0.252</td>
</tr>
<tr>
<td>Median Income $0-$15,000</td>
<td>0.053</td>
</tr>
<tr>
<td>Median Income $15-$20,000</td>
<td>0.173</td>
</tr>
<tr>
<td>Median Income $20-$25,000</td>
<td>0.257</td>
</tr>
<tr>
<td>Median Income $25-$30,000</td>
<td>0.207</td>
</tr>
<tr>
<td>Median Income $30-$35,000</td>
<td>0.118</td>
</tr>
<tr>
<td>Median Income $35-$40,000</td>
<td>0.065</td>
</tr>
<tr>
<td>Median Income $40-$45,000</td>
<td>0.026</td>
</tr>
<tr>
<td>Median Income $45,000+</td>
<td>0.021</td>
</tr>
<tr>
<td>income missing</td>
<td>0.080</td>
</tr>
<tr>
<td>Age &lt;20</td>
<td>0.246</td>
</tr>
<tr>
<td>Age 20-25</td>
<td>0.371</td>
</tr>
<tr>
<td>Age 25-30</td>
<td>0.222</td>
</tr>
<tr>
<td>Age 30-35</td>
<td>0.111</td>
</tr>
<tr>
<td>Age 35-40</td>
<td>0.041</td>
</tr>
<tr>
<td>Age 40-45</td>
<td>0.008</td>
</tr>
<tr>
<td>Age 45+</td>
<td>0.001</td>
</tr>
<tr>
<td>Year 92</td>
<td>0.248</td>
</tr>
<tr>
<td>Year 91</td>
<td>0.222</td>
</tr>
<tr>
<td>Year 90</td>
<td>0.225</td>
</tr>
<tr>
<td>Year 89</td>
<td>0.170</td>
</tr>
<tr>
<td>Year 88</td>
<td>0.135</td>
</tr>
<tr>
<td>Arizona</td>
<td>0.021</td>
</tr>
<tr>
<td>California</td>
<td>0.331</td>
</tr>
<tr>
<td>Colorado</td>
<td>0.022</td>
</tr>
<tr>
<td>Florida</td>
<td>0.283</td>
</tr>
<tr>
<td>Iowa</td>
<td>0.051</td>
</tr>
<tr>
<td>Illinois</td>
<td>0.099</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>0.037</td>
</tr>
<tr>
<td>New Jersey</td>
<td>0.025</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>0.044</td>
</tr>
<tr>
<td>Washington</td>
<td>0.027</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>0.059</td>
</tr>
<tr>
<td>Previous Cesarean</td>
<td>0.093</td>
</tr>
<tr>
<td>Breech Presentation</td>
<td>0.029</td>
</tr>
<tr>
<td>Maternal Distress</td>
<td>0.0001</td>
</tr>
<tr>
<td>Fetal Distress</td>
<td>0.094</td>
</tr>
<tr>
<td>Table 1B</td>
<td>Medicaid Fee Difference</td>
</tr>
<tr>
<td>----------</td>
<td>-------------------------</td>
</tr>
<tr>
<td></td>
<td>(Cesarean Fee) - (Vaginal Fee)</td>
</tr>
<tr>
<td>CA</td>
<td>305.6</td>
</tr>
<tr>
<td>CO</td>
<td>72</td>
</tr>
<tr>
<td>FL</td>
<td>0</td>
</tr>
<tr>
<td>IA</td>
<td>253.66</td>
</tr>
<tr>
<td>IL</td>
<td>75</td>
</tr>
<tr>
<td>NJ</td>
<td>0</td>
</tr>
<tr>
<td>PA</td>
<td>146.5</td>
</tr>
<tr>
<td>WA</td>
<td>61.82</td>
</tr>
<tr>
<td>WI</td>
<td>180.31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 1C</th>
<th>Medicaid Fee Difference Divided by Private Vaginal Fee</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(((Cesarean Fee) - (Vaginal Fee))/ (Private Vaginal Fee))</td>
</tr>
<tr>
<td>CA</td>
<td>0.15</td>
</tr>
<tr>
<td>CO</td>
<td>0.04</td>
</tr>
<tr>
<td>FL</td>
<td>0</td>
</tr>
<tr>
<td>IA</td>
<td>0.35</td>
</tr>
<tr>
<td>IL</td>
<td>0.06</td>
</tr>
<tr>
<td>NJ</td>
<td>0</td>
</tr>
<tr>
<td>PA</td>
<td>0.12</td>
</tr>
<tr>
<td>WA</td>
<td>0.04</td>
</tr>
<tr>
<td>WI</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Age 20-25</td>
<td>-0.035</td>
</tr>
<tr>
<td></td>
<td>(0.0131)</td>
</tr>
<tr>
<td>Age 25-30</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>(0.0146)</td>
</tr>
<tr>
<td>Age 30-35</td>
<td>0.115</td>
</tr>
<tr>
<td></td>
<td>(0.0176)</td>
</tr>
<tr>
<td>Age 35-40</td>
<td>0.353</td>
</tr>
<tr>
<td></td>
<td>(0.0245)</td>
</tr>
<tr>
<td>Age 40-45</td>
<td>0.559</td>
</tr>
<tr>
<td></td>
<td>(0.0496)</td>
</tr>
<tr>
<td>Age 45+</td>
<td>0.546</td>
</tr>
<tr>
<td></td>
<td>(0.1719)</td>
</tr>
<tr>
<td>White</td>
<td>-0.041</td>
</tr>
<tr>
<td></td>
<td>(0.0208)</td>
</tr>
<tr>
<td>Black</td>
<td>-0.109</td>
</tr>
<tr>
<td></td>
<td>(0.0246)</td>
</tr>
<tr>
<td>Not White or Black</td>
<td>-0.155</td>
</tr>
<tr>
<td></td>
<td>(0.0241)</td>
</tr>
<tr>
<td>Teaching Hospital</td>
<td>-0.243</td>
</tr>
<tr>
<td></td>
<td>(0.0148)</td>
</tr>
<tr>
<td>Private, for profit</td>
<td>0.081</td>
</tr>
<tr>
<td></td>
<td>(0.0215)</td>
</tr>
<tr>
<td>Public</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>(0.0127)</td>
</tr>
<tr>
<td># of discharges</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
</tr>
<tr>
<td>Median Income $0-$15,000</td>
<td>-0.115</td>
</tr>
<tr>
<td></td>
<td>(0.0316)</td>
</tr>
<tr>
<td>Median Income $15-$20,000</td>
<td>-0.046</td>
</tr>
<tr>
<td></td>
<td>(0.0301)</td>
</tr>
<tr>
<td>Median Income $20-$25,000</td>
<td>-0.019</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
</tr>
<tr>
<td>Median Income $25-$30,000</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>(0.0294)</td>
</tr>
<tr>
<td>Median Income $30-$35,000</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>(0.0309)</td>
</tr>
<tr>
<td>Median Income $35-$40,000</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>(0.0337)</td>
</tr>
<tr>
<td>Median Income $40-$45,000</td>
<td>0.082</td>
</tr>
<tr>
<td></td>
<td>(0.0404)</td>
</tr>
<tr>
<td>Median Income $45,000+</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>(0.0433)</td>
</tr>
<tr>
<td>Previous cesarean</td>
<td>2.973</td>
</tr>
<tr>
<td></td>
<td>(0.00136)</td>
</tr>
<tr>
<td>Breech Presentation</td>
<td>3.720</td>
</tr>
<tr>
<td></td>
<td>(0.0261)</td>
</tr>
<tr>
<td>Maternal Distress</td>
<td>1.035</td>
</tr>
<tr>
<td></td>
<td>(0.3568)</td>
</tr>
<tr>
<td>Fetal Distress</td>
<td>1.821</td>
</tr>
<tr>
<td>Description</td>
<td>Estimate</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Fee Differential</td>
<td>0.0005</td>
</tr>
<tr>
<td></td>
<td>(.000065)</td>
</tr>
<tr>
<td>Fee Differential Normalized by Private Vaginal Fee</td>
<td>1.024</td>
</tr>
<tr>
<td></td>
<td>(.1348)</td>
</tr>
</tbody>
</table>

Note: Standard errors are in parentheses, N=365,942
Chapter 2
Impact of Proprietary Positions and Equity Interest in the Pricing of Network ATM Services

2.1 Introduction

In recent years, ATM access has become one of the key services consumer’s demand from their retail banks. According to the Bank Network News, annual ATM transaction volume has nearly doubled from 5.7 billion in 1990 to 10.9 billion in 1997, corresponding to roughly five monthly transactions per customer. Approximately half of these transactions are network transactions—i.e., a depositor from bank x uses an ATM machine belonging to a different bank y. This statistic demonstrates the significant utility consumer derive from the shared nature of these ATM networks. From the bank’s perspective, Steven Felgran of the Federal Reserve of Boston writes, “providing ATM access is a competitive necessity for financial institutions seriously considering a retail business” (1985). The Antitrust Division of the Department of Justice claims that “A bank’s ability to offer its depositors access to other bank’s ATMs, and thereby to offer its depositors convenient access to their accounts, is in most bankers’ view necessary to attract and retain deposits.” In the eyes of consumers, the effective price\(^8\) charged for ATM access has emerged as a key element of differentiation among retail banks. Certain banks have responded to this need—most noticeably in the Boston area, US Trust has highlighted in their advertising campaigns that they offer free unlimited usage of any bank’s ATM who is on the same network as US Trust. Thus whereas a larger institution like BankBoston can

---

\(^8\) Since consumers are charged for using other bank’s ATMs but not for their own, the effective price charged by bank x is best described as a price bank x charges for using other banks’ ATMs adjusted by the extent of bank x’s proprietary network.
provide a similar offering with the extensive reach of their own proprietary network, smaller banks rely heavily on the shared ATM networks to provide this crucial service. As banks find it harder and harder to differentiate themselves along traditional dimensions, ATM access will likely become an even more important aspect of their service offerings.

ATM access is a two-tiered industry- in the upstream market, the network sells to each retail bank access to foreign ATMs. Retail banks use it as an input into a more general bundle of services it sells to consumers. The increased importance of ATM access downstream to consumers has coincided with a tremendous amount of upstream consolidation among ATM networks. It is now estimated that the ten largest regional networks account for 80 percent of all regional ATM activity (Baker, 1996). Despite this horizontal integration, the Department of Justice and the Board of Governors have embraced a relatively lenient policy towards these network mergers. For a generation, nearly every ATM network merger reviewed went forward, even when the result was a regional monopoly (Baker, 1996). A number of recent authors have criticized this laissez-faire approach by the authorities. In particular, McAndrews and Rob (1996) offer a theoretical model which leads to marginal cost pricing at the upstream market, but the retail prices charged for ATM access to consumers are set at monopoly levels. Their basic point is that a thorough antitrust analysis of the problem must look jointly at the upstream and downstream markets and determine if retail banks, end-consumers, or both are the affected parties. In the current literature, no one has taken an integrated look at the way incentives to the relevant players (the network and the downstream banks) and the non-standard pricing structure for upstream ATM access affect both the upstream and
downstream pricing. This paper addresses these issues by formally modeling the unique nature of these interactions as a two-stage game.

2.2 Background on ATM Networks

In the early 1980s, banks had their own proprietary network of ATM machines which gave customers 24 hour access to cash, information on account balances, and other relevant information. These were typically set up outside of an existing branch location. Soon thereafter, banks began to place ATM in non-bank locations such as malls, convenience stores and other retail areas. Banks and other financial institutions soon learned that there were both supply and demand side gains to be had from sharing one another’s ATMs. There would be enhanced demand from consumers due to greater access to ATM services and, the costs associated with operating the system could be consolidated in an efficient manner. Thus a unique situation arose where multiple firms who compete in the retail market joined forces in the upstream input market. The backward integration itself was not novel- the avoidance of double marginalization is a well-documented motivation for this action. However, such arrangements typically involve one upstream firm and one downstream firm, not a group of downstream competitors owing one upstream firm. A legal entity called “the network” was created which set the regulations governing how fees are transferred each time a transaction occurs. Several different organization forms exist today for the network: some (e.g., PULSE) are run on a non-profit basis\(^9\), some are owned by a third party non-bank- typically a switch operator, and others are owned by some subset of the banks who participate in the downstream retail competition. It is this latter structure which is becoming increasingly common and also
poses the most interesting theoretical problem. It is therefore this structure which I focus on in this paper.

I should distinguish between the two kinds of ATM transactions performed today— if an individual who is a member of bank $x$ goes to an ATM machine owned by bank $x$, this is considered an on-us transaction. There is generally no exchange of fees between depositor and bank for this kind of transaction. In addition, most banks provide this as a free service to their depositors. On the other hand, if a member of bank $x$ goes to an ATM machine owned by bank $y$, this is what I will refer to as a network ATM transaction. Associated with such a transaction is a web of fees exchanged among the various participants in the transaction. The relevant parties in such a transaction are the network, the consumer and the two banks—bank $x$ is called the issuing bank and bank $y$ is known as the acquiring bank. This transaction has a variety of fees associated with it which are described in the diagram below:

![Diagram showing ATM transaction fees]

9 Non-profit status is stated in the charter, but my experience is that it may not formally occur in practice.
The issuing bank pays two fees: the interchange fee is paid to the acquiring bank, and the switch fee is paid to the network. The interchange fee provides an incentive for the various banks to join the network and give access to their machines for other bank’s depositors. The switch fee is set either for the network to break even, or in the case of a for-profit network to maximize some profit function. The issuing bank typically charges its consumers a fee for each network ATM transaction called the foreign fee, and sometimes there is also a surcharge fee charged from the acquiring bank to the consumer.\textsuperscript{10} In summary, the legal structure gives the typical bank three different roles in relation to the network. First, they are a purchaser of the service provided by the network (access to other bank’s ATM’s) for which they pay out the interchange and switch fee. Secondly, they are an input into the network’s production process and receive the interchange fee as payment. Thirdly, they are residual claimant to any profits generated by the network through the switch fee. In addition, the structure naturally leads to essentially three different types of retail banks: net issuers, net acquirers and network neutral banks. Net issuing banks are defined as banks who’s depositors use other banks’ machines more than other banks’ depositors use their machines, net acquiring banks are the reverse, and network neutral are those banks who play the role of acquiring bank as often as they play the role of issuing bank.

The purpose of this paper is to model the competition between banks that have integrated their proprietary ATM machines to form the network entity but are competitors in the downstream retail banking business. I have constructed an analytically tractable 2-

\textsuperscript{10} The surcharge fee has been an issue of much interest in the press lately but will not be addressed in this paper. Many states are looking to ban surcharging by acquiring banks.
stage model where the network sets interchange and switch fees in stage 1 and the banks compete Bertrand in stage 2. The approach yields closed form solutions for the choice variables, but because the solutions are algebraically cumbersome to deal with, the comparative statics entail simulations of their behavior over a reasonable domain space for the parameters. There is a natural economic interpretation for each "region" in our domain space and so this approach seems to be best. Earlier works have (McAndrews, 1996) have modeled these interactions similarly, but what distinguishes this paper is its focus on how each member banks' equity interest in the network and the sizes of the proprietary network affect the optimal upstream and downstream prices. Therefore the interesting comparative statics involve an examination of how the equilibrium interchange fees, switch fees, and downstream retail banking prices charged to consumers vary with these two parameters. As a result, this paper also addresses the question of under what conditions is it likely that an ATM network will have the ability to exercise market power in the upstream input market as well as examining the potential impact on downstream retail banking prices.

The basic results of this model are the following: the upstream price banks pay for access to other institutions' ATMs (the interchange fee plus the switch fee) is minimized when the equity holders of the upstream network entity are \textit{network neutral}. In other words, banks whose customers use foreign ATMs as much as other banks' customers use their ATMs. These banks have nothing to gain by manipulating interchange fees or switch fees since they pay them out when they play the role of issuing bank as often as they receive them as an acquiring bank and as a shareholder in the network. When banks that are either net issuers or net acquirers in the system begin to obtain the majority equity interest, the upstream price for access to foreign ATMs increases. The increase is larger
when net acquiring banks are also majority shareholders in the network. As for the
downstream price levels, the model predicts that prices are highest when upstream access
prices are lowest. That is, when the equity in the network belongs to network neutral
banks, upstream costs are low but downstream prices are high. The intuition behind this
result is that the network neutral banks choose to keep access prices low upstream and earn
the majority of their profits downstream from consumers. However, when net issuing
banks or net acquiring banks secure a majority interest, they decide instead to earn their
profit at the upstream level through interchange and switch fees from their competitors,
and subsequently lower prices downstream to stimulate end-customer demand for network
transactions.

The remainder of this paper is organized as follows: Section 2.3 briefly discusses
some of the earlier literature comprising both theoretical approaches to ATM access
pricing as well as articles which review the antitrust aspects of these markets. Section 2.45
lays out the basic model and discuss the results, first for the upstream prices and then the
downstream prices. Section 2.5 gives an outline of the policy implications, and section 2.6
concludes with discussions about how the model might be improved.

2.3 Previous Literature

Theoretical Literature
One of the earlier theoretical papers dealing specifically with ATM network
economics is by Matutes and Padilla (1994). Their work examines the incentives of banks
to share their ATM machines when they are imperfect competitors in the downstream retail
banking market. They investigate the decision to share ATM machines as a tradeoff
between the following two effects: the network effect and the substitution effect. The
network effect refers to the fact that consumers value each banks' services more when they offer the greater locational convenience of access to competitor banks' ATMs. The substitution effect refers to the fact that banks become better substitutes when they decide to share their machines since an individual can simultaneously benefit from the convenient location of bank x's ATM machine and the other banking services offered by a competitor. In equilibrium, either a strict subset of banks decide to share their networks, or no banks share their networks. The reason why full compatibility does not prevail is that no bank obtains a network advantage over its rival, but all banks simply become better substitutes for one another. Competition is made tougher, and profits are lower than under incompatibility.

McAndrews and Rob (1996) construct a model with similar features to the one I present in this paper. Their interest also centers on the legal structure in which the network returns its profits to the member banks. They present a two-stage game where n firms compete symmetric Cournot in the downstream retail banking market, and two networks compete upstream for banks to join their networks. The essence of their model is to compare pricing when the two upstream entities are solely owned, i.e. they both do not have any direct interest in the downstream market, and when one of the entities is solely owned and the other is jointly owned by the member banks. Their primary result is that the latter structure leads to monopoly pricing for ATM access at the downstream level. Thus when the member banks are the residual claimants to the profits of the network, they have the incentive to extract monopoly rents from final consumers instead of other member banks. This result is a special case in the model I present, and it calls into question the
view that networks jointly owned by member banks results in maximization of their members' objectives rendering monopoly pricing a non-issue.

In summary, the Matutes-Padilla paper addresses compatibility of rival bank's ATM machines—an issue which I take as a given in this paper. The evidence to date indicates that it is fair to say that the decision of whether or not to make ATMs compatible has become a non-issue in favor of "yes". The McAndrews-Roberts paper is closer to mine, but they fail to consider the importance of proprietary position and the member banks' equity stake in the network. Their downstream game is symmetric Cournot, and they also ignore the unique nature of upstream pricing for foreign ATM access.

**Antitrust Literature**

The authors in the antitrust literature have highlighted the lack of opposition by the governing authorities to the recent wave of network ATM mergers as well as the absence of guiding principles for analyzing these mergers (Baker, 1996). The last decade has seen the following mergers approved without much resistance from the government. The 1988 MAC-Cashstream acquisition rendered control of nearly all branded ATMs in Pennsylvania to the newly formed entity; the 1994 merger of NYCE, Yankee and Citibank's ATM network created a dominant regional network in the New England area; and the 1996 merger between Honor and Most created a dominant network in the Southeastern United States. The most extensive attempt to date at providing some sort of analytical framework has come from the approval of the merger between NYCE and Yankee 24 forming an entity called InfinNet Payment Services. The rationale presented by the government for permitting the merger to go through centered on the network operating rules which permit 1) third-party switch processors to participate, 2) members of a given
network to participate in other networks, 3) card issuers to determine the routing of network requests\textsuperscript{11}, and 4) institutes to participate on a non-discriminatory basis. The Fed’s reliance on these operating rules as having significant weight in the approval suggests a willingness to accept certain types of operating rules as sufficient to mitigate concerns about market power. Also important to networks considering consolidation was the Fed’s recognition of efficiencies in the form of economies of scale and reduced costs. While this decision does spell out a number of factors, it is specific to this case and fails to provide any usable framework for analysis. Thus Guerin-Calvert (1996) states, “Two areas on which guidance from the Department and the Fed would prove useful for assessing current policy toward ATM mergers are:

1) Under what conditions does an ATM network have market power, and

2) What types of operating rules raise (or resolve) concerns about market power.”

This paper goes a long way towards fulfilling these needs by formally addressing the first of the two above questions.

2.4 The Basic Model

There are three entities in the basic model: the network and the two downstream banks. The interactions between them is modeled as a two-stage game. In the upstream market (stage 1) the network sets the interchange and switch fees thereby determining the price banks pay for access to competitors’ ATMs in the city. In the downstream market (stage 2), the retail banks use ATM access as an input into a more complete bundle of services made available to the public. This downstream offering is generically defined as

\textsuperscript{11} This is relevant when a bank is a member of multiple networks, hence allowing the lowest cost network to be the route of choice to process a transaction.
“banking services” and is composed of the variety of things consumers demand from their commercial banks. These include check writing, interest on short-term deposits, interest on longer term deposits, financial planning etc. The two banks’ offerings are differentiated along some subset of these dimensions, as will be explained in greater detail in the following sections. For now, it suffices to say that each bank has some local monopoly power afforded to it by this differentiation. Rather than having the banks choose a separate price for each of these services, I abstract it all into a single price firms charge for the full bundle. This price can be interpreted as an overall price index for retail banking services. Under this setup, the standard approach for solving the equilibrium in a two-stage game is used. The stage two downstream prices are solved for as a function of the interchange and switch fees, and then network profits are maximized after substituting the optimal values for the stage 2 choice variables into the stage 1 objective function. Thus the primary output of the model will be statements about how the three variables of interest, downstream prices, the interchange fee and the switch fee vary with the equity and proprietary network sizes.

**Consumer Demand**

As aforementioned, each bank sells one unit of an aggregate good which is differentiated in some sense from its competitor. ATM access is just one part of this bundle of services individuals demand. However, it is important to note that the ATM access is NOT what drives the differentiation in the two banks’ service offerings. There is some structure I place on the nature of consumer demand for ATM access which illustrates this fact.
I first assume that it is common knowledge to both of the banks that every consumer will need to conduct some predetermined number of ATM transactions \( n \)- a fixed parameter in the model. This can be thought of as the average number of transactions an individual does per month or so, and it is constant for all consumers in the model. The division between transactions done on bank 1 machines and those done on bank 2 machines is determined in the following way. Individuals are constantly moving about the city, and when they need access to an ATM, they have an \( \alpha_1 \) probability of being closest to a bank 1 machine and an \( \alpha_2 \) probability of being closest to a bank 2 machine. \( \alpha_1 \) and \( \alpha_2 \) can be thought of as proxies for the size of each bank's proprietary network. The greater is a bank's presence, the more likely a consumer will be near its machine. As shown in the earlier diagram, most institutions will charge a foreign fee each time a customer conducts a network ATM transaction in order to recoup some of the fees they pay out to the network and acquiring bank for the transaction. Therefore, we might expect individuals to be somewhat conscientious about whether they are performing network ATM transactions or on-us transactions. However, when consumers in the model need to conduct a transaction and are next to a foreign ATM, I assume that at that moment the valuation net of the foreign fee always exceeds the valuation net of the travel cost associated with finding one of their own ATMs. \textit{ATM demand is therefore presumed inelastic at the point of transaction.}^{12} In summary while the total number of transactions is fixed ex-ante, the relative share done on each bank's machines is completely determined by the relative sizes of each bank's proprietary network.

---

\(^{12}\) While this fact enters the model as an assumption, the empirical evidence indicates that it is probably also be true in equilibrium. Foreign fees typically range from $1.00-$2.00, and the average withdrawal amount from an ATM is closer to $50. Since the fee is so small relative to the amount of cash procured, it does not appear to be a tremendous deterrent to conducting network ATM transactions.
This setup pins down each bank’s entire expected cost of providing ATM access for each of their customers. Since the cost to the bank each time a customer conducts a network ATM transaction is the interchange fee, \( r \), plus the switch fee, \( s \), bank \( j \) knows that it will pay out this cost \( (n^*\alpha_i) \) times for each customer. He pays out nothing when his consumers use his own machines, and so the marginal cost for bank \( j \) of providing ATM access per customer is \( (r+s)^* (n^*\alpha_i) \). The composite price each banks charges for its total bundle of services reflects this cost.

To summarize, I have used the assumption that all individuals perform a common number of ATM transactions and that individual demand for ATM access is inelastic at the point of sale to fix the entire marginal cost firms face for providing full ATM access to their customers. In addition, this structure renders ATM access a non-differentiated good since consumers pay a single price for the liberty to use any bank’s ATM, and the service they receive is identical from each. As a result, product differentiation must be driven by other dimensions in this bundle of services in order to utilize Hotelling’s approach to modeling consumer demand, which I will turn to now.

*Downstream Competition*

Now that the structure of consumer demand is set, I move onto describing the downstream competition. The stage two game is similar in spirit to Hotelling’s differentiated product price-setting game played between two firms on the opposite ends of the unit interval. The two firms offer one unit of an indivisible good, and the travel costs reflect the fact that the firms offer differentiated products which suit some customer’s preferences better than others. Consumers buy one unit of the good from one of the firms or they buy nothing, and they buy if their valuation net of the travel cost and the price for
the good is greater than zero. Applying this notion to the retail banking setting, customers choose a single bank to place their deposits, and demand in return a variety of services. ATM access is just one of the many services banks offer which is captured in this composite good; it also includes things like check writing, interest on deposits, credit cards, financial planning, 24-hour computer access to accounts over the internet, etc. The notion is that the two banks will differentiated themselves along some or all of these dimensions\(^\text{13}\); however, for the sake of simplicity, it is all collapsed along one dimension. In addition, it is assumed that the valuation for these services is high enough such that in equilibrium, all consumer choose one of the two banks.

The firm’s strategic variable is the price for this bundle of services. Using the standard Hotelling approach of locating the indifferent consumer on the unit interval where consumers valuation of the product is \(v\) and per unit travel cost is \(t\) gives \(v-\alpha x-p_1=v-(1-x)t-p_2\), which yields the familiar \(x=q_1=1/2-p_1+p_2\) for \(t=1/2\). However, I impose a slight variation on this formulation and write customer demand as a function of prices as

\[
q_i = \frac{1}{2} - p_i + \gamma p_j
\]

where \(\gamma < 1\). The reason an additional coefficient of \(\gamma\) is placed in front of the other firm’s price is purely technical. If we use the standard formulation given by \(\gamma=1\), then when inserting the optimal prices from stage two into the stage one maximization problem, the objective function becomes linear in the choice variables. Clearly, this results in the

\(^{13}\) For example, bank 1 may have better interest rates on savings, and bank 2 better interest on more illiquid accounts such as CDs. Depending on an individual’s income, liquidity needs and time preference, one institution may serve his needs better than the other. In addition, bank 1 may offer WWW access to accounts,
function being maximized at a value of positive or negative infinity for the choice variables, i.e. no equilibrium. This only occurs when the coefficients on both prices in the demand functions are identical. The slight variation noted above doesn’t change the substance of the model, but does gives us a tangible result.

The two firms differ in their equity interests and the total number of ATM machines they contribute to the network. We let $\sigma_i$ and $\sigma_j$ represent each firm’s equity stake in the network, and $\alpha_i$ and $\alpha_j$ serve as proxy variables for the size of each bank’s network. For algebraic simplicity, I define $(\sigma_i + \sigma_j) = 1 \text{ and } (\alpha_i + \alpha_j) = 1$. Since there are only two banks who comprise this network, it makes sense for their equity stakes to sum to unity. As for the proxy for each bank’s network, one can think of this as bank 1 having $N$ machines and bank 2 possessing $M$ machines, $\alpha_i = N/(N+M)$ and $\alpha_j = M/(N+M)$\textsuperscript{14}. Under this formulation, the implied domain space of $\sigma_i$ and $\alpha_i$ is $[0,1]$. No economic content is lost in this formulation since with only two firms, we’re interested in how the relative equity stakes and relative proprietary network sizes affect the upstream and downstream prices. However, much is gained in analytical simplicity.

The equilibrium is solved using a simple, Stackleberg-like backward induction. Hence in stage 2, downstream retail prices are chosen given the interchange and switch fee, and in stage one, the interchange and switch fee are set with the knowledge of how foreign fees will adjust to changes in interchange and switch fees. The stage 2 profit functions of the two banks are as follows:

\begin{align*}
\text{bank 1:} & \quad \pi_1 = \alpha_1 \pi_1^{\text{merch}} + (1-\alpha_1) \pi_1^{\text{int}} \\
\text{bank 2:} & \quad \pi_2 = \alpha_2 \pi_2^{\text{merch}} + (1-\alpha_2) \pi_2^{\text{int}}
\end{align*}

which suits the computer literate public better, while bank 2 may have more branch offices, favoring those who prefer doing their banking face-to-face or are computer illiterate.

\textsuperscript{14} More accurately, $\alpha_i$ and $\alpha_j$ can be thought of as indicators of presence.
\( \pi_1 = (q_1 (p_1 - (r+s)n\alpha_2 - c)) + \{r n \alpha_1 q_2 \} + \{\sigma_1 s(n \alpha_1 q_2 + n \alpha_2 q_1)\} \)  

(ii)

\( \pi_2 = (q_2 (p_2 - (r+s)n\alpha_1 - c)) + \{r n \alpha_2 q_1 \} + \{\sigma_2 s(n \alpha_2 q_1 + n \alpha_1 q_2)\} \)  

(iii)

There are three components to each firm’s profit function. The first term is banking profits with the total marginal cost per consumer for bank \( i \), \( \{c + (r+s)\alpha_i\} \), represented as the marginal cost of providing full ATM access plus the constant marginal cost \( c \) of providing all other banking services. \( p_i \) and \( p_j \) are each firm’s representative price index for their services. The second term represents interchange fee profits arising from transactions which bank \( j \)'s customers perform on bank \( i \)'s machines, \( q_i \) is the number of consumers times the switch fee \( r \) times the number of transactions \( n \alpha_i \). The third term represents each bank's share of the profits as a shareholder in the network.

To solve for the stage 2 Nash equilibrium, each firm maximizes profits with respect to its price for ATM access.\(^{15}\) The first order conditions for each of the banks are as follows:

\[
\frac{\partial \pi_1}{\partial p_1} = (1 - p_1^* + \frac{1}{2} p_2) - (p_1^* - (r + s - s\sigma_1)\alpha_2) + \frac{1}{2} \alpha_1 (r + s\sigma_1) = 0
\]

\[
\frac{\partial \pi_2}{\partial p_2} = (1 - p_2^* + \frac{1}{2} p_1) - (p_2^* - (r + s - s\sigma_2)\alpha_1) + \frac{1}{2} \alpha_2 (r + s\sigma_2) = 0
\]

The second order conditions for a maximum are satisfied since each bank’s profit functions are strictly concave in own prices. The above two first order conditions define a system of equations which is linear in the two prices for ATM access. After some rearranging, the equilibrium values turn out to be
\[ p_1^*(r,s) = \frac{1}{3}(1 + 2c) + r\left(\frac{3}{5} - \frac{1}{5}\alpha_i\right) + s\left(\frac{3}{5}(1 - \alpha_i - \sigma_1) + \alpha_i\sigma_1\right) \]  \hspace{1cm} (iv)

\[ p_2^*(r,s) = \frac{1}{3}(1 + 2c) + r\left(\frac{3}{5} - \frac{1}{5}\alpha_i\right) + s\left(\frac{3}{5}(1 - \alpha_i - \sigma_2) + \alpha_i\sigma_2\right) \]  \hspace{1cm} (v)

For simplicity sake, I substitute in \((1 - \sigma_i)\) for \(\sigma_i\) and \((1 - \alpha_i)\) for \(\alpha_i\), so the optimal prices can be written as

\[ p_1^*(r,s) = \frac{1}{3}(1 + 2c) + r\left(\frac{3}{5} - \frac{1}{5}\alpha_i\right) + s\left(\frac{3}{5}(1 - \alpha_i - \sigma_1) + \alpha_i\sigma_1\right) \]  \hspace{1cm} (vi)

\[ p_2^*(r,s) = \frac{1}{3}(1 + 2c) + r\left(\frac{2}{5} + \frac{1}{5}\alpha_i\right) + s\left(\frac{2}{5}(1 - \alpha_i - \sigma_1) + \alpha_i\sigma_1\right) \]  \hspace{1cm} (vii)

As expected in the equilibrium of a Bertrand game, the prices are a weighted average of the three components of cost- \(c\), \(r\) and \(s\). Both prices are strictly increasing in aggregate service cost, the interchange fee, and switch fees. \(\{(3/5) - (1/5)\alpha_i\},\)

\(\{(2/5) + (1/5)\alpha_i\}, \quad \{(3/5)(1 - \sigma_i - \alpha_i) + \sigma_i\alpha_i\},\) and \(\{(2/5)(1 - \sigma_i - \alpha_i) + \sigma_i\alpha_i\}\) are all always \(\geq 0\) for the possible values of \(\sigma_i\) and \(\alpha_i\). My first observation is the difference in the sensitivity of the downstream prices to changes in \(c\), \(r\) and \(s\), i.e. the derivatives of the downstream prices with respect to the three components of cost. As expected, downstream prices are always most sensitive to \(c\). At all times, both firm’s prices are at least as sensitive to the interchange fee as they are to the switch fee, i.e. \(\{(3/5) - (1/5)\alpha_i\} \geq \{(3/5)(1 - \sigma_i - \alpha_i) + \sigma_i\alpha_i\}\) and \(\{(2/5) + (1/5)\alpha_i\} \geq \{(2/5)(1 - \sigma_i - \alpha_i) + \sigma_i\alpha_i\}\). Only when \(\sigma_i\) and \(\alpha_i\) jointly move towards 1 (or zero) does the downstream price sensitivity to the switch and interchange fees become more and more equal. This reason is quite intuitive if one recalls the structure

\[ n=1 \text{ and } \gamma=1/2 \text{ for simplicity sake} \]
of fee transfer for a network ATM transaction outlined in the earlier diagram. For a given firm \( i \), the cost component attributable to the switch fee can be mitigated in two ways- by its equity stake in the network \((\sigma)\), or by possessing a large proprietary network \((\alpha)\). A large equity stake feeds most of the switch fee payments back to the payer, and a large proprietary network means the switch fee is paid out less frequently since customers do more on-us transactions than network transactions. However, the costs attributable to the interchange fee can only be mitigated by \((\alpha)\). At the two extremes, \((\sigma=\alpha=1, \text{ or } \sigma=\alpha=0)\) each component of cost is born equally by the two firms, and hence the downstream price is equally sensitive to both of these prices. This greater sensitivity of downstream prices to interchange fee movements versus switch fee movements is a point I will return to later when we examine the full equilibrium after the stage one maximization has been solved.

*Upstream Maximization*

In stage one, the network sets the switch and interchange fees to maximize its profits knowing the stage two Nash equilibrium in prices. Of the three possible legal arrangements for the network, I assume an equity structure such that the member banks of the network are its sole stockholders. The stage one network objective function is therefore not just the network profits, but the equity-weighted sum of the individual bank’s downstream profits as well as each bank’s share of the network profits since the banks themselves are residual claimant to the profits of the network. This formulation may not be exactly correct-for example, it may be that a single bank with a simple majority of all voting shares has complete autonomy in the determination of switch and interchange fees. This case notwithstanding, the equity-weighted sum formulation assumes that the
shareholding banks in the network negotiate amongst themselves and set prices jointly with each bank's equity stake reflected in his "say" in the final decision making process.

Having solved for the stage two Nash Equilibrium, we now proceed to the stage one problem. The network itself now maximizes profits given the stage two equilibrium. Because member banks are shareholders in the network, we define the network objective function as the equity-adjusted sum of the individual bank's objective functions. Hence,

\[
\max \pi = \sigma_1 \pi_1 + \sigma_2 \pi_2 \tag{viii}
\]

\[
\frac{d\pi}{dr} = \sigma_1 [\frac{\partial \pi_1}{\partial r} + \frac{\partial \pi_1}{\partial p_1^*} + \frac{\partial \pi_1}{\partial p_2^*}] + \sigma_2 [\frac{\partial \pi_2}{\partial r} + \frac{\partial \pi_2}{\partial p_2^*} + \frac{\partial \pi_2}{\partial p_1^*}] = 0
\]

\[
\frac{d\pi}{ds} = \sigma_1 [\frac{\partial \pi_1}{\partial s} + \frac{\partial \pi_1}{\partial p_1^*} + \frac{\partial \pi_1}{\partial p_2^*}] + \sigma_2 [\frac{\partial \pi_2}{\partial s} + \frac{\partial \pi_2}{\partial p_2^*} + \frac{\partial \pi_2}{\partial p_1^*}] = 0
\]

By the envelope theorem, the middle terms inside each of the four brackets drop out since the derivative of each bank's profit function with respect to his own downstream price is zero at the optimal downstream price. The above first order conditions are also a system of equations linear in \(r\) and \(s\). The matrix is invertible, and a solution exists. The equilibrium interchange and switch fees are as follows:

\[
r^*(\alpha_1, \sigma_1) = \frac{(c - l)[33(1 - \alpha_1) + \sigma_1 (134\alpha_1 - 142) + \sigma_1^2 (193 - 134\alpha_1) - 84\sigma_1^3]}{3\alpha_1 (1 - \alpha_1)(2\sigma_1 - 1)(48\sigma_1^2 - 48\sigma_1 + 11)} \tag{ix}
\]

\[
s^*(\alpha_1, \sigma_1) = \frac{(c - l)[-33 + 41\alpha_1 + \sigma_1 (109 - 168\alpha_1) + \sigma_1^2 (168\alpha_1 - 84)]}{3\alpha_1 (1 - \alpha_1)(2\sigma_1 - 1)(48\sigma_1^2 - 48\sigma_1 + 11)} \tag{x}
\]
The effective cost of a network ATM transaction, \((r+s)\) is

\[
(r + s)^* = T C^* = \frac{(c - l)[8\alpha_1 - \sigma_1 (33 + 34\alpha_1) + \sigma_1^2 (109 + 34\alpha_1) - 84\sigma_1^3]}{\alpha_1 (1-\alpha_1)(2\sigma_1 -1)(48\sigma_1^2 - 48\sigma_1 +11)}
\]  

(xi)

The optimal values for the interchange fee and switch fee have closed-form solutions but are somewhat impractical to deal with. Instead of trying to take derivatives of the optimal values with respect to \(\sigma_1\) and \(\alpha_1\), I chose to simulate the results over the domain space of the variables. Because the domain space of \((\sigma_1, \alpha_1)\) is limited to \([0,1] \times [0,1]\), this was the most efficient way of extracting meaning out of these results. Some clear patterns emerged during these simulations which I will discuss here.

2.6 Discussion of Results

The focus of this paper is how the proprietary network positions and equity stakes of the member banks affect the upstream and downstream prices. Before starting this discussion, in order to provide a more intuitive framework for interpreting the results, I will divide up the domain space into four quadrants as shown in the diagram below. \(\sigma_1\) is placed on the horizontal axis and moves from 0 to 1, and \(\alpha_1\) is placed on the vertical axis and moving from 0 to 1.
Region I corresponds \( \{ \sigma, \alpha \} \) belonging to the \([0, \frac{1}{2}] \times [0, \frac{1}{2}] \), region II to \( \{ \sigma, \alpha \} \) belonging to the \([\frac{1}{2}, 1] \times [0, \frac{1}{2}] \), region III to \( \{ \sigma, \alpha \} \) belonging to the \([0, \frac{1}{2}] \times [\frac{1}{2}, 1] \) and region IV to \( \{ \sigma, \alpha \} \) belonging to the \([\frac{1}{2}, 1] \times [\frac{1}{2}, 1] \). Region IV is roughly where firm 1 has both a larger proprietary network and higher equity stake in the network- in some sense a dominant position. Region I represents the situation where firm 2 is in a dominant position. Regions II and III are the hybrid cases, Region III is where firm 1 has a large proprietary network but low equity stake, and region II is where firm 1 has a large equity stake in the network but small proprietary network. I conduct the analysis on a region by region basis since the behavior of the optimal values while consistent within a given region appears to change as we move across regions. However, the substance of the interpretation does not change. It should be clear that because changes in \( \sigma \), and \( \alpha \), represent a redistribution of equity and proprietary positions between the two firms, raising \( \alpha \), in region IV is identical to lowering \( \alpha \), in region I. Both scenarios describe situations where the dominant firm acquires a stronger proprietary position. Likewise raising \( \alpha \), in region III is identical to lowering \( \alpha \), in region II. Both describe situations where the firm in the weaker equity position acquires a stronger proprietary position. Similar arguments can be made for changes in \( \sigma \). Therefore, examining the movement of the optimal values in any two adjacent regions captures the substance of the model and so for simplicity’s sake, I focus on region II and IV and conduct the analysis from firm 1’s perspective.

**Optimal Interchange and Switch Fees**

The first issue to be addressed is the sign of our choice variables in equilibrium.

The switch fee values are always positive, and the interchange fee values are always
negative. This runs contrary to what is commonly observed in today’s networks, in particular, both interchange and switch fees are usually positive. The following drives the divergence here: in reality, the network must provide an incentive for various banks to take ATM cards from individuals at competitor’s banks. The interchange fee must be positive so acquiring banks receive the payment. In the model, it is assumed that since both parties have already agreed to integrate upstream, the “stage 0” choice of whether or not to make one’s ATM machines compatible with the network (addressed in Matutes-Padilla) is bypassed. The need to incent banks to open up their machines to depositors of other banks is removed. The general implication of this result is that as the decision to share ATMs becomes relatively less important, interchange fees will presumably decline. In the limit, it may very well be possible that interchange fees may become negative, i.e. the issuing bank will receive the interchange fee in order to compensate him for the larger switch fee which gets paid out to the network.

A logical place to begin the marginal analysis is at the intersection of the four regions, i.e. the center of the unit square where the proprietary positions and the equity stakes are ½ for the two banks. I will refer to this point as the symmetric case. It turns out that the total cost, \((r+s)\), is minimized at the point where both firms have approximately ½ equity interest and equal size ATM networks. This makes sense since neither firm has an advantage over his competitor, and so they agree to minimize the upstream cost of access to one another’s machines. As we move away from this point total cost will rise. This is true whether we hold \(\sigma_i\) (or \(\alpha_i\)) constant or change both simultaneously. In general, switch fee rises and the interchange fee falls as we move away from this point, but because the increase in the switch fee outweighs the decrease in the interchange fee, total cost will
generally rise. The basic intuition is that as we move from the symmetric equilibrium, one of the firms acquire an advantage over the other either in their equity stakes, their proprietary positions, or both. Thus the firm with the greater controlling interest in the network will have a greater say in the interchange/switch fee setting process and use it to their advantage. I will turn now to a more detailed analysis of how fees move with respect to $\sigma$, and $\alpha$.

*Changes in $r^*$ and $s^*$ with respect to $\alpha_1$*

I first examine how the total upstream cost changes while holding equity stakes constant and altering the relative proprietary positions. The derivative of total cost with respect to $\alpha$, is negative in region II. Another way of stating this fact is that when the firm with the larger equity stake and a smaller proprietary network gradually acquires a weaker proprietary position, the interchange fee decreases but the switch fee rises at an even faster rate resulting in a net increase in total costs. This result is consistent with what one might expect in this situation- the optimal decision for a firm with a large equity stake and an increasingly weak proprietary position is to lower the interchange fees and raise the switch fees. The fees he pays as a net issuing bank are reduced and his profits as a shareholder in the network rise.

In region IV, as $\alpha_1$ increases away from $\frac{1}{4}$, firm 1 is dominant and acquiring an even stronger proprietary position. The same results obtain here- the interchange fee declines and the switch fee rises with the increase in the switch fee far exceeding the decrease in the interchange fee. One might expect both the interchange fee and the switch fee to rise in this case since net acquiring banks and who are large shareholders will benefit both from higher interchange fees as well as higher switch fees. But the reason why this
does not occur is the following- recall the results of the stage 2 Bertrand game and formulation of downstream demand. Suppose both the interchange and switch fees rise-both firms' prices rise when the interchange/switch fees increase. The fall in firm i's customer demand driven by firm i's price increase exceeds the increased demand driven by firm j's price increase. The net effect of reduced customer demand impacts all three components of the banks' profit functions and outweighs the potential gain from increased revenue as a net acquiring bank and a shareholder in the network. Thus raising both switch and interchange fees is a poor decision. Instead, because the dominant firm receives more benefit from increasing switch fees rather than interchange fees (he recoups the switch fee on all transactions but the interchange fee only when the competitor's depositors conduct transactions on the dominant firm's machines) the smarter move is to raise the switch fee and lower the interchange fee to maintain customer demand.

Changes in \( r^* \) and \( s^* \) with respect to \( \sigma_1 \)

I move now to examine how the upstream values change when \( \alpha_1 \) is held constant and \( \sigma_1 \) is allowed to vary. In region IV, where firm 1 is dominant, for any given value of \( \alpha_1 \), total costs are minimized where \( \sigma_1 \) is roughly \( \frac{1}{2} \). As we increase \( \sigma_1 \) away from \( \frac{1}{2} \), or as the dominant firm acquires a larger equity stake, the switch fee rises and the interchange fee falls with the increase in the switch fee again exceeding the fall in the interchange fee. Again, though we might expect the dominant firm to increase both the switch and interchange fees, he is reluctant to do so because of the effect on the downstream profitability. The reasoning is analogous to those laid out in the previous section- the fall in end consumer demand driven by the higher input prices makes raising both switch and interchange fees a poor decision. Given the choice of raising one or the other, he chooses.
to raise the switch fee. In region II, it turns out that total cost actually falls as the firm with
the weaker proprietary position increases its equity stake. In region II, in a neighborhood
around \( \sigma_l = 1/2 \), total cost falls as \( \sigma_l \) rises. The intuition behind this result is that if a firm
with a small proprietary network gains a proportionally larger equity stake, he will do his
best to lower the total cost at first since he is a net issuer in the downstream game. He
actually lowers the switch fee since his ability to recoup these losses as a shareholder is
outweighed by the frequency with which they are paid out as a net issuing bank. However
as his equity stake continues to grow, total costs rise again since his profits as a
shareholder begins to exceed the losses at the retail level. The basic results are summarized
in the following table:

\[
\begin{align*}
\text{Region IV} & \quad \frac{dr^*}{d\sigma_l} < 0, \quad \frac{ds^*}{d\sigma_l} > 0, \quad \frac{dT^*}{d\sigma_l} > 0 \\
\text{Region II} & \quad \text{for } \sigma_l \text{ in a neighborhood around } \frac{1}{2}, \quad \frac{dr^*}{d\sigma_l} > 0, \quad \frac{ds^*}{d\sigma_l} < 0, \quad \frac{dT^*}{d\sigma_l} < 0 \\
& \quad \text{for } \sigma_l > .55, \quad \frac{dr^*}{d\sigma_l} < 0, \quad \frac{ds^*}{d\sigma_l} > 0, \quad \frac{dT^*}{d\sigma_l} > 0 \\
\frac{dr^*}{d\alpha_1} < 0, \quad \frac{ds^*}{d\alpha_1} > 0, \quad \frac{dT^*}{d\alpha_1} > 0 \\
\frac{dr^*}{d\alpha_1} > 0, \quad \frac{ds^*}{d\alpha_1} < 0, \quad \frac{dT^*}{d\alpha_1} < 0
\end{align*}
\]

Downstream Results
As for the downstream price levels, I first observe that they fall as we move away
from the area where equity shares are roughly equally divided. If we recall the earlier
upstream results, this is the opposite from when total costs at the upstream level are lowest.
Our model therefore predicts a divergence of what is thought to be good at the promoting
competition at the retail bank level, and what is good at the end-consumer welfare level.
The reason for this result is the following. As mentioned earlier, the downstream access prices are generally more sensitive to how the interchange fee moves than how the switch fee moves. As $\alpha$ moves away from $\frac{1}{2}$ either towards 1 or 0, the switch fee will increase, and the interchange fee declines. The switch fee rises faster than the interchange fee declines since the firm who prefers a higher switch fee has increasingly more say in the fee setting process as his equity stake increases. Hence while the upstream price of providing ATM access $(r+s)$ may rise, since prices are more sensitive to the interchange fee, the final price charged by the firms goes down. In addition because the downstream prices move together in a Bertrand game, the firm with the greater equity stake will find an even greater incentive to lower prices because the competition will also lower prices in response to the given firms lower price. The point raised earlier applies again here, the net effect of both firm’s decreasing their prices is an increase in overall customer demand. The benefit to these lower prices is the additional customers you attract for greater overall banking profits as well as the additional ATM transactions that both members of your own bank and the competition will undertake. The more transactions, the more profits he earns as a shareholder in the network. Overall, it appears that the bank with the controlling equity interest decides to sacrifice “banking profits” in favor of profits as a shareholder in the network.

*Equilibrium Profit Functions*

A final result of interest in this model is that in equilibrium, the profit functions for the individual firms are independent of the proprietary positions. When the upstream and downstream prices are solved for and substituted back into $\pi$, and $\pi_s$, the $\alpha$, term disappears. The model therefore implies that if there were a stage 0 where the firms could
decide to make investments in their proprietary networks, neither would find it desirable. Thus rather than proprietary positions being something which is fixed *ex-ante*, it is instead endogenous in the model. This seems consistent with what we are observing in today’s ATM sector. The investments in proprietary networks occurred up to the early 1990’s, but the recent industry trend indicates that mergers and acquisitions are the more effective way to extend the reach of a network. In addition, newer entrants are finding it more advantageous to participate in a shared ATM network to give their customers ATM access as opposed to building a proprietary network themselves.

2.7 Policy Implications

The upstream policy implications of this model are relatively clear- if shareholders of the network are composed of banks with both large and small proprietary networks, or predominantly network neutral banks, the potential for anti-competitive harm at the upstream level decreases. The simple intuition for this result is that each type of bank’s interests are represented when interchange and switch fees are set. As we move towards a situation where the shareholders of the network become firms with dominant proprietary networks, there is no “policing” done by the smaller proprietary network banks. Total costs for access to foreign ATMs defined as \((r+s)\) rise since the shareholding banks bear less and less of the costs they are initiating. Total costs will also rise if we move towards a situation where a subset of banks with smaller proprietary networks retain controlling interest in the network. However, the increase in total cost is smaller than when we move towards a situation with dominant firm.

Suppose we are at a situation where the shareholders are banks with large proprietary networks. Then if the objective is to create a more level playing field for
smaller banks without the extensive proprietary network (or a substantial equity stake in the network) to compete, there are several options available. The above results show that one can either redistribute the voting shares from the large proprietary network banks to the small ones, or force the banks with extensive networks to divest some of their ATMs to the smaller ones. It turns out that the redistribution of equity is a more effective instrument of getting total upstream costs down. At nearly every point in the domain space, the drop in total cost for a unit change in the equity redistribution (towards ½) is greater than the drop in total cost for a unit change in the proprietary positions (towards ½). As stated earlier, the upstream fees are lower when the equity is equally divided among net issuing banks, net acquiring banks, and network neutral banks. This describes the scenario outlined by Charles Rule, acting assistant attorney for the antitrust division of the Dept. of Justice, when he stated that joint ownership as an institution coped with the monopoly pricing problem. As the model shows, only if the equity stakes represents the interests of the different types of banks does this hold true.

The fact that downstream and upstream prices move in opposite directions is somewhat problematic. The ultimate objective is generally to maximize consumer welfare, and this is achieved when upstream prices are highest. Because the model contains only two stages, it does not consider the long-run implications of market where a single type of firm possesses the majority of the equity. If consumer welfare is maximized today and monopoly rents are extracted from the banks as opposed to the customer, the long-run effect could be exit from the industry by smaller banks and a subsequent rise in the index for retail banking prices. The level of analysis required to answer this question goes beyond the scope of this paper.
2.8 Conclusions

This paper sheds light on the nature of competition between banks who are competitors in the downstream market for retail banking yet joint owners in the network entity who provides ATM access, a critical input into the downstream service offering. This model is most applicable in situations where there is a single dominant regional ATM network, the number of total machines in the area has reached some critical mass, and the strategic focus centers on upstream and downstream pricing. The simplifying assumptions is that the banks have already agreed ex-ante to share their ATMs and consumer demand for ATM transactions is fixed and inelastic at the point of transaction. The current industry trend toward consolidation of regional networks into a single entity and the mature nature of this industry indicates that these are reasonable assumptions. Indeed most banks today do participate in a shared ATM network, and at current prices, it is reasonable to think that demand for an ATM access is inelastic.

The primary results are the following: upstream prices for access to another bank’s ATM are lowest when the equity holders of the network are neither net issuing nor net acquiring banks. As expected they have little to gain from raising upstream prices. As one type of bank begins to acquire a majority interest in the network, switch fees steadily rise and interchange fees fall (the net effect is an increase in r+s). Secondly, downstream prices are more sensitive to interchange fee movements than switch fee and so as we move away from the symmetric situation, upstream prices rise but downstream prices fall. In a static sense, this result indicates that there may be a divergence between the anticompetitive consequences in the upstream versus the downstream market. The basic decision the
network faces is whether or not it will extract the monopoly rents from consumers or other member banks. In the symmetric case, they choose to earn their profits in the downstream market; as we move away from this point, they earn more of their profits upstream.

The alternative legal structures discussed in the introduction, i.e. networks operating either as a non-profit or as an entity which does not participate in the downstream retail banking market yield trivial results in our model. In particular, if the switch fee must be set so the network breaks even, then the sign and the type of bank who owns the majority equity interest determines magnitude of the interchange fee. Simply put, net acquirers will likely impose large, positive interchange fees, whereas net issuers will impose large, negative. Network neutral banks will likely be indifferent. The network will optimize where the gains from large switch fees (if a net acquirer is the equity holder) offset the reduced downstream network ATM demand. If the network is operated by some alternative non-bank entity, he will probably raise switch fees to increase his profitability and make interchange fees large and negative to induce network ATM usage at the downstream level. In this case, there is clearly no equilibrium since the firm can increase the switch fee indefinitely while lowering the interchange fee indefinitely to maximize profits.

This model was certainly structured in some sense in order to achieve closed-form solutions. As with most applied theory papers, there is a tradeoff between greater generality/endogeneity versus analytical tractability. I certainly opted for the latter. If one were to pursue the other end of the spectrum, an alternative way to formulate the model is for firms to choose foreign fees, making the number of on-us transactions versus network ATM transactions endogenous. Individuals could be differentiated in their location relative
to each bank's ATMs as well as their valuation for ATM access. This greater flexibility comes at a great expense in terms of complexity since the objective function becomes a cubic or fourth degree polynomial in the choice parameters, rendering the first order conditions a quadratic or cubic system of equations. However, modern computing technology makes simulating the results feasible, and this may prove to be a useful exercise in the future.
Chapter 3
Pricing of High-Tech Software Startups: Art or Science?

3.1 Introduction

The initial public offering of Netscape Communications Corp. in August of 1995 marked the beginning of revolution in information technology. The Internet, a network based on the TCP/IP and HTTP protocol, and its graphical user interface portion known as the World Wide Web (WWW), put eighty million Americans in touch with individuals and computers all over the world. Information dissemination reached a new level of simplicity as text, images, audio and video became accessible to anyone with a browser and connection into the underlying backbone of the Internet. What was once a loose confederation of academic institutions sharing ideas over e-mail and virtual bulletin boards became a top priority of business IT agendas due to its significant economic advantages over traditional client/server and mainframe computing. These advantages include open standards/platform independence\(^{16}\), low cost of deployment/maintenance, and the simple universal client known as the browser. Everyone from the small family business to the Fortune 500 has felt the impact of this new medium for conducting business.

The release of Netscape’s flagship product Navigator also sent major shockwaves through the incumbent software giant in the industry, Microsoft. Microsoft had enjoyed enormous prosperity driven by owning the operating system (the gateway for end-user applications) to 85\% of the Personal Computers in the world. The Internet and subsequent technologies such as the Java programming language posed an immediate and long-term

\(^{16}\) Whether an individual is running a PC with Microsoft Windows, Unix or Macintosh as their primary system platform is irrelevant.
threat to this dominance. Microsoft was forced to draw deeply into its 8-9 billion dollar cash reserve in order to play “catch up” to companies espousing the new technologies based on Web protocols. While their turnaround has been nothing short of impressive, a basic point was made very clear: with such low barriers to entry into the software industry, the opportunities for a small high-tech startup to reap huge financial gain and become the next dominant player in software were abundant. As one might expect, this reality has spurred a tremendous amount of entry into industry. In the last several years, places like Silicon Valley in California and Rte. 128 in Massachusetts has witnessed the birth of an unprecedented number of new companies trying to find niche opportunities in software as well as come up with the next big idea.

The purpose of this paper is to analyze the financing aspect of these technology ventures, much of which has been done by the venture capitalist community. In particular, the basic question I try to answer is how do venture capitalists determine the price of a high-tech startup? The importance of this question is simple: since nearly all early stage financing is done in exchange for preferred equity (as opposed to debt), the pre-money price sets the terms for how much ownership is exchanged for a given dollar investment. Thus if a company receives a pre-money valuation of $8 million and a venture capitalist makes a $2 million investment, he essentially receives 20% ownership stake in the company in exchange for the cash infusion. The tradeoff most entrepreneurs must deal with when securing investors is the desire to retain equity among founders and making sure the venture does not run out of money. The pre-money valuation along with the size of the investment determine how much equity will be forsaken.

---

17 On a network based on Internet protocols, the browser represents an alternative gateway to end-user applications written in Java.
Many individuals in the industry have argued that pricing these companies is more art than science. After all, there are few tangible assets and no history of cash flow/earnings on which to base traditional approaches such as book value or discounted cash flow. Others have argued that the single most important factor in determining a firm’s price may very well be each investor’s subjective belief in the entrepreneur’s ability (Muzyka, Birley and Leleux, 1995). However, my analysis indicates that there are some concrete factors playing a significant role in determining a company’s valuation. In particular, how close a company is to a completed product appears to be an extremely important factor in how a venture capitalist will price a firm. In addition, the results suggest that factors such as prior managerial/technical experience and investors’ bullishness on the technology sector also play a significant role in the pricing of these early stage deals. This paper puts some structure on these claims by estimating the relationship between the aforementioned factors and a firm’s valuation.

The rest of this paper proceeds as follows: section 3.2 contains a brief background on high-tech startups and a review of the existing literature on the relationship between the venture capitalist and the entrepreneur. Section 3.3 gives a description of the data-set and my empirical strategy. Section 3.4 discusses the results, and Section 3.5 has concluding thoughts.

3.2 Background on High-Tech Startups

I will give here a brief description of the lifecycle of a high-tech startup. A more detailed depiction can be found in Sahlman (1990). The story typically begins with one or several individuals having an inspiration for a product. Often times, it is an offshoot of
expertise built up in a given area while working at a prior job. The early stages (known as
the “garage phase”) usually entail individuals working at their existing jobs full-time and
meeting at nights or weekends to hammer out some structure to the product as well as
determining a product’s market potential. After a clear product definition is reached, the
firm initiates a round of financing with the venture community\(^{18}\). The early round of
financing goes to support product development costs, early equipment purchases, rent and
utilities, and the salaries of the founders. The primary purpose of this first round of funds is
to provide the firm with enough cash to produce a product demo illustrating proof of
concept. When a \(\frac{3}{4}\) functional product is complete\(^{19}\), the firm looks for partners to begin
testing of the product to work out the coding bugs and possibly add additional
functionality. After testing is complete, the firm is ready to begin shipping product. A new
round of financing dollars is needed again to ramp up the organization. Outside expertise
(such as a CEO, Channel Manager, Salesman) is secured in areas such as marketing and
PR to facilitate this ramp-up. If the company is able to reach profitability, there is then the
possibility of going public and arrange for another influx of cash to get the company to the
next level. This process from startup to profitability can range anywhere from three to five
years (Sahlman, 1990).

Previous Literature

By there very nature, high-tech ventures carry high levels of risk for those
involved. Unlike mature capital markets which revolve around a wealth of published data

---

\(^{18}\) If possible, many entrepreneurs will go to “Angel Investors”. These are typically high net worth
individuals who are willing to invest at a stage which might be too early for a venture capitalist, or to firms
who’s capital requirements fall below the minimum investment necessary for a venture capitalist ($1
million).

\(^{19}\) Early versions of the product are known as Alpha, or Beta versions.
about companies and their employees, investors in early stage ventures do not have reliable information about the firm's managers or prospects. While raising debt is an effective method for disciplining an unproven entrepreneur, this is an extremely unlikely event since early stage software companies have few tangible assets and no real cash flow. Investors therefore rely almost exclusively on some form of equity financing. Much of the current literature on venture capital financing has centered on mechanisms to resolve the agency problem between the venture capitalist and the entrepreneur. Economists have examined the vast array of monitoring tactics, contracting arrangements, and financial instruments utilized to deal with the information asymmetry inherent in this industry. I will review some of this literature here.

The most comprehensive work to date on venture capital organizations belongs to Sahlman (1990). Through extensive field research, Sahlman explores both the relationship between the investors and the venture capitalist as well as the venture capitalist and the entrepreneur. In reference to the second relationship, Sahlman argues that the venture capitalist attacks the agency problem between manager and shareholder in four ways. First, they structure the investments as staged capital infusions which are generally contingent on successfully reaching agreed upon milestones. The staging of capital injections allows venture capitalists to gather information and monitor the progress of firms while maintaining the option to periodically abandon a project. The earlier description given on the life-cycle of a high-tech startup illustrates this concept. Second, they devise compensation schemes that provide venture managers with appropriate incentives. Third, they become actively involved in managing the companies they fund. Venture capitalists sit on boards of directors, help recruit key individuals, manage relationships with
suppliers/customers, and even help map out business strategy. Fourth, they preserve mechanisms to make their investments liquid.

A significant literature has developed analyzing the first two methods Sahlman puts forth-the staging of capital infusions and devising compensation schemes to align the incentives of the venture capitalist and entrepreneur. Gompers (1995) provides empirical support to hypothesis that increased agency costs leads to greater likelihood of staged infusions by regressing the frequency of funding (i.e. the number of stages) on several proxies for agency costs. Agency costs increase as the tangibility of the assets decline, the share of growth options in the firm value rises, and asset specificity grows. He finds that the frequency of financing rounds as well as the amount of cash injected at any given stage decreases as agency costs rise. In addition, ventures which are particularly R&D intensive lead to shorter funding duration.

In regards to devising compensation schemes which provide entrepreneurs the appropriate incentives, Berglof (1994) and Landskrouner and Paroush (1995) model the optimal contract when the mutual goal of the relationship is a public offering (or other sale of business) at a later date. Their models explain the prevalence of convertible preferred stock to the venture capitalist and common stock for the entrepreneur. Common stock is an effective mechanism since the entrepreneur receives his payoff only if the company creates value. This forces him to bear some of the risk and encourages him to give high effort. The venture capitalist typically receives preferred equity thereby retaining the high-level privileges such as the right to sell the company.

3.3 Data and Empirical Strategy
The data for this study comes from a database known as VentureOne. This data-base tracks detailed information on companies who have entered the venture capitalist community for financing. A company is inserted into this database when it receives its first round of venture financing, and is tracked all the way to IPO, acquisition, failure, or whatever fortunes may befall it. A typical example of a company profile is inserted in the appendix. For each company, VentureOne begins with a basic company overview— the date of founding, industry the company is in, number of employees, public or privately held, etc. A list of investors and a detailed history of the financing (date money taken in, amount of money raised, post-money valuation, and stage of product) is the key piece of information which the database provides. Additional information on the executives and board members (including backgrounds and when they joined on) is sometimes available, and it concludes with some general information about the industry, competitors and market potential of the product. The data I use for this study begins with firms who entered the process as early as 1990, and as late as fourth quarter 1997. Although the dataset covers a variety of industries from hardware, software, biotech, information services, to health care, only firms who were classified as part of the software industry are used in this paper. A unit of observation in the data-set is an instance of a firm going out for a round of financing; a snapshot of the firm’s status is taken at this time and recorded. Thus a given firm may appear multiple times if it receives multiple rounds of financing. There are a total of 323 observations in this sample.

Construction of Variables

There are three basic sets of regressors I construct for the empirical analysis. They are the product stage, a measure of experience, and overall investor sentiment on high-tech
stocks captured by the Nasdaq composite index. I give an explanation for each of the regressor here.

Product stage is the most natural way to partition the sample space of startups. The industry has chosen the following six stages as its benchmarks for monitoring a company's growth process: startup, product development, beta/testing, shipping product, profitable and IPO. The definitions of each of these stages are as follows:

**STARTUP**- a product notion, preliminary scoping of product's market potential.

**PRODUCT DEVELOPMENT**- product clearly defined with requirements analysis, and high-level design and architecture complete. Business side (market potential, distribution, pricing, etc.) well under way and the core team of engineers is established.

**BETA/TESTING**- ¾-functional product available. Potential customers and partners are using product and providing feedback with changes incorporated into first release of software.

**SHIPPING PRODUCT**- First release of product ready to be sold. Full-fledged demo shown at conventions and trade shows.

**PROFITABLE**- Company reaches stage where it has the ability (although it may choose not to) to finance projects internally and is earning net profit. Subsequent versions of the initial product are being released at this time.

**IPO**- Initial Public Offering. Companies who are going public may actually be either in shipping or profitable. For a company to be ready for IPO, it must also reach some critical mass of size.

For a given firm, successfully reaching each of these milestones increases valuation since the underlying riskiness of the venture decreases with the passing of each stage. The fact that valuations appear to center on product stage implies that at the early stages, the greatest source of risk is whether or not the company can deliver a product with the stated functionality on time. Once a company is at the shipping product stage, other factors like distribution/adoptions risk become paramount. In the regressions, I insert a dummy for each product stage. The categories are obviously mutually exclusive.
Experience

Experience is another factor which venture companies will look for in order to determine company valuations. I have essentially defined two categories for experience. The first is whether or not an individual who is on the executive team is a previous founder of a high-tech startup. Members of the venture capitalist community stated in interviews that an individual who has gone through the process from startup to profitable distribution brings a tremendous amount of positive signaling to the negotiations of price. Often times, a successful past relationship is the best way to resolve uncertainty about the entrepreneur’s type. The second kind of experience is when an individual was previously an executive at a major high-tech firm or some other Fortune 500 firm. I observe that a great deal of startups are founded by ex-VPs of established technology companies like Oracle, Microsoft, Sun, IBM or Digital Equipment. In addition, there are many cases where someone with general business experience (not specific to high-tech) is asked to join a team and help with distribution, marketing, or some other standard business activity. The relevant trait I am trying to capture here is prior decision-making authority and accountability for one’s actions. As a final category, if the company executives do not have any prior experience defined above, they fall into the “no experience category”. If a firm has individuals in one of the first two categories, they are defined as having experience. Otherwise, they are put into the non-experienced bin.

Time Trend: Return on the Nasdaq Composite Index

The final regressor is the return on the Nasdaq Composite index for the calendar year prior to the one in which a software firm receives its funding. The Nasdaq Composite
Index measures all Nasdaq domestic and non-US based common stocks listed on The Nasdaq Stock Market, Inc. The constituent companies in this index are part of one of the following eight industries: bank, biotechnology, computer, industrial, insurance, other finance, telecommunications, and transportation. I insert this on the right hand side because the return on the index provides a proxy of how "bullish" investors are on software technology firms. The notion is that a good year for the Nasdaq composite would likely filter down to the venture financing market and increase the valuation of early stage firms. One might argue that the composite index is not an ideal measure for this purpose since the performance of firms in the transportation sector should not necessarily effect the outlook for software startups. An index composed solely of software companies would certainly be a more "pure" measure of the outlook for the industry in general. In fact, there is a more specific index called the Nasdaq Computer Index composed only of firms in the computer hardware and software industries. However, the data on this index is only available for 1994 onward rendering the Composite Index a better candidate for the analysis. The table below presents the annual returns to the index from 1989 to 1996.

<table>
<thead>
<tr>
<th>Year</th>
<th>Return</th>
</tr>
</thead>
<tbody>
<tr>
<td>1988</td>
<td>12.7%</td>
</tr>
<tr>
<td>1989</td>
<td>20.1%</td>
</tr>
<tr>
<td>1990</td>
<td>-18.6%</td>
</tr>
<tr>
<td>1991</td>
<td>57.5%</td>
</tr>
<tr>
<td>1992</td>
<td>15.4%</td>
</tr>
<tr>
<td>1993</td>
<td>15.6%</td>
</tr>
<tr>
<td>1994</td>
<td>-2.4%</td>
</tr>
<tr>
<td>1995</td>
<td>41.5%</td>
</tr>
<tr>
<td>1996</td>
<td>22.0%</td>
</tr>
</tbody>
</table>
Part of the strong performance of the index in 1995 and 1996 can be attributed to the arrival of the Internet and Web-based technologies. From 1994 to 1996, the Nasdaq Computer Index had annual returns of 14.5%, 61.5% and 38.6%. Most consider the initial public offering of Netscape Communications Corp. in August of 1995 as the unofficial beginning of the Internet explosion.

**Dependent Variable**

The dependent variable in these regressions is the natural log of the firm’s pre-money valuation. The reason I choose the log pre-money valuation specification as opposed to just the pre-money valuation itself is straightforward. The effect of the Nasdaq composite and the experience measure in absolute dollar terms may vary by product stage and is therefore better expressed as percentage increases over some baseline valuation for each product stage. For example, more experience may lead to an extra $20 million in pre-money valuation at the profitable stage but only an additional $3 million at the startup phase. Similarly, positive investor sentiment on high-tech stocks driven by higher returns on the Nasdaq may boost valuations of a company going public by $30 million, but a product development company by only $5 million.

The Venture capitalists typically determine the value of a company $v$, they then make a decision to invest $x$ dollars, and proceed to acquire $x/(v+x)$ equity stake in the company. Because the pre-money valuation is not explicitly available in the data-set, I back it out by subtracting the size of the investment made from the post-money valuation, both of which are available in the data.

The summary statistics for the VentureOne variables are provided in table 2.
TABLE 2  Means of Dataset

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Startup</td>
<td>0.096</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Product Development</td>
<td>0.207</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Beta/Testing</td>
<td>0.059</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Shipping Product</td>
<td>0.471</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Profitable</td>
<td>0.065</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>IPO</td>
<td>0.102</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Executive is previous founder</td>
<td>0.127</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Executive has general business background</td>
<td>0.836</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Executives have no experience</td>
<td>0.074</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Pre-Money Valuations (in millions of $'s)</td>
<td>32.43</td>
<td>$61</td>
<td>.24</td>
<td>590</td>
</tr>
<tr>
<td>Amount of $ Raised (in millions of $'s)</td>
<td>8.03</td>
<td>$14</td>
<td>.15</td>
<td>180</td>
</tr>
</tbody>
</table>

Most firms go out for financing at the product development stage or when shipping product. The shipping stage is when most firms need to ramp-up and bring on additional people and engage in downstream distribution, so that explains why nearly half the sample is accounted for by firms at this stage. Nearly 84% of the firms in my sample had an executive on board with previous experience in a high-tech company or a major general business, while only 7% have no prior experience. The mean pre-money valuation and the amount of money raised are not as telling as the significant variation associated with each variable. Startup phase firms can be priced under $1 million while companies which are going public range in the several hundred millions of dollars range. Similar ranges are present for the amount of money raised by each firm.

One final point on the data to be addressed is the issue of selection bias. In this data set, firms remain in the sample even if they are no longer in operation. Thus a firm which received financing at the startup or product development stage and then went out of business prior to reaching the next stage remains in the sample. This data set is not limited to companies who have gone public/acquired, or are still in production. Firms are inserted
as soon as they have taken a round of venture capitalist money, and are followed through no matter what fate they meet.

Empirical Specification
I run OLS regressions of the form:

\[ \log(\text{Pre-Money Valuation}) = \alpha + \beta_1(\text{Product Stage}) + \beta_2(\text{measure of experience}) + \beta_3(\text{Nasdaq Index}) + \varepsilon \]

In the first specification, I regress log pre-money valuation only on product stage with the startup stage as the omitted category. In specification two, I include the return on the Nasdaq Composite index for the year prior to the one in which the firm receives its funding. In the third specification, I include the dummy indicating whether or not members of the executive team have prior experience. Finally, the fourth regression includes all the available regressors.

3.4 Results
The results of the regressions are shown in table 3. In specification one, the coefficients on product stage show the expected increase in pre-money valuation as a company proceeds through its development. In addition, the estimates themselves are all significant at conventional levels. Generally speaking the \(t\)-statistics increase with product stage (with the exception of the beta/testing stage) indicating that a firm’s valuation becomes more precise as they become closer and closer to its IPO. This result is consistent with the notion that it becomes much easier to pin down the potential cash flows of a company at the later stages as the execution risk associated with the venture declines. The product development and beta/testing coefficients are both approximately 1, the shipping
product coefficient roughly 2, the profitable coefficient about 3, and IPO about 4. These results also provide estimates of the valuation path a startup will follow contingent on reaching each successive product stage. Roughly speaking, successfully passing each of the aforementioned milestones will lead to increased valuations on the order of 100% per stage.

The first hypothesis I formally test is whether or not the coefficients on the product stages are really different from one another. I run a series of Wald tests and conclude the following: the product development and beta/testing coefficients are not significantly different from one another, but the later stage coefficients on shipping product, profitable and IPO are sufficiently distinct from the one another. An F-test on the null hypothesis that the coefficient on the product development dummy and the beta/testing dummy are indistinguishable yields a test statistic of .92 and an associated p-value of .34. Thus we fail to reject. However, the F-statistic on the null that the beta/testing dummy and the shipping product are identical is 7.5 with a p-value of .0065. In short, these pair-wise F-tests reject the null of identical coefficients for all adjacent product stage pairings with the exception of the product development and beta/testing dummies. This latter result appears to be driven by the lack of firms in the sample (5.9%) who received financing at the beta/testing phase.

In specification two, the Nasdaq Composite Indicator is inserted into the right-hand side. The coefficients and standard errors on our product stage dummies are essentially unaltered, and the estimate on the Nasdaq Composite Index is .566 and significant at the 5% level. This result confirms our expectation that better performance by the Nasdaq produces higher pre-money valuations for firms at early stage financing. Investors will be
overall more bullish on tech stocks thus driving their prices up even at the early stage level. In addition, the magnitude of the Nasdaq effect is reasonable. If the return on the Nasdaq in year $t$ is 50% and then increases to 100% in year $(t+1)$, this translates into a 28% increase in the valuation of software ventures for year $(t+1)$ over year $t$.

In specification three, I introduce the experience effect. This is also significant at the 5% level, with a point estimate of .39. This effect is quite large- an experienced firm will receive nearly a 40% higher valuation than a non-experienced firm. This suggests that in the minds of the venture capitalist, the market signal provided by prior experience in the technology or general business sector is an informative one. First-time entrepreneurs are penalized heavily for the paucity of information available concerning the entrepreneur's type. Prior experience and the references associated with them appears to be a significant factor for an early stage investor. Given the previous literature on the topic, this result is also not surprising. Much of the earlier papers in this field discuss the mechanisms utilized to resolve this information asymmetry between venture capitalist and entrepreneur. Apparently, the know-how brought on by having gone through the process before from startup to acquisition/IPO goes a long way to achieving this objective.

In a separate regression not reported in table 3, I regress the log pre-money valuation on product stages and interact the experience dummy with the product stage. This allows me to test the hypothesis that the effect of experience depends on the product stage the firm is at when it goes for financing. The estimates for experience interacted with startup and product development are insignificant at conventional levels. The coefficients and standard errors are (.59, .64) and (-.09, .33), respectively. The estimates on the later stages, however, are marginally significant. For the shipping and profitable stages, the coefficients
and standard errors are (.41, .26) and (1.35, .90). This is consistent with our expectation—since the early stages are typically “in the garage” there is less management expertise required than at shipping and profitability. Later stages are characterized by managing a distribution channel, conducting marketing campaigns, and providing customer service and support. The increased number of outside help needed to run a business requires an experienced management team who has done the routine before. While the results do suggest some differential effect of experience across product stages, we fail to reject the $F$-test on the null that the four coefficients are jointly all zero. The $F$-statistic for this test is 1.46 which yields a $p$-value of .21.

The final specification inserts product stage, the Nasdaq Composite Index and the experience dummy as regressors. The estimates come out roughly similar to those for the prior three regressions, and we reject the null that the coefficients on the experience dummy and the Nasdaq Composite are jointly zero. In summary, product stage seems to be the dominant factor in determining the valuation of an early stage venture. The return on the Nasdaq Composite in the year prior and prior experience are statistically significant but small relative to the coefficients on the product stage dummies.

3.5 Conclusions

Information technology has become an increasingly important part of our economy. In the past, it was relegated to the back-office, decision-support domain but it is now making its way into the forefront of modern business. The high-tech software startup in Silicon Valley, California and Rte. 128, Massachusetts has established itself as the entity most responsible for pushing the envelope on innovation further and faster than ever
before. The venture capitalist is the party most responsible for providing financing, discipline and ultimate viability of these nascent companies.

The large information asymmetry between the investor and entrepreneur renders the due diligence and selection of a successful venture an extremely difficult process. In addition, conditional upon selection, determining a consensus price for a high-tech deal is shrouded in uncertainty. Anecdotally, many claim that pricing a deal is a simple case of willing buyer meeting willing seller rendering consensus market price a non-viable concept. However, in this paper, I find that several factors go a long way towards determining the price of a company. They are the product stage of a given venture, experience of the management/technical team, and overall investor sentiment on high-tech captured by the return on the Nasdaq Composite Index. The result on the experience regressor indicates that past history is an extremely effective way to resolve the well-documented information asymmetry problem between the entrepreneur and the venture capitalist. There are surely other important factors affecting the pricing of early stage software ventures, but they were not made available to me in this data set.

An interesting direction to take this line of research might be to ascertain if there are systematic differences in the performance of startups financed by certain venture capitalists over others. A firm-specific venture capitalist effect was left out of the analysis, so this paper does not address the issue. However, it has traditionally been claimed that true value-added of a venture capitalist is not really the money they provide, but access to the network into potential personnel and buyers of a startup’s product. A separate question one may try to answer with this data is whether or not startups who are financed by certain venture capitalists outperform those financed by others.
Table 3
OLS Regression Results
Dependent Variable is Log(Pre-Money Valuation)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.76</td>
<td>0.68</td>
<td>0.39</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>(0.16)</td>
<td>(0.16)</td>
<td>(0.24)</td>
<td>(0.24)</td>
</tr>
<tr>
<td>Product Development</td>
<td>0.95</td>
<td>0.95</td>
<td>1.01</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>(0.19)</td>
<td>(0.19)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>Beta/Testing</td>
<td>1.21</td>
<td>1.17</td>
<td>1.18</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>(0.26)</td>
<td>(0.26)</td>
<td>(0.26)</td>
<td>(0.26)</td>
</tr>
<tr>
<td>Shipping Product</td>
<td>1.80</td>
<td>1.78</td>
<td>1.81</td>
<td>1.79</td>
</tr>
<tr>
<td></td>
<td>(0.18)</td>
<td>(0.18)</td>
<td>(0.18)</td>
<td>(0.17)</td>
</tr>
<tr>
<td>Profitable</td>
<td>2.88</td>
<td>2.89</td>
<td>2.88</td>
<td>2.89</td>
</tr>
<tr>
<td></td>
<td>(0.25)</td>
<td>(0.25)</td>
<td>(0.25)</td>
<td>(0.25)</td>
</tr>
<tr>
<td>IPO</td>
<td>3.85</td>
<td>3.79</td>
<td>3.84</td>
<td>3.79</td>
</tr>
<tr>
<td></td>
<td>(0.22)</td>
<td>(0.22)</td>
<td>(0.22)</td>
<td>(0.22)</td>
</tr>
<tr>
<td>Return on Nasdaq Composite</td>
<td>0.57</td>
<td>0.54</td>
<td>0.39</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>(0.25)</td>
<td></td>
<td>(0.19)</td>
<td></td>
</tr>
<tr>
<td>Experience Measure</td>
<td></td>
<td>0.37</td>
<td>0.37</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>0.55</td>
<td>0.56</td>
<td>0.56</td>
<td>0.56</td>
</tr>
</tbody>
</table>

Note: Standard errors are in parentheses, N=323
Chapter 4
Disaggregation of the Software Value Chain

4.1 Introduction

The last several years have witnessed dramatic changes in the information technology sector. First and foremost, the pace of innovation in both hardware and software has reached unprecedented heights. Object-oriented and Internet-based technologies have reduced software product life-cycles to a matter of months, and processor speeds continue to double according to Moore’s Law\(^{20}\). Secondly, integration of competing standards and technologies offered by different vendors has increased the technical complexity of today’s information systems. It is not uncommon for a single company’s information system to be composed of databases from the legacy era, peripherals\(^{21}\) sharing using client/server technologies, and front-end applications accessed through a browser based on the protocols of the World Wide Web.

Perhaps the most significant implication of these changes for the rest of the economy is the following: the increase in the pace of innovation represents an opportunity for the early adopters of cutting edge technologies to differentiate themselves from the competition and earn the rents associated with being a first mover. Differentiation can take place both at the demand side through enhanced product/service offerings as well as on the supply side via greater productive efficiency. A simple example of superior information technology utilized as a method for winning consumers is commercial banks establishing a Web presence to offer customers access to accounts over the Internet. Knowledge workers

\(^{20}\) Processor Speeds will double every eighteen months.
\(^{21}\) This includes machines like printers and scanners.
who spend considerable amounts of time on their computers appreciate the convenience and visual representation of their account information made possible through a browser. The telephone does not provide the same visual element, and the ATM/bank branch does not have the locational convenience. The traditional example of information technology as a cost cutting mechanism is automated workflow software which reduces and even replaces the amount of human input necessary to carry out standard back-office tasks. A third example of information technology contributing to productivity is the use of computer-aided design (CAD) software to enhance the work of engineers. Thus unlike many other business processes which are considered ancillary to a company’s core operations and therefore outsourced to specialty firms, information technology has the potential to transform organizations along many dimensions.

From an industrial organization perspective, one of the more interesting developments has been the software industry’s structural response to this rapid pace of innovation. While the changes I discuss here are just recently emerging and far from complete, there is a discernible industry trend best described as one of disaggregation of the software value-chain. What I mean by this is the following: the kinds of firms engaging in the upstream innovation end of the value chain are very different from the firms who carry out the implementation of the technology at the end-user level. In addition, both large and small corporations are beginning to outsource the vast majority of their information system needs to companies who specialize in information technology, as opposed to developing these systems in-house. Thus the creator, the distributor, and the end-user of software are all distinct entities in the modern information economy.
This is somewhat of a recent trend- in the earlier days of client/server and mainframe computing, it was not unusual for a major corporation to maintain in-house a full-time staff of programmers and tech gurus who fulfilled all the IT needs of the company. These individuals created from scratch the entire information system and customized software applications tailored to the specific needs of individuals at all levels in the organization. Network maintenance, software upgrades, and education on how to use the software was also done internally. These days it is more common for the major corporations to outsource a significant portion of their technology needs to a large systems integrator such as IBM, Digital Equipment and Electronic Data Services (EDS). In addition, these systems integrators are creating less and less of the software for their customers from scratch and are instead buying technologies from startups and other entities who specialize in software innovation. The systems integrator primarily performs end-user level customization of software to specific business needs, as well as customer support and maintenance. In summary, in-house information system delivery is being replaced by outsourcing, and systems integrators are favoring purchasing of software technologies over internal development.

The objective of this essay is to explain these recent shifts using modern industrial organization theory as the framework for analysis. As a brief background and introduction to the various players in the industry, I will give a concrete example of the value chain in action. In the downstream market, the buyer of information technology represented by firm $I$ produces some good or service $y$. As aforementioned, IT can be used both to increase productive efficiency as well as enhance a firm's product/service offering. Firm 2,
represented by a systems integrator (SI), value-added reseller (VAR), or an IT consulting firm provides a one-stop shop for all of firm 1’s IT needs. Firm 1 will explain its business processes and IT needs to firm 2, who then proceeds to implement a solution. This entails finding the necessary hardware/software components available and providing customization and/or configuration of the components. Firm 2 will then work closely with firm 1 to explain how the system works and educating firm 1’s staff on how to use the final applications. Once everything is in place and workers are using the system, there is still the need to provide customer support, upgrades of software/hardware components, and basic maintenance of the network. Customer support in particular is a very time-intensive process requiring knowledge not only of the technology but how the average non-tech savvy worker responds to it. In short, there is a principle-agent relationship with firm 1 relying heavily on the knowledge and advice of firm 2.

In the upstream market firm 2 will purchase the underlying technology from startups, or packaged software vendors such as Lotus and Netscape. This configuration/customization of the software and customer support are the main value-added services offered by firm 2. Typically, a client/server software package such as Lotus’ Notes program requires extensive customization as well as education on its functionality prior to being an effective tool for usage. Firm 2 provides these services necessary to bring the technology to the end-user.

4.2 Downstream Disaggregation: Long Term Contracts

With that as the background, the first issue I address is a major trend we see nowadays- that of corporations outsourcing their IT needs to a systems integrator and
locking that integrator in for the long run in 5-10 year contracts. Some of the more recent high-profile examples of this trend include the following:

- Swiss Bank Corp. signed up Perot Systems Corp. for $250 million a year to run its worldwide IT operations for at least ten years.

- J.P. Morgan & Co. agrees to pay a hand-picked quartet of outsourcers- Andersen Consulting, CSC, AT&T and Bell Atlantic Corp- as much as $2 billion to build and maintain its networks and create customized applications for the next seven years.

- Chemical giant DuPont Co. contracts with Andersen Consulting and CSC at $4 billion plus for 10 years of services.

(Computer World, July 28, 1997)

The presence of the long-term contract is interesting. To explain this, we need not look any further than the rationale put forth in Joskow (1987). This paper provided empirical support to the notion that as investments become increasingly relationship specific, long-term contracts become the chosen mechanism to avoid the consequences of the hold-up problem. This is exactly the situation we have between the firm 1 and 2. After firm 2 implements a complete IT solution, he has firm 2 at his mercy. If the knowledge captured in the setup and configuration of an information system for a given business were to unexpectedly go on strike, many modern corporations would be in disarray. Today’s information systems provide not just decision support but operation support as well, hence many firms would not be able to operate without their IT at maximal efficiency. Firm 2 therefore has the incentive midway into the relationship to hold out and take advantage of
the tremendous switching cost associated with the corporation having to find a new outsourcing partner. The best way for firm \( I \) to protect themselves against this happening is to sign firm 2 to a long-term contract which has provisions for upgrades, maintenance and extensive customer support. \(^{23}\)

Long term contracts act as a substitute for vertical integration. The natural question following from the above analysis is why not full integration? The two reasons I give here can again be traced to the rapid pace of innovation in this sector. First, an SI is responsible for providing an integrated information technology solution incorporating the latest, best of breed technologies. It is extremely expensive to keep in-house the kind of expertise and knowledge necessary to stay up to date with the latest advancements, since the outside opportunities (represented as offering the same service to other corporations) to individuals with that kind of skill are extremely profitable. There simply are not enough people with this kind of knowledge for everyone to keep them in-house. Secondly, as outsourcing information technology becomes increasingly popular, more firms will enter the industry and this puts the SI in an increasingly competitive environment. Competitive forces will ensure that in order to continue winning new long-term contracts, the SI must demonstrate that it is indeed always on the cutting edge. A fully integrated information technology division would be more inclined to "rest on their laurels", but an SI firm who is constantly

\(^{22}\) Concrete examples of this are customer support at a credit card company and the trading floor at an investment bank.

\(^{23}\) An important consequence of this contractual arrangement is that businesses like firm 2 become the gateway into the corporation for software and hardware products. As the agent in the principle-agent relationship, firm 2 has the authority to decide which vendor's software, and hardware is used to meet the client's needs. The corporation is effectively a captive consumer given the tremendous switching costs associated with finding a new outsourcing partner. Players in the upstream market must go through companies like firm 2 in order to initiate adoption of their technologies. This trend becomes important when we examine the dynamics of the upstream market.
trying to win more contracts would have the greater incentive to always be a step ahead of their competitors.

In summary, the bidding for long-term outsourcing contracts has emerged as an effective mechanism for dealing with the potential hold-up problem between the SI and the end-user. In addition, the competitive pressures associated with winning new contracts gives an SI the incentive to stay on the cutting edge of technology. It is worth addressing at this point how the reasons for outsourcing IT differ from the rationale behind other forms of business process outsourcing (BPO). Take for example one of the most commonly outsourced functions—payroll. Payroll is substantially different from IT along several dimensions: the first is that unlike IT, payroll is a non-mission-critical application. Secondly, the corporation does not need to make the same degree of relationship specific investments to the firm they outsource payroll to the way they do to their SI shop. Thus there is less potential for hold-up. Consequently, the traditional rationale behind outsourcing payroll and other similar areas of operations lies in the ability for firms to focus on core competencies when such areas are outsourced. However, industries in which IT is a mission critical application with high levels of relationship specific investments by the member firms (such as banking, insurance and financial services) outsource IT for reasons put forth earlier. According to leading industry analyst Gartner Group, when the ten largest SIs were asked to list their top three vertical markets served, seven of the ten wrote banking, insurance and financial services. In addition, three of the five largest VARs put banking, insurance and financial services as their most intensively served vertical market. This suggests that firms for whom the best information technology is a necessary component to remain viable do indeed outsource to the leading SIs and VARs.
4.3 Upstream Disaggregation: Spot Market for Technologies

The other interesting trend we observe in the industry today is the separation of software creation from the downstream processes of software delivery and support. Recent data from International Data Corp. on operating budgets of the major SIs, VARs and IT Consultants demonstrate that funds are migrating away from internal R&D and moving more towards marketing, customer support, and consulting. I infer from this that companies like firm 2 are focusing more on “owning the customer” through maintenance of customer relations and less on providing the underlying technological innovation. These companies are increasingly looking upstream to acquire (or secure exclusive licensing contracts with) the startup firms instead of trying to come up with the innovation on their own. They are finding that the smaller, more nimble startup is better suited to the process of innovation, and that it is cheaper to “purchase technologies” than to try and develop them in-house.

Why won’t an SI, or IT consultant spend money on internal R&D? There are two primary reasons for this: the first is that incentive to innovate is significantly higher for a startup. The technologies being developed in the startup community are in some sense, business unaware. By this, I mean that the technology is not specific to the business process of a particular vertical industry, or the needs of any given department within a firm (such as accounting, human resources, etc.). By their very nature, many Internet/Web-related technologies are business unaware. Since most SIs specialize their efforts on the technology needs of a specific vertical industry or particular departments within a company, the market potential of these products exceed the customer base of any one SI.
Therefore, the startup is better able to extract the rents to the innovation by selling their business unaware technologies to a variety of distributors who focus on different vertical markets. Secondly and perhaps more significantly, the SI is not in the best position to finance the lengthy and risky process of internal R&D to come up with next-generation technologies. Who then is in the better position to bear this risk? The venture capitalist community. Because these technology funds range in the hundreds of million to billion dollar range, they are able to diversify away a good portion of the risk by investing in large numbers of startups all producing competing technologies. The majority of downstream players with their far more limited budgets are unable to take on this kind of risk. Hence the venture capitalist industry together with the high-tech entrepreneur has emerged as the R&D arm of the software industry. Once a technology is fully developed and ready, the downstream players purchase outright or license the technologies he needs to meet his specific customers needs. A recent example of this process in action was Netscape Communications Corp. purchase of a two-year old startup called Kiva Software. Kiva produced a Web middleware technology providing complementary services to Netscape’s core technology offering.

A final major reason why creation is separate from distribution has to do with fear of product/technology bias on the part of the SI. If a company knows that an SI develops all the software in-house, then it’s only natural for the buyer to concern itself with whether or not he is really receiving the best of breed software technology available. For example, in response to the recent Computer Associates bid for CSC, CSC customers voiced their concern over the possibility of CSC using the CA’s flagship software development platform even when a competitive platform might be superior. Again, with the pace of
innovation moving at such breakneck speeds, corporations are finding it necessary to have
the cutting edge technology in order to get a step ahead or simply remain competitive in
their respective industries. They do not want to be concerned about their systems integrator
delivering a sub-par technology simply because it's been developed in-house- corporations
demand the state-of-the-art no matter which entity is providing the technology. With
software creation being handled by startups and other entities not focused on downstream
distribution, the SI can simply objectively select the best technology and concern
themselves only with customization, maintenance and support.

4.4 Conclusions

This essay has examined the forces which have initiated the disaggregation of the
software value chain. By no means is the trend complete- in fact, we can point to many
firms which participate in two or more parts of the value chain. Most ostensibly, Microsoft
Corp. does a tremendous amount of internal research as well as having one of the most
extensive distribution channels in the industry. In addition, firms like IBM perform a
significant amount of both software and hardware innovation while remaining the largest
systems integrator in the country in terms of revenue. But we feel that as the pace of
innovation spurred by Internet related technologies continues, the segmentation will
become more pronounced. Business will increasingly outsource their IT needs to the
service/consulting based companies like CSC and EDS. The primary value these
companies bring will be the services and support for the end-user. These large systems
integrators will become the effective distribution channel into the end-user. On the
upstream side, the small, flexible startup financed by the venture capitalist will become the
bastion of innovation. Licensing agreements and the purchasing of companies through the M&A market will be the mechanism to transfer technology from the innovator to the distributor.
BIBLIOGRAPHY


