ESSAYS ON TECHNOLOGY, FINANCE, AND MACROECONOMICS

By

Lihong Han

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Economics

August, 2008

Nashville, Tennessee

Approved:

Professor Peter L. Rousseau

Professor Robert A. Driskill

Professor Kevin Huang

Professor David C. Parsley

ACKNOWLEDGMENTS

My special gratitude goes to my advisor Professor Peter L. Rousseau, who gives me enormous help and support with my research. I would also like to give special thanks to my committee members—Professor Robert A. Driskill, Professor Kevin Huang and Professor David C. Parsley for their insightful comments and useful discussions.

I also have been benefited from the excellent comments from my presentations at Vanderbilt University, the 7th EWC International Graduate Student Conference, and the 2nd Small Open Economies in a Globalized World Conference. In particular, I thank professors Jeremy Atack, Eric Bond, Yanqin Fan, James Foster, Hui He, and Boyan Jovanovic.

I would like to thank as well the School of Arts and Science, Graduate School and the Department of Economics at Vanderbilt University for their generous financial support of my graduate studies and research. In addition, I am thankful to Vanderbilt University Center for Ethics for its financial support and the dissertation writing group organized under this center for the discussion and editing on the second chapter of my dissertation. I also want to thank Ms. Kathleen Finn for her kindness and professional assistance.

Finally, I want to thank my parents, Jianxin Han and Yufen Li, as well as my young brother, Lixiang Han, for their endless and unconditional love. And most importantly, I would like to thank my dearest husband, Chih-Wei Wang, who gives me the endless support and provides me a warm and happy environment that enables me to finish my work.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS iii

LIST OF TABLES vi

LIST OF FIGURES ix

Chapter

I INTRODUCTION 1

II TECHNOLOGICAL DISTANCE, TOBIN'S Q, AND THE PROPENSITY TO MERGE 3

Introduction 3

Model 7

Data Construction and Description 16

Econometric Method 19

Efficiency of the Quasi-bootstrap Logistic Regressions 19

Bootstrap Tests 21

Correction for the Estimates 23

Evidence 24

Basic Model Assumptions and their Implications for the Data 24

Relation between Q and TFP 27

Effects of q_r and λ on the Likelihood of Mergers by Controlling for Alternative Explanations 28

Robustness Tests 34

Conclusion 35

III A SIMULATION-BASED METHOD FOR RARE EVENTS DATA 39

Introduction 39

The Choice-Based Sample Probability Bias 43

The Relation between the Sample-Based and Population-Based Merger Probability 43

Prior Correction of the Estimates from Logit Model 48

A Simulation-Based Method 50

Efficiency of the Quasi-bootstrap Logistic Regressions 51

Bootstrap Tests 53

Correction for the Estimates 55

Conclusion 56

IV INTERNATIONAL TECHNOLOGY DIFFERENCES AND INCOME INEQUALITY: EVIDENCE FROM CROSS-COUNTRY HISTORICAL DATA 58

Introduction 58

Data 62

Empirical Evidence 63
Conclusion 81

BIBLIOGRAPHY 99

LIST OF TABLES

Table Page

1 Summary statistics of observations in each industry 18

2 Characteristics of non-merging, target and acquiring firms before matching 25

3 Summary statistics of z_r and q_r in each industry 26

4 Pooled regressions of $q_{i,t}$ on $z_{i,t}$, 1986-2005.. 28

5 Quasi-bootstrap logistic regressions for specifications that control for k while also allowing z_r and q_r to enter separately with both linear and quadratic terms.......................29

6 Quasi-bootstrap logistic regressions for specifications that control for k while also allowing λ to enter with both linear and quadratic terms 32

7 SUR logistic regressions on the Kronecker Products of ($\log k, (\log k)^2, \log q_r, (\log q_r)^2$) and industry dummies 37

8 SUR logistic regressions on the Kronecker Products of ($\log k, (\log k)^2, \log \lambda$) and industry dummies... 38

9 Technology variables and their measurement 83

10 (Continued) Technology variables and their measurement 84

11 (Continued) Technology variables and their measurement 85

12 Invention data and β convergence rate across all countries, DCs and LDCs for technologies based on per-capita or per-unit-real-GDP... 86

13 (Continued) Invention data and β convergence rate across all countries, DCs and LDCs for technologies based on per-capita or per-unit-real-GDP.................................... 87

14 (Continued) Invention data and β convergence rate across all countries, DCs and LDCs for technologies based on per-capita or per-unit-real-GDP.................................... 88

15 Invention data and β convergence rate across all countries, DCs and LDCs for technologies measured in share 89

16 Summary of kernel density estimates and their feature tests of technologies measured on per capita basis.. 90

17 (Continued) Summary of kernel density estimates and their feature tests of technologies

measured on per capita basis 91
18 Summary of kernel density estimates and their feature tests of technologies measured on share and on unit real GDP based.......... 92

19 Each technology β convergence rates in 1919-1938, 1946-1972 and 1973-2000 across all countries 93

20 (Continued) Each technology β convergence rates in 1919-1938, 1946-1972 and 1973-2000 across all countries.......... 94

21 (Continued) Each technology β convergence rates in 1919-1938, 1946-1972 and 1973-2000 across all countries.......... 95

22 Average β convergence speed of technologies and income per capita in 1919-1938, 1946-1972 and 1973-2000 across all countries, DCs as well as LDCs 96

23 Income per capita β convergence rate and average β convergence rate in each department in 1919-1938, 1946-1972 and 1973-2000 across all countries 96

24 Income per capita β convergence rate and average β convergence rate in each department in 1919-1938, 1946-1972 and 1973-2000 across DCs.......... 97

25 Income per capita β convergence rate and average β convergence rate in each department in 1919-1938, 1946-1972 and 1973-2000 across LDCs.......... 97

26 Dispersion in technology adoption relative to dispersion in income per capita 98

LIST OF FIGURES

Figure Page

1 Tobin's Q for subgroups of Compustat firms, 1986 - 2005 4

2 Gains from mergers 13

3 q_r, z_r and λ's effect on the probability of mergers 31

4 q_r and λ's effect on the probability of mergers... 33

5 The densities of Y=1 and Y=0 separately. 40

6 Kernel density estimates and their corresponding contour plots of income per capita and electricity production in year 1950, 1960, 1970, 1980, 1990 and 2000………………...104

7 Significance in Scale-Space analysis of the joint density of income per capita and electricity production at year 1950...106

8 Significance in Scale-Space analysis of the joint density of income per capita and electricity production at year 1960...107

9 Significance in Scale-Space analysis of the joint density of income per capita and electricity production at year 1970...108

10 Significance in Scale-Space analysis of the joint density of income per capita and electricity production at year 1980...109

11 Significance in Scale-Space analysis of the joint density of income per capita and electricity production at year 1990...110

12 Significance in Scale-Space analysis of the joint density of income per capita and electricity production at year 2000...111

CHAPTER I

INTRODUCTION

My dissertation evaluates technology adoption and transfer both theoretically and empirically, with the goal of providing new insights into the consequences of technology adoption – an area that remains understudied in the macroeconomic literature.

I begin the investigation from the perspective of firms. The major aim of my second chapter is to analyze the mechanism of technology adoption across firms under a framework in which technology is transferred through mergers and acquisitions. In this chapter I present a model that incorporates the cost of converting one firm's specific capital into that of another firm. I show that merger activity involves a pattern in which firms that have high-market valuations with respect to the book value of their assets (i.e., Tobin's Q) will merge with firms that have lower yet not the lowest valuations. I also show that the ratio of bidder to target Qs and the size differential between acquirer and target have an inverted U-shaped effect on the probability of two firms being involved in a merger, and that the likelihood of a merger is positively and linearly related to the relative potential growth between acquirer and its target. In terms of potential growth, the typical merger pattern is 'high buys low'. Based on data for mergers among US firms available from the Securities Data Corporation from 1986 to 2005, a series of bootstrap logistic regressions of the probability of an actual merger on the ratio of bidder to target Qs, the two firms' size differential, their relative potential growth, and other controls bear out the main implications of the model.

The third chapter provides a theoretical justification for the bootstrap logistic regressions, a new simulation-based method, for rare events data in which the binary dependent variables have dozens to thousands of times fewer ones (events, such as mergers) than zeros ("nonevents", such as pseudo mergers). The essentials of this method include the following: First, the "nonevents" (pseudo mergers) are randomly selected to match the events (mergers) and the logistic regression is applied. This procedure is then repeated hundreds of times. We also construct the bootstrap standard errors and p values of the estimates.

The last chapter analyzes the Cross-Country Historical Adoption of Technology (CHAT) data set, which covers the diffusion of about 110 technologies in over 150 countries since 1820. We estimate and compare the convergence speed of each technological adoption and that of income per capita across all countries, and then across the developed and developing countries (DCs and LDCs). We then document six general facts about cross-country technology adoption and income inequality that emerge from these data: (i) Though DCs always adopt a new technology earlier than LDCs, on average the convergence speed of technology adoption across LDCs is faster than that across DCs. (ii) Most technological adoptions among poorer economies cluster in a lower level than those among richer economies. (iii) The convergence speed of the adoption of most technologies is non-monotone. (iv) The invention of the computer and the internet has not increased the average convergence speed of other technological adoptions. (v) The relation between the average convergence speed of technological adoptions and that of per capita income is negative across all countries and across LDCs, but is positive across DCs in the post-WWII period. (vi) The dispersion in technology adoption for individual technology is 3 - 5 times larger than the dispersion in income per capita both across DCs and LDCs.

CHAPTER II

TECHNOLOGICAL DISTANCE, TOBIN'S Q, AND THE PROPENSITY TO MERGE

Introduction

According to the Securities Data Corporation (SDC), more than 3,700 firms were involved in domestic within-industry corporate mergers between 1987 and 2006 with a total transaction value in excess 4.7 trillion constant 2005 dollars. At the same time, the vast majority of US companies were not involved in mergers. Accordingly, the topic of who merges, and with whom do they merge, has received a great deal of attention in recent economic and financial research.

One of the most well-established stylized facts about the pattern of mergers and acquisitions (M&A) is that firms with high values of Tobin's Q (defined as the ratio of a firm's market value to the replacement cost of its capital) usually buy firms with low Qs. Manne (1965), for example, argues that low value, badly-managed firms will be bought by better-managed firms. Servaes (1991) finds that the total takeover returns, which measure the abnormal increase in the combined values of the merging parties, are larger when the bidder has a higher Q than its target. Andrade, Mitchell and Stafford (2001), report that roughly two-thirds of mergers since 1973 involve an acquiring firm with a higher Q than its target. Jovanovic and Rousseau (2002, 2007) provide a Q-theory of mergers to capture these stylized facts.

Another group of researchers thinks that valuation errors affect merger activity. For example, Rhodes-Kropf and Viswanathan (2004) and Shleifer and Vishny (2003) provide theories suggesting that misvaluations drive mergers. Rhodes-Kropf, Robinson and

Viswanathan (2005) find strong empirical support for the prediction that increasing market misvaluation of a firm increases the probability of being the acquirer when a merger occurs.

In this paper, we take a fresh look at who merges with whom. Our study is motivated by Jovanovic and Rousseau (2002, 2007), who describe a theory of mergers in which firms with high Qs acquire firms with low Qs, since the most value is created when the worst performing assets are paired with the best managers. They argue that mergers are a way for acquirers to pass their better technology to targets, or to substitute the target's poor management or inappropriate use of assets with superior management and direction. Synergies are created in all of these cases. In their papers, they assume that capital is firm specific and a cost is needed to put new and used capital in place. When the conversion cost is infinite, there would be no net gains from a merger and thus no merger regardless of the difference between the counter parties' Qs.

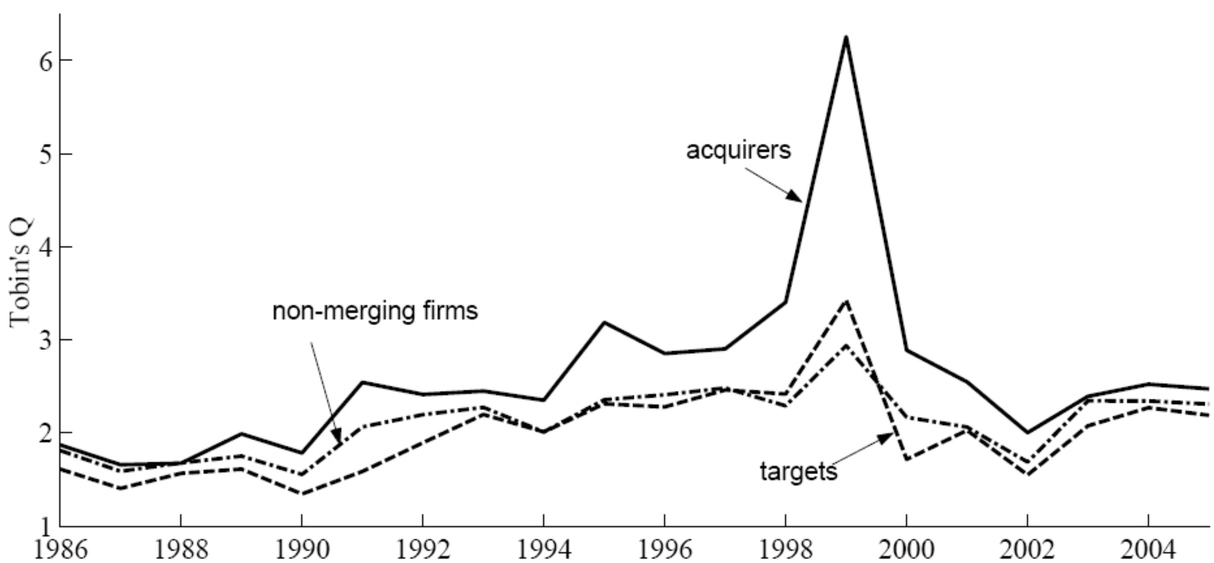


Figure 1. Tobin's Q for subgroups of Compustat firms, 1986 - 2005

As Figure 1 shows, however, the Tobin's Q's of firms that become acquirers, while larger than those of their targets, on average do not exceed those of the targets more than they exceed the Q's of firms not involved in mergers at all.[1] This suggests that a high Q firm will purchase a firm with a lower, yet not the lowest Q, considering the conversion cost. This phenomenon is the focus of our paper.

The 'high buys less high' merger pattern has been observed by Rhodes-Kropf, Robinson and Viswanathan (2005). Instead of explaining this pattern, Rhodes-Kropf and Robinson (2007) suggest an even stronger pattern of 'like buys like', arguing that mergers reflect a desire to place complementary assets under common control more than the substitution of badly performing assets with better ones.

Keeping the substitution assumption, however, we argue that a 'high buys lower yet not the lowest' pattern in terms of Q emerges from the firm specificity of capital and the costs associated with converting a target's capital into a form usable by the acquirer. We build a model based on three assumptions. First, we assume there is a positive technology shock which affects one group of firms more than the others. After this shock, two kinds of firms will coexist. Second, like Jovanovic and Rousseau (2002, 2007), we assume that mergers are a channel through which capital flows from low technology projects to high technology ones. Finally, we assume that the total conversion costs are a convex function of the technological "distance" between the acquirer and its target. Firms negotiate to determine how they will share the surplus generated by the merger. If a high-tech and low-tech firm can make higher profits under common control than they can separately, they will merge. Guided by the model developed in the paper, we confirm that merger activity presents a pattern as 'high buys lower yet not the lowest' in terms of Q, but 'high buys low' in terms of relative growth potentials. We also prove the following results:

[1] The annual averages presented in Figure 1 all constructed for US firms listed on Standard & Poor's Compustat database from 1986 to 2005. We identify mergers among these firms using information from the Securities Data Corporation (SDC).

(1) The ratio of bidder to its target Qs has an inverted U-shaped effect on the probability of the two firms being involved in a merger. This means the bidder may not purchase the lowest Q firm that it can find, but rather a firm with a lower Q.

(2) The likelihood of a merger is positively and linearly related to the relative potential growth between an acquirer and its prospective target, which means that the likelihood of a merger is larger when the potential growth of the acquirer is sizeable relative to its target. Therefore, the relative potential growth is a more potent variable than Q in characterizing what drives merger activity.

We restrict our data to US domestic merger activities reported in the SDC and US exchange companies from 1986 to 2005. We pool a dataset with 1,317 merger pairs among 3,050,489 observations. Our data support the merger pattern of 'high buys lower yet not the lowest' in terms of Q. At the same time, we find the average total factor productivity (i.e., TFP, measured as the ratio of net sales to assets) of the targets is greater than that of the acquirers and that the average TFP of non-merging firms exceeds that of the targets, with both of these differences statistically significant at the one percent level. This implies that the non-merging firms are the most productive on average, while the target firms are second and the acquirers last, which runs counter to the implication of Jovanovic and Rousseau (2002, 2007). Further, we find that the number of acquirers and targets are not the same, as many potential mergers with multiple bidders appear in the SDC data but are not completed. These findings lead us to focus on the differences between the ratio of bidder to target Qs and the two firms' technological distance, considering only actual mergers and potential within-industry pairings.

Using the data, we construct a series of quasi-bootstrap logistic regressions of the probability of an actual merger on the ratio of bidder to target Qs, the two firms' size differential, their relative potential growth, and other controls. We then provide bootstrap tests for all estimates. These regressions bear out the main implications of the model.

The remainder of this paper is organized as follows. In Section 2, we develop a Nash bargaining solution model that incorporates a cost of converting capital. In Section 3, we describe the construction of the data. In Section 4, we introduce a new econometric method. In Section 5, we make an empirical assessment of our theoretical model. In Section 6, we conclude.

Model

The model is based on Gort (1969) as well as Jovanovic and Rousseau (2002, 2007). In an economy, some firms are well positioned to take advantage of a shock, while others are not. Hence, after a positive technology shock, some firms are more productive than others. We model this as two kinds of technology with a distinct type of capital. Capital is technology-specific, as in Hulten (1992) and Greenwood et al. (1997, 2000). Given that the high and the low technology firms face the same output price, the high technology firms make better use of the assets they control and thus they have a higher Q than the low technology firms. As a result, the firms with high technology have a desire to expand their market share and find it optimal to acquire plants from less productive firms in the industry, even when it entails costs to convert the capital associated with low technology. By the same token, a positive shock in an industry increases the opportunity cost of operating as an inefficient producer in that industry. In a sense, merger and acquisition (M&A) are often the least-cost means for industry structure to respond to the changes brought about by economic shocks. Thus, a positive industry shock alters the value of the assets and creates incentives for transfers to more productive users through M&A.[2]

All firms in the model are assumed to be price-takers, to produce a homogenous output, to be endowed with technology-specific production assets, and to have the same technology initially. At time 0, there is a positive technology shock that affects one group of firms more than the other. Hence, after the shock, there are two kinds of technology, each associated with one kind of assets. High and low technology are respectively represented by z_h and z_l, with $z_h > z_l$. The high technology asset is denoted as K_h, while the low technology asset is denoted as K_l. Since the technology-specific assets K_i can be directly used only by the firms with technology z_i, a cost is needed to convert low technology assets to high technology uses.[3] In this model the cost is assumed to be a convex function of the technological distance between these two firms. Namely, the larger the technological distance, the higher is the cost incurred to convert the capital associated with the low technology. After merging, the combined firm inherits the technology from high productivity firm and also its latent intangibles, so the combined firm has the same Q as the high productivity firm. It is also assumed that the acquirer and its target arrive at a Nash bargaining solution to share the rents.[4]

Under a rational stock market, a firm's value can be written as

$$V_i = q_i K_i,\ i = h \text{ or } l, \tag{II.1}$$

where V_i is the market value of the firm, and q_i is the Tobin's Q of firm i that assigns a market value to a given replacement cost of all its assets.

[2]Mitchell and Mulherin (1996) document that the rate of corporate takeovers and restructurings within industries during 1980s is directly related to the economic shocks.
[3]See Jovanovic and Rousseau (2002, 2007), and Jovanovic (2006).
[4]Different choices for negotiating profits will not change the main results of the model.

After merging, the new firm's value function can be written as

$$V_M = (1 - C)q_h K_l + q_h K_h, \tag{II.2}$$

where V_M is the market value of the combined firm's and q_h is its Tobin's Q. C is the per unit cost to convert K_l into K_h, which satisfies $C' > 0$, $C'' > 0$, $C(1) = C'(1) = 0$ and $\lim_{\frac{z_h}{z_l} \to +\infty} C(\frac{z_h}{z_l}) = +\infty$.

We use a model of negotiations to determine how pre-merger firms share the surplus generated by the merger. There are different choices for the model of negotiations, but the simplest one is the Nash bargaining solution, which in this case is just the solution to

$$\begin{aligned} W(V_h, V_l) &= \max_{\Pi_{hM}, \Pi_{lM}} (\Pi_{hM} - V_h)^{\sigma} (\Pi_{lM} - V_l)^{(1-\sigma)} \qquad \text{(II.3)} \\ s.t.\ V_M &= \Pi_{hM} + \Pi_{lM} \end{aligned}$$

where W is the joint welfare of the aquirer and its target, Π_{iM} is the merger share to firm i, and $\sigma \in [0, 1]$ is the acquirer's bargaining weight. The larger σ implies a lower bargaining power of the target.

Lemma 1 *In equilibrium the resulting merger share for high-tech firm merging with low-tech firm is*

$$\Pi_{hM} = \sigma(V_M - V_h - V_l) + V_h, \tag{II.4}$$
$$\Pi_{lM} = (1 - \sigma)(V_M - V_h - V_l) + V_l. \tag{II.5}$$

Proposition 1 *Assume at time 0, the firms affected by a positive technology shock adopt a high technology immediately, while the others use the low technology. If the combined market value of the total assets of a high-tech firm and a low-tech firm is higher under common control than it is separately, they will merge immediately at time 0.*

Proof. If the high-tech and low-tech firms merge at time s, then firm h's market value at time 0 is

$$\begin{aligned}
V_{hs} &= \int_0^s (e^{-rt} r V_h) dt + e^{-rs}\Pi_{hM} \\
&= V_h - e^{-rs} V_h + e^{-rs}[\sigma(V_M - V_h - V_l) + V_h] \\
&= V_h + e^{-rs}\sigma(V_M - V_h - V_l) \text{ (since } \sigma(V_M - V_h - V_l) > 0) \\
&< V_h + \sigma(V_M - V_h - V_l) = V_{h0} \text{ (when s=0)}. \\
\text{Therefore, } V_{hs} &< V_{h0}.
\end{aligned}$$

With the same logic, we can prove that $V_{ls} < V_{l0}$.

Since both firms are worth more by merging at time 0 rather than waiting until a later time s, they merge at time 0. ■

The proposition shows that mergers occur in waves and that technological shocks drive industry merger waves. Several papers provide strong empirical supports for this proposition, such as Faria (2003) and Harford (2005).

If the high technology and the low technology firms merge at time 0, the gain for each firm from the merger is[5]

$$\begin{aligned}
G_h &= \Pi_{hM} - V_h = \sigma(V_M - V_h - V_l) + V_h - V_h = \sigma(V_M - V_h - V_l), \\
G_l &= \Pi_{lM} - V_l = (1-\sigma)(V_M - V_h - V_l) + V_l - V_l = (1-\sigma)(V_M - V_h - V_l).
\end{aligned}$$

The two firms will merge if and only if $G_i > 0$, which is equivalent to $V_M - V_h - V_l > 0$, or

$$G_h + G_l = (1-C)q_h K_l + q_h K_h - q_h K_h - q_l K_l > 0, \tag{II.6}$$

[5]Since merger will happen right after the technolgy shock, we disregard time index.

After rearranging,

$$\frac{q_h}{q_l} > \frac{1}{1 - C(\frac{z_h}{z_l})}. \tag{II.7}$$

However, since C is a convex function of $\frac{z_h}{z_l}$, z_i is a technology parameter that represents the total factor productivity (TFP), and q_i is actually a function of z_i, this inequality cannot be directly used to predict whether a merger will occur. TFP reflects a firm's current and past performance, while Tobin's Q is an expected profitability based on a firm's ex post and current performance. In other words, Tobin's Q not only measures the relation between TFP and market value, but also measures the relation between latent intangibles and market value.[6] If there are no latent intangibles, Q and TFP are equivalent, which is what Jovanovic and Rousseau (2002, 2007) implicitly assume. In our model, however, the high-tech firm is better equipped to adopt a new technology than the low-technology firm. This means that high-technology firms have a greater latent ability to adopt an advanced technology than low-technology firms, and therefore have higher growth potential. Q includes this latent ability (e.g. perhaps more flexible management), while TFP does not. Therefore, Q and TFP are not equivalent. Nevertheless, Q and TFP are positively correlated. Indeed, Dwyer (2001) shows that the plant-level productivity and the market value of a firm are positively related, and a manufacturing technique with high productivity acts as an intangible asset for the firm that owns it. In this paper, we introduce a new variable λ_i which is a firm's Q deflating by its TFP. This should measure the firm's latent adoptability of a new technology and also its growth potential. We define

[6] Griliches (1981) and Cockburn & Griliches (1988) report there is a significant relation between the market value of a firm and its unanticipated intangible capital.

$$\lambda_i = \frac{q_i}{z_i},\ i = h \text{ or } l, \tag{II.8}$$

where a higher λ_i implies a higher growth potential. Next we define $\lambda = \frac{\lambda_h}{\lambda_l}$ as the relative growth potentials of two firms. When the dispersion between λ_h and λ_l is larger, λ is larger.

Define the ratio of an acquirer to a target Qs, $\frac{q_h}{q_l}$, as q_r, with $q_r \geq 1$, and the ratio of bidder to target productivity, $\frac{z_h}{z_l}$, as z_r, with $z_r \geq 1$. $G_h + G_l$ normalized by the value of low technology firm can be simplified as

$$g(q_r) = \frac{G_h + G_l}{K_l \times q_l} = q_r(1 - C) - 1 > 0. \tag{II.9}$$

Thus $g(q_r)$ measures the gain from merging as a share of the low technology firm's pre merger value.

When $g(q_r) > 0$, the two firms will merge. When $g(q_r) \leq 0$, the two firms will not merge regardless of the difference between the acquirer and target Qs.

Proposition 2 *Given $q_r \geq 1$ and $\lambda > 0$: The distance between the potential acquirer and target Qs has an inverted U-shaped effect on the probability of a merger.*

Proof. Since $\frac{z_h}{z_l} \geq 1$ and $\frac{z_h}{z_l} = \frac{q_r}{\lambda}$, C can be written as a function of $\frac{q_r}{\lambda}$ and $q_r \geq \lambda$. The two firms will merge if and only if $g(q_r) > 0$. We can calculate

$$g(1) = 0, \tag{II.10}$$

$$g(\lambda) = \lambda - 1, \tag{II.11}$$

$$g'(q_r) = 1 - C - \frac{q_r}{\lambda}C', \tag{II.12}$$

$$g'(\lambda) = 1 > 0, \tag{II.13}$$

$$g''(q_r) = -2\frac{1}{\lambda}C' - \frac{q_r}{\lambda^2}C''. \tag{II.14}$$

Since $C' > 0$ and $C'' > 0$ when $q_r > \lambda$, $g''(q_r) < 0$ holds. Consequently, $g(q_r)$ is a concave function of q_r.

When $q_r \to +\infty$, $C(\frac{q_r}{\lambda}) \to +\infty$ given λ. Therefore $g(q_r) \to -\infty$ and $g'(q_r) \to -\infty$ when $q_r \to +\infty$.

Since $g'(\lambda) > 0$, $g'(q_r) \to -\infty$ when $q_r \to +\infty$ and $g''(q_r) < 0$, there exists a q_r^*, such that $g'(q_r^*) = 0$, and for any $q_r \in [\lambda, q_r^*)$, $g'(q_r) > 0$, for any $q_r \in (q_r^*, +\infty)$, $g'(q_r) < 0$. Hence $g(q_r^*) = \max_{q_r \in [\lambda, +\infty)} g(q_r)$. When $g(q_r^*) \leq 0$, no merger occurs. When $g(q_r^*) > 0$, the difference between the potential acquirer's and target's Qs has an inverted U-shaped effect on the probability of a merger, since $g''(q_r) < 0$. ■

For example, when $C = c(\frac{z_h}{z_l} - 1)^2$ we can solve for q_r as follows:

$$q_r \in \left(\frac{2\lambda}{3} - \frac{\lambda}{3}\sqrt{1 + \frac{3}{c}}, \frac{2\lambda}{3} + \frac{\lambda}{3}\sqrt{1 + \frac{3}{c}} \right) \iff g'(q_r) > 0. \tag{II.15}$$

In equation II.15, $g(q_r)$ increases until $q_r^* = \frac{2\lambda}{3} + \frac{\lambda}{3}\sqrt{1 + \frac{3}{c}}$, which corresponds to the maximum value of $g(q_r)$. Figure 2 shows that the relative Q between bidder and its target has an inverted U-shaped effect on the potential gains from a merger.

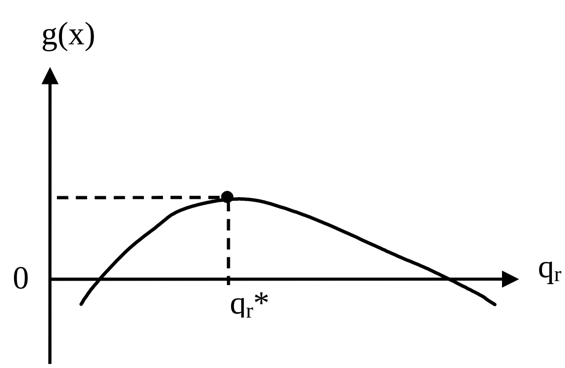


Figure 2. Gains from mergers

If $g(q_r^*) < 0$ (i.e., lies beneath the horizontal axis in Figure 2), no merger occurs regardless of the difference between the two firms' Qs. For instance, if c is too large, $g(q_r^*) < 0$ will hold.

If $g(q_r^*) > 0$, a merger occurs. There are two cases:

Case 1: $\frac{2\lambda}{3} - \frac{\lambda}{3}\sqrt{1+\frac{3}{c}} \leq 1 < q_r^*$.

For any $q_r \in [1, q_r^*)$, $g'(q_r) > 0$. This means $g(q_r)$ is increasing in the range of $[1, q_r^*)$.

For any $q_r \in [q_r^*, +\infty)$, $g'(q_r) < 0$ and $\lim_{q_r \to +\infty} g(q_r) = -\infty$. This means $g(q_r)$ is decreasing in the range of $[q_r^*, +\infty)$. Hence, there exists q_r^{**} such that $g(q_r^{**}) = 0$ and $g(q_r)$ will be negative when $q_r > q_r^{**}$. Consequently, we have a range $(1, q_r^{**})$ such that for any $q_r \in (1, q_r^{**})$, $g(q_r) > 0$ also holds, and the two firms will merge.

Case 2: $1 < \frac{2\lambda}{3} - \frac{\lambda}{3}\sqrt{1+\frac{3}{c\lambda}} < q_r^*$.

Since $g(1) = 0$ and $g'(q_r) < 0$ for any $q_r \in \left(1, \frac{2\lambda}{3} - \frac{\lambda}{3}\sqrt{1+\frac{3}{c}}\right)$, $g\left(\frac{2\lambda}{3} - \frac{\lambda}{3}\sqrt{1+\frac{3}{c}}\right) < 0$. Using the same logic as in case 1, we can show there exists q_r^{21} and q_r^{22} such that $q_r^{21} \in \left(\frac{2\lambda}{3} - \frac{\lambda}{3}\sqrt{1+\frac{3}{c}}, q_r^*\right)$, $q_r^{22} \in (q_r^*, \infty)$, $g(q_r^{21}) = 0$, and $g(q_r^{22}) = 0$. Then, for any valid $q_r \in (q_r^{21}, q_r^{22})$, it follows that $g(q_r) > 0$ and the two firms merge.

This proposition is striking precisely because it demonstrates that the distance between the acquirer and its target Qs is non-monotonically related to the likelihood of a merger. This inverted U-shape stands in contrast to the Q-theory of mergers, which suggests high market-to-book firms simply acquire those with low market-to-book values. Thus, in our model, an acquirer may not purchase the lowest Q firm that it can find, but rather a firm with a lower Q. The model also implies that the probability of being involved in a merger depends on λ. A firm with greater ability to adopt new techniques and a higher growth potential is likely to hold more capital and to acquire other firms. A firm with less ability to adopt new techniques and a lower growth potential will be in an inferior position in future competition, so it will be more likely to be acquired by high productivity firms. Thus, when the relative potential growth between the acquirer and its target is high, the probability of a merger is high. The following proposition states how mergers are affected by the firms' size differential and their relative growth potential.

Proposition 3 *Given $q_r \geq 1$, and $\lambda > 0$: the probability of merger is positively and linearly related to λ.*

Proof. The maximum value of function g can be simplified as follows.

Define $G = \max g(q_r) = q_r^* - q_r^* C(\frac{q_r^*}{\lambda}) - 1$. A larger value of G implies a larger likelihood of a merger, since there is a greater probability that $g(q_r)$ exceeds zero.

The effect of λ on G is

$$\frac{dG}{d\lambda} = \frac{\partial G}{\partial q_r^*}\frac{dq_r^*}{d\lambda} + \frac{(q_r^*)^2}{\lambda^2}C', \tag{II.16}$$

$$\begin{aligned}
\frac{d^2G}{d\lambda^2} &= \frac{\partial G}{\partial q_r^*}\frac{d^2q_r^*}{d\lambda^2} + \frac{\partial^2 G}{\partial (q_r^*)^2}\left(\frac{dq_r^*}{d\lambda}\right)^2 + \frac{2q_r^*}{\lambda^2}C'\frac{dq_r^*}{d\lambda} + \frac{(q_r^*)^2}{\lambda^3}C''\frac{dq_r^*}{d\lambda} \\
&\quad + \frac{2q_r^*}{\lambda^2}C'\frac{dq_r^*}{d\lambda} + \frac{(q_r^*)^2}{\lambda^3}C"\frac{dq_r^*}{d\lambda} - \frac{2(q_r^*)^2}{\lambda^3}C' - \frac{(q_r^*)^3}{\lambda^4}C'' \\
&= \frac{\partial G}{\partial q_r^*}\frac{d^2q_r^*}{d\lambda^2} + \frac{\partial^2 G}{\partial (q_r^*)^2}\left(\frac{dq_r^*}{d\lambda}\right)^2 - 2\frac{\partial^2 G}{\partial (q_r^*)^2}\frac{q_r}{\lambda}\frac{dq_r^*}{d\lambda} + \frac{\partial^2 G}{\partial (q_r^*)^2}\left(\frac{q_r}{\lambda}\right)^2,
\end{aligned} \tag{II.17}$$

since

$$\begin{aligned}
\frac{\partial G(q_r^*)}{\partial q_r^*} &= 1 - C - \frac{q_r^*}{\lambda}C', \\
\frac{\partial^2 G}{\partial (q_r^*)^2} &= -2\frac{1}{\lambda}C' - \frac{q_r^*}{\lambda^2}C''.
\end{aligned}$$

According to the envelope theorem, $\frac{dG}{d\lambda} = \frac{(q_r^*)^2}{\lambda^2}C' > 0$. And since

$$g'(q_r^*) = 1 - C\left(\frac{q_r^*}{\lambda}\right) - \frac{q_r^*}{\lambda}C'\left(\frac{q_r^*}{\lambda}\right) = 0,$$

we have

$$\begin{aligned}\frac{dg'(q_r^*)}{d\lambda} &= 0 = \left(-\frac{1}{\lambda}C' - \frac{q_r^*}{\lambda^2}C''\right)\left(\frac{dq_r^*}{d\lambda} - \frac{q_r^*}{\lambda}\right) \\ &= \frac{\partial^2 G}{\partial (q_r^*)^2}\left(\frac{dq_r^*}{d\lambda} - \frac{q_r^*}{\lambda}\right).\end{aligned}$$

Since $\frac{\partial^2 G}{\partial (q_r^*)^2} < 0$, $\frac{dq_r^*}{d\lambda} = \frac{q_r^*}{\lambda}$ holds. Substituting $\frac{dq_r^*}{d\lambda} = \frac{q_r^*}{\lambda}$ into (16), we get $\frac{d^2G}{d\lambda^2} = 0$.

Since $\frac{\partial G}{\partial \lambda} > 0$ and $\frac{d^2G}{d\lambda^2} = 0$, G is a linear and increasing function of λ. This means that the likelihood of a merger rises with λ. ■

Proposition 3 highlights the factors which affect mergers. Differences in the growth potentials have a significant effect on the propensity to merge. Propositions 2 and 3 show that while Q is not a linear factor affecting mergers, λ is. Because the conversion cost, a convex function of the relative technological distance between the two merging firms, drives the high productivity firm to purchase a firm with lower but not the lowest productivity, and because Q positively relates to TFP, a high Q firm buys a lower yet not the lowest Q firm. However, λ, which is Q deflated by TFP, is positively and linearly related to the likelihood of a merger. Consequently, λ is a more potent variable than Q in characterizing what drives merger activity.

In light of the theoretical observations on who merges with whom and which factors affect mergers, our next task is to test whether the data on US mergers are consistent with the theory.

Data Construction and Description

We restrict the empirics to data on domestic mergers available from Thompson's Securities Data Corporation (SDC) for firms traded on US stock exchanges. Since we compare actual mergers with potential ones, we proceed to construct a dataset for empirical testing in five steps.

First, we collect annual data items from 1986 to 2005 for the 22,888 firms listed on the 2006 version of the Compustat database, excluding firms with less than two years of sales data, and classify them according to the twelve industry groupings defined by Fama & French.[7] We then measure Tobin's Q as the ratio of market-to-book values for each firm.[8] Since total factor productivity (TFP) is a measure of the economic efficiency of a firm's operations, we measure it as the ratio of net sales to the book value of total assets (Compustat item 12 divided by item 6).[9] We refer to this dataset of listed firms and their accounting data as the "Compustat" file.

Second, we select US domestic mergers listed on the SDC's Mergers and Acquisitions Database from 1987 to 2006, excluding repurchases and leveraged buyouts. To avoid double-counting multiple announcements of the same merger, we work with only one observation per calendar year for each unique acquirer-target pair. We then separate the targets and their corresponding merger information from the acquirers and their corresponding information. Table 1 provides summary statistics for the merger data that we use. From Columns 4 and 5 in Table 1, we can see that the acquirers and targets are not matched.

Third, we match each target with its Compustat accounting data from the end of the fiscal year preceding the merger announcement (from step 1), and refer to the resulting

[7]The 12 Industry Portfolios classfication based on the four-digit SIC code can be found through the following link: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[8]Following Jovanovic and Rousseau (2007), the numerator of Q is the sum of common equity at current share prices (the product of Compustat items 24 and 25), the book values of preferred stock (item 130) and short- and long-term debt (items 34 and 9). The denominator is the sum of the book values of common equity (item 60), preferred stock (item 130), and short- and long-term debt (items 34 and 9). We omitted Q's for firms with negative values for net common equity since they imply negative market to book ratios, and eliminated observations with market-to-book values in excess of 100, since many of these were likely to be serious data errors.

[9]The average ratio of the acquirer's asset divided by the number of its employees to that of its target is 1.93, the median is 1.2, and the standard deviation is 3.32. For acquirers, the correlation between assets and the ratio of asset to the number of employees is 0.33, and that for targets is 0.34.

Table 1. Summary statistics of observations in each industry

Industry	No. observations per year			Merger activity		
	Mean	Min.	Max.	Acquirers	Targets	Total
(1)Consumer nondurables	280	200	343	646	340	986
(2)Consumer durables	133	91	175	280	139	419
(3)Manufacturing	548	397	711	1,344	663	2,007
(4)Energy	125	88	159	455	155	610
(5)Chemicals	101	79	117	255	128	383
(6)Computers,software,etc.	803	465	1,172	2,966	1,050	4,016
(7)Telephone and TV	145	72	241	512	241	753
(8)Utilities	177	118	225	250	123	373
(9)Wholesale	492	348	640	1,057	567	1,624
(10)Medical	342	173	442	1,095	503	1,598
(11)Finance	643	380	985	2,135	839	2,974
(12)Everything else	580	422	808	1,455	717	2,172

Note: Industry definitions are taken from Fama and French. Merger activity is measured by the number of firms involved in mergers in each industry.

dataset as the "SDC targets".

Fourth, we take the Cartesian product of the "Compustat" file (with acquirers flagged) and the SDC targets file (from step 3) in each year and for each industry to create a database of all possible pairs. In other words, our final data set contains each observation from the SDC target file merged with every observation in the same year and the same industry from the flagged Compustat file. (i.e., each target is paired with its actual acquirer as well as its potential acquirers.) This is important because our model analyzes the case where the target's technology is substituted with a better technology, and if two firms are not in the same industry, their assets are more likely to be complements than substitutes.

Finally, we identify whether each pair is the actual within-industry merging pair or the pseudo merging pair. The final dataset includes 1,317 actual merger pairs and 3,054,479 potential transactions.

Econometric Method

In this paper, we are discussing the probability of two firms being involved in a merger, the value of which should be between 0 and 1, so logistic regression is desirable. After pooling all actual and pseudo mergers and running a logistic regression, however, we can not get convergent results. In order to get convergent and efficient estimates, the quasi-bootstrap logistic regressions, a new econometric method, is constructed. The procedure includes four steps. Step 1, we randomly select 1,317 pseudo-merger pairs with replacement for each logistic regression, making sure that the number of actual and pseudo-mergers from each industry are the same (as in Rhodes-Kropf and Robinson 2007, which is called RR henceforth). Step 2, we match the randomly selected pseudo-transactions to the actual merger sample, and then do a logistic regression. In the third step, we repeat Steps 1 and 2 for 100 times, and then report the mean and the standard deviation of the estimated coefficients. In the last step, we use bootstrap to test the statistical significance of each estimate. Since we randomly select 1,317 pseudo-merger pairs with replacement for each logistic regression, our selected sample is choice-based. Thereby, we correct our estimates from the quasi-bootstrap logistic regressions. The remainder of this section is organized as follows. First, we provide a theoretical justification to explain the efficiency of our estimates. We then present the bootstrap tests for the statistical significance of each estimate in detail. In addition, we provide the justification for the correction of the estimates.

Efficiency of the Quasi-bootstrap Logistic Regressions

Assume the set N includes all the dependent variable $Y_{i0} = 0$, $i = 1, 2, ..., n_0$, and the set M includes all the dependent variable $Y_{j1} = 1$, $j = 1, 2, ..., n_1$. It is also assumed that $n_0 >> n_1$, which means that n_0 is dozens to thousands of times more than n_1. And then the fraction of ones in the population, τ, equals $\frac{n_1}{n_0+n_1}$. Given some regressors x_i, the goal is to estimate $P(Y_{i1} = 1|\ x_i)$, as this is the full conditional distribution. We assume that the underlying distribution of the dependent variable is logistic, $P(Y_{i1} = 1|x_i)$ can be expressed as:

$$P(Y_{i1} = 1|x_i) = \frac{1}{1 + e^{-x_i'\beta}},$$

where β is the true parameters for the choice-based sample.

We construct a new set A_t, $t = 1, 2, ..., T$, which contains n_1 observations randomly selected with replacement from N. And then we run a logistic regression using all the observations from A_t and M, $t = 1, 2, ..., T$. From this procedure, we can get T estimates of β, which are $\left\{\hat{\beta}_t, t = 1, 2, ..., T\right\}$.

For some $T \times T$ weight matrix W> 0, let

$$J_T(\beta) = T \begin{pmatrix} (\hat{\beta}_1 - \beta)/T \\ \vdots \\ \vdots \\ (\hat{\beta}_T - \beta)/T \end{pmatrix}' W \begin{pmatrix} (\hat{\beta}_1 - \beta)/T \\ \vdots \\ \vdots \\ (\hat{\beta}_T - \beta)/T \end{pmatrix} \tag{II.18}$$

We use the minimum distance method (MDM) to define an estimator which minimizes $J_T(\beta)$. We set W as the identity matrix. The solution of β for minimizing $J_T(\beta)$ is the mean of $\left\{\hat{\beta}_t, t = 1, 2, ..., T\right\}$. We define $\overline{\hat{\beta}}$ is the MDM estimate,

$$\overline{\hat{\beta}} = \frac{\sum_{t=1}^{T} \hat{\beta}_t}{T}. \tag{II.19}$$

Since all the estimators $\left\{\hat{\beta}_t, t = 1, 2, ..., T\right\}$ share some dependent variables M, $\left\{\hat{\beta}_t, t = 1, 2, ..., T\right\}$ are not independent from each other. To get more efficient estimate, we better set W as the variance-covariance of $\{\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_T\}$, which is unknown and can not be constructed easily. That means, the estimate $\overline{\hat{\beta}}$ is not the most efficient, when W is the identity matrix. However, it still has the following asymptotic properties.

Claim 1 *The asymptotic properties of* $\overline{\hat{\beta}}$ *are:*
(1) $\overline{\hat{\beta}} \rightarrow_p \beta$.
(2) Under $H_0 : \beta = 0$, $\sqrt{T}(\hat{\beta}) \rightarrow_d N(0, \sigma^2)$, *where* σ^2 *is unknown.*
(3) $\overline{\hat{\beta}}$ *is more efficient than the estimator from RR one time logistic regression.*

Since σ^2 is unknown, we can not directly test the statistical significance of $\overline{\hat{\beta}}$. Instead, we use a simulation method called the bootstrap. From the bootstrap samples, we perform bootstrap tests on the basis of bootstrap P values.

Bootstrap Tests

To obtain the bootstrap samples, we use four steps:

Step 1. Draw with replacement n_1 observations from M;

Step 2. Draw with replacement n_1 observations from A_t, and combine them with the sample we obtain in Step 1;

Step 3. Run logistic regression using each combined sample from Step 2;

Step 4. Repeat Steps 1-3 B times.[10]

Hence, we define set N_{tb} that includes n_1 observations randomly drew with replacement from each A_t, and set M_{tb} that includes n_1 observations randomly drew with replacement from M, where $t = 1, 2, ..., T$ and $b = 1, 2, ..., B$. Next, we run a logistic regression using all the observations from sets N_{tb} and M_{tb}, and denote the estimator as

[10] According to Davidson and MacKinnon (2004), if we will perform a bootstrap test at level α, then B should be choosen to satisfy the condition that $\alpha(B + 1)$ is an integer.

$\hat{\beta}^*_{tb}$.

We define that

$$\overline{\hat{\beta}}^*_b = \frac{\sum_{t=1}^T \hat{\beta}^*_{tb}}{T}, \ b = 1, 2, ..., B, \tag{II.20}$$

which is constructed in the same way as that of $\overline{\hat{\beta}}$. And then the standard deviation of $\left\{\overline{\hat{\beta}}^*_b, \ b = 1, 2, ..., B\right\}$ will be the standard error of our quasi bootstrap estimates, which is called quasi bootstrap standard error in this paper. Since the mean of $\overline{\hat{\beta}}^*_b$ is $\overline{\hat{\beta}}$ and the null hypothesis H_0 is $\beta = 0$, we take $\overline{\hat{\beta}}^*_b - \overline{\hat{\beta}}$ as the simulated test statistics. There are two cases to construct the empirical distribution function (EDF) based on the one-sided test.

If the alternative hypothesis H_1 is $\beta > 0$, then the EDF is

$$\hat{F}^*(\overline{\hat{\beta}}) = \frac{1}{B}\sum_{b=1}^{B} I(\overline{\hat{\beta}}^*_b - \overline{\hat{\beta}} \le \overline{\hat{\beta}}). \tag{II.21}$$

Our estimate of the true P value for this case is therefore

$$\hat{p}^*(\overline{\hat{\beta}}) = 1 - \hat{F}^*(\overline{\hat{\beta}}) = 1 - \frac{1}{B}\sum_{b=1}^{B} I(\overline{\hat{\beta}}^*_b - \overline{\hat{\beta}} < \overline{\hat{\beta}}) = \frac{1}{B}\sum_{b=1}^{B} I(\overline{\hat{\beta}}^*_b > 2\overline{\hat{\beta}}). \tag{II.22}$$

The last equality in II.22 means that the true P value is approximated by the proportion of simulations, in which $\overline{\hat{\beta}}^*_b$ is greater than $2\overline{\hat{\beta}}$. For example, if $B = 599$, and 25 of all the $\overline{\hat{\beta}}^*_b$ are greater than $2\overline{\hat{\beta}}$, then $\hat{p}^*(\overline{\hat{\beta}}) = 25/599 = 0.042$. As a result in this example, we would reject the null hypothesis that $\beta = 0$ at 5 percent statistic significant level.

If the alternative hypothesis H_1 is $\beta < 0$, then the EDF is

$$\hat{F}^*(\overline{\hat{\beta}}) = \frac{1}{B}\sum_{b=1}^{B} I(\overline{\hat{\beta}}^*_b - \overline{\hat{\beta}} \ge \overline{\hat{\beta}}). \tag{II.23}$$

Our estimate of the true P value is

$$\hat{p}^*(\bar{\hat{\beta}}) = 1 - \hat{F}^*(\bar{\hat{\beta}}) = 1 - \frac{1}{B}\sum_{b=1}^{B} I(\bar{\hat{\beta}}_b^* - \bar{\hat{\beta}} \geq \bar{\hat{\beta}}) = \frac{1}{B}\sum_{b=1}^{B} I(\bar{\hat{\beta}}_b^* < 2\bar{\hat{\beta}}). \tag{II.24}$$

If B is infinitely large, the EDF converges to the true conditional distribution function (CDF). Consequently, our procedure would yield an exact test and the outcome of the test would be the same as the P value computed by using the conditional distribution function (CDF) of $\bar{\hat{\beta}}$.

Correction for the Estimates

Our above method use a "matched-pairs" design, which results in a sample proportion of merged pair firms of 0.50. This type of sampling implies typically that the proportion of merged pairs in the sample is much larger than the proportion of such companies in the grant population of all pairs (merged and non-merged). This "matched-pairs" design causes a "choice-based sample bias" of the constant and the coefficients in estimated standard probit/logit models, in turn meaning that the probabilities being assessed in such models are more or less biased. Hence, we provide the correction for the estimates from the above quasi bootstrap logistic regressions.

Since the fraction of ones in the population, τ, is known, we can use the prior correction for the logit model (see King and Zeng, 2001). For the logit model, in any of the above sampling designs, the estimated parameters except the constant item, $\bar{\hat{\beta}}_0$, is statistically consistent estimates of the true parameters, but the unbiased estimate $\bar{\hat{\beta}}_0^c$ for the constant item is

$$\bar{\hat{\beta}}_0 - \ln\left[(\frac{1-\tau}{\tau})\right], \tag{II.25}$$

because $\overline{\hat{\beta}}_0 = \frac{\sum_{t=1}^{T} \hat{\beta}_{t0}^*}{T}$, and the correction for the constant item $\hat{\beta}_{t0}^*$ of each logistic regression is $\hat{\beta}_{t0}^* - \ln\left[\left(\frac{1-\tau}{\tau}\right)\right]$.[11]

Evidence

In this section we begin by checking the basic model assumptions and their implications for the data. Next, we investigate the relation between $\log q$ and $\log z$ that emerges from the data. Then, using the econometric method constructed in Section 4, we ask whether the 'high buys lower but not the lowest' pattern that we observe for Q is an artifact of another phenomenon such as firms' relative sizes or relative cash holdings by including these indicators as controls in a multiple regression framework. We also investigate the high-buys-low pattern that we observe for relative growth by testing the sensitivity of the likelihood of a merger to λ after including these same control variables. Finally, we repeat the analysis allowing a common covariance structure across industries to test the robustness of our findings.

Basic Model Assumptions and their Implications for the Data

Using the sample of 1,317 actual merger transactions from the SDC database, we first investigate whether acquirers typically absorb targets with lower market-to-book ratios, while at the same time avoiding potential targets with the lowest market-to-book ratios in the economy. Table 2 includes pooled summary statistics of book values, the market value of all financial assets, and the market value of common equity in millions of 2005 US dollars

[11] From King and Zeng (2001), if the propotion of Y=1 in the selected sample is κ, the corrected estimate is consistent for constant item β_0 :

$$\hat{\beta}_0 - \ln\left[\left(\frac{1-\tau}{\tau}\right)\left(\frac{\kappa}{1-\kappa}\right)\right].$$

for acquirers, targets, and firms not involved in a merger for each year from 1986 to 2005, along with the average values of Tobin's Q and z.[12] The table shows that the average Q of acquirers is nearly 13 times greater than that of targets but that we can not distinguish targets' Qs from those of non-merging firms. This finding that acquirers usually have higher Qs than their targets is consistent with Jovanovic and Rousseau (2002, 2007).

Table 2. Characteristics of nonmerging, target and acquiring firms before matching

Variable	Nonmerging firms	Target	Acquirer	t(T-N)	t(A-T)
Observations	69,598	5,465	12,450		
Market value	4,286	2,293	8,427	-6.90***	12.68***
Book value	2,738	1,565	4,388	-5.25***	8.81***
Market equity	2,652	1,286	5,764	-8.52***	14.29***
Tobin's Q	2.149	2.096	2.837	-1.54	12.86***
z	1.088	1.051	0.941	-2.98***	-8.53***

Note: The data are from the Compustat database, and are pooled observations from 1986 to 2005. Acquirers and targets are not matched. "Market value" is measured as the sum of the value of common stock (the product of items 24 and 25), preferred stock (item 130), and short and long-term debt (items 34 and 9), with the latter three components as book values. "Book value" replaces the market value of common stock in the above calculation with its book value (item 60). "Market equity" is simply the market value of common stock. All measures are constructed from end-of-year data and are converted to millions of 2005 dollars. T statistics for the differences in means across groups appear in columns 4 and 5, with *** representing statistical significance at 1 percent level.

At the same time, row 6 of Table 2 reports that the average z (measured as the ratio of net sales to assets) of the targets is greater than that of the acquirers and that the average z of non-merging firms exceeds that of the targets, with both of these differences statistically significant at the one percent level. This implies that the non-merging firms are the most productive on average, while the target firms are second and the acquirers last, which runs counter to the model's implication. At the same time, the average z for the three groups are actually quite close. Further, as Table 1 shows, the number of acquirers and targets are not the same, as many potential mergers with multiple bidders appear in

[12] We restrict the sample to include only those firms with market values of equity that exceed $10 million.

the SDC data but are not completed. For this reason, and because the model focuses on the relative growth potential of horizontal merger pairs, Table 3 focuses on the differences in the model's key parameters (z_r and q_r) at the industry level, considering only actual mergers and potential within-industry pairings.

Table 3. Summary statistics of z_r and q_r in each industry

Industry	z_r			q_r		
	Pseudo-merger mean	Merger mean	t(M-N)	Pseudo-merger mean	Merger mean	t(M-N)
(1)Consumer nondurables	1.331	1.023	-2.11**	2.260	1.661	-2.77***
(2)Consumer durables	1.149	0.891	-1.39	1.916	1.719	-0.74
(3)Manufacturing	1.260	1.010	-2.63***	1.925	1.554	-2.58**
(4)Energy	1.785	1.241	-1.70*	1.679	1.253	-3.44***
(5)Chemicals	1.199	1.035	-1.05	2.115	1.401	-1.32
(6)Computers,software,etc.	2.007	1.262	-1.85**	3.347	2.081	-5.44***
(7)Telephone and TV	4.763	1.120	-0.79	2.163	1.126	-4.33***
(8)Utilities	1.247	1.012	-1.47	1.215	1.096	-2.36**
(9)Wholesale	1.410	1.023	-2.54**	2.376	1.656	-3.48***
(10)Medical					None	
(11)Finance	3.766	0.972	-1.31	1.841	1.145	-6.29***
(12)Everything else	4.335	1.033	-0.53	2.596	1.465	-3.40***

Note: t(M-N) in columns 4 and 7 is the t staitistics for the differences of z_r and q_r in means across groups, respectively. *, **, and *** represent statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

Columns 1 and 2 in Table 3 respectively report the average of z_r of the pseudo-merging pairs and the actual merging pairs in each industry. Nine of the eleven averages listed in column 2 are greater than 1, which highlights that on average the acquirer is more efficient than its target when a merger actually occurs. At the same time, column 1 shows that the z_r associated with potential yet non-merging pairs are higher in all instances than those associated with actual mergers. Column 3 shows that the difference in mean z_r between merging firms and potential mergers is statistically significant at conventional levels for 5 of the 11 industries. This means that on average the technological distance between the actual merging pairs is less than that of the pseudo merging pairs. The results displayed in columns 1 - 3 of the table also suggest that the pooled (i.e., unmatched) summary statistics in Table 2 are misleading.

The last three columns of Table 3 reports the average q_r of the pseudo-merging and actual merging pairs by industry, and shows that acquirers have higher Qs on average than their targets in each industry when the merger occurs. Once again, however, the t-statistics for the null hypothesis that actual mergers have the same q_r as potential mergers are all negative and for the most part statistically significant. This indicates that on average the relative Qs of the actual merging pairs is lower than that of the pseudo pairs in each industry. That is to say, a high Q firm purchases a firm associated with a lower but not the lowest Q.

All the observations from Table 3 are consistent with the theoretical implications that a high Q firm will buy a firm with a lower yet not the lowest Q.[13] In the next section, we test the implications of our model and explore the high-buys-lower pattern in terms of relative potential growth.

Relation between Q and TFP

Table 4 reports results from the pooled regressions of $q_{i,t}$ on $z_{i,t}$ from 1986 to 2005. Column 1 reports the baseline OLS regression. The estimated coefficient of $\log z_{i,t}$ is 0.076 and is statistically significant at the one percent level. Columns 2 and 3 show that this result is robust to the inclusion of fixed effects for industries and time. Since current technology is highly correlated with that of the previous year ($\rho = .88$), columns 4 and 5 repeat the results from two-step least squares (2SLS) regressions, in which five annual lags of the firm's TFP and 12-industry dummies are used as instruments. In both regressions

[13] Even after excluding potential pairs where the pseudo-target has a higher Q than its pseudo-acquirer, we still find that on average the actual mergers' x is significantly smaller than that of the pseudo mergers in each industry. We do not report these results in this paper, but they are available upon request.

Table 4. Pooled regressions of $q_{i,t}$ on $z_{i,t}$, 1986-2005

	OLS(1)	OLS(2)	OLS(3)	IV(1)	IV(2)
Const.	0.565***	0.519***	0.643***	0.445***	0.581***
	(0.002)	(0.006)	(0.011)	(0.006)	(0.011)
$\log(z_{i,t})$	0.076***	0.072***	0.077***	0.101***	0.111***
	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)
Industry Effect		Yes	Yes	Yes	Yes
Year Effect			Yes		Yes
Obs.	86,863	86,863	86,863	40,573	40,573
R^2	0.151	0.114	0.153	0.100	0.123

Note: The dependent variable is $\log(q_{i,t})$. The IV(1) and IV(2) regressions use five lags of z_i as instruments in the first stage regression. Standard errors are in parentheses. *** represents statistical significance at the 1 percent level.

the correlation coefficient between $\log q_{i,t}$ and $\log z_{i,t}$ is about 10 percent. Thus, Table 4 indicates that Q and TFP are positively correlated, which implies that productivity has an implicit value, yet because the correlation between $\log q_{i,t}$ and $\log z_{i,t}$ is far below 1 they can not be regarded as equivalent.[14]

Effects of q_r and λ on the Likelihood of Mergers by Controlling for Alternative Explanations

Our methodology compares mergers which actually occurred with mergers that might have occurred but did not. This follows Rhodes-Kropf and Robinson (2007), which randomly pairs any two non-merging firms to create a matched sample of pseudo-transaction in which the sample of pseudo-mergers must have the same number of observations as the sample of actual mergers. Our next step then estimates a series of logistic regressions of the probability of being involved in a merger on combinations of z_r, q_r, λ, and other control variables. To obtain efficient estimates from the logistic, we adopt the econometric method presented in Section 4 and we set $T = 100$ and $B = 599$. That is to say, we

[14]Dwyer (2001) finds a positive relation between plant-level productivity and the market value of a firm that is consistent with our data.

repeat randomly selecting pseudo-transactions with replacement to match the actual merger sample and running a logistic regression for 100 times, and then report the mean and the standard deviation of the estimated coefficients as well as the bootstrap p values, and also correct the constant item by subtracting 7.749, which equals $\ln\left(\frac{1-\tau}{\tau}\right)$ and in our data $\tau = \frac{n_1}{n_0+n_1} = \frac{1{,}317}{3{,}054{,}479+1{,}317} = 0.0004$.

Table 5. Quasi-bootstrap logistic regressions for specifications that control for k while also allowing q_r and z_r to enter separately with both linear and quadratic terms

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Const.	-8.162***	-7.977***	-7.398***	-7.211***	-8.042***	-7.849**	-7.283***	-7.090
	(0.041)	(0.053)	(0.043)	(0.060)	(0.044)	(0.059)	(0.046)	(0.066)
logk	0.889***	0.900***	0.803***	0.813***	0.875***	0.883***	0.765***	0.771***
	(0.043)	(0.045)	(0.046)	(0.046)	(0.043)	(0.044)	(0.044)	(0.044)
$(\text{logk})^2$	-0.115***	-0.116***	-0.091***	-0.092***	-0.112***	-0.113***	-0.088***	-0.088***
	(0.009)	(0.009)	(0.010)	(0.010)	(0.009)	(0.009)	(0.008)	(0.009)
$\log q_r$	0.612***	0.613***	0.613***	0.613***				
	(0.058)	(0.058)	(0.060)	(0.060)				
$(\log q_r)^2$	-0.405***	-0.425***	-0.412***	-0.429***				
	(0.043)	(0.046)	(0.046)	(0.049)				
$\log z_r$					-0.066	-0.072	-0.059	-0.065
					(0.059)	(0.061)	(0.061)	(0.064)
$(\log z_r)^2$					-0.367***	-0.378***	-0.366***	-0.376***
					(0.058)	(0.061)	(0.055)	(0.058)
log_c_r	No	No	0.106***	0.107***	No	No	0.135***	0.136***
			(0.033)	(0.033)			(0.034)	(0.034)
$(\text{log\_c\_r})^2$	No	No	-0.029***	-0.029***	No	No	-0.135***	-0.136***
			(0.006)	(0.006)			(0.006)	(0.006)
Ind. FE	No	Yes	No	Yes	No	Yes	No	Yes
Obs.	2634	2634	2602	2602	2634	2634	2602	2602
Bootstrap	100	100	100	100	100	100	100	100

Note: The reported estimates are the mean of each quasi bootstrap estimation. The quasi bootstrap standard error is reported in parentheses. * and *** represent statistical significance at the 10 percent and 1 percent levels, respectively.

Table 5 presents estimates from these quasi-bootstrap logistic regressions (with standard deviation of the estimated coefficients) for specifications that control for firms' size differential (i.e., k) while also allowing q_r and z_r to enter separately with both linear and quadratic terms. For example, in the first column of the table, the coefficient on $\log k$ is 0.889 and that of $(\log k)^2$ is -0.115, with both statistically significant at the one percent level. This indicates that the probability of a merger is a nonlinear function of the size differential between bidder and its target. The estimated coefficients for $\log q_r$ and $(\log q_r)^2$ are 0.612 and -0.405 respectively, and both are also statistically significant at one percent level. This finding shows that the distance between acquirer's Q and target's Q has a strongly inverted U-shaped effect on the probability of the two firms being involved in a merger, even after controlling for the relative sizes of the acquirer and its target. Columns 2 - 4 of the table show that this finding is robust to the inclusion of dummy variables for industries and after controlling for the level of cash balances (log_c_r). Interestingly, and broadly consistent with Jensen's (1986) "free cash" hypothesis, cash-rich firms are also more likely to engage in mergers with firms that are cash-poorer but not the poorest, though the effect of the cash holdings is small relative to that of k and q_r. Based on the regression results in column 3, for example, when $q_r = k =$c_r=5, the partial effect of $\log q_r$ on the probability of a merger is -0.001, that of $\log k$ is 0.001, and that of log_c_r is only 0.0002.[15] Based on the results in Table 5, we conclude that the merger pattern of 'high-buys-lower-yet-not-the-lowest' in terms of Q is reasonable and not driven by some alternative explanation.

Columns 5 - 8 in Table 5 indicate that the technological distance (z_r) between an acquirer and its target also has a nonlinear effect on the likelihood of a merger. Another observation from Table 5 is that $(\log z_r)^2$ and $(\log q_r)^2$ have very similar effects on the probability of two firms involved in a merger because the estimated coefficients are very close. For example, the estimated coefficient of $(\log z_r)^2$ is -0.367 in column 5, while that

[15]Cash_r is the ratio of acquirer to its target cash holdings and log_cash_r is the log of cash_r.

of $(\log q_r)^2$ is -0.405 in column 1. Their absolute difference is only 0.038. Therefore, from the definition of λ, which is $\log \lambda = \log q_r - \log z_r$, we can say that $\log \lambda$ has almost no quadratic effect on the probability of a merger. This means that $\log \lambda$ affects mergers in a predominately linear fashion. We obtain similar results after controlling for industry fixed effects and the relative cash holdings of acquirers and their potential and actual target and report these results in columns 6 - 8.

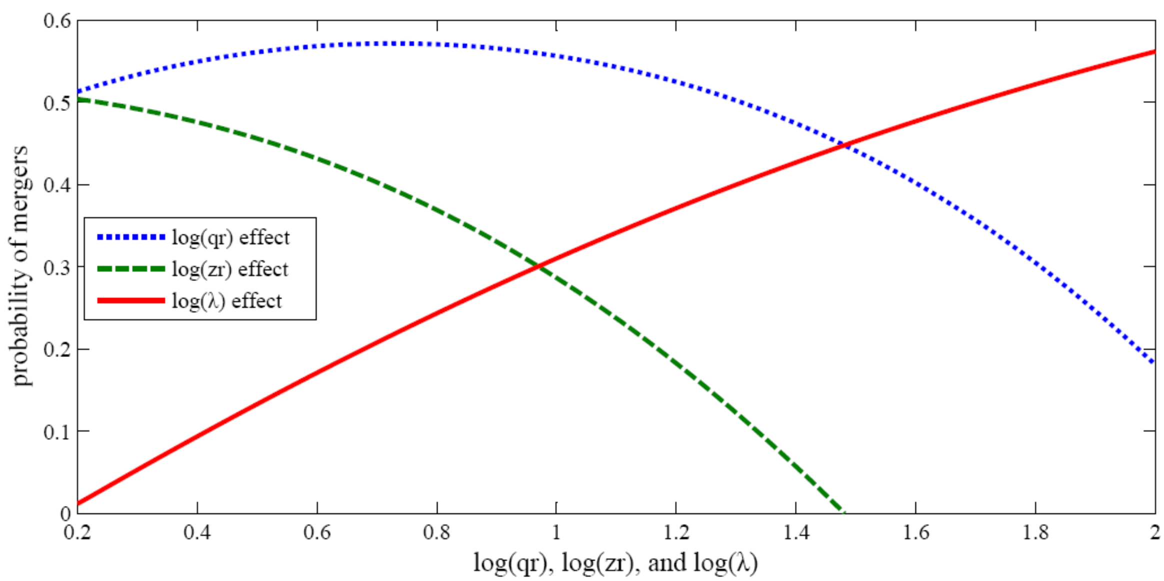


Figure 3. qr, zr and λ's effect on the probability of mergers

To illustrate our findings, we simulate the effects of $\log q_r$, $\log z_r$ and $\log \lambda$ on the likelihood of a merger using the estimated coefficients listed in column 1 and column 5 in Table 5 and present them in Figure 3.[16] This figure shows that $\log q_r$ and $\log z_r$ have an inverted U-shaped effect on mergers, but their difference (i.e., $\log \lambda$) affects mergers positively and linearly.[17] Thus, in terms of λ, the merger pattern is that firms with high

[16]Since the actural likelihood of a merger is close to zero, we simulate the probability of a merger with uncorrected coefficients and set $\log k = 1.5$.

[17]Indeed, the "q_r" effect in Figure 3 is the empirical analog of Figure 1.

growth potential tend to acquire firms with low potential, or simply put, that 'high buys low'.

Table 6. Quasi-bootstrap logistic regressions for specifications that control for k while also allowing λ to enter with both linear and quadratic terms

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Const.	-8.261***	-8.089***	-7.504***	-7.329***	-8.047***	-7.868**	-7.291***	-7.111
	(0.038)	(0.050)	(0.039)	(0.056)	(0.045)	(0.058)	(0.047)	(0.063)
logk	0.881***	0.891***	0.773***	0.783***	0.857***	0.868***	0.766***	0.777***
	(0.042)	(0.043)	(0.044)	(0.044)	(0.043)	(0.044)	(0.045)	(0.046)
$(\text{logk})^2$	-0.121***	-0.122***	-0.098***	-0.098***	-0.112***	-0.113***	-0.090***	-0.091***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)	(0.009)	(0.010)	(0.010)
$\log\lambda$	0.132***	0.133***	0.120***	0.121***	0.382***	0.391***	0.376***	0.382***
	(0.023)	(0.023)	(0.024)	(0.024)	(0.059)	(0.060)	(0.060)	(0.062)
$(\log\lambda)^2$					-0.300***	-0.308***	-0.297***	-0.302***
					(0.047)	(0.048)	(0.047)	(0.048)
log_c_r	No	No	0.131***	0.1318***	No	No	0.112***	0.112***
			(0.033)	(0.033)			(0.034)	(0.034)
$(\text{log\_c\_r})^2$	No	No	-0.029***	-0.030***	No	No	-0.028***	-0.028***
			(0.006)	(0.006)			(0.006)	(0.006)
Ind. FE	No	Yes	No	Yes	No	Yes	No	Yes
Obs	2634	2634	2602	2602	2634	2634	2602	2602
Bootstrap	100	100	100	100	100	100	100	100

Note: The reported estimates are the mean of each quasi bootstrap estimation. The quasi bootstrap standard error is reported in parentheses. *, ** and *** represent statistical significance at the 10 percent, 5 percent and 1 percent levels, respectively.

In addition, in Table 6, we directly estimate the effects of $\log\lambda$ and $(\log\lambda)^2$ on the probability of a merger, which control for k and the log of cash holdings. Columns 1 - 4 indicate that $\log\lambda$ has a positive effect on mergers, in which the estimated coefficients of $\log\lambda$ are close to 0.13 and all are statistically significant. When including $(\log\lambda)^2$ in our regressions, we get a positive estimate on $\log\lambda$ but a negative one for $(\log\lambda)^2$ on the probability of an actual merger, with both statistically significant at the one percent level. However, the absolute values of the estimates of $\log\lambda$ and $(\log\lambda)^2$ are smaller than that of

$\log q_r$ and $(\log q_r)^2$ respectively. This indicates that the effect of $\log \lambda$ on mergers has less curvature than that of $\log q_r$, and therefore that λ more succinctly explains merger activity than q_r. For example, in column 5 of Table 6, the estimated coefficient of $\log \lambda$ is 0.382 and that of $(\log \lambda)^2$ is -0.3, while the estimated coefficient of $\log q_r$ is 0.612 and that of $(\log q_r)^2$ is -0.405 in column 1 of Table 5.

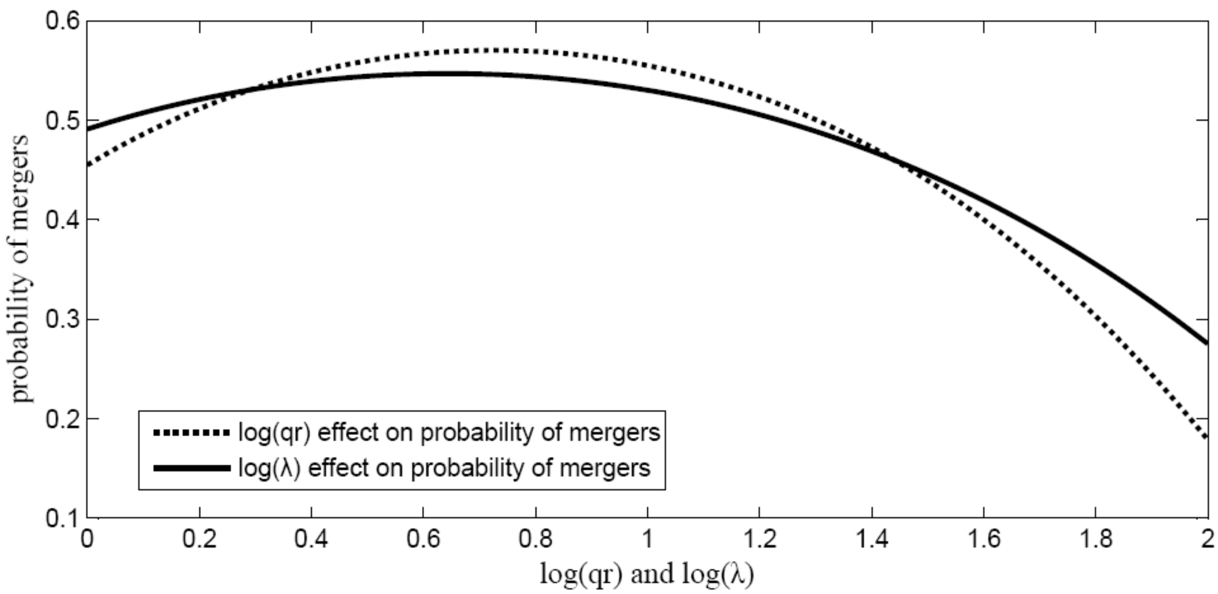


Figure 4. qr and λ's effect on the probability of mergers

Based on the above estimates, Figure 4 displays the simulation results of the effects of $\log q_r$ and $\log \lambda$ on the likelihood of a merger.[18] In Figure 4, the curve representing the effect of $\log \lambda$ with its quadratic term has less curvature than that representing the effect of $\log q_r$ with its quadratic term. This indicates that when both the linear and quadratic terms of $\log q_r$ and $\log \lambda$ are considered, $\log \lambda$ affects the probability of an actual merger more linearly than $\log q_r$. Indeed, though some curvature is apparent in the relationship between the probability of merger and λ when λ enters the empirics explicitly, it remains

[18]Once again, we simulate the probability of a merger with uncorrected coefficients and set $\log k = 1.5$.

quite nearly linear for ranges of λ that we would most likely encounter in the data (say $\log \lambda < 1$, i.e., $\lambda < 2.73$). In a word, the simulation results also support our theoretical implication that $\log \lambda$ is a more potent variable than $\log q_r$ in characterizing what drives mergers.

Robustness Tests

For robustness, we revisit the questions of who purchases whom and which factors drive mergers using Seemingly Unrelated Regression (SUR) methods.[19] The results are reported in Tables 7 and 8. We use the same strategy as Section 4 randomly selecting 1,317 pseudo mergers from all possible mergers on each of 100 iterations of an SUR logistic regression, with Table 7 showing the means of the estimated coefficients from the 100 logistic regressions of the probability of an actual merger on the Kronecker Products of ($\log k$, $(\log k)^2$, $\log q_r$, $(\log q_r)^2$) and industry dummies. Table 8 provides the means of the estimated coefficients from these logistic regressions of the probability of an actual merger on the Kronecker Products of ($\log k$, $(\log k)^2$, $\log \lambda$) and industry dummies. In Tables 7 and 8, all the estimates of $\log k$ are positive and that of $(\log k)^2$ are negative, and almost all are statistically significant. This finding strongly supports that firms' relative size has a nonlinear effect on the propensity to merge. In columns 3 and 4 of Table 7, almost all the estimated coefficients of $\log q_r$ are positive and those on $(\log q_r)^2$ are negative, while more than half of these estimates are statistically significant. The last two columns in Table 7 indicate that the ratio between acquirer to target Qs also has nonlinear effect on the likelihood of the two firms being involved in a merger. In the last column of Table 8, seven of the eleven estimated coefficients of $\log \lambda$ are positive and statistically significant. This

[19] Since our model is talking about the within-industry mergers, it is necessary and important to distinguish industries.

suggests that the likelihood of a merger is larger when the relative potential growth between acquirer and its target is sizeable. As a result, these two tables also provide empirical support that our nonlinear Q theory on M&A is robust to disaggregation by sector.

Conclusion

The key assumptions in our model are that a firm's capital is firm specific and a cost is needed to convert target's capital into a form usable by the acquirer. We show that mergers present a pattern in which 'high buys lower yet not the lowest' in terms of Q, but 'high buys low' in terms of the firms' growth potentials. Therefore, the relative growth potential of a prospective merger pair is more suitable to explain whether they merge than the ratio of the bidder to target's Q. From our model, we document the following findings about technology and the propensity to merge: (1) The distance between acquirer and its target Qs has an inverted U-shaped effect on the probability of the two firms being involved in a merger. This means the acquirer may not purchase the lowest Q firm that it can find, but rather a firm with a lower Q. (2) The likelihood of a merger is positively and linearly related to the relative growth potentials of the acquirer and its target. This means that the likelihood of a merger is larger when the relative potential growth between acquirer and its target is high.

Using data for mergers among US firms available from the Securities Data Corporation from 1986 to 2005, we construct a series of quasi-bootstrap logistic regressions of the probability of an actual merger on the ratio of bidder to target Qs, the two firms' size differential, their relative adoptability of a new technology, and other controls. The empirical evidence supports the main implications of the model.

There are four possible extensions. One might be to construct more accurate measurements of Q than current proxies in the literature, which might in turn lead to more accurate estimates of their effects on mergers.[20] The second one would be to expand the empirics to consider international and cross-border merger activity. Third, to get more efficient estimator, we better set the weight matrix in Section 4 as the variance-covariance matrix of the estimates from the logistic regressions. It might also improve the empirical results, if we use the non-parametric method to estimate the effects of q_r and λ on the likelihood of mergers, since this method has no problem to deal with the huge unbalance between actual mergers and pseudo-mergers.

[20] The accuracy of measures of Q has been discussed in Erickson and Whited (2000, 2006), in which they argue that most proxies for Q are poor.

Table 7. SUR logistic regressions on the Kronecker Products of ($\log k$,$(\log k)^2$, $\log q_r$,$(\log q_r)^2$) and industry dummies

Industry Name	logk	$(\text{logk})^2$	$\log q_r$	$(\log q_r)^2$
(1)Consumer nondurables	0.627***	-0.102***	0.361***	-0.329***
	(0.012)	(0.003)	(0.025)	(0.022)
(2)Consumer durables	0.693	-0.164	0.645	0.164
	(7.334)	(2.431)	(12.931)	(10.920)
(3)Manufacturing	0.575***	-0.094***	0.265***	-0.064***
	(0.010)	(0.003)	(0.019)	(0.016)
(4)Energy	0.917***	-0.174***	0.271***	-0.568***
	(0.184)	(0.035)	(0.258)	(0.232)
(5)Chemicals	0.572**	-0.075**	0.499**	-0.536**
	(1.484)	(0.292)	(0.834)	(1.103)
(6)Computers,software,etc.	0.452***	-0.034***	0.390***	-0.259***
	(0.007)	(0.002)	(0.006)	(0.004)
(7)Telephone and TV	0.543***	-0.072***	-0.072**	-0.300***
	(0.013)	(0.003)	(0.024)	(0.025)
(8)Utilities	0.504***	-0.089***	1.342***	-0.543*
	(0.021)	(0.007)	(0.119)	(0.398)
(9)Wholesale	0.519***	-0.069***	0.513***	-0.347***
	(0.010)	(0.003)	(0.021)	(0.017)
(11)Finance	0.574***	-0.082***	0.435***	-0.866***
	(0.006)	(0.001)	(0.015)	(0.022)
(12)Everything else	0.562***	-0.083***	0.492***	-0.364***
	(0.010)	(0.003)	(0.018)	(0.017)
Bootstrap	100	100	100	100

Note: The mean of the 100 constants is -7.963 with one percent statistically significant and the standard deviation is 0.003. The reported estimators are the mean of the 100 times regressions of each coefficient. The standard deviations are reported in parentheses. *, **, and *** represent statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

Table 8. SUR logistic regressions on the Kronecker Products of ($\log k$, $(\log k)^2$, $\log \lambda$) and industry dummies

Industry Name	logk	$(\text{logk})^2$	$\log\lambda$
(1)Consumer nondurables	0.631***	-0.105***	0.121***
	(0.012)	(0.003)	(0.014)
(2)Consumer durables	0.616	-0.155	0.785
	(6.402)	(3.799)	(39.221)
(3)Manufacturing	0.598***	-0.095***	0.135***
	(0.010)	(0.003)	(0.014)
(4)Energy	0.892***	-0.173***	0.004
	(0.026)	(0.009)	(0.025)
(5)Chemicals	0.551***	-0.078*	0.094**
	(0.069)	(0.014)	(0.048)
(6)Computers,software,etc.	0.434***	-0.037***	0.144***
	(0.007)	(0.002)	(0.004)
(7)Telephone and TV	0.528***	-0.073***	0.075***
	(0.012)	(0.003)	(0.014)
(8)Utilities	0.560***	-0.090***	-0.175***
	(0.025)	(0.008)	(0.039)
(9)Wholesale	0.527***	-0.081***	0.262***
	(0.010)	(0.003)	(0.011)
(11)Finance	0.589***	-0.085***	-0.129***
	(0.006)	(0.001)	(0.006)
(12)Everything else	0.558***	-0.094***	0.113***
	(0.010)	(0.003)	(0.010)
Bootstrap	100	100	100

Note: The mean of the 100 constants is -8.063 with one percent statistically significant and the quasi bootstrap standard error is 0.003. The reported estimators are the mean of the 100 times regressions of each coefficent. The quasi standard deviations are reported in parentheses. ** and *** represent statistical significance at the 10 percent and 5 percent, respectively.

CHAPTER III

A SIMULATION-BASED METHOD FOR RARE EVENTS DATA

Introduction

Over the years a considerable number of empirical investigations of Mergers and Acquisitions (M&A) have been carried out. Various types of statistical methods have been used, ranging from fairly ad hoc applications of regression analysis to more sophisticated variants of discriminant or probit/logit analysis. However, most of the methods may not provide unbiased estimates for the probability of two firms being involved in a merger. This bias is due to the data structure, which contains binary dependent variables with dozens to thousands of times fewer ones (merger pairs) than zeros (nonmerger pairs). In my second chapter, for example, I collect 1,317 actual within-industry merger pairs but construct more than 3 million nonmerger pairs. This kind of data is termed rare events data, and the difficulties in explaining it as well as predicting it have already been discussed in the literature. The problems in the statistical analysis of rare events data are reviewed and a new simulation-based method is provided in this paper.

It has been documented that there are three problems for explaining and predicting rare events. First, the probability of rare events can be sharply underestimated by most popular statistical procedures, such as probit and logit regressions. Furthermore, the commonly used data collection strategies are far more efficient. In addition, even when all observations and necessary variables are collected, it is sometimes difficult for the the probit and logit regression to reach convergent results. Detailed explanations for the three problems are discussed in the following paragraphs.

First of all, the statistical properties of the regressions for the binary dependent variable models are not invariant to the (unconditional) mean of the dependent variable. In the binary dependent variable models, the mean of the dependent variable is the relative frequency of events in the data, which, in addition to the number of observations, constitutes the information content of the data set.

For the rare events data, a "matched-pairs" design has been often used, resulting in a sample proportion of merger pairs of 0.50. This type of sampling typically implies that the proportion of merger pairs in the sample is much larger than the proportion of such pairs in the population of all actually merged and pseudo merged pairs. In a word, the "matched-pairs" design causes a "choice-based sample bias" of the coefficients in estimated standard probit/logit models, in turn meaning that the probabilities being assessed in such models are more or less biased. Usually the merger probability will be underestimated, and hence the nonmerger probability will be overestimated.

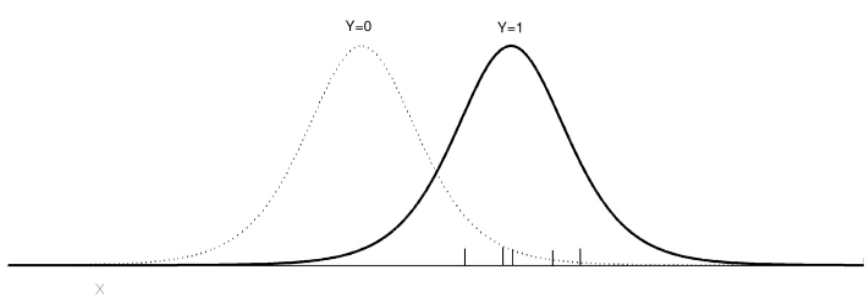


Figure 5. The densities of Y=1 and Y=0 separately.

To explain intuitively why the probability of rare events will be misestimated, King and Zeng (2001) provide a simplified case with one explanatory variable illustrated in Figure 5, where X denotes the explanatory variable and Y represents the dependent variable. This figure draws the densities of Y=1 and Y=0 separately, where the observations are arrayed horizontally according to the value of X. Here, Y=1 means that two selected firms are merged, and Y=0 indicates that they are not involved in a merger. In figure 5, the left dotted curve demonstrates the distribution of Y=0 and the right solid one exhibits the distribution of Y=1. In their example, King and Zeng assume that there are only five observations of Y=1, which are displayed by the short vertical lines. They argue that the dotted density curve can be estimated essentially without error because of the large number observations of Y=0, but any estimate of the solid density curve from the mere five data points will be very poor, and the estimate of the density of Y=1 will be systematically biased toward tails. The cutting point, which maximally distinguishes the zeros and ones, will be too far to the right since no information exists about the left end of the solid density. As a result, Pr(Y=1) will be underestimated and then Pr(Y=0) will be overestimated.

In addition, data collection causes another difficulty in analyzing rare events. There is a trade-off between gathering more observations and including better or additional variables when resources are limited. King and Zeng (2001) mention that approximately 99% of the costs in the data collection can be used to add new variables to an existing collection. For fear of null observation of rare events, some researchers usually choose very large numbers of observations with few, and in most cases poorly measured, explanatory variables, which turns out inefficient data collection strategies. Because it is believed that the real information in the data lies much more with the ones than the zeros, the strategies to have more observations have been criticized for spending time in analyzing very crude measures on many observations almost all of which contain no relevant information. To address the controversy in selecting the dependent variable, some advice has been given in King and Zeng (2001). For choosing the ones, it is best to collect all available observations.

For choosing the zeros, it depends on the cost: If it is costless to collect the zeros, the researcher should collect as many as he or she can. If it is not costless but not more expensive than in collecting the ones, the researcher should collect two to five times more zeros than ones. This sampling method is known in econometrics as choice-based or endogenous stratified sampling.

Besides, there may still have management and estimation issue when all observations and necessary variables are collected. In my second chapter, for example, there are 1,317 actual within-industry mergers and more than 3 million nonmergers. Data sets of this size are not uncommon, but they make data management difficult and statistical analyses time consuming. This outcome has already been shown in my second chapter. After pooling all actual and pseudo mergers and running a probit or logit regression to analyze the M&A activities, however, convergent results cannot be reached numerically by using the built-in program in STATA or SAS[1].

To provide a consistent and efficient estimate for the probability of rare events, several methods have already been constructed in the literature. The first one is prior correction, which involves computing the usual logit regression maximum likelihood estimation (MLE) and correcting the estimates based on prior information about the fraction of ones in the population and the observed fraction of ones in the sample. It is easy to use, but requires knowledge of the fraction of ones in the population, which may not be available in some cases. Another disadvantage of prior correction is that if the model is misspecified, the estimates will be biased. The last but not the least issue of prior correction is how to randomly select the zeros. The second method is the weighted exogenous sampling maximum-likelihood method, which is proposed by Xie and Manski (1989). The essential

[1] To get a convergent result, this kind of data need more iteration times than those set in the logistic and probit program in STATA and SAS.

component of this method is to weight the data to compensate for differences in the sample and population fractions of ones induced by choice-based sampling. However, it is very difficult to apply, since it requires specialized software for estimation. Combining the above two methods, King and Zeng (2001) build a corrected version of weighting with rare event corrections.

Given that the underlying distribution of the dependent variable is logistic, this paper proposes a new simulation-based econometric method to estimate the probability of a merger, which increases the efficiency of prior correction. We randomly select the "nonevents" (pseudo merger pairs) to match the events (mergers) and apply the logit regression, and then repeat the procedure a thousand times. We also construct the variance and covariance matrix of the estimates and identify their limiting distributions, and compare the efficiency of our method with Rhodes-Kropf and Robinson (2007), who adopt the "matched-pairs" design.

The remainder of this paper is organized as follows. In section 2, we discuss the choice-based sample probability bias. In section 3, we present a simulation-based method for the rare events data, and analyze the efficiency and the asymptotic properties of the estimates from this method. Section 4 concludes our analysis.

The Choice-Based Sample Probability Bias

The Relation between the Sample-Based and Population-Based Merger Probability

From Compustat and the Securities Data Corporation databases, we collect the companies' name and their accounting information, and also identify which two firms have been merged in a given year. If two firms in the same industry have not been involved in a merger in a given year, they may have the propensity to merge. In this sense, we take them as a potential merger pair. However, in this setting, the number of actual merger pairs is very small relative to the number of potential mergers. In the empirical work, we expect to identify the factors that could affect two firms' propensity to merge. After pooling all actual and pseudo mergers, we can not reach a convergent result numerically by using probit and logistic regressions.

One way to get convergent results numerically is to reduce the number of pseudo merger pairs. However, if the statistical methods have not been adjusted in this context, the estimated coefficients should have been arbitrarily affected by the chosen sample compositions. That means, the expected probability of two firms to be merged based on the exogenous stratified sample strongly differs from the corresponding probability in the whole population.

In order to analyze the chioce-based sample bias, the following notation is introduced:

$\{Y\}$ = set of dependent variables from the whole population, which is dummy variables. $Y = 1$ means that the selected two firms are merged, and $Y = 0$ means that the selected two firms are not merged.

$\{y\}$ = set of dependent variables from the selected sample, which is dummy variables. $y = 1$ means that the selected two firms are merged, and $y = 0$ means that the selected two firms are not merged.

$\{X\}$ = set of independent variables from the whole population.

$\{x\}$ = set of independent variables from the selected sample.

τ $= P(Y = 1)$, proportion of merger pairs in the grand population, where $0 <$

$\tau < 1$.

$P_m^\tau = P(Y = 1|\{X\})$, population-based probability of two firms involving in a merger.

$p = P(y = 1)$, proportion of merger pairs in the estimation sample, where $0 < p < 1$.

$\hat{P}_m^p = P(y = 1|\{x\})$, unadjusted estimated probability of two firms involving in a merger based on a selected estimation sample with a proportion of merger pairs equals to p.

If the prediction model has been estimated with a sample proportion of merger pairs p is the same as the population proportion of merger pairs τ, the estimated merger probability $\hat{P}_m^p$ from the sample is an unbiased probability assessment of the corresponding population-based probability P_m^τ. Otherwise, the calculated probability $\hat{P}_m^p$ will no longer be an unbiased estimate of P_m^τ, but they are correlated.

Proposition 4 *If $P(\{X\}|Y = 1) = P(\{x\}|y = 1)$, and $P(\{X\}|Y = 0) = P(\{x\}|y = 0)$, the relation between the merger probability $\hat{P}_m^p$ from the sample and the corresponding population-based probability P_m^τ can be expressed as*

$$\hat{P}_m^p = \left[1 + \left(\frac{1-p}{p}\right)\left(\frac{\tau}{1-\tau}\right)\left(\frac{1-P_m^\tau}{P_m^\tau}\right)\right]^{-1}. \tag{III.1}$$

Proof. Since $P(Y = 1) = \tau$, then $P(Y = 0) = 1 - \tau$.

Assuming $P_m^\tau > 0$, the unbiased probability P_m^τ can be unfolded according to Bayes' theorem:

$$\begin{aligned} P_m^\tau &= P(Y = 1|\{X\}) \\ &= \frac{P(Y = 1, \{X\})}{P(\{X\})} \\ &= \frac{P(\{X\}|Y = 1)P(Y = 1)}{P(\{X\}|Y = 1)P(Y = 1) + P(\{X\}|Y = 0)P(Y = 0)} \\ &= \frac{P(\{X\}|Y = 1)\tau}{P(\{X\}|Y = 1)\tau + P(\{X\}|Y = 0)(1-\tau)}, \end{aligned} \tag{III.2}$$

Similarly, we get

$$\hat{P}_m^p = \frac{P(\{x\}|y=1)p}{P(\{x\}|y=1)p + P(\{x\}|y=0)(1-p)}. \tag{III.3}$$

Equations III.2 and III.3 can be simplified as III.4 and III.5 correspondingly.

$$\frac{P(\{X\}|Y=1)}{P(\{X\}|Y=0)} = \left((P_m^\tau)^{-1} - 1\right)\frac{\tau}{1-\tau}, \tag{III.4}$$

$$\frac{P(\{x\}|y=1)}{P(\{x\}|y=0)} = \left(\left(\hat{P}_m^p\right)^{-1} - 1\right)\frac{p}{1-p}. \tag{III.5}$$

Since $P(\{X\}|Y=1) = P(\{x\}|y=1)$, and $P(\{X\}|Y=0) = P(\{x\}|y=0)$, we have

$$\frac{P(\{X\}|Y=1)}{P(\{X\}|Y=0)} = \frac{P(\{x\}|y=1)}{P(\{x\}|y=0)}. \tag{III.6}$$

Therefore,

$$\left((P_m^\tau)^{-1} - 1\right)\frac{\tau}{1-\tau} = \left(\left(\hat{P}_m^p\right)^{-1} - 1\right)\frac{p}{1-p}, \tag{III.7}$$

After simplification, equation III.7 can be written as

$$\hat{P}_m^p = \left[1 + \left(\frac{1-p}{p}\right)\left(\frac{\tau}{1-\tau}\right)\left(\frac{1-P_m^\tau}{P_m^\tau}\right)\right]^{-1}. \tag{III.8}$$

■

From equation III.8, we can see that the sample-based probability of a merger is not only a function of the unbiased probability P_m^τ, but also depends on the proportion of merged pairs in the population (τ) and the proportion of merged pairs in the sample (p).

To better understand the association between $\hat{P}_m^p$ and P_m^τ, τ as well as p, we provide the following corollaries.

Corollary 1 *If and only if $p = \tau$ or $P_m^\tau = 1$, then $\hat{P}_m^p = P_m^\tau$.*

Proof. $\Rightarrow$ When $p = \tau$, equation III.8 is

$$\hat{P}_m^p = \left[1 + \left(\frac{1 - P_m^\tau}{P_m^\tau}\right)\right]^{-1} = P_m^\tau. \tag{III.9}$$

When $P_m^\tau = 1$, then the right side of equation III.8 is 1. That means $\hat{P}_m^p = P_m^\tau = 1$.

$\Leftarrow$ when $\hat{P}_m^p = P_m^\tau$, equation III.8 can be rewritten as

$$\begin{aligned} \hat{P}_m^p &= \left[1 + \left(\frac{1-p}{p}\right)\left(\frac{\tau}{1-\tau}\right)\left(\frac{1 - P_m^\tau}{P_m^\tau}\right)\right]^{-1}, \\ &= \frac{P_m^\tau}{P_m^\tau - \left(\frac{1-p}{p}\right)\left(\frac{\tau}{1-\tau}\right)P_m^\tau + \left(\frac{1-p}{p}\right)\left(\frac{\tau}{1-\tau}\right)}, \end{aligned} \tag{III.10}$$

Since $\hat{P}_m^p = P_m^\tau$, we have

$$P_m^\tau - \left(\frac{1-p}{p}\right)\left(\frac{\tau}{1-\tau}\right)P_m^\tau + \left(\frac{1-p}{p}\right)\left(\frac{\tau}{1-\tau}\right) = 1, \tag{III.11}$$

Equation III.11 is equivalent to

$$\left(1 - \left(\frac{1-p}{p}\right)\left(\frac{\tau}{1-\tau}\right)\right)P_m^\tau = 1 - \left(\frac{1-p}{p}\right)\left(\frac{\tau}{1-\tau}\right), \tag{III.12}$$

Equation III.12 holds only if $p = \tau$ or $P_m^\tau = 1$.

Therefore, $\hat{P}_m^p = P_m^\tau$, if and only if $p = \tau$ or $P_m^\tau = 1$. $\blacksquare$

Corollary 1 indicates that the sample-based probability of a merger $\hat{P}_m^p$ will differ from an unbiased estimate of the merger probability P_m^τ, when $p \neq \tau$ and $P_m^\tau \neq 1$. The following corollary discusses the factors that affect $\hat{P}_m^p$.

Corollary 2 *The sample-based probability of a merger $\hat{P}_m^p$ is positively related to the unbiased merger probability P_m^τ and the proportion of merged pairs in the sample (p), but negatively related to the proportion of merged pairs in the population (τ).*

Proof. When $p \neq \tau$, from equation III.8, we can calculate

$$\frac{\partial(\hat{P}_m^p)}{\partial(P_m^\tau)} = \left[1 + \left(\frac{1-p}{p}\right)\left(\frac{\tau}{1-\tau}\right)\left(\frac{1 - P_m^\tau}{P_m^\tau}\right)\right]^{-2}(P_m^\tau)^{-2} > 0, \tag{III.13}$$

When $p = \tau$, we have

$$\frac{\partial(\hat{P}_m^p)}{\partial(P_m^\tau)} = 1 > 0, \tag{III.14}$$

Inequalities III.13 and III.14 show that $\hat{P}_m^p$ is positively associated with the unbiased merger probability P_m^τ.

The first order conditions of p and τ on $\hat{P}_m^p$ are

$$\frac{\partial(\hat{P}_m^p)}{\partial p} = \left[1 + \left(\frac{1-p}{p}\right)\left(\frac{\tau}{1-\tau}\right)\left(\frac{1-P_m^\tau}{P_m^\tau}\right)\right]^{-2} p^{-2} > 0, \tag{III.15}$$

$$\frac{\partial(\hat{P}_m^p)}{\partial \tau} = -\left[1 + \left(\frac{1-p}{p}\right)\left(\frac{\tau}{1-\tau}\right)\left(\frac{1-P_m^\tau}{P_m^\tau}\right)\right]^{-2} (1-\tau)^{-2} < 0, \tag{III.16}$$

Inequalities III.15 and III.16 imply that $\hat{P}_m^p$ is an increasing function in p, but a decreasing function in τ. ■

Equation III.8 demonstrates $\hat{P}_m^p$ as a function of P_m^τ, p and τ. In a decision context, however, one might primarily be interested in an assessment of the unbiased probability P_m^τ, and identifying the factors, which affect this probability. Though we have problem to directly estimate P_m^τ, we can estimate $\hat{P}_m^p$ first, and then approximate P_m^τ through equation III.8. We rewrite it as

$$\hat{P}_m^\tau = \left[1 + \left(\frac{1-\tau}{\tau}\right)\left(\frac{p}{1-p}\right)\left(\frac{1-\hat{P}_m^p}{\hat{P}_m^p}\right)\right]^{-1}, \tag{III.17}$$

where $\hat{P}_m^\tau$ represents an estimate of P_m^τ.

A necessary condition for $\hat{P}_m^\tau$ to be an unbiased estimate of P_m^τ is that the selected sample should be a random drawing from a sub-population of the whole sample. After estimating $\hat{P}_m^p$ from the selected sample, we can calculate $\hat{P}_m^\tau$ with some correction if p and τ are given.

Prior Correction of the Estimates from Logit Model

Assume $\hat{P}_m^p = P(y_i = 1|\{x_i\}) = P(x_i\beta)$, where β are the unknown parameter vector $(\beta_0, \beta_1^{'})^{'}$, a $k \times 1$ vector. Here, β_0 is the scalar constant term and β_1 is a vector with elements corresponding to the explanatory variables. The parameters can be estimated by maximum likelihood method, where the likelihood function is formed by assuming independence over the observations. The likelihood function is $L(\beta|y) = \Pi_{i=1}^n [P(x_i\beta)]^{y_i} [1 - P(x_i\beta)]^{1-y_i}$. By taking logs, the log-likelihood simplifies to

$$\ln L(\beta|y) = \sum_{\{y_i=1\}} \ln P(x_i\beta) + \sum_{\{y_i=0\}} \ln [1 - P(x_i\beta)] . \tag{III.18}$$

The estimate of β, labeled as $\hat{\beta}$, gives the maximum value of function III.18. As a consequence, $\hat{P}_m^p = P(x_i\hat{\beta})$.

According to equation III.17, we can approximate P_m^τ as a function of $\hat{P}_m^p$, τ and p. In most general formulation of $\hat{P}_m^p$, the prior correction of $\hat{P}_m^\tau$ is consistent but not necessary feasible to apply. Fortunately, in the logit special form of $\hat{P}_m^p$, the prior correction is not only consistent and fully efficient, but also easy to apply.

Proposition 5 *In the logit model if* $\hat{P}_m^p = \frac{1}{1+e^{-x\hat{\beta}}}$, *then*

$$\hat{P}_m^\tau = \frac{1}{1 + e^{-x\hat{\beta}+\ln\left[\left(\frac{1-\tau}{\tau}\right)\left(\frac{p}{1-p}\right)\right]}}. \tag{III.19}$$

Proof. *When* $\hat{P}_m^p = \frac{1}{1+e^{-x\hat{\beta}}}$, *equation III.8 can be written as*

$$\frac{1}{1 + e^{-x\hat{\beta}}} = \left[1 + \left(\frac{1-p}{p}\right)\left(\frac{\tau}{1-\tau}\right)\left(\frac{1-\hat{P}_m^\tau}{\hat{P}_m^\tau}\right)\right]^{-1},$$

$$
\begin{aligned}
\Leftrightarrow \quad & 1+e^{-x\hat{\beta}} = 1+\left(\frac{1-p}{p}\right)\left(\frac{\tau}{1-\tau}\right)\left(\frac{1-\hat{P}^{\tau}_{m}}{\hat{P}^{\tau}_{m}}\right), \\
\Leftrightarrow \quad & e^{-x\hat{\beta}} = \left(\frac{1-p}{p}\right)\left(\frac{\tau}{1-\tau}\right)\left(\frac{1}{\hat{P}^{\tau}_{m}}-1\right), \\
\Leftrightarrow \quad & \frac{1}{\hat{P}^{\tau}_{m}}-1 = \left(\frac{1-\tau}{\tau}\right)\left(\frac{p}{1-p}\right)e^{-x\hat{\beta}}, \\
\Leftrightarrow \quad & \frac{1}{\hat{P}^{\tau}_{m}} = 1+e^{-x\hat{\beta}+\ln\left[\left(\frac{1-\tau}{\tau}\right)\left(\frac{p}{1-p}\right)\right]}, \\
\Leftrightarrow \quad & \hat{P}^{\tau}_{m} = \frac{1}{1+e^{-x\hat{\beta}+\ln\left[\left(\frac{1-\tau}{\tau}\right)\left(\frac{p}{1-p}\right)\right]}}.
\end{aligned}
$$

■

The above proposition implies that for the estimate of P^{τ}_{m}, parameters $\hat{\beta}_1$ need not be changed, and only the constant term $\hat{\beta}_0$ should be corrected by subtracting out the bias factor, $\ln\left[\left(\frac{1-\tau}{\tau}\right)\left(\frac{p}{1-p}\right)\right]$. Furthermore, the estimate after correction is consistent. It has been proved that the prior correction is identical to the conditional maximum likelihood estimate proposed by Manski et al. (1981) and Amemiya et al. (1987) when the model is logistic. And it is also equivalent to the generalized method of moments estimate by Imbens (1992), Cosslett (1981a, b), and Lancaster et al. (1996a, b) when the functional form of $\hat{P}^{p}_{m}$ is logistic and the sampling probability, $E(p)$, is unknown[2].

However, as indicated in the previous section, corrected $\hat{P}^{\tau}_{m}$ is an unbiased estimate of P^{τ}_{m} only when the sample of merged pairs constitutes a random drawing from the sub-population of merged pairs and the sample of pseudo merged pairs are randomly drew from all possible merger pairs. It will be an issue for randomly selecting the sample, especially for rare events data, in which the binary dependent variables with dozens to thousands of times fewer ones than zeros. This paper adopts the bootstrap method to select the sample and construct a simulation-based method.

[2]See King and Zeng (2001).

A Simulation-Based Method

Many papers, such as Rhodes-Kropf and Robinson (2007), have already studied the probability of two firms being involved in a merger. Since probit and logistic regressions may not reach convergent results numerically from all observations, Rhodes-Kropf and Robinson (2007) adopt the "matched-pairs" design, but they do not collect their estimates. As a result, their estimates are biased. In order to get unbiased estimates, we provide a new econometric method, in which we randomly select the “nonevents” (pseudo mergers) to match the events (mergers), and then repeat the procedure one hundred times. As a summary, for each time selection, we match the zeros with the ones and keep all ones, since it is believed that the real information in the data lies much more with the ones than the zeros.

The remainder of this section is organized as follows. First, we provide a theoretical justification to explain the efficiency of our estimates. Then, we present the bootstrap tests for the statistical significance of each estimate in detail. Moreover, we discuss how to correct the estimates.

Efficiency of the Quasi-bootstrap Logistic Regressions

Assume the set N includes all the dependent variable $Y_{i0} = 0$, $i = 1, 2, ..., n_0$, and the set M includes all the dependent variable $Y_{j1} = 1$, $j = 1, 2, ..., n_1$.It is also assumed that $n_0 >> n_1$, which means that n_0 is dozens to thousands of times more than n_1[3]. And then the fraction of ones in the population, τ, equals $\frac{n_1}{n_0+n_1}$. Given some regressors x_i, the goal is to estimate $P(Y_{i1} = 1|\ x_i)$, as this is the full conditional distribution. We assume that the underlying distribution of the dependent variable is logistic, $P(Y_{i1} = 1|\ x_i)$ can be

[3] Here, 1 represents the actual merger pairs and 0 represents the pseudo-merger pairs. Therefore, $N_1 = 1,317$ and $N_0 = 3,054,479$.

expressed as:

$$P(Y_{i1} = 1|x_i) = \frac{1}{1 + e^{-x_i'\beta}},$$

where β is the true parameters for the choice-based sample.

We construct a new set A_t, $t = 1, 2, ..., T$, which contains n_1 observations randomly selected with replacement from N. And then we run a logistic regression using all the observations from A_t and M, $t = 1, 2, ..., T$. From this procedure, we can get T estimates of β, which are $\left\{\hat{\beta}_t, t = 1, 2, ..., T\right\}$.

For some $T \times T$ weight matrix $W > 0$, let

$$J_T(\beta) = T \begin{pmatrix} (\hat{\beta}_1 - \beta)/T \\ \vdots \\ \vdots \\ (\hat{\beta}_T - \beta)/T \end{pmatrix}' W \begin{pmatrix} (\hat{\beta}_1 - \beta)/T \\ \vdots \\ \vdots \\ (\hat{\beta}_T - \beta)/T \end{pmatrix} \tag{III.20}$$

We use the minimum distance method (MDM) to find an estimate, which minimizes $J_T(\beta)$. For simplicity, we set W as the identity matrix. And then the solution of β for minimizing $J_T(\beta)$ is the mean of $\left\{\hat{\beta}_t, t = 1, 2, ..., T\right\}$. We define $\bar{\hat{\beta}}$ is the MDM estimate,

$$\bar{\hat{\beta}} = \frac{\sum_{t=1}^{T} \hat{\beta}_t}{T}. \tag{III.21}$$

Since all the estimates $\left\{\hat{\beta}_t, t = 1, 2, ..., T\right\}$ share some dependent variables M, $\left\{\hat{\beta}_t, t = 1, 2, ..., T\right\}$ are not independent from each other. To get more efficient estimate, we better set W as the variance-covariance of $\{\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_T\}$, which is unknown and can not be constructed easily. That means, the estimate $\bar{\hat{\beta}}$ is not the most efficient, when W is the identity matrix. However, it still has the following asymptotic properties.

Claim 2 *The asymptotic properties of* $\bar{\hat{\beta}}$ *are:*
(1) (1) $\bar{\hat{\beta}} \rightarrow_p \beta$.
(2) Under $H_0 : \beta = 0$, $\sqrt{T}(\hat{\beta}) \rightarrow_d N(0, \sigma^2)$, *where* σ^2 *is unknown.*
(3) $\bar{\hat{\beta}}$ *is more efficient than the estimate from Rhodes-Kropf and Robinson (2007) "matched-pairs" design probit regression.*

Since σ^2 is unknown, we can not directly test the statistical significance of $\bar{\hat{\beta}}$. Instead, we use the bootstrap, a simulation based method. From the bootstrap samples, we perform bootstrap tests on the basis of bootstrap P values.

Bootstrap Tests

To obtain the bootstrap samples, we use four steps:

Step 1. Draw with replacement n_1 observations from M;

Step 2. Draw with replacement n_1 observations from A_t, and combine them with the sample we obtain in Step 1;

Step 3. Run logistic regressions using each combined sample from Step 2;

Step 4. Repeat Steps 1-3 B times.[4]

Hence, we define set N_{tb} that includes n_1 observations randomly drew with replacement from each A_t, and set M_{tb} that includes n_1 observations randomly drew with replacement from M, where $t = 1, 2, ..., T$ and $b = 1, 2, ..., B$. Next, we run a logistic regression using all the observations from sets N_{tb} and M_{tb}, and denote the estimate as $\hat{\beta}^*_{tb}$.

We define that

[4]According to Davidson and MacKinnon (2004), if we will perform a bootstrap test at level α, then B should be choosen to satisfy the condition that $\alpha(B + 1)$ is an integer.

$$\overline{\hat{\beta}}_b^* = \frac{\sum_{t=1}^{T} \hat{\beta}_{tb}^*}{T}, \ b = 1, 2, ..., B, \tag{III.22}$$

which is constructed in the same way as that of $\overline{\hat{\beta}}$. And then the standard deviation of $\left\{\overline{\hat{\beta}}_b^*, \ b = 1, 2, ..., B\right\}$ will be the standard error of our quasi bootstrap estimates, which is called quasi bootstrap standard error in this paper. Since the mean of $\overline{\hat{\beta}}_b^*$ is $\overline{\hat{\beta}}$ and the null hypothesis H_0 is $\beta = 0$, we take $\overline{\hat{\beta}}_b^* - \overline{\hat{\beta}}$ as the simulated test statistics. There are two cases to construct the empirical distribution function (EDF) based on the one-sided test.

If the alternative hypothesis H_1 is $\beta > 0$, then the EDF is

$$\hat{F}^*(\overline{\hat{\beta}}) = \frac{1}{B} \sum_{b=1}^{B} I(\overline{\hat{\beta}}_b^* - \overline{\hat{\beta}} \leq \overline{\hat{\beta}}). \tag{III.23}$$

Our estimate of the true P value for this case is therefore

$$\hat{p}^*(\overline{\hat{\beta}}) = 1 - \hat{F}^*(\overline{\hat{\beta}}) = 1 - \frac{1}{B} \sum_{b=1}^{B} I(\overline{\hat{\beta}}_b^* - \overline{\hat{\beta}} < \overline{\hat{\beta}}) = \frac{1}{B} \sum_{b=1}^{B} I(\overline{\hat{\beta}}_b^* > 2\overline{\hat{\beta}}). \tag{III.24}$$

The last equality in equation III.24 means that the true P value is approximated by the proportion of simulations, in which $\overline{\hat{\beta}}_b^*$ is greater than $2\overline{\hat{\beta}}$. For example, if $B = 599$, and 25 of all the $\overline{\hat{\beta}}_b^*$ are greater than $2\overline{\hat{\beta}}$, then $\hat{p}^*(\overline{\hat{\beta}}) = 25/599 = 0.042$. As a result in this example, we would reject the null hypothesis that $\beta = 0$ at 5 percent statistic significant level.

If the alternative hypothesis H_1 is $\beta < 0$, then the EDF is

$$\hat{F}^*(\overline{\hat{\beta}}) = \frac{1}{B} \sum_{b=1}^{B} I(\overline{\hat{\beta}}_b^* - \overline{\hat{\beta}} \geq \overline{\hat{\beta}}). \tag{III.25}$$

Our estimate of the true P value is

$$\hat{p}^*(\bar{\hat{\beta}}) = 1 - \hat{F}^*(\bar{\hat{\beta}}) = 1 - \frac{1}{B}\sum_{b=1}^{B} I(\bar{\hat{\beta}}_b^* - \bar{\hat{\beta}} \geq \bar{\hat{\beta}}) = \frac{1}{B}\sum_{b=1}^{B} I(\bar{\hat{\beta}}_b^* < 2\bar{\hat{\beta}}). \quad \text{(III.26)}$$

If B is infinitely large, the EDF converges to the true conditional distribution function (CDF). Consequently, our procedure would yield an exact test and the outcome of the test would be the same as the P value computed by using the conditional distribution function (CDF) of $\bar{\hat{\beta}}$.

Correction for the Estimates

Though a bootstrap has been adopted in the above method, it is still a "matched-pairs" design, which results in a sample proportion of merged pairs of 0.50. This type of sampling typically implies that the proportion of merged pairs in the sample is much larger than the proportion of such pairs in the grant population of all pairs (merged and non-merged). This design causes a "choice-based sample bias" of the constant and the coefficients in the standard logit models, in turn meaning that the probabilities being assessed in such models are more or less biased. Hence, we provide the correction for the estimates from the above quasi bootstrap logistic regression.

Since we know the fraction of ones in the population, τ, which equals $\frac{n_1}{n_0+n_1}$, we can use the prior correction for the logit model. For each logistic regression above, the constant item $\hat{\beta}_{t0}^*$ should be corrected by subtracting out the bias factor, $\ln\left[\left(\frac{1-\tau}{\tau}\right)\right]$, and other parameters are statistically consistent[5]. The final corrected estimate $\bar{\hat{\beta}}_1^c$ is the same as $\bar{\hat{\beta}}_1$ and the final corrected estimate $\bar{\hat{\beta}}_0^c$ for the constant item $\bar{\hat{\beta}}_0$ is

[5]Here, $y = 1/2$, so $\frac{y}{1-y} = 1$.

$$\overline{\hat{\beta}}_0 - \ln\left[\left(\frac{1-\tau}{\tau}\right)\right], \tag{III.27}$$

because $\overline{\hat{\beta}}_1 = \frac{\sum_{t=1}^{T}\hat{\beta}_{t1}^{*}}{T}$ and $\overline{\hat{\beta}}_0 = \frac{\sum_{t=1}^{T}\hat{\beta}_{t0}^{*}}{T}$. Subsequently, we have the following corollary.

Corollary 3 *The corrected estimate* $\left(\overline{\hat{\beta}}_0^c,\ \overline{\hat{\beta}}_1^c\right)$ *has the same asymptotic properties as* $\overline{\hat{\beta}}$.

Conclusion

This paper reviews three problems in explaining and predicting rare events data. First, the probability of rare events can be sharply underestimated by most popular statistical procedures, such as probit and logit regressions. In addition, the commonly used data collection strategies are far more efficient. Further, even when collecting all observations and necessary variables, the logistic and probit regressions may sometimes not reach numerically to a convergent result.

Although there are some problems for estimating rare events data, several methods have been constructed to provide consistent estimates under some conditions, such as the prior correction, weighting, and the corrected version of weighting with rare event corrections. In addition, this paper contributes a new simulation-based econometric method to the literature. The method adopts bootstrap to select the data. The procedure goes as the following. First, we randomly select the "nonevents" (pseudo merger pairs) to match the events (mergers) and run a logistic regression, and then repeat this procedure hundreds of times. We also construct the bootstrap standard errors and p values of the estimates.

In this paper, we set the identity matrix as the weighted matrix to estimate the parameters with the minimum distance method. With this setting, the simulation-based method bring consistent but not the most efficient estimates, since all the intermediate estimates $\left\{\hat{\beta}_t, t = 1, 2, ..., T\right\}$ from each logistic regression share some dependent variables M and then independent from each other. To get more efficient estimates, however, we better set the variance-covariance matrix of the estimates $\{\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_T\}$ from the logistic regressions as the weighted matrix, which we would like to explore further in the future.

CHAPTER IV

INTERNATIONAL TECHNOLOGY DIFFERENCES AND INCOME INEQUALITY: EVIDENCE FROM CROSS-COUNTRY HISTORICAL DATA

Introduction

Today, we live in a world with significant differences in living standards and production across countries. According to the Penn World Table project conducted by Robert Summers and Alan Heston, the richest countries such as the United States and Switzerland are about thirty times richer than the poorest countries in Africa and South Asia. Production per person in the wealthiest economy is thirty times more per person than in the poorest economy. Why are some economies so much richer than others? What accounts for sustained economic growth in these countries? Although there are a few different voices, most economists agree that the most important driver of economic growth is progress in technology. Klenow and Rodriguez-Clare (1997) as well as Hall and Jones (1999) show that differences in total factor productivity (TFP) account for a majority of the gap in income per capita between rich and poor countries. Prescott (1998) argues that differences in physical or human capital are not, in themselves, sufficient to account for the large international income differences. He emphasizes the need for developing a theory of TFP. "To account for sustained growth," Robert E. Lucas Jr. has written in the 2003 Annual Report of the Federal Reserve Bank of Minneapolis, "the modern theory needs to postulate continuous improvement in technology or in knowledge ..." In focusing on this area of the literature, we reexamine these fundamental questions by analyzing the relation between cross-country technology differences and income inequity.

A number of researchers have devoted much effort to proposing answers to the question: What are the determinants of cross-country disparities in technology? Most macroeconomists use the vintage capital models to understand the adoption of new technologies, such as Johansen (1959), Solow (1960), Gilchrist and Williams (2001) and Laitner and Stolyarov (2002). However, most vintage capital models assume that firms or countries invest only in frontier technology and the old vintages decrease because of depreciation, which violates the fact that investment in non-frontier technologies is an important empirical reality in cross-country technological adoption. Vintage human capital theory explains the technology adoption delays only for those technologies that are associated with technology-specific skills, such as Chari and Hopenhayn (1991), Brezis et al. (1993), Jovanovic and Nyarko (1996) and Jovanovic (2006b). Jones (1994) argues that different policies result in different technologies across countries. Barro and Sala-i-Martin (1997) and Eeckhout and Jovanovic (2002) provide imitator-innovator models to explain the fact that leaders tend to innovate and to be the first to adopt new technologies while the lagging countries mostly imitate. Parente and Prescott (1999) argue that poor countries cannot adopt better technologies because of monopoly rights that are protected via regulation. The poor country uses inefficient technology, and therefore, remains poor. Basu and Weil (1999) introduce an appropriate technology model, in which new technologies can only be implemented successfully by countries with the appropriate endowments. Khan and Ravikumar (2002) develop a model of technology adoption incorporating an irrecoverable and fixed cost. They show that there is a unique threshold level of wealth, which depends on technology parameters. If a country is richer than the threshold, it will adopt the new technology, otherwise it will retain the outdated technologies. Comin and Hobijn (2003) show that a country's human capital endowment, type of government, degree of openness to trade, and adoption of predecessor technologies are the most important determinants of the technology adoption speed. Antras (2005) argues that goods are initially manufactured in the country where they are first introduced because of contractual frictions, and then shifted to a low-wage foreign location when they become sufficiently standardized and need less management.

Some researchers focus on technology adoption and inequality. Matsuyama (2002) shows that income inequality induces technological adoption lags between rich and poor countries. He argues that each new product is bought first by the rich and then by the poor, and the diffusion lags are determined by income dispersion. Rogers (2003) concludes that the consequences of the diffusion of innovations usually widen the socio-economic gap between the earlier and later adopting categories in a system. Jovanovic (2006) states that a new technology adoption will first slow down economic growth and then speed it up. Lahiri and Ratnasiri (2007) use a dynamic general equilibrium model with household specific costs of technology adoption to catch the negative correlation between the degree of technology adoption and income inequality within countries.

However, most of the literature is short of measurement, which can capture the extensive and intensive margin of the technological adoptions, and then may only explain some specific anecdotes about the cross-country technology adoption patterns rather than general facts. It is therefore worthwhile to dig out the adoption processes and their relation to cross-country income inequality, which occur for most major technologies and most countries.

This paper re-investigates the Cross-Country Historical Adoption of Technology data set (CHAT) introduced by Comin and Hobijn (2006). This dataset covers about 110 technologies in over 150 countries over the last 200 years. Comin and Hobijn (2006) have summarized five facts about the historical technology usage: (1) Once the intensive margin is measured, technologies do not diffuse in a logistic way. (2) Within a typical technology, the dispersion in the adoption levels across countries is about 5 times larger than the cross-country dispersion in income per capita. (3) The rankings of countries by level of technology adoption are very highly correlated across technologies. (4) Within a typical technology, there has been convergence at an average rate of 4 percent per year. (5) The convergence rates of the technologies developed since 1925 have been almost three times higher than those of the technologies developed before 1925.

In this paper, we focus on the relation between technological adoptions and income dispersion across countries. We follow the example of the convergence of income per capita literature and estimate measures of absolute β-convergence. We also study the kernel joint density estimate of income and each technology, and test that most of the estimates are multimodal by means of the graphical technique of Significance in Scale Space, a visualization method based on the gradient direction. Besides these findings from Comin and Hobijn (2006), we explore CHAT further and document the following facts:

(i) Though DCs always adopt a new technology earlier than LDCs do, the convergence speed of technology adoption across LDCs is faster than that across DCs after this technology is also adopted by LDCs.

(ii) Most technological adoptions among poorer economies cluster in a lower level than those among richer economies.

(iii) The convergence speed of each technology adoption is non-monotone over time rather than accelerating.

(iv) Computer and internet invention have not increased the convergence speed of other technologies adoption across all countries, across DCs or across LDCs.

(v) The relation between the convergence speed of technology adoption and that of per capita income is negative across all countries and across LDCs, but is positive across DCs from the period 1946 - 1972 to the period 1973-2000.

(vi) The dispersion in technology adoption for individual technologies is 3 - 5 times larger than the dispersion in income per capita both across DCs and LDCs.

The remainder of this paper is organized as follows. In section 2, we describe the CHAT data set. In section 3, we present the empirical evidence. Our conclusions are presented in section 4.

Data

This paper reanalyzes the Cross-Country Historical Adoption of Technology (CHAT) data set, introduced by Comin and Hobijn (2006). This data set is an unbalanced panel, which contains historical information on the adoption of about 110 technologies in the past 200 years across over 150 countries. The technologies in the data set can be classified into nine groups: (1) Agriculture, (2) Finance, (3) Health, (4) Steel, (5) Telecommunication, (6) Textiles, (7) Tourism, (8) Transportation, and (9) General technologies.[1] CHAT also has the records of these countries' population from 1750 to 2000 and real GDP from 1820 to 2000 that is measured in million of 1990 international Stone-Geary dollars.

To capture both the extensive and the intensive margins of diffusion, we follow the measurement introduced by Comin and Hobijn (2003) to approximate the level of technology adoption. There are five different proxies. First, some technologies are measured as share of output produced by various production technologies, such as percent of irrigated land out of cultivated land in agriculture, percent of children aged 12 - 23 months who received a measles immunization before the age of one year in health, percent of steel produced by the acid

[1] The general technologies include three technologies, namely electricity production, the number of computers and the number of internet users.

Bessemer method and so on. Second, technologies in textiles and shipping are measured as capital share, which means a technology is approximated by the fraction of capital stock to be made up of equipment that embodies a particular technology. Third, we use production to real GDP to measure some production technologies for which CHAT does not have capital stock data but only data on output produced. We have four technologies measured in this way: civil aviation ton-KM of cargo carried on scheduled services by companies registered in the country concerned, ton-KM of freight carried on railways, metric tons of freight carried on railways and geographical length of line open at the end of the year. The forth proxy, for passenger cars, mobile phones, ATMs and so on, is defined as capital stocks per capita. The last but not the least measure is consumption per capita used for mail, telegrams, cheques issued, debit and credit card transactions as well as passenger transportation variables. The five proxies can be classified into three groups: on per capita basis, on unit real GDP basis, and the share of output produced with the technique. Technology variables description and their measurements are listed in Tables 9 - 11.

Empirical Evidence

Following the literature about the convergence of income per capita, we estimate the measures of absolute β-convergence. We estimate the speed of β-convergence of technology i by running the following regression for technologies measured in log-per-capita terms or in log-unit-real-GDP terms:

$$\ln Y_{ij,t} - \ln Y_{ij,t-1} = \alpha + c_j - (1 - e^{-\beta_i}) \ln Y_{ij,t-1} + u_{ij,t}; \tag{IV.1}$$

For technologies measured as shares, we estimate β-convergence from:

$$y_{ij,t} - y_{ij,t-1} = \alpha + c_j - (1 - e^{-\beta_i}) y_{ij,t-1} + u_{ij,t}, \tag{IV.2}$$

where α is constant, c_j is country j fixed effect, and $u_{ij,t}$ is the residual.

Following equation IV.1, we estimate βs across all countries, across developed countries (DCs) as well as across developing countries (LDCs) for these technologies measured in log-per capita or log-unit-real-GDP terms over all available time.[2] The results are reported in Tables 12 - 14.[3] It indicates that, among the 92 technologies, 22 of them experience convergence rates 0 - 0.5 times faster across LDCs than across DCs, 13 of them experience convergence rates 0.5 - 1 times faster within LDCs than within DCs, 21 of them have at least one time faster convergence speeds within LDCs than within DCs, 5 of them converge within LDCs but diverge within DCs, and only 20 of them have lower convergent speeds within LDCs. Half of the 20 technologies are from health industry and another half includes ship_sail, TV, txtlamat_totalraw, telephone, steel_stainless, railp, railPKM, railTKM, txtlmat_synth, and shipton_total.

Following equation IV.2, we estimate βs across all countries, across DCs as well as across LDCs for the 21 technologies measured in shares over all available time. Table 15 presents the results. Among the 21 technologies, one of them is convergent across LDCs but divergent across DCs, 12 of them converge faster across LDCs than across DCs, only five of them have lower convergence speed across LDCs than across DCs, and others have no enough information.

Fixing the physical time, we get very similar results as those from equation IV.1 and equation IV.2. Since World War I took place between 1914 and 1918, World War II

[2]We exclude Israel when we divide the countries into developed countries and developing countries.

[3]We measure steel and ship technologies both on log-per capita basis and as the share of output producted with the technique.

happened from 1939 to 1945, and personal computer and internet were invented on 1973, we divide the time series into three part: 1919 - 1938, 1946 - 1972 and 1973 - 2000.[4] In period 1919 - 1938, 6 percent out of the total 110 technologies have higher convergence speeds across DCs than across LDCs, 13 percent technologies have lower convergence speeds across DCs than across LDCs, and others do not have enough observations. In period 1946 - 1972, there are 9 percent technologies having higher convergence speeds across DCs than across LDCs, and 19 percent technologies with higher convergence speeds across LDCs than across DCs. In period 1973 - 2003, only 16 percent technologies that DCs have higher convergence speed than LDCs do, but 48 percent technologies that DCs have lower convergence speed than LDCs, and 4 percent technologies diverge in DCs but converge in LDCs.

One may argue that LDCs adopt lower quality technologies, and then have lower cost than DCs. Consequently, it takes less time for LDCs converging to their equilibrium than DCs. However, there are some technologies experiencing higher convergence speed across LDCs than across DCs, such as credit and debit card transactions as well as steel produced by electric arc furnaces, the measure of which is relatively homogenous both over time and across countries. For some technologies, such as cars, computers and TVs, though they are constantly reinvented and have important differences in the quality of the object measured over time and across countries, the measure we apply partially reflects the cross-country and time-series variation in the quality of technologies and also keep as homogenous as possible of these variant technologies, resulting from the positive correlation between demand and the quality of a technology moderating in part the differences in quality. The more important fact is that this kind technologies with different quality over time and across countries are not the majority which experiences a higher convergence speed across LDCs

[4]We think the inventions of internet and computer are special, because after that, low-cost information, data storage and transmission technologies are in general use, leading the way to deep changes in all fields of life.

than across DCs. As a result, the quality difference may not be the main reason to result a higher convergence speed in LDCs.

For robustness of the finding that technologies converge faster within LDCs than within DCs, we run the following regression for technologies measured in log-per-capita terms or in log-unit-real-GDP terms:

$$\ln Y_{ij,t} - \ln Y_{ij,t-1} = \alpha + c_j - (1 - e^{-\beta_i}) \ln Y_{ij,t-1} + \gamma_j \ln Y_{ij,t-1} \times dc + u_{ij,t}, \tag{IV.3}$$

while for technologies measured as shares we estimate β-convergence from:

$$y_{ij,t} - y_{ij,t-1} = \alpha + c_j - (1 - e^{-\beta_i}) y_{ij,t-1} + \gamma_j y_{ij,t-1} \times dc + u_{ij,t}, \tag{IV.4}$$

where dc is a dummy. When country j belongs to developed countries, $dc = 1$; otherwise $dc = 0$. If $\gamma_j > 0$, then β_j for LDCs is larger than DCs. If $\gamma_j = 0$, they have the same β_j. Otherwise, DCs have a larger β_j than LDCs. During the whole time series, 59 percent of technologies have a positive γ, and 63 percent out of them are significant. While 25 percent of technologies have a negative γ, but only 28 percent out of them are significant. We also fix the physical time. From 1919 to 1938, 13 percent of technologies have a positive γ, and 54 percent out of them are significant. While 8 percent of technologies have a negative γ, but only 38 percent out of them are significant. From 1946 to 1972, 28 percent of technologies have a positive γ, and 54 percent out of then are significant. While 10 percent of technologies have a negative γ, but only 10 percent out of them are significant. From 1973 to 2000, 59 percent of technologies have a positive γ, and 63 percent out of them are significant. While 25 percent of technologies have a negative γ, but only 28 percent out of them are significant. Consequently, based on these facts, we conclude our first finding as the following.

Fact 1: Though developed countries always adopt a new technology earlier than developing countries, the convergence speed of technology adoption across developing countries is faster than that across developed countries after this technology is adopted by developing countries as well.

In growth economics, β-convergence has been widely used to test whether each country converges to its own equilibrium, but does not provide much information on how economies perform relative to each other. A process of either technology homogenization or persistent gaps can be manifested in the shape of the joint density distribution of income per capita and the level of each technology adoption along the time. As Bianchi (1997), we also let the data speak for themselves through the nonparametric density estimators. To test whether the observed features from the nonparametric joint density estimators are "really there", we use the graphical technique of significance in scale space, a visualization method proposed by Godtliebsen, Marron and Chaudhuri (2002).

The kernel joint density estimate of the log of income per capita and technological adoption at a given year can be expressed as the following:

$$\hat{f}_h(x, y) = N^{-1} \sum_{k=1}^{N} K_h(x - X_k, y - Y_k), \tag{IV.5}$$

where X_k is the log of income per capita in country k at a given year, Y_k is the log (or the share) of the adoption level of technology Y in country k at a given year, $k = 1, ..., N$; K is the kernel function and h is the bandwidth. After K is taken to be a spherically symmetric Gaussian density, equation IV.5 can be written as the following product form,

$$K_h(x - X_k, y - Y_k) = \varphi_{h_x}(x - X_k)\varphi_{h_y}(y - Y_k), \tag{IV.6}$$

where φ_{h_i} denotes the rescaling

$$\varphi_{h_i}(\cdot) = \frac{1}{h_i}\varphi(\frac{\cdot}{h_i}),\ i = x \text{ or } y, \tag{IV.7}$$

where φ is the standard Gaussian density.

To investigate the features of the joint distribution of income per capita and technology adoptions, we use the "normal reference rule" to select the optimal bandwidth, which minimizes the mean integrated squared errors (MISE). The optimal bandwidth can be approximated by h_i,

$$h_i = \sigma_i n^{-1/6},\ i = x \text{ or } y,$$

where σ_i is the standard deviation of the ith variate and can be replaced by its sample estimator in practical implementations.[5]

Under these settings, we estimate the joint distribution of income per capita and each technology. These estimates can tell us which joint density is bimodal, and in turn tell us which technological adoption clusters in a lower level at LDCs. Because of the "curse of the dimension", kernel density estimates may miss important structure via oversmoothing, or may find unimportant spurious structure via undersmoothing, though it is a good smoothing method which can show structure in data. Even we apply the 'normal

[5]See Scott (1992), Bowman and Azzalini (1997), in which $h_i = \sigma_i\{\frac{4}{(d+2)n}\}^{1/(d+4)}$, for $i = 1, 2, ..., d$ and d is the dimension. In our paper, d=2.

reference rule', a data-based bandwidth selection, to select the bandwidth, we may not get good estimates of the joint density of income and each technological adoption, because the underlying distribution of our data may be non-Gaussian. The streamline version of Significance in Scale Space, which simultaneously study a very wide range of bandwidth, avoid bandwidth selection and test whether the observed features from the kernel density estimates are "really there". The streamlines are curves essentially indicating the gradient direction and therefore indicating the statistically significant structure. In others words, these curves indicate the direction that a drop of water would follow as it moves down the surface.

To construct these streamlines, we randomly pick up a point where the gradient is significant, and then end the streamlines when there is no significant gradient in this region or a peak/vally or a boundary is reached. The main idea is to test the significancy of the gradient of the kernel estimates.

Without specifying the bandwidth, as Godtliebsen et al. (2002), we bin our data to an equally spaced rectangular grid,

$$\{(x_i, y_j) : x_i = L_x + i\Delta_x, y_j = L_y + j\Delta_y, \quad i = 0, ..., n, \ j = 0, ..., m\}. \tag{IV.8}$$

The bin is a rectangular lattice, in which x_i is equally spaced over $[L_x + n\Delta_x]$ and y_j is equally spaced over $[L_y + m\Delta_y]$. Then the mapped points are counted to give a matrix C of bin counts, whose ith, jth entry is

$$c_{i,j} = \sharp(\text{data points assigned to bin } i, j). \tag{IV.9}$$

The kernel density estimate $\hat{f}_h$ of x and y can be approximated by

$$\tilde{f}_h = N^{-1}(C * \tilde{K}_h), \tag{IV.10}$$

where $*$ denotes bivariate discrete convolution, and $\tilde{K}_h$ is a matrix of evaluation of the kernel function K_h. The partial derivatives of $\hat{f}_h$ can be approximated by

$$D\tilde{f}_h = N^{-1}(C * D\tilde{K}_h), \tag{IV.11}$$

where D denotes various partial derivative operators, including $(\partial/\partial x), (\partial/\partial y), (\partial^2/\partial x^2), (\partial^2/\partial x \partial y)$ and $(\partial^2/\partial y^2)$.

At a given location, the gradient of the underlying density f is

$$G(f) = [(f_x)^2 + (f_y)^2]^{1/2},$$

where $f_x = (\partial/\partial x)f$ and $f_y = (\partial/\partial y)f$. The estimate of $G(f)$ is

$$\hat{G}(f) = [(\tilde{f}_{h,x})^2 + (\tilde{f}_{h,y})^2]^{1/2},$$

where $\tilde{f}_{h,x}$ and $\tilde{f}_{h,y}$ come from formula IV.11. The null hypothesis is

$$H_0 : G(f) = 0.$$

The null distribution of this test is based on the bivariate Gaussian distribution

$$\begin{pmatrix} \tilde{f}_{h,x} \\ \\ \tilde{f}_{h,y} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{\sigma_x^2}{8\pi h^4}, & 0 \\ \\ 0, & \frac{\sigma_y^2}{8\pi h^4} \end{pmatrix}\right).$$

Because of this Gaussian distribution, we have

$$\frac{(\tilde{f}_{h,x})^2}{\sigma_x^2/8\pi h^4} + \frac{(\tilde{f}_{h,y})^2}{\sigma_y^2/8\pi h^4} \sim \chi_2^2.$$

That means, we will reject the null hypothesis when

$$\frac{(\tilde{f}_{h,x})^2}{\hat{\sigma}_1^2} + \frac{(\tilde{f}_{h,y})^2}{\hat{\sigma}_2^2} > q_{\chi_2^2}(\alpha'), \tag{IV.12}$$

where $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are the estimates of the sample variance

$$\widehat{var}[D\hat{f}_h(x,y)] = \frac{1}{N-1}\left\{N^{-1}\sum_{k=1}^{N}(DK_h(x-X_k, y-Y_k))^2 - \left(D\tilde{f}_h(x,y)\right)^2\right\}. \tag{IV.13}$$

A new binned approximation for the first term inside the braces of the expression of $\widehat{var}[D\hat{f}_h(x,y)]$ is

$$N^{-1}\left\{C * [(D\tilde{K}_h)^2]\right\}.$$

In equation IV.12, α' is the significance level for confidence intervals, which cover the effective independent points of each bin at a desired overall significance level α.[6] The relation between α' and α is

[6] In the empirical work , we set $\alpha = 0.05$.

$$\begin{aligned}\alpha &= P\{k\text{th confidence interval not covering, } k = 1, ..., l\} \\ &= 1 - P\{\text{confidence interval covers}\}^l \\ &= 1 - (1-\alpha')^l,\end{aligned}$$

where l is the average number of the effective independent data points. According to "kernel weighted count" of the number of points in each bin, the effect sample size l in each bin is defined as

$$l = \sum\nolimits_{i=1}^{n} \sum\nolimits_{j=1}^{m} \mathrm{ESS}_{i,j} = \sum\nolimits_{i=1}^{n} \sum\nolimits_{j=1}^{m} \frac{\sum_{k=1}^{N} K_h(x_i - X_k, y_j - Y_k)}{K_h(0,0)} \approx \sum\nolimits_{i=1}^{n} \sum\nolimits_{j=1}^{m} \frac{C * \tilde{K}_h}{K_h(0,0)},$$

where ESS stands for effect sample size for each point.

After introducing the theories, we put them with the data. Since we will test that technological adoptions among poorer economies cluster in a lower level than that among richer economies, we only investigate these technologies that have observations in at least 65 countries in some years. Totally 32 technologies satisfy our conditions, out of which 24 are measured on per capita, 4 are measured in share, and the last 4 are measured in unit real GDP. To save space, we only present the kernel density estimates and their corresponding contour plots of the electricity production and income at year 1950, 1960, 1970, 1980, 1990 and 2000 in figure 6. From the kernel density estimates in figure 6, we can see that except in year 1950, the joint density of income and electricity production is multi-modal. From the contour plots, we can see that all the data are along the 45 degree line. It seems the LDCs' electricity production cluster in a lower level than that of DCs.

To test whether the LDCs experience a lower electricity production level, we draw the streamlines for each kernel density estimate. To save space again, only 4 scales, i.e.

bandwidth, are shown in figures 7 - 12, in which we present h=5, 5.3, 5.6, and 6.[7] In these figures, gray level images are shown, which are adjusted for maximum contrast. That is, the color white is used for all regions where the density is higher than 20% of its maximum and the color black is used for the minimum. The green "streamlines" indicate the gradient direction and therefore the structure of a surface. The purple lines are contour lines, which are orthogonal to the green ones. Statistical significant cluster will be highlighted by a purple circle surrounding it, because it will be a hill of high density (light gray). In these figures, the gray regions are around the 45 degree line, which means richer countries have higher electricity production. According to the light gray region, streamlines and purple lines show that the joint densities of income and electricity production are bimodal for year 1970, 1980, 1990 and 2000, but not for year 1950 and 1960. Although Streamlines and purple lines will give us strong evidence of presence of a feature, lack of streamlines and purples only indicates the evidence is not strong and does not prove absence of a cluster. Subsequently, we can not tell whether the bimodal structure is really not there at year 1950 and 1960.

Though we do not present every kernel density estimate of each technology in this paper, we summarize the result in Tables 16 - 18.[8] Tables 16 and 17 shows that 18 out of 24 technologies measured in per capita basis have multi-modal joint density with income per capita. In Table 18, 3 out of 4 technologies measured in unit real GDP and 1 out of 4 technologies measured in share have multi-modal joint densities with income per capita. This observation implies that technological adoptions in poor countries may converge to a lower level than that in rich countries, which is consistent with 'club convergence' theory in growth economics. The 'club convergence' theory says that developing countries dis-

[7]The full scale space can be shown in a movie, which is available upon request.

[8]Each kernel density estimate is available upon request.

play faster growth rates in terms of β-convergence but can not catch up with developed economies, such as Durlauf and Johnson (1995), Quah (1996a, b).

Based on our kernel density estimates and the streamline version of Significance in Scale Space, we conclude the following fact:

Fact 2: Most technological adoptions among developing economies cluster in a lower level than those among developed economies.

Comin and Hobijn (2006) document a fact that the speed of technological convergence across countries has accelerated over time. They also document that across all countries, the median speed of convergence for technologies invented before 1925 is 2 percent per year, that for technologies invented between 1925 and 1950 is 5.5 percent per year, and that for technologies invented since 1950 is about 6 percent per year. Besides, from Tables 12 - 15, we observe that some technologies invented in 1960s have very high convergence speeds, which result in a high median speed of convergence for technologies invented since 1950: the convergence speeds for transplant_lung invented in 1963, pctimmunizmeas invented in 1964, med_mammograph invented in 1966, transplant_liver invented 1967, transplant_heart invented in 1967, and atm invented 1967 correspondingly are 11.8 percent, 11.9 percent, 10.4 percent, 11.2 percent, 9.3 percent and 14.8 percent. There are some technologies invented after 1950 having very low convergence rate. For example, the convergence speeds of eft invented in 1979 and med_lithotriptor invented in 1980 are only 2 percent. From these tables, we also observe that technologies invented after 1973 have lower β than those invented in 1960s.

Looking at the technological adoption convergence speed across LDCs and DCs separately in Tables 12 - 15, we observe that across DCs the median speed of convergence for technologies invented before 1925 is 2.4 percent per year, that for technologies invented between 1925 and 1950 is 8.5 percent per year and that for technologies invented since 1950 is about 6.6 percent per year. Across LDCs the median speed of convergence for technologies invented before 1925 is 4.4 percent per year, that for technologies invented between 1925 and 1950 is 11.4 percent per year and that for technologies invented since 1950 is about 9.5 percent per year. The observations do not support the Comin and Hobijn's argument that the speed of convergence of technology across countries has accelerated over time.

Tables 19 -21 report the βs in 1919 - 1938, 1946 - 1972 and 1973 - 2000 across all countries estimated according to equations IV.1 and IV.2. From period 1919 - 1938 to period 1946 - 1972, among 110 technologies shown in this table, 16 technologies' convergence rates decrease but only 4 technologies' convergence rates increase, and from period 1946 - 1972 to period 1973 - 2000, 19 technologies' β decrease but 15 technologies' β increase. On average, the convergence speed of technologies adoption in period 1919-1938 across all countries is 0.111, that in period 1946 - 1972 is 0.137 and that in period 1973-2000 is 0.115, which are presented in Table 22. Table 22 also reports the βs across DCs and across LDCs in the three periods. We find the βs are increasing from period 1919-1938 to period 1946-1972, but decline from period 1946-1972 to 1973-2000.

The change of the cost to adopt a new technology may contribute to the non-monotone change of technological adoption convergence rate among these three periods. The cost of technological adoptions decrease in 1940s but increase in 1970s, which has the opposite trend as the change of technological adoption convergence rates.

Schmookler (1954) document that the patent-invention ratio fell after 1940, which implies a lower cost of adopting technologies. There are two reasons for the fall of patent-invention ratio: (1) As science and technology grew in complexity, the sheer intellectual difficulty of advancing beyond the existing boundaries probably increased. (2) The rise in the size of the average firm and the growth of scientific management probably increased the relative attention devoted to unpatentable small improvements.

The Patent Cooperation Treaty concluded in 1970 which covers 137 countries and the European Patent Convention to be effective in 1973 result the rising of the cost to adopt a new technology innovated by other countries after 1970s. The transition to information technology in the middle of 1970s cause the adjustment costs more than doubled, which is found by Bessen (2001).

From our exploration of the cross-country technology adoption, we can summarize our third finding as the following.

Fact 3: The convergence speed of each technology adoption is non-monotone along the time series.

After the inventions of internet and computer, low-cost information, data storage and transmission technologies are in general use, leading the way to deep changes in all fields of life. The Solow's paradox—that "we can see the computer age everywhere except in the productivity statistics" is still a puzzle to economists. Gordon (2000) summarizes five reasons to explain why the internet has not had a significant effect on productivity growth outside of durable manufacturing. First, since consumer time is limited, the internet use only substitutes for other forms of entertainment and information gathering. Second, much investment in Internet web sites and infrastructure competes for market share by redistributing sales rather than creating them. Third, internet provides preexisting information cheaply and conveniently rather than creates truly new products and activities. Fourth, much web site development duplicates existing forms of commerce and information which results higher costs. Fifth, a large fraction of consumption activity on the web takes place at the office, which decreases the efficiency of working. Another reason, not mentioned in

Gordon (2000), is the limited human being who grasp the knowledge to use PC and internet. For PC use and Internet access, a more differentiated and complex set of skills is required. However, the Eurobarometer survey on "ICT and work" indicates that most workers have no formal qualifications and many have had no training. The number of ICT trained workers is still limited and relative human capital does not have a significant increase after the invention of Computer and Internet.

Our regression results indicate that computer and internet invention have not increased the convergence speeds of other technologies invented after 1973, which are consistent with Solow's paradox. Across all countries, the convergence speeds for eft invented on 1979, for med_lithotriptor invented on 1980, for surg_corstent invented on 1980 and for med_mriunit invented on 1981 are 2 percent, 2 percent, 5.9 percent and 5.4 percent separately, some of which are even lower than the βs for technologies invented before 1925. Across DCs, the convergence speeds for eft invented on 1979, for med_lithotriptor invented on 1980, for surg_corstent invented on 1980 and for med_mriunit invented on 1981 are 6.9 percent, 12 .9 percent, 12 percent and 5.8 percent separately. Across LDCs, the convergence speeds for eft invented on 1979, for med_lithotriptor invented on 1980 and for med_mriunit invented on 1981 are 8.8 percent, 27.7 percent, and 5.4 percent separately.

Table 22 shows that βs increase from period 1919 - 1938 to period 1946 - 1972, but decline from period 1946 - 1972 to 1973 - 2000 across all countries, across DCs or across LDCs. In a words, the convergence speed of each technology is not significantly affected by the invention of computer and internet.

Therefore, we reach the following conclusion:

Fact 4: Computer and internet invention have not increased the convergence speed of other technologies adoption across all countries, across developed countries or across *developing countries.*

Klenow and Rodriguez-Clare (1997) and Hsieh and Klenow (2003) argue that differences in technology are the main determinant of income dispersion across countries. Rogers (2003) think that the diffusion of innovations generally causes wider socioeconomic gaps within an audience. One of his arguments is that by adopting innovations relatively sooner than others in their system, innovators and early adopters achieve windfall profits, thereby widening the socioeconomic gap between these earlier adopting categories versus later adopting categories. Thus, the earlier adopters get richer and the later adopters' economic gain is comparatively less.

From CHAT, however, we can not observe monotone relation between technological adoption and income growth. Across all countries and across LDCs, the convergence speed of technology adoption and that of per capita income is negative , while across DCs, it is negative from period 1919 - 1938 to period 1946 - 1972 and positive from period 1946 - 1972 to period 1973 - 2000. Table 7 reports the average β-convergence speed of technologies and income per capita in 1919 - 1938, 1946 - 1972 and 1973 - 2000 across all countries, DCs as well as LDCs. Across all countries or developing countries, the β-convergence of income per capita decrease from period 1919 - 1938 to period 1946 - 1972 but increase from period 1946 - 1972 to 1973 - 2000, while the average β-convergence of technologies adoption increase from period 1919 - 1938 to period 1946 - 1972 but decrease from period 1946 - 1972 to 1973 - 2000. Across developed countries, the β-convergence of income per capita decrease along the time series, while the average β-convergence of technologies adoption keep the same trend as those across all countries and across LDCs.

Across all countries in period 1946 - 1972, the technologies that have the highest convergence speed are spindle_ring, txtlmat_otherraw and txtlmat_otherraw. Their convergence speeds are 1.26, 0.942 and 0.6 respectively. To study whether the outliers dominate the average technological adoption convergence speed, we exclude the three technologies and recalculate the average convergence speed. After excluding them, however, technology adoption and income across all countries still experience a non-positive relation. Across DCs in period 1946 - 1972, the top three highest convergence speed is: that of spindle_ring is 0.777, that of txtlmat_otherraw is 1.173 and that of steel_acidbess is 0.600. Excluding the three technologies, the relation between technological adoption and income across DCs is still non-positive. Across LDCs in period 1946 - 1972, after spindle_ring and txtlmat_otherraw are excluded, which experience highest convergence speed, we still get a non-positive relation between technological adoption and income per capita.

To investigate whether the characteristics of each technology play an important role in stimulating its adoption speed, we study the average convergence speed for each industry. Table 23 reports the income per capita β-convergence and the average β-convergence of technologies in each industry in period 1919-1938, 1946-1972 and 1973-2000 across all countries. Table 24 reports the income per capita β-convergence and the average β-convergence of technologies in each industry in period 1919-1938, 1946-1972 and 1973-2000 across DCs. Table 25 reports the income per capita β-convergence and the average β-convergence of technological adoptions in each industry in period 1919 - 1938, 1946 - 1972 and 1973 - 2000 across LDCs. These tables show that most industries experience a decreasing convergence rates and then increasing among the three periods.

Since the cross-sectional dispersion falls with β, the negative relation between the dispersion of income and technology adoption implies enlarging dispersion of income per capita across countries when the dispersion of technology adoption declines. From period 1919 - 1938 to period 1946 - 1972, since a new technology was adopted by DCs countries in very different date, the income inequality enlarges. Now because of the globalization, a new technology is adopted by each developed country quickly, which results a positive relation between the dispersion of income per capita and that of technological adoption among the DCs from period 1946 - 1972 to period 1973 - 2000. For the developing countries, however, there are still huge gaps among the dates to adopt a new technology in period 1973 - 2000. As a result, the dispersion of per capita income and that of technology adoption are still negative related across LDCs from period 1946 - 1972 to period 1973 - 2000. Since there are 137 developing countries and only 28 developed countries in CHAT data set, the relation between income and technological adoption across all countries is dominated by developing countries.

Hence, we conclude with another finding of our analysis:

Fact 5: The relation between the convergence speed of technology adoption and that of per capita income is negative across all countries and across LDCs, but is positive across DCs from period 1946 - 1972 to period 1973 - 2000.

Comin and Hobijn (2006) find that the cross-country dispersion in technology adoption for individual technology is 3 - 5 times larger than cross-country dispersion in income. Using the same measurement provided by Comin and Hobijn (2006), we explore the relation between the dispersion of the technological adoption and that of income per capita across DCs and across LDCs.

For the technologies measured in log per capita or in log per real GDP, we compute the cross-country variance of 5-year moving averages of each technology adoption level after the year when it has been adopted in some countries. For the technologies measured as shares, we compute the cross-developed-country (cross-developing-country) coefficient of variance of 5-year moving averages of each technology level for which we have data.

We then compare the cross-developed-country (cross-developing-country) dispersion of each technology with the cross-developed-country (cross-developing-country) dispersion of either the log of income per capita (for log per capita and log per real GDP) or income per capita (for shares) over 5 years for each interval over the same time period. We aggregate all information across technologies both weighted by the length of our time series (measured by the number of five year periods for which we have data) and un-weighted. The results are shown in Table 26, which indicates that the relation between the variation in technology adoption and income across DCs and LDCs is consistent with the finding documented by Comin and Hobijn (2006). It can be summarized as follows:

Fact 6: the dispersion in technology adoption for individual technologies is 3 - 5 times larger than the dispersion in income per capita both across DCs and LDCs.

Conclusion

This paper reanalyzes the Cross-Country Historical Adoption of Technology (CHAT) data set. We estimate and compare the convergence speed of each technological adoption and that of income per capita across all countries, and then across the developed and developing countries (DCs and LDCs). We then document six new general facts about cross-country technology adoption and income inequality that emerge from these data: (i) Though DCs always adopt a new technology earlier than LDCs, on average the convergence speed of technology adoption across LDCs is faster than that across DCs. (ii) Most technological adoptions among poorer economies cluster in a lower level than those among richer economies. (iii) The convergence speed of the adoption of most technologies is non-monotone. (iv) The invention of the computer and the internet has not increased the average convergence speed of other technological adoptions. (v) The relation between the average convergence speed of technological adoptions and that of per capita income is negative across all countries and across LDCs, but is positive across DCs in the post-WWII period. (vi) The dispersion in technology adoption for individual technology is 3 - 5 times larger than the dispersion in income per capita both across DCs and LDCs. In addition, our paper is the first one in economics literature to adopt the graphical technique of Significance in Scale Space, a visualization method based on the gradient direction. This method avoids the bandwidth selection when studying the joint kernel density. In future analysis, we will try to provide a theoretical model to explain these findings.

Table 9. Technology variables and their measurement

Variable	Variable Description	Measurement
fert_total	Fertilizer consumed, total	on per capita basis
ag_harvester	Harvesters	on per capita basis
irrigatedarea	Irrigated area	on per capita basis
ag_milkingmachine	Milking machines	on per capita basis
pctirrigated	Percent of irrigated land out of cultivated land	share
pest_total	Pesticide consumed, total	on per capita basis
ag_tractor	Tractors	on per capita basis
atm	ATMS	on per capita basis
cheque	Cheques issued	on per capita basis
creditdebit	Debit and credit card transactions	on per capita basis
eft	Electronic funds transfers	on per capita basis
pos	Points of service for debit/credit cards	on per capita basis
elecprod	Electricity production	on per capita basis
internetuser	Internet users	on per capita basis
computer	Personal computers	on per capita basis
surg_appendectomy	Appendectomies	on per capita basis
bed_acute	Beds: in-patient acute care	on per capita basis
bed_longterm	Beds: in-patient long-term care	on per capita basis
bed_hosp	Beds: total hospital	on per capita basis
transplant_bonemarrow	Bone marrow transplants	on per capita basis
surg_breastcnsv	Breast conservation surgeries	on per capita basis
surg_csection	Caesarean sections	on per capita basis
surg_cardcath	Cardiac catheterizations	on per capita basis
surg_cataract	Cataract surgeries	on per capita basis
surg_cholecyst	Cholecystectomies	on per capita basis
surg_lapcholecyst	Cholecystectomies, laparoscopic	on per capita basis
med_catscanner	Computed tomography (CAT) scanners	on per capita basis
surg_corbypass	Coronary bypasses	on per capita basis
surg_corinterven	Coronary Interventions, Percutaneous	on per capita basis
surg_corstent	Coronary stenting procedures	on per capita basis
kidney_dialpat	Dialysis patients	on per capita basis
kidney_homedialpat	Dialysis patients, home	on per capita basis
transplant_heart	Heart transplants	on per capita basis
surg_hernia	Hernia procedures, inguinal and femoral	on per capita basis
surg_hipreplace	Hip replacement surgeries	on per capita basis
surg_hysterectomy	Hysterectomies	on per capita basis
transplant_kidney	Kidney transplants	on per capita basis
surg_kneereplace	Knee replacement surgeries	on per capita basis
med_lithotriptor	Lithotripters	on per capita basis
transplant_liver	Liver transplants	on per capita basis
transplant_lung	Lung transplants	on per capita basis
med_mammograph	Number of dedicated mammographs machines	on per capita basis

Table 10. (Continued) Technology variables and their measurement

Variable	Variable Description	Measurement
surg_mastectomy	Num. of mastectomies performed	on per capita basis
med_mriunit	Num. of MRI units	on per capita basis
surg_pacemaker	Pacemaker surgical procedures	on per capita basis
pctimmunizdpt	Per immunized for DPT, children 12-23 months	share
pctimmunizmeas	Per immunized for measles, children 12-23 months	share
pctdaysurg_cataract	Per of cataract surgeries done as day cases	share
pctdaysurg_lapcholecyst	Per of cholecystectomies done as day cases	share
pctdaysurg_cholecyst	Per of cholecystectomies done as day cases	share
pcthomedialysis	Per of dialysis patients at home	share
pctdaysurg_hernia	Per of hernia procedures done as day cases	share
pctdaysurg_tonsil	Per of tonsillectomies done as day cases	share
pctdaysurg_varicosevein	Per of varicose veins proc. done as day cases	share
surg_prostatextrans	Number of non-trans. prostatectomies performed	on per capita basis
surg_prostatetrans	Number of trans. prostatectomies performed	on per capita basis
med_radiationequip	Radiation therapy equipment	on per capita basis
surg_tonsil	Numer of tonsillectomies performed	on per capita basis
surg_varicosevein	Number of varicose vein corr. proc. performed	on per capita basis
steel_stainless	Stainless steel production	on per capita basis
steel_acidbess	Steel prod. by the acid Bessemer method	share
steel_basicbess	Steel prod. by the basic Bessemer method	share
steel_bof	Steel production in blast oxygen furnaces	share
steel_eaf	Steel production in electric arc furnaces	share
steel_ohf	Steel production in open hearth furnaces	share
steel_other	Steel production by other methods	share
cabletv	Cable television subscribers	on per capita basis
cellphone	Number of users of portable cell phones	on per capita basis
mail	Number of items mailed	on per capita basis
newspaper	Number of newspaper copies circulated daily	on per capita basis
radio	Number of radios	on per capita basis
telegram	Number of telegrams sent	on per capita basis
telephone	Number of mainline telephone lines	on per capita basis
tv	Number of television sets in use	on per capita basis
loom_auto	Looms: automatic	on per capita basis
loom_total	Looms: total	on per capita basis
spindle_mule	Number of mule spindles in place at year end	on per capita basis
spindle_ring	Number of ring spindles in place at year end	on per capita basis
txtlmat_artif	Spindle raw materials weight: artifical fibers	share
txtlmat_otherraw	Spindle raw materials weight: other	share
txtlmat_synth	Spindle raw materials weight: synthetic	share
visitorbed	Number of visitor beds available in hotels	on per capita basis
visitorroom	Number of visitor rooms available in hotels	on per capita basis

Table 11. (Continued) Technology variables and their measurement

Variable	Variable Description	Measurement
aviationPKM	Aviation: passenger kilometers	on per capita basis
aviationTKM	Aviation: freight ton-kilometers	on unit real GDP basis
railTKM	Railroads: freight ton-kilometers	on unit real GDP basis
railT	Railroads: freight ton-kilometers	on unit real GDP basis
railline	Railroads: length of line open	on unit real GDP basis
railP	Railroads: passenger journeys	on per capita basis
railPKM	Railroads: passenger kilometers	on per capita basis
ship_motor	Number of motor ships in use at midyear	share
ship_sail	Number of sail ships in use at midyear	share
ship_steam	Num. of steam ships in use at midyear	share
ship_steammotor	Num of steam&motor ships in use at midyear	share
shipton_motor	Ton of motor ships in use at midyear	share
shipton_sail	Ton of sail ships in use at midyear	share
shipton_steam	Ton of steam ships in use at midyear	share
shipton_steammotor	Ton of steam&motor ships in use at midyear	share
vehicle_com	Num of comm. vehicles, including buses & taxis	on per capita basis
vehicle_car	Num of passenger cars in use	on per capita basis

Table 12. Invention data and β convergence rate across all countries, DCs and LDCs for technologies based on per-capita or per-unit-real-GDP

Variable	Invention date	β				
		All	DCs	se	LDCs	se
ship_sail	pre-1500	0.02	0.001	0.006	-0.003	0.012
surg_appendectomy	1885	0.037	0.139	0.095***	0.007	0.190
kidney_dialpat	1943	0.065	0.043	0.006***	0.009	0.028
surg_corbypass	1953	0.036	0.104	0.030***	0.048	0.046**
tv	1924	0.009	0.076	0.005***	0.042	0.003***
txtlamat_totalraw	1884	0.134	0.164	0.040***	0.104	0.022***
telephone	1876	0.041	0.009	0.001***	0.006	0.002***
med_radiationequip	1900	0.083	0.087	0.042***	0.066	0.054**
steel_stainless	1913	0.022	0.248	0.043***	0.189	0.099***
railP	1825	0.013	0.015	0.002***	0.011	0.004***
railPKM	1825	-0.002	0.015	0.003***	0.011	0.005***
railTKM	1825	0.008	0.018	0.006***	0.014	0.006***
txtlmat_synth	1884	0.126	0.148	0.021***	0.115	0.022***
surg_hipreplace	1938	0.025	0.129	0.052***	0.114	0.306
transplant_bonemarrow	1956	0.043	0.066	0.017***	0.060	0.046***
transplant_liver	1967	0.112	0.095	0.018***	0.087	0.090*
shipton: total	pre-1500	0.009	0.031	0.006***	0.029	0.005***
med_mriunit	1981	0.054	0.058	0.016***	0.054	0.089
surg_pacemaker	1952	0.028	0.066	0.080*	0.063	0.277
med_catscanner	1972	0.037	0.061	0.012***	0.058	0.051**
vehicle_com	1885	0.024	0.021	0.002***	0.021	0.003***
shipton_steammotor	1788	0.02	0.026	0.005***	0.027	0.005***
vehicle_car	1885	0.055	0.013	0.002***	0.015	0.003***
loom_auto	1801	0.001	0.156	0.041***	0.176	0.027***
bed_hosp	pre-1500	0.082	0.061	0.022***	0.070	0.025***
irrigatedarea	pre-1500	0.012	0.026	0.005***	0.030	0.004***
aviationTKM	1903	0.033	0.031	0.005***	0.036	0.007***
bed_longterm	pre-1500	0.011	0.049	0.012***	0.058	0.072*
transplant_heart	1967	0.093	0.157	0.023***	0.189	0.132**
steel_eaf	1900	0.048	0.051	0.009***	0.065	0.012***
cellphone	1947	0.033	0.038	0.006***	0.048	0.008***
kidney_dialpat	1943	0.069	0.073	0.014***	0.093	0.084**
eft	1979	0.02	0.069	0.020***	0.088	0.053***
telegram	1835	0.001	0.037	0.006***	0.048	0.009***

Note: *** means significant at 1%, ** means significant at 5% and * means significant at 10%.

Table 13. (Continued) Invention data and β convergence rate across all countries, DCs and LDCs for technologies based on per-capita or per-unit-real-GDP

Variable	Invention date	β				
		All	DCs	se	LDCs	se
steel_bof	1950	-0.027	0.135	0.011***	0.177	0.018***
radio	1901	0.004	0.019	0.002***	0.025	0.002
pos	1950	0.148	0.095	0.021***	0.129	0.030***
ag_milkingmachine	1870	0.006	0.016	0.005***	0.022	0.007***
computer	1973	0.082	0.040	0.005***	0.057	0.006***
loom_total	1785	0.123	0.101	0.034***	0.145	0.026***
steel_ohf	1867	0.025	-0.030	0.021***	-0.044	0.030
surg_csection	pre-1500	0.03	0.052	0.046**	0.078	0.195
cabletv	1949	0.015	0.085	0.011***	0.132	0.021***
pest_total	1939	0.164	0.315	0.073***	0.490	0.051***
aviationPKM	1903	0.047	0.016	0.003***	0.026	0.004***
visitorbed	pre-1500	0.011	0.042	0.015***	0.070	0.012***
ship_steammotor	1788	0.002	0.018	0.007***	0.031	0.006***
fert_total	1815	0.004	0.045	0.013***	0.078	0.007***
transplant_kidney	1951	0.229	0.060	0.010***	0.106	0.038***
visitorroom	pre-1500	0.035	0.028	0.017***	0.050	0.010***
ag_tractor	1868	0.007	0.017	0.003***	0.032	0.003***
creditdebit	1950	-0.002	0.037	0.015***	0.067	0.065**
elecprod	1882	0.01	0.009	0.001***	0.016	0.002***
surg_hysterectomy	1843	0.028	0.207	0.079***	0.391	0.108***
ag_harvester	1850	0.008	0.012	0.003***	0.025	0.004***
med_lithotriptor	1980	0.02	0.129	0.029***	0.277	0.147***
mail	pre-1500	0.02	0.007	0.001***	0.015	0.004***
internetuser	1973	0.078	0.049	0.009***	0.107	0.010***
surg_cholecyst	1882	0.002	0.081	0.106	0.183	0.108**
ships: total	pre-1500	0.033	0.017	0.006***	0.046	0.007***
txtlmat_artif	1884	0.073	0.082	0.047***	0.223	0.115***
atm	1967	0.148	0.049	0.014***	0.148	0.023***
surg_tonsil	pre-1500	0.023	0.203	0.113***	0.648	0.459
newspaper	1606	0.041	0.010	0.006***	0.032	0.005***
shipton_steam	1788	0.001	0.018	0.006***	0.058	0.037***
surg_hernia	pre-1500	0.122	0.169	0.110***	0.589	0.123***
surg_mastectomy	pre-1500	-0.002	0.165	0.089***	0.726	0.248***
surg_cardcath	1941	0.044	0.040	0.040**	0.195	0.210

Note: *** means significant at 1%, ** means significant at 5% and * means significant at 10%.

Table 14. (Continued) Invention data and β convergence rate across all countries, DCs and LDCs for technologies based on per-capita or per-unit-real-GDP

Variable	Invention date	β				
		All	DCs	se	LDCs	se
surg_cataract	1748	0.072	0.040	0.040**	0.195	0.210
med_mammograph	1966	0.104	0.022	0.021**	0.117	0.081***
txtlmat_otherraw	1884	0.230	0.088	0.064***	0.607	0.103***
railT	1825	-0.001	0.001	0.003	0.009	0.005***
surg_kneereplace	1970	0.016	0.057	0.036***	0.663	0.564
ship_steam	1788	-0.002	0.004	0.005*	0.055	0.025***
surg_prostatextrans	1883	0.024	0.044	0.064	0.667	0.525
spindle_ring	1828	0.015	0.033	0.032**	0.657	0.064***
spindle_mule	1779	0.000	0.000	0.029	0.003	0.015
shipton_motor	1897	0.039	0.004	0.006***		
ship_motor	1897	0.024	0.009	0.008***		
steel_basicbess	1855	0.082	0.119	0.075***		
surg_corstent	1980	0.059	0.120	0.066***		
surg_prostatetrans	1931	0.051	0.145	0.070***		
surg_breastcnsv	pre-1500	0.022	0.160	0.087***		
surg_lapcholecyst	1901	0.039	0.180	0.051***		
steel_acidbess	1855	0.08	0.192	0.065***		
transplant_lung	1963	0.118	0.204	0.033***		
surg_corbypass	1953	0.033				
surg_varicosevein	pre-1500	0.033	-0.041	0.121	0.279	0.227
shipton_sail	pre-1500	0.027	-0.002	0.004	0.010	0.016
railline	1825	0.004	-0.004	0.002	0.009	0.005***
bed_acute	pre-1500	0.035	-0.015	0.010***	0.025	0.017***
cheque	pre-1500	0.059	-0.073	0.032***	0.022	0.112

Note: *** means significant at 1%, ** means significant at 5% and * means significant at 10%.

Table 15. Invention data and β convergence rate across all countries, DCs and LDCs for technologies measured in share

Variable	Invention date	β				
		All	DCs	se	LDCs	se
pctdaysurg_cataract	1748	0.039	-0.019	0.026	0.366	0.120***
pctdaysurg_varicosevein	pre-1500	0.007	0.031	0.037*	-0.215	0.746
pctimmunizdpt	1927	0.086	0.158	0.024***	0.082	0.009***
pctloom_auto	1801	0.063	1.275	0.055***	0.742	0.038***
pctirrigated	pre-1500	-0.005	0.015	0.006***	0.011	0.003***
pctimmunizmeas	1964	0.119	0.110	0.018***	0.102	0.010***
pctshipton_steammotor	1788	0.007	0.014	0.004***	0.016	0.004***
pctsteel_ohf	1855	0.014	0.009	0.010**	0.010	0.016
pctship_steammotor	1788	0.004	0.007	0.005***	0.009	0.002***
pctsteel_bof	1950	0.061	0.058	0.012***	0.081	0.016***
pctship_sail	pre-1500	0.002	0.002	0.003	0.003	0.009
pctsteel_eaf	1900	0.015	0.027	0.011***	0.048	0.011***
pctspindle_ring	1828	0.025	0.019	0.020**	0.049	0.036***
pctdaysurg_tonsil	pre-1500	0.001	0.223	0.069***	0.598	0.177**
pctship_steam	1788	0.051	0.039	0.012***	0.198	0.064***
pctshipton_sail	pre-1500	0.008	0.006	0.003***	0.041	0.015***
pctshipton_steam	1788	0.025	0.019	0.008***	0.174	0.067***
pctdaysurg_hernia	pre-1500	0.009	0.024	0.034	1.880	0.433*
pctdaysurg_cholecyst	1882	-0.145	0.010	0.042		
pctdaysurg_lapcholecyst	1901	-0.133	0.042	0.053*		
pcthomedialysis	1943	0.021			4.074	0.000***
pctship_motor	1897	0.009	0.009	0.008***		
pctshipton_motor	1897	0.004	0.004	0.006*		
pctsteel_acidbess	1855	0.276	0.597	0.078***		
pctsteel_basicbess	1855	0.023	0.002	0.039		
pctsteel_other	1855	0.04	0.173	0.072***		
pctsteel_stainless	1913	-0.005				

Note: *** means significant at 1%, ** means significant at 5% and * means significant at 10%.

Table 16. Summary of kernel density estimates and their feature tests of technologies measured on per capita basis

Technologies	Year	Observations	Modality	Multi-modality test	Bandwidth for test
Bed_hosp	1960	97	2	Not clear	5-6
	1970	97	2	Clear	5-6
	1980	69	2	Clear	5-6
	1990	120	2	Clear	5-6
visitorbed	1980	98	2	Clear	5-6
	1990	94	3	Clear	5-6
	2000	88	2	Clear	5-6
visitorroom	1980	99	2	Clear	5-6
	1990	95	3	Clear	5-6
	2000	95	2	Clear	5-6
Newspaper	1960	91	2	Clear	5-6
	1970	101	3	Clear	5-6
	1980	115	3	Clear	5-6
	1990	115	2	Clear	5-6
	1999	102	2	Clear	5-6
Railp	1960	85	3	Clear	5-6
	1970	84	3	Clear	5-6
	1980	72	3	Clear	5-6
Tractors	1970	108	2	Clear	5-6
	1980	118	3	Clear	5-6
	1990	119	3	Clear	5-6
	2000	109	2	Clear	5-6
Telephones	1960	98	2	Not clear	5-6
	1970	101	2	Clear	5-6
	1980	103	3	Clear	5-6
	1990	86	2	Clear	5-6
Elecprod	1950	94	1	Not clear	5-6
	1960	106	2	Not clear	5-6
	1970	115	2	Clear	5-6
	1980	118	2	Clear	5-6
	1990	120	3	Clear	5-6
	2000	75	2	Clear	5-6
Vehicle_car	1960	101	2	Clear	5-6
	1970	107	2	Clear	5-6
	1980	106	2	Clear	5-6
	1990	120	2	Clear	5-6
Vehicle_com	1960	103	2	Not clear	5-6
	1970	106	2	Not clear	5-6
	1980	105	2	Clear	5-6
	1990	94	3	Clear	5-6

Table 17. (Continued) Summary of kernel density estimates and their feature tests of technologies measured on per capita basis

Technologies	Year	Observations	Modality	Multi-modality test	Bandwidth for test
Radios	1960	94	3	Clear	5-6
	1970	112	3	Clear	5-6
	1980	117	3	Clear	5-6
	1990	119	3	Clear	5-6
	1999	115	3	Clear	5-6
Aviationpkm	1960	74	2	Clear	5-6
	1970	92	2	Clear	5-6
	1980	96	3	Clear	5-6
	1990	97	3	Clear	5-6
TVs	1970	90	2	Not clear	5-6
	1980	113	2	Clear	5-6
	1990	139	1	Clear	5-6
Cell phones	1995	105	2	Clear	5-6
	2000	110	2	Clear	5-6
Computers	1995	92	2	Clear	5-6
	2000	103	3	Clear	5-6
Internet users	1995	96	2	Clear	5-6
	2000	110	2	Clear	5-6
Textlmat_artif	1975	70	2	Clear	6-7
	1979	70	1	Not clear	6-7
Txtlmat_synth	1975	71	2	Clear	6-7
	1979	73	1	Not clear	6-7
Steel_EAF	1995	82	2	Clear	6-7
	2000	69	2	Clear	6-7
Irrigated area	1970	114	2	Not clear	7-8
	1980	115	2	Not clear	7-8
	1990	116	1	Not clear	7-8
	2000	108	2	Not clear	7-8
Fert	1970	113	2	Clear	7-8
	1980	117	3	Clear	7-8
	1990	117	2	Not clear	7-8
	2000	109	1	Not clear	7-8
Railpkm	1960	76	2	Clear	7-8
	1970	83	2	Clear	7-8
	1980	80	3	Clear	7-8
	1989	71	2	Clear	7-8
Harvesters	1970	87	2	Clear	7-8
	1980	87	2	Clear	7-8
	1990	88	2	Clear	7-8
	2000	90	1	Not clear	7-8
Cable TV	2000	72	2	Not clear	7-8

Table 18. Summary of kernel density estimates and their feature tests of technologies measured on share and on unit real GDP based

Technologies	Year	Observations	Modality	Multi-modality test	Bandwidth for test
In unit real GDP					
Railline	1960	94	1	Not clear	7-8
	1970	93	1	Not clear	7-8
	1980	87	2	Not clear	7-8
	1990	87	2	Not clear	7-8
RailT	1960	88	2	Clear	7-8
	1970	88	2	Clear	7-8
	1980	73	2	Clear	7-8
RailTKM	1960	83	2	Not clear	7-8
	1970	88	2	Clear	7-8
	1980	86	2	Clear	7-8
	1990	82	2	Clear	7-8
Aviationtkm	1970	89	3	Clear	7-8
	1980	95	1	Not clear	7-8
	1990	92	2	Clear	7-8
In share					
pctirrigated	1970	114	1	Not clear	5-6
	1980	115	2	Not clear	5-6
	1990	116	2	Clear	5-6
	2000	108	1	Not clear	5-6
pct_txtlmat_artif	1975	70	1	Not clear	5-6
	1979	70	1	Not clear	5-6
pct_txtlmat_synth	1975	71	2	Clear	5-6
	1979	73	2	Not clear	5-6
pct_steel_eaf	1995	82	3	Clear	7-8
	2000	69	3	Clear	7-8

Table 19. Each technology β-convergence rates in 1919-1938, 1946-1972 and 1973-2000 across all countries

Variable	1919-1938		1946-1972		1973-2000	
	β	se	β	se	β	se
Agriculture						
fert_total			0.095	0.016***	0.126	0.011***
harvester			0.050	0.013***	0.019	0.004***
irrigated area			0.052	0.014***	0.045	0.006***
milking machines			0.062	0.016***	0.016	0.008***
pctirrigated			0.054	0.014***	0.022	0.004***
pest_total					0.475	0.042***
tractors			0.065	0.008***	0.029	0.004***
Financial						
ATM					0.090	0.017***
cheque						0.031
creditdebit					0.067	0.032***
eft					0.073	0.023***
pos					0.094	0.023***
General						
elecprod	0.030	0.008***	0.024	0.004***	0.042	0.005***
internetuser					0.098	0.010***
computer					0.057	0.005***
Health						
appendectomies					0.206	0.097***
bed_acute					-0.009	0.010**
bed_longterm			0.073	0.031***	0.043	0.018***
bed_hosp					0.079	0.021***
transplant_bonemarrow					0.064	0.017***
surg_breastcnsv					0.171	0.098***
surg_csection					0.062	0.049***
surg_cardcath					0.080	0.034***
surg_cataract					0.052	0.041***
surg_cholecyst					0.130	0.112**
surg_lapcholecyst					0.215	0.051***
med_catscanner					0.060	0.020***
surg_corbypass					0.075	0.024***

Note: *** means significant at 1%, ** means significant at 5% and * means significant at 10%.

Table 20. (Continued) Each technology β-convergence rates in 1919-1938, 1946-1972 and 1973-2000 across all countries

Variable	1919-1938		1946-1972		1973-2000	
	β	se	β	se	β	se
surg_corinterven					0.087	0.029***
surg_corstent					0.107	0.075***
Dialysis patients					0.033	0.007***
Dialysis patients, home					0.086	0.017***
heart transplante					0.163	0.026***
surg_hernia					0.353	0.097***
surg_hipreplace					0.130	0.062***
surg_hysterectomy					0.272	0.080***
kidney transplants			0.106	0.078***	0.079	0.014***
surg_kneereplace					0.067	0.043***
lithotripters					0.145	0.032***
liver transplants					0.096	0.020***
lung transplant					0.221	0.036***
mammographs					0.079	0.042***
surg_mastectomy					0.200	0.090***
MRI units					0.061	0.024***
surg_pacemaker					0.115	0.093**
pctimmunizdpt					0.092	0.009***
pctimmunizmeas					0.107	0.010***
pctdaysurg_cataract					-0.030	0.032**
pctdaysurg_lapcholecyst					0.062	0.067*
pctdaysurg_cholecyst						
pcthomedialysis					0.098	0.028***
pctdaysurg_hernia						
pctdaysurg_tonsil					0.271	0.062***
pctdaysurg_varicosevein						
surg_prostatextrans						
surg_prostatetrans					0.185	0.080***
med_radiationequip					0.072	0.044***
surg_tonsil					0.228	0.104***
surg_varicosevein						
steel						
steel stainless					0.215	0.050***
steel_acidbess			0.600	0.081***		
steel_basicbess			0.072	0.060**	0.158	0.032*
steel_bof					0.091	0.014***
steel_eaf			0.088	0.03***	0.045	0.009***
steel_ohf			0.014	0.018*	0.015	0.016**
steel_other			0.113	0.058***		

Note: *** means significant at 1%, ** means significant at 5% and * means significant at 10%.

Table 21. (Continued) Each technology β-convergence rates in 1919-1938, 1946-1972 and 1973-2000 across all countries

Variable	1919-1938		1946-1972		1973-2000	
	β	se	β	se	β	se
telecommunications						
cabletv					0.103	0.012***
cell phones					0.046	0.007***
mail	0.100	0.020***	0.058	0.010***	0.055	0.014***
newspaper			0.066	0.007***	0.036	0.007***
radio	0.083	0.028***	0.036	0.004***	0.028	0.004***
telegram	0.563	0.033***	0.040	0.012***		
telephone	0.049	0.012***	0.020	0.005***	0.008	0.005***
TV			0.091	0.008***	0.042	0.004***
Textiles						
looms: automatic			0.211	0.033***	0.577	0.046***
looms: total			0.085	0.032***	0.480	0.043***
spindle_mule	0.125	0.093***	0.147	0.084***		
spindle_ring	0.135	0.069***	1.260	0.055***		
txtlmat_artif			0.288	0.039***	0.389	0.047***
txtlmat_otherraw			0.943	0.180***	0.139	0.091***
txtlmat_synth			0.120	0.040***	0.188	0.041***
Tourism						
visitorbed					0.075	0.011***
visitorroom					0.051	0.010***
Transportation						
aviationPKM	0.085	0.031***	0.057	0.007***	0.073	0.011***
aviationtkm			0.059	0.009***	0.092	0.012***
railtkm	0.207	0.034***	0.052	0.011***	0.049	0.012***
railt	0.077	0.021***	0.039	0.010***	0.054	0.013***
railline	0.058	0.018***			0.035	0.008***
railp	0.089	0.016***	0.050	0.011***	0.040	0.014***
railpkm	0.059	0.022***	0.058	0.011***	0.027	0.013***
ship_motor			0.023	0.021**		
ship_sail	0.041	0.020***	0.055	0.027***		
ship_steam	0.170	0.082***	0.093	0.027***	0.308	0.092***
ship_steammotor	0.049	0.024***	0.088	0.015***		
shipton_motor			0.013	0.013**		
shipton_sail	0.075	0.026***	0.232	0.018***	0.116	0.082***
shipton_steam	0.037	0.034**	0.032	0.017***	0.182	0.051***
shipton_steammotor	0.151	0.030***	0.135	0.012***	0.125	0.007***
vehicle_com	0.077	0.010***	0.045	0.006***	0.051	0.009***
vehicle_car	0.073	0.010***	0.028	0.005***	0.044	0.007

Note: *** means significant at 1%, ** means significant at 5% and * means significant at 10%.

Table 22. Average β-convergence speed of technologies and income per capita in 1919-1938, 1946-1972 and 1973-2000 across all countries, DCs as well as LDCs

	1919-1938	1946-1972	1973-2000
All countries	β	β	β
Per capita Income	0.068	0.010	0.028
Average β of tech_adoption	0.111	0.137	0.115
Developed countries	β	β	β
Per capita Income	0.074	0.010	0.006
Average β of tech_adoption	0.109	0.128	0.099
Developing countries	β	β	β
Per capita Income	0.063	0.011	0.036
Average β of tech_adoption	0.124	0.196	0.158

Table 23. Income per capita $\beta-$convergence rate and average $\beta-$convergence rate in each department in 1919-1938, 1946-1972 and 1973-2000 across all countries

	1919-1938	1946-1972	1973-2000
All countries	β	β	β
Income per capita	0.068	0.010	0.028
Agriculture		0.063	0.105
Finance			0.081
Electricity	0.031	0.024	0.042
Internet			0.098
Computer			0.057
Bed_longterm		0.073	0.043
Kidney_transplants		0.106	0.079
Health		0.090	0.118
Steel		0.177	0.105
Telecommunication	0.199	0.052	0.046
Textile	0.130	0.436	0.354
Tourism			0.063
Transportation	0.089	0.066	0.092

Table 24. Income per capita $\beta-$convergence rate and average $\beta-$convergence rate in each department in 1919-1938, 1946-1972 and 1973-2000 across DCs

Developed countries	1919-1938 β	1946-1972 β	1973-2000 β
Income per capita	0.074	0.010	0.006
Agriculture		0.060	0.069
Finance			0.045
Electricity	0.025	0.011	0.021
Internet			0.039
Computer			0.041
Bed_acute		0.027	-0.014
bed_longterm		0.071	0.040
Kidney_transplants		0.101	0.060
Health		0.066	0.111
Steel		0.177	0.132
Telecommunication	0.215	0.053	0.026
Textile	0.105	0.384	0.314
Tourism			0.037
Transportation	0.085	0.060	0.078

Table 25. Income per capita $\beta-$convergence rate and average $\beta-$convergence rate in each department in 1919-1938, 1946-1972 and 1973-2000 across LDCs

Developing countries	1919-1938 β	1946-1972 β	1973-2000 β
Income per capita	0.063	0.011	0.036
Agriculture		0.066	0.107
Finance			0.158
Electricity	0.035	0.026	0.045
Internet			0.125
Computer			0.062
Health			0.229
Steel		0.811	0.087
Telecommunication	0.194	0.058	0.051
Textile	0.214	0.502	0.407
Tourism			0.065
Transportation	0.093	0.078	0.060

Table 26. Dispersion in technology adoption relative to dispersion in income per capita

DCs	Average dispersion			Percentage of instances with ratios>1		
	Log per capita	Share	All	Log per capita	Share	All
Weighted by # of 5-Year Intervals	3.97	2.36	3.99	100	79	93
Un-weighted	4.15	3.54	3.98	99	85	96

LDCs	Average dispersion			Percentage of instances with ratios>1		
	Log per capita	Share	All	Log per capita	Share	All
Weighted by # of 5-Year Intervals	3.48	1.43	3.12	98	51	90
Un-weighted	5.21	1.55	4.45	92	57	85

BIBLIOGRAPHY

[1] Alvarez, F., Lucas, R.E., Jr. 2005. "General Equilibrium Analysis of the Eaton-Kortum Model of International Trade." NBER Working Paper.

[2] Amemiya, T. and Q. H. Vuong, 1987, "A Comparison of Two Consistent Estimators in the Choice-Based Sampling Qualitative Response Model." Econometrica 55(3), 699-702.

[3] Andrade, G., M. Mitchell and E. Stafford, 2001, "New Evidence and Perspectives on Mergers." Journal of Economic Perspectives, 15(2), 103-120.

[4] Antras, P., 2005. "Incomplete Contracts and the Product Cycles." American Economic Review 95(4), 1054-1073.

[5] Backus, D. K., Kehoe, P. J., kydland, F. E., 1994. "Dynamics of the Trade Balance and the Terms of Trade: the J-curve?" American Economic Review 84(1), 84-103.

[6] Baily, M.N., Hulten, C., Campbell, D., 1992. "Productivity Dynamics in Manufacturing Plants." Brookings Papers on Economic Activity. Microeconomics, 187-267.

[7] Bernard, A.B., Eaton, J., Jensen, J.B., Kortum, S., 2002. "Plants and Productivity in International Trade." Working Paper.

[8] Bessen, J. 1998. "Discovery, Learning and Adoption of New Techniques: Choosing specialization to optimize technical progress." Working paper.

[9] Bianchi, M. 1997. "Testing for convergence: evidence from non-parametric multimodality tests." Journal of Applied Econometrics, 12, 393-409.

[10] Bowman, A.W., and A. Azzalini, 1997. "Applied Smoothing Techniques for Data Analysis." London: Oxford University Press.

[11] Chari, V. and H. Hopenhayn, 1991."Vintage Human Capital, Growth, and the Diffusion of New Technology." Journal of Political Economy, 99, 1142-1165.

[12] Cockburn, I. and Z. Griliches, 1988, "Industry Effects and Appropriability Measures in the Stock Market's Valuation of R&D and Patents." American Economic Review, 78 (2), 419-423.

[13] Comin, D. and B. Hobijn, 2004. "Cross-country Technology Adoption: Making the Theories Face the Facts." Journal of Monetary Economics, 51, 39-83.

[14] Comin, D. and B. Hobijn, 2007. "Implementing Technology." Working paper.

[15] Comin, D., B. Hobijn and E. Rovito, 2006a, "Five Facts You Need to Know About Technology Diffusion." Working paper.

[16] Comin, D., B. Hobijn and E. Rovito, 2006b, "World Technology Usage Lags." Working paper.

[17] Cosslett, S. R., 1981a, "Maximum Likelihood Estimator for Choice-Based Samples." Econometrica 49(5), 1289-1316.

[18] Cosslett, S. R., 1981b, "Efficient Estimation of Discrete-Choice Models." In Structural Analysis of Discrete Data with Econometric Applications, eds. Charles F. Manski and Daniel McFadden. Mit Press. MA: Cambridge.

[19] Davidson, R. and J. G. MacKinnon, 2004, "Econometric Theory and Methods." Oxford University Press.

[20] Dollar, D., Kraay, A., 2004. "Trade, Growth, and Poverty." Economic Journal, 114, F42-F49.

[21] Durlauf, S. and P. Johnson, 1995, "Multiple regimes and cross-country growth behaviour." Journal of Applied Econometrics, 10, 365-384.

[22] Dwyer, D. W., 2001, "Plant-level Productivity and The Market Value of A Firm." Working Paper.

[23] Eaton, J., Kortum, S., 1999. "International Technology Diffusion: Theory and Measurement." International Economic Review, 40, 537-570.

[24] __________________,2001. "Technology, Trade, and Growth: A Unified Framework." Working Paper.

[25] __________________, 2002. "Technology, Geography, and Trade." Econometrica, 70(5), 1741-1779.

[26] Erickson, T., and T. M. Whited, 2000. "Measurement Error and the Relationship between Investment and q." Journal of Political Economy, 108, 1027-1057.

[27] Erickson, T., and T. M. Whited, 2006. "On the Accuracy of Different Measures of q." Financial Management, 35, 5-33.

[28] Faria, A., 2002, "Mergers and the Market for Organization Capital." University of Chicago, Mimeo.

[29] Gilchrist, S. and J. Williams, 2001. "Transition Dynamics in Vintage Capital Models: Explaining the Postwar Catch-up of Germany and Japan." Finance and Economics Discussion Series, 2001-7, Federal Reserve Board of Governors.

[30] Godtliebsen, F., J.S. Marron and P. Chaudhuri, 2002. "Significance in Scale Space for Bivariate Density Estimation." Journal of Computational and Graphical Statistics, 11(1), 1-21.

[31] Gordon, R.J., 2000. "Does the ‘New Economy’ Measure up to the Great Inventions of the Past?" Journal of Economic Perspectives, 4(14), 49-74.

[32] Gort, M., 1969, "An Economic Disturbance Theory of Mergers." Quarterly Journal of Economics, 83, 624-642.

[33] Greenwood, J., Z. Hercowitz and P. Krusell, 1997, "Long-Run Implications of Investment-Specific Technological Change." The American Economic Review, 87(3), 342-362.

[34] Greenwood, J., Z. Hercowitz and P. Krusell, 2000, "The Role of Investment-Specific Technological Change in The Business Cycle."European Economic Review, 44, 91-115.

[35] Griliches, Z., 1981, "Market Value, R&D and Patents," Economics Letters 7, l83-l87.

[36] Harford, J., 2005, "What Drives Merger Waves?" Journal of Financial Economics, 77, 529-560.

[37] Helpman, E., 1981. "International Trade in the Presence of Product Differentiation, Economies of Scale and Monopolistic Competition." Journal of International Economics, 11, 305-340.

[38] Hulten, C. R., 1992, "Growth Accounting When Technical Change Is Embodied in Capital." The American Economic Review, 82(4), 964-980.

[39] Imbens, G., 1992, "An Efficient Method of Moments Estimator for Discrete Choice Models with Choice-Based Sampling." Econometrica 60(5), 1187-1214.

[40] Jensen, M. C., 1986, "The Market for Corporate Control: Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers," The American Economic Review, (Papers and Proceedings), 76(2), 323-329.

[41] Johansen, L., 1959. "Substitution Versus Fixed Production Coefficients in the Theory of Economic Growth: a Synthesis." Econometrica, 27, 157-176.

[42] Jovanovic, B., and P. Rousseau, 2002, "The Q-Theory of Mergers." The American Economic Review, (Papers and Proceedings), 92(2), 198-204.

[43] Jovanovic, B., and P. Rousseau, 2007, "Mergers as Reallocation." Review of Economics and Statistics, forthcoming.

[44] Jovanovic, B., 2006a, "Asymmetric Cycles." Review of Economic Studies, 73, 145-162.

[45] Jovanovic, B., 2006b, "The Technology Cycle and Inequality." Working paper.

[46] Khan A. and Ravikumar B, 2002. "Costly Technology Adoption and Capital Accumulation." Review of Economic Studies, 5, 489-502.

[47] King, G. and L. Zeng, 2001a, "Logistic Regression in Rare Events Data." The Society for Political Methodology, 137-163.

[48] King, G. and L. Zeng, 2001b, "Explaining Rare Events in International Relations." International Organization 55(3), 693-715.

[49] Kortum, S., 1997. "Research, Patenting, and Technological Change." Econometrica, 65(6), 1389-1419.

[50] Kremer, M., 1993. "Population Growth and Technological Change: One Million B.C. to 1990." Quarterly Journal of Economics, 108, 681-716.

[51] Kuznets, S., 1966. "Economic Growth and Structure: Selected Essays by Simon Kuznets."

[52] Lahiri, R. and and S. Ratnasiri, 2007. "Concerning Technology Adoption and Inequality." Working paper.

[53] Laitner, J. and D. Stolyarov, 2002. "Technological Change and the Stock Market." Mimeo, University of Michigan.

[54] Lancaster, T., and G. Imbens, 1996a, "Case-Control with Contaminated Controls." Journal of Econometrics, 71, 145-160.

[55] Lancaster, T., and G. Imbens, 1996b, "Efficient Estimation and Stratified Sampling." Journal of Econometrics, 74, 289-318.

[56] Lawson, A.B. and D. G. Denison, 2002. "Spatial Cluster Modelling." Chapman & Hall/CRC press.

[57] Manski, C. F., 1999, "Nonparametric Identification Under Response-Based Sampling." In Nonlinear Statistical Inference: Essays in Honor of Takeshi Amemiya, eds. C. Hsiao, K. Morimune, and J. Powell. New York: Cambridge University Press.

[58] Manski, C. F. and S. R. Lerman, 1977, "The Estimation of Choice Probabilities from Choice Based Samples." Econometrica 45(8), 1977-1988.

[59] Manski, C. F. and D. McFadden, 1981, "Alternative Estimators and Sample Designs for Discrete Choice Analysis." In Structural Analysis of Discrete Data with Econometric Applications, eds. Charles F. Manski and Daniel McFadden. Cambridge: MA: MIT Press.

[60] Matsuyama, K., 2002. " The Rise of Mass-Consumption Societies." Journal of Political Economy, 110(5), 1035-1070.

[61] Mitchell, M.L. and J.H. Mulherin, 1996, "The Impact of Industry Shocks on Takeover and Restructuring Activity." Journal of Financial Economics, 41, 193-229.

[62] Parente, S.L. and E. Prescott, 2000. "Barriers to Riches." MIT Press.

[63] Pritchett, L., 1997. "Divergence, Big Time." Journal of Economic Perspectives, 11, 3-18.

[64] Quah, D.T., 1996a. "Empirics for Economic Growth and Convergence." European Economic Review, 40(6), 1353-75.

[65] __________, 1996b. "Convergence Empirics across Economies with (some) Capital Mobility." Journal of Economic Growth,1 95-124.

[66] Rhodes-Kropf, M. and D. Robinson, 2007, "The Market for Mergers and the Boundaries of the Firm." Journal of Finance, forthcoming.

[67] Rhodes-Kropf, M., D. Robinson, and S. Viswanathan, 2005,"Valuation Waves and Merger Activity: the Empirical Evidence." Journal of financial Economics, 77, 561-603.

[68] Rhodes-Kropf, M. and S. Viswanathan., 2004, "Market Valuation and Merger Waves." Journal of finance, 59, 2685-2718.

[69] Rogers, E.M., 2003. "Diffusion of Innovations." Free Press.

[70] Schmitz, J.A., 2005. "What Determines Productivity? Lessons from the Dramatic Recovery of the U.S. and Canadian Iron Ore Industries Following Their Early 1980s Crisis." Journal of Political Economy, 113, 582-625.

[71] Schmookler, J., 1954. "The Level of Inventive Activity." The Review of Economics and Statistics, 36, 183-190.

[72] Scott, D.W., 1992. "Multivariate Density Estimation: Theory, Practice, and Visualization." New York: John Wiley.

[73] Servaes, H., 1991, "Tobin's Q and the Gains from Takeovers." The Journal of Finance, 46(1), 409-419.

[74] Shleifer, A. and R. Vishny, 2003, "Stock Market Driven Acquisitions." Journal of Financial Economics, 70, 295-311.

[75] Skogsvik, K., 2005, "On the Choice-Based Sample Bias in Probabilistic Business Failure Predecition." Working Paper.

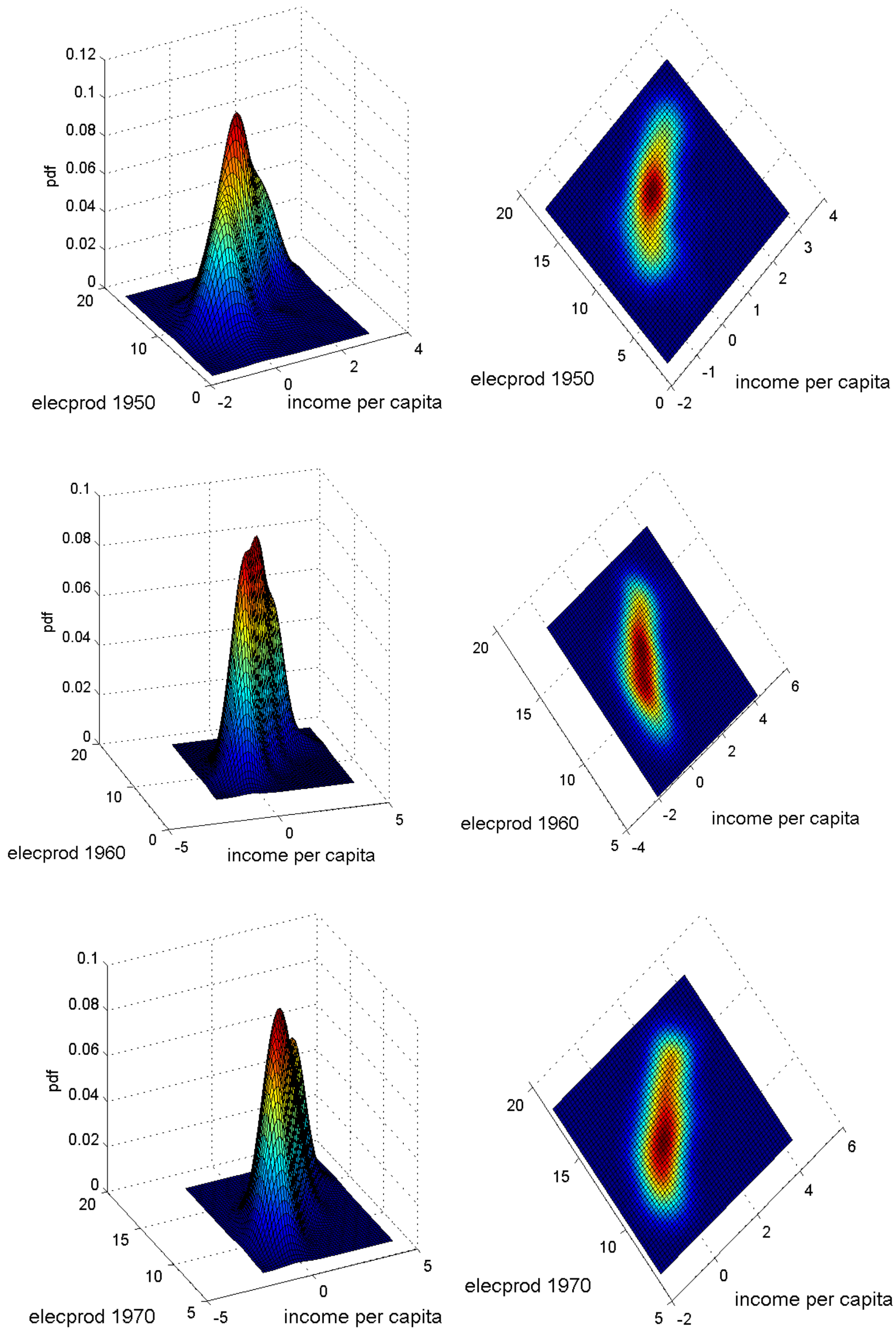
[76] Solow, R., 1960. "Investment and Technical Progress." In: Arrow, K. (Ed.), Mathematical Methods in the Social Sciences, Stanford University Press.

[77] Takii, K., 2006. "The Persistence of Differences in Productivity, wages, Skill Mix and Profits Across Firms." Working paper.

[78] Wilson, C.A., 1980. "On the General Structure of Ricardian Models with Continuum Goods: Application to Growth, Tariff Theory, and Technical Change." Econometrica, 48, 1675-1702.

[79] Xie, Y. and C. F. Manski, 1989, "The Logit Model and Response-Based Samples." Sociological Methods and Research 17(3), 283-302.

Figure 6: Kernel density estimates and their corresponding contour plots of income per capita and electricity production in year 1950, 1960, 1970, 1980, 1990 and 2000



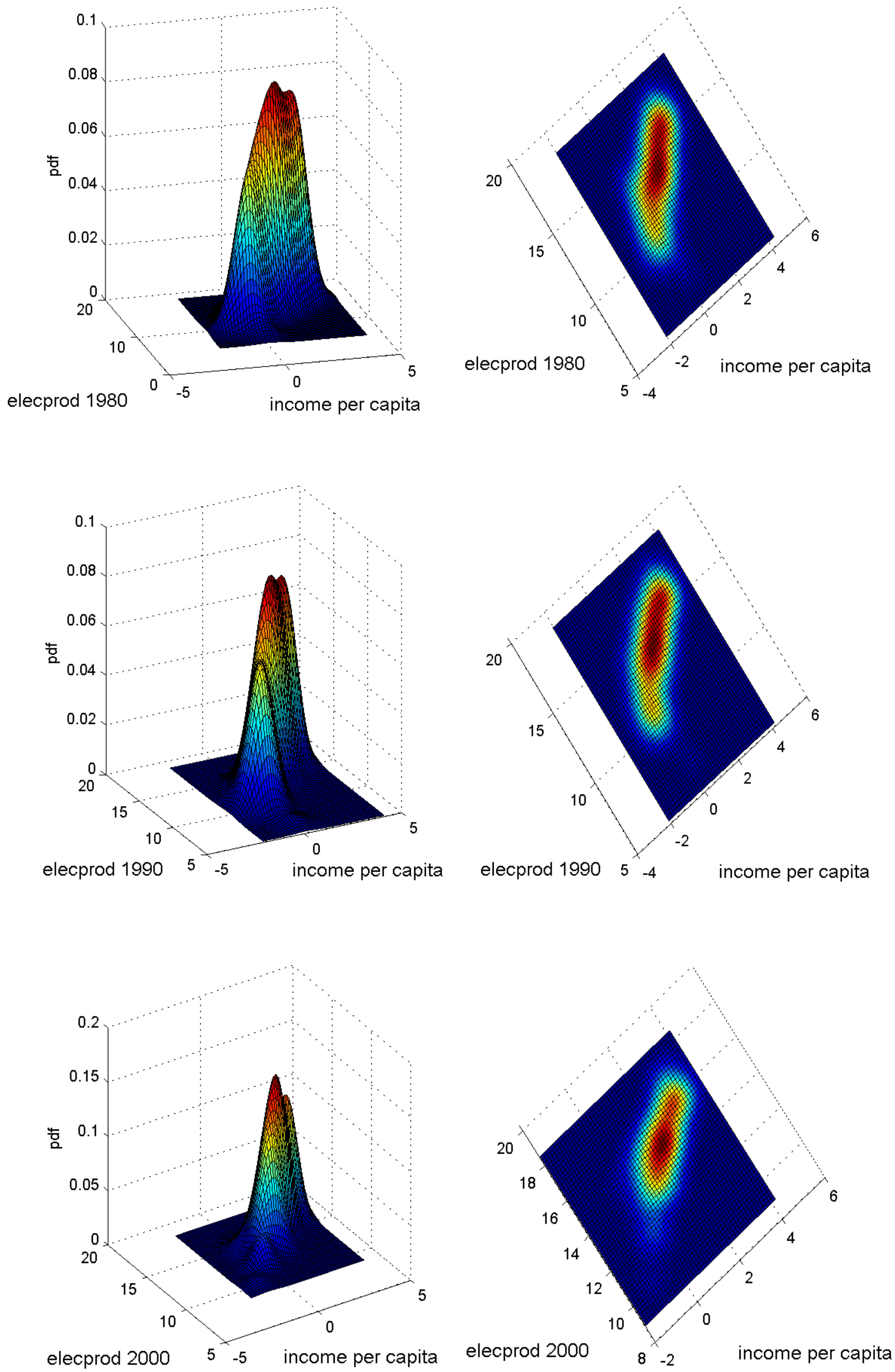
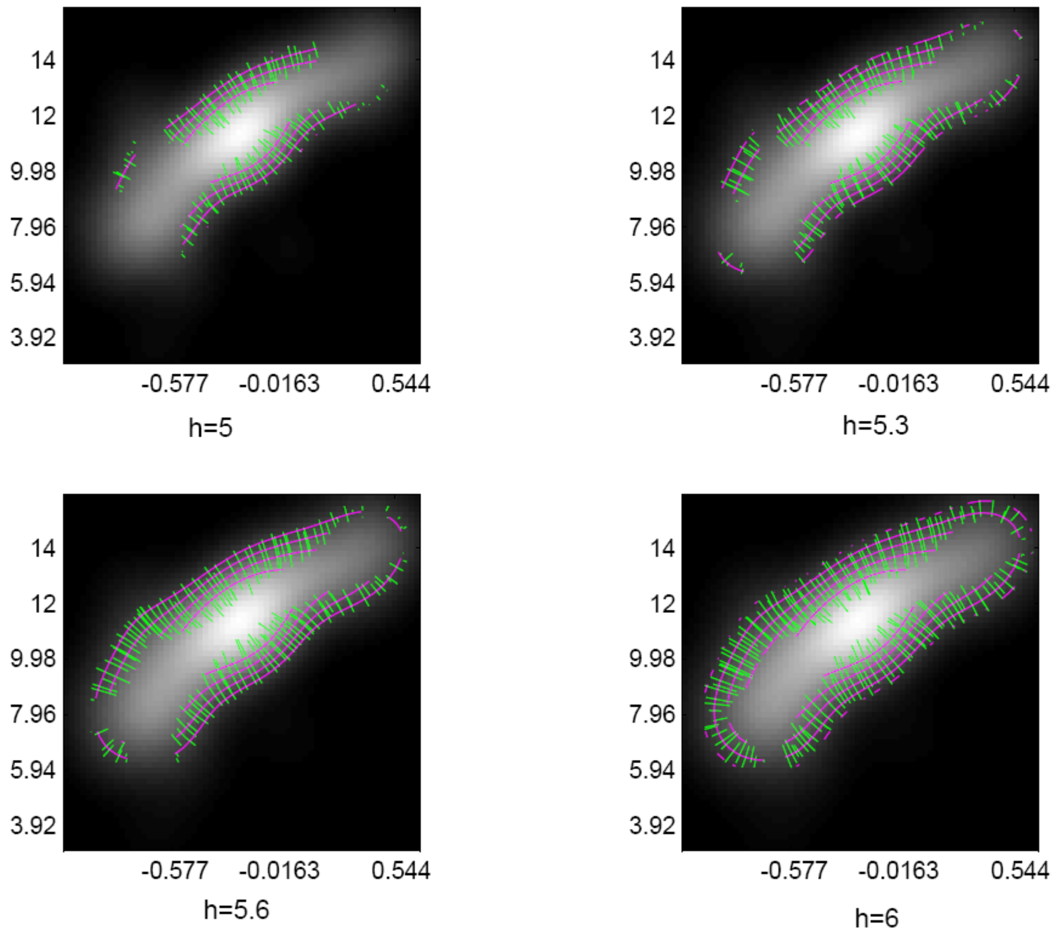

pdf
elecprod 1980
income per capita
elecprod 1980
income per capita
pdf
elecprod 1990
income per capita
elecprod 1990
income per capita
pdf
elecprod 2000
income per capita
elecprod 2000
income per capita

Figure 7: Significance in Scale-Space analysis of the joint density of income per capita and electricity production at year 1950



Note: Figures 7 – 12 show the streamlines for each corresponding kernel density estimate. The x-axis represents income per capita, the y-axis represents electricity production, and h represents bandwidth. The color white is used for all regions where the density is higher than 20% of its maximum and the color black is used for the minimum. The green ‘streamlines’ indicate the gradient direction and therefore the structure of a surface. The purple lines ate contour lines, which are orthogonal to the green gradient lines. Statistical significant cluster will be highlighted by a purple circle surrounding it, because it will be a hill of high density.

Figure 8: Significance in Scale-Space analysis of the joint density of income per capita and electricity production at year 1960

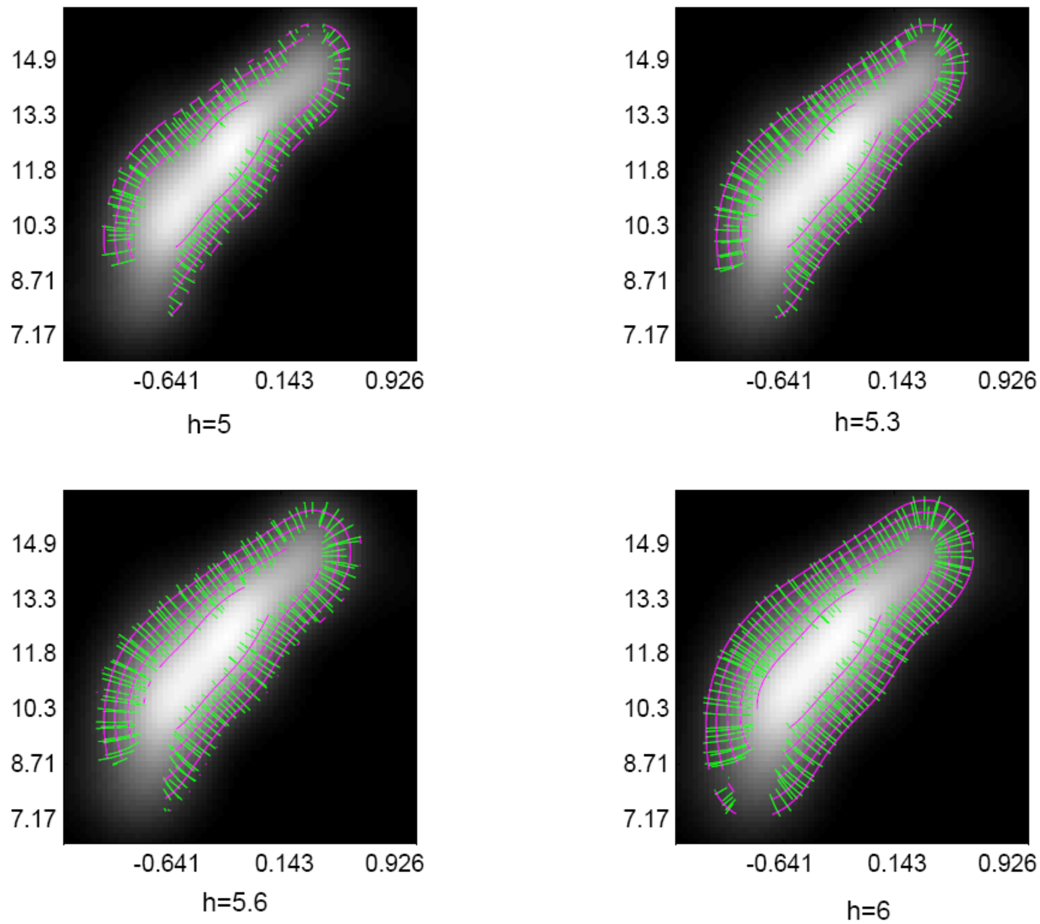


Figure 9: Significance in Scale-Space analysis of the joint density of income per capita and electricity production at year 1970

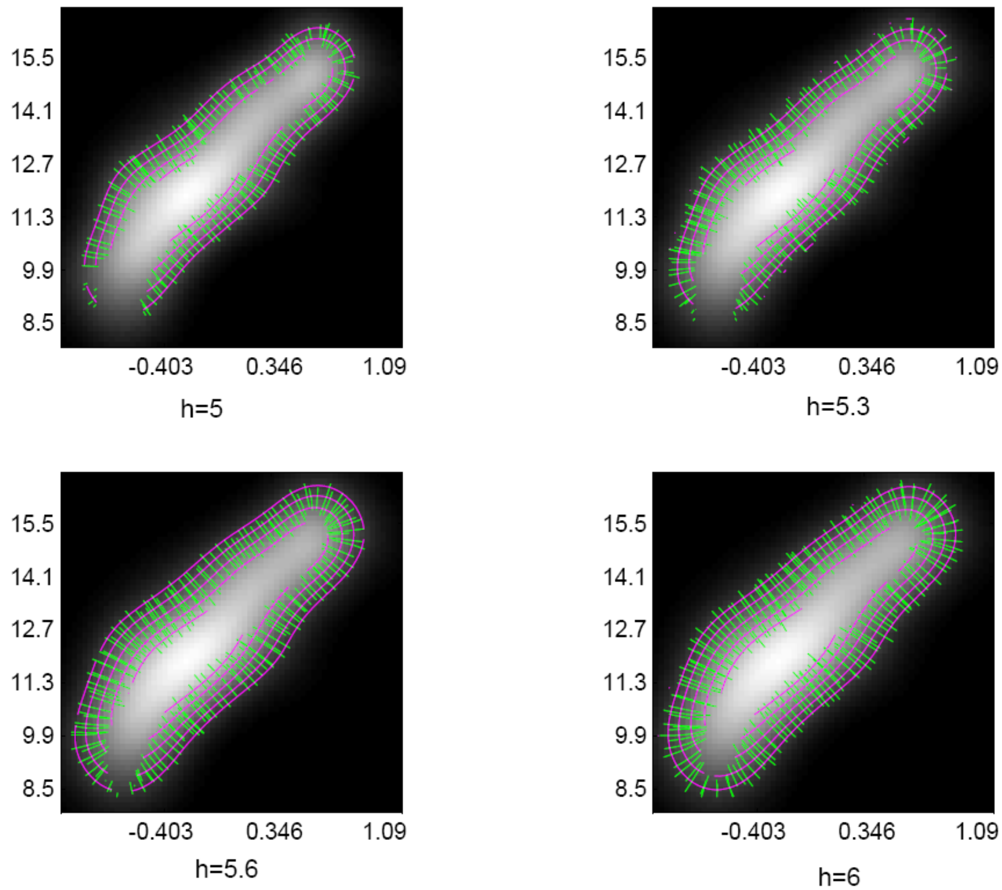


Figure 10: Significance in Scale-Space analysis of the joint density of income per capita and electricity production at year 1980

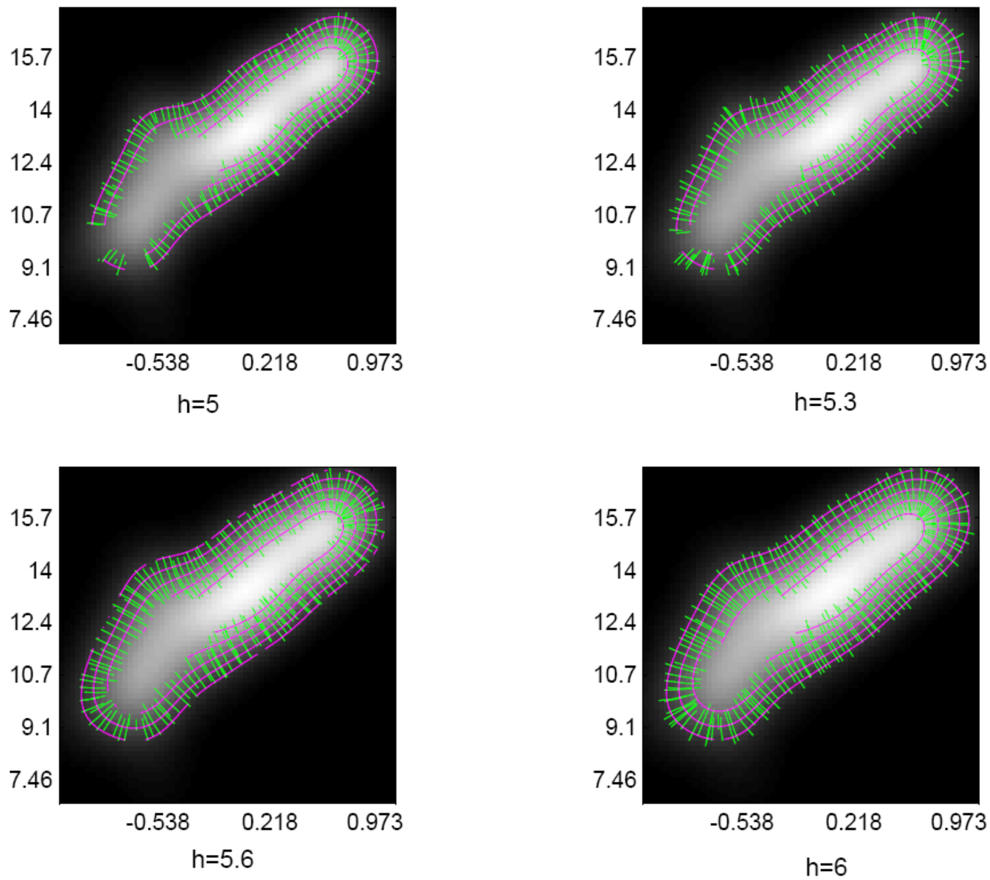


Figure 11: Significance in Scale-Space analysis of the joint density of income per capita and electricity production at year 1990

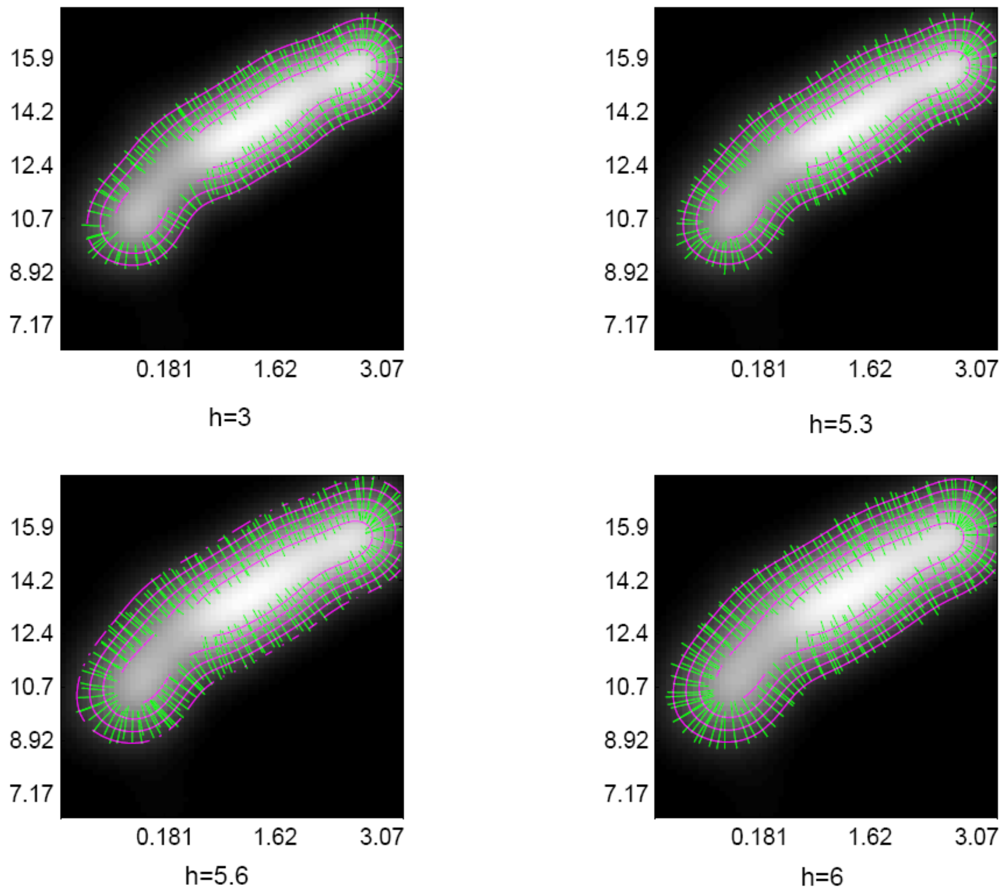


Figure 12: Significance in Scale-Space analysis of the joint density of income per capita and electricity production at year 2000