Learning, Diffusion and Industry Life Cycle

Zhu Wang*

Updated: January 15, 2006

Payments System Research Department
Federal Reserve Bank of Kansas City
Working Paper 04-01

Abstract

An industry typically experiences initial mass entry and later shakeout of producers over its life cycle. It can be explained as a competitive equilibrium outcome driven by dynamic interactions between technology progress and demand diffusion. After a new product is introduced, technology improves with cumulative production and *S*-shaped diffusion is generated as the product penetrates a positively skewed income distribution. Eventually fewer new adopters are available and the net number of producers starts to decline. It is shown that faster technological learning, higher mean income or larger market size contributes to faster demand diffusion and earlier industry shakeout. Comparative studies on the US and UK television industries and evidence from ten other US industries support the theoretical findings.

Keywords: Learning by Doing, Demand Diffusion, Industry Shakeout
JEL Classification: D30, O30, L10

---

* Mailing Address: Federal Reserve Bank of Kansas City, 925 Grand Blvd., Kansas City, MO 64198. Email: zhu.wang@kc.frb.org. I am indebted to my thesis advisors Boyan Jovanovic (Chair), Gary S. Becker, Robert E. Lucas, Jr., and Chad Syverson for invaluable guidance and support. I thank Fernando Alvarez, Susanto Basu, Jeremy Fox, Lars Hansen, Narayana Kocherlakota, Hanno Lustig, Kiminory Matsuyama, Casey Mulligan, Nancy Stokey, John Van Reenen and in particular Steven Klepper for helpful comments. Feedback from seminar participants at Chicago, Rochester, USC, KC Fed, Emory, Arkansas, NBER Summer Institute Productivity and Macroeconomics Meeting, and financial support from John M. Olin Foundation Dissertation Fellowship are greatly appreciated.

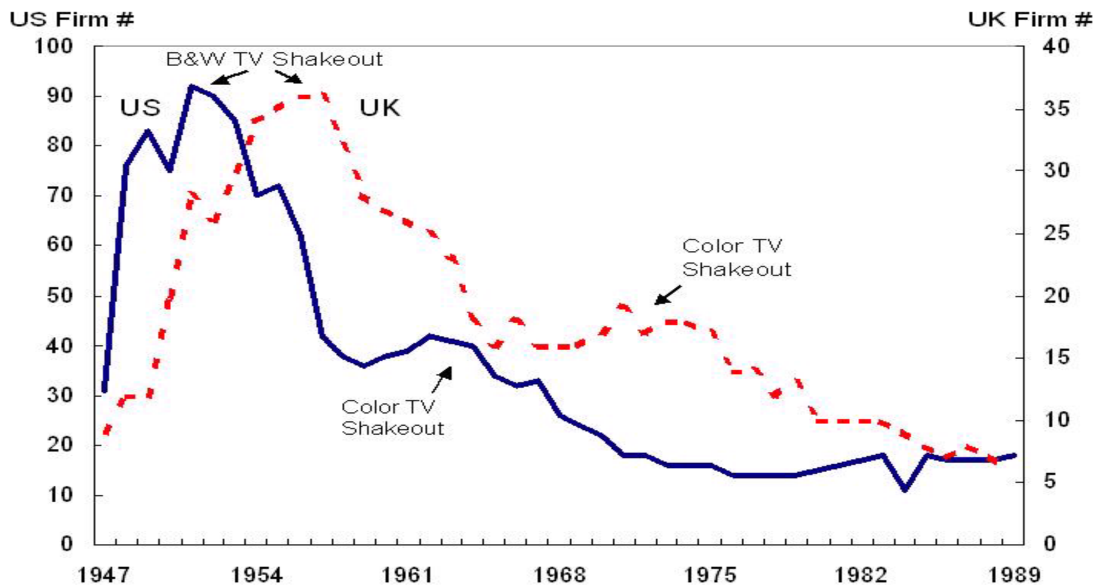

Figure 1: TV Industry Shakeout: US vs. UK

1 Introduction

1.1 Questions on Industry Life Cycle

As a new industry evolves from birth to maturity it is typically observed that price falls, output rises, and the net number of firms initially rises and later falls (Gort and Klepper 1982, Klepper and Graddy 1990). In particular, the nonmonotonic time path of firm numbers, termed as "shakeout", has been the focus of many recent studies of industry economics. The big question is why there is a shakeout and when it occurs.

To answer that question, most existing theories emphasize supply-side factors, particularly the inter-firm differences in technology. It is shown that shakeout can be triggered by "emergence of dominant design" (Utterback and Suárez 1993), "race of innovation" (Jovanovic and MacDonald 1994, Wang 2005), or "scale economy of R&D" (Klepper 1996, Klepper and Simons 2000). Some other explanations appeal to uncertainties in new product markets. For example, "uncertain profit" (Horvath, Schivardi and Woywode 2001) or “uncertain market size” (Rob 1991, Zeira 1987, 1999) can also result in a mass entry and later shakeout.

Though these theories have contributed to a much improved understanding of industry shakeout, some important issues are still underexplored. In particular, the impacts of demand characteristics on industry life cycle dynamics are largely overlooked.[1] As a result, it remains difficult to explain certain empirical facts. For example, Figure 1 plots the firm numbers in the US and UK television industries.[2] The television was commercially introduced into the US and the UK at the same time after the WWII, and the two markets were segmented for the following two decades because of their different technical standard.[3] This natural experiment shows that the patterns of industry evolution were very similar across countries, but the mass entry and shakeout of TV producers were uniformly lagged behind in the UK.

What can explain this cross-country difference of timing of shakeout? The existing theories may apply to some extent, but not quite enough. On one hand, industry studies (e.g. Arnold 1985, LaFrance 1985, Klepper and Simons 1996) document numerous technological changes in the TV industry that may have a major cumulative effect on inter-firm heterogeneity. However, they do not directly explain the timing of shakeout, let alone the cross-country difference. On the other hand, the market uncertainty could not have caused the shakeout repeatedly since at least the UK producers could easily learn from the US market experience. Moreover, the TV shakeouts, especially the Black & White TV shakeouts, had little to do with foreign competition since the import and export were very insignificant at that time.[4]

[1] An exception is Hopenhyan (1993), where some demand issues are briefly discussed.

[2] Data source: Simons (2002), *Television Factbook* (various issues).

[3] The UK adopted the 405-line screen standard in 1943, but other nations proceeded to adopt standard with higher resolutions. The UK standard remain anomalous through 1964, when some UK broadcasts began using the internationally common PAL 625-line color standard. Hence through 1964 and even later, UK market was isolated from the foreign competition. See Levy (1981).

[4] In 1950s, US and UK were the two largest TV producers in the world, but imports and exports were nil for both nations. Imports started to increase in 1960s as Japan took off, but did not

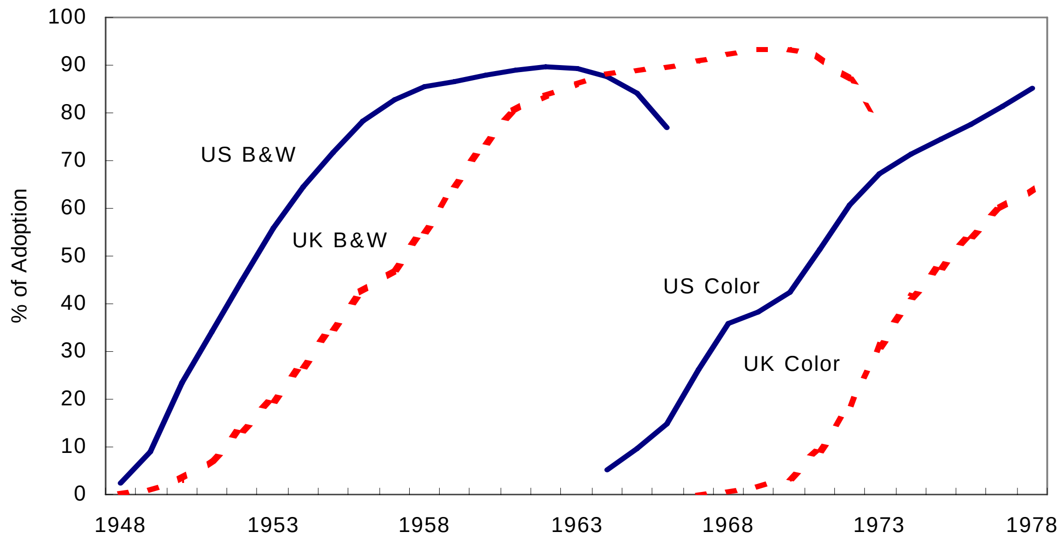

Figure 2: Household Adoption of TV: US vs. UK

1.2 New Hypothesis

One potential problem is that the aforementioned theories may have overlooked the importance of demand-side issues. A new product, over its life cycle, typically experiences strong demand growth at the early stage but the growth eventually diminishes as the market reaches maturity. As a result, the demand characteristics can have influential, sometimes critical, effects on the evolution of industry. Without taking that into account, the analyses on industry life cycle would be incomplete.

In this paper, we consider explicitly the roles that demand plays in driving the industry life cycle, and reveal how country-specific demand factors, the income distribution and market size in particular, shape this process. When a new product is introduced, high-income consumers tend to adopt it first. The technology then improves with cumulative production (Learning by Doing) and S-shaped diffusion is

reach 10% of domestic production until 1965 in US, and until 1970 in UK. Data sources: *Television Factbook* (US), *Monthly Digest of Statistics* (UK).

generated as the product penetrates a positively skewed income distribution (Trickle Down Effect and Income Growth Effect). Eventually fewer new adopters are available and the number of firms starts to decline. It is shown that faster technological learning, higher mean income or larger market size contributes to faster demand diffusion and earlier industry shakeout. This new theory therefore offers additional demand-side explanations for the varying pattern of industry evolution across products and countries. For example, it suggests it was the lower per capita income and smaller market size in the UK that led to a slower diffusion of TV (Figure 2)[5] and hence a lagged industry shakeout.

1.3 Relation to Other Research

The main purpose of this paper is to complement existing theories of industry life cycle by exploring the previously largely unexplored demand side. In doing that, it also links the studies of industry economics with research in several separate fields. For example, in the marketing literature, the diffusion of new products has been studied for the purpose of demand forecast and monopoly pricing (Bass 1969, Kalish 1983, Horsky 1990). In the growth literature, the learning by doing is one of the most important sources of technology progress (Arrow 1962, Lucas 1993, Jovanovic & Rousseau 2002, Matsuyama 2003). In the international trade literature, the most celebrated "product cycle theory" claims that the demand for new consumer goods is initially greatest in high-income countries, and diffuse to lower-income countries later on (Vernon 1966, Stokey 1991 and Grossman & Helpman 1991). In the technology adoption literature, it has been found that per capita GDP is one of the key variable that explains the diffusion rate of new technology across nations (Comin & Hobijn 2004). Those studies have so far been unconnected with the study of industry evolution. This paper is a first step to fill this gap, and show that the typical pattern of industrial evolution is closely related and consistent with the findings in those fields.

[5] Data source: *Television Factbook* (various issues), Bowden & Offer (1994).

1.4 Road Map

The paper is organized as follows. Section 2 studies the supply structure of a competitive industry and reveals the comovement between the firm numbers and the relative industry GDP. Section 3 endogenizes the logistic diffusion curves by modeling explicitly consumers' heterogeneity of income and preference. Section 4 puts together the supply and demand, and introduces the law of motion for technology and income. Section 5 characterizes the industry dynamics and discusses the time paths of key industry variables. Section 6 extends the model to durable goods. Section 7 estimates our model using data of the US and UK TV industries as well as ten other US industries. Section 8 contains concluding remarks.

2 The Supply Structure

In this section, we model the supply structure of a competitive industry and reveal some important relationship among key industry variables, namely the comovement between the firm numbers and the relative industry GDP.

2.1 The Model

Assume a competitive industry produces one homogenous product. There are M potential producers that differ in their production efficiency $\theta \in (0, \infty)$ for participating in this industry. The efficiency θ is distributed with cdf function $S(\theta)$. Each period, a firm that actively produces in this industry incurs an opportunity cost C, which corresponds to the foregone earnings of human capital needed to run the firm.[6] For

[6]Our model of firm is a simplified treatment as in Lucas (1978), but can be interpreted broadly. For example, the parameter θ may include any firm-specific factors that affect production efficiency, e.g. management ability, physical location, industry experience and etc. Also, C captures the foregone earnings of firm human capital that are compensated from the production residual, e.g. the management group and R&D team. For simplicity, C is assumed identical across firms. However,

a typical firm, x and y respectively denote the input and output. The production function is assumed to be $y = \theta A x^{\alpha}$ where A is the technology and $0 < \alpha < 1$ is the "span of control" parameter.[7] Let P denote the price of output, w the price of input. We also assume that firms can enter and exit freely.[8]

An individual firm, indexed by its efficiency θ, has zero measure. Each period, firm θ takes the market price P as given to maximize the profit:

$$\pi_\theta = \max_{y_\theta, x_\theta} \; P y_\theta - w x_\theta - C \qquad s.t. \quad y_\theta = \theta A x_\theta^{\alpha}$$

and gets the following solution:

$$\begin{aligned} y_\theta^* &= \left(\frac{\alpha P A^{\frac{1}{\alpha}} \theta^{\frac{1}{\alpha}}}{w}\right)^{\frac{\alpha}{1-\alpha}}; \qquad x_\theta^* = \left(\frac{\alpha P A \theta}{w}\right)^{\frac{1}{1-\alpha}}; \\ \pi_\theta &= (1-\alpha)(\alpha)^{\frac{\alpha}{1-\alpha}} w^{\frac{-\alpha}{1-\alpha}} (P A \theta)^{\frac{1}{1-\alpha}} - C. \end{aligned}$$

The free entry and exit condition ensures that the marginal firm $\tilde{\theta}$, the lowest-efficiency player allowed in the industry, breaks even. The market price is then determined as

$$\pi_{\tilde{\theta}} = 0 \Longrightarrow P = \frac{C^{1-\alpha} w^{\alpha}}{(1-\alpha)^{1-\alpha} (\alpha)^{\alpha} A \tilde{\theta}} \tag{1}$$

and we can solve more explicitly each firm's choice:

$$y_\theta^* = \left(\frac{\alpha C}{(1-\alpha) w}\right)^{\alpha} A \theta^{\frac{1}{1-\alpha}} \tilde{\theta}^{\frac{\alpha}{\alpha-1}}; \qquad x_\theta^* = \frac{\alpha C}{(1-\alpha) w} \theta^{\frac{1}{1-\alpha}} \tilde{\theta}^{\frac{1}{\alpha-1}}; \tag{2}$$

$$\pi_\theta = [(\theta/\tilde{\theta})^{\frac{1}{1-\alpha}} - 1] C. \tag{3}$$

The total market supply Y is the sum of incumbent firms' outputs:

$$Y = M \int_{\tilde{\theta}}^{\infty} y_\theta^* dS(\theta) = \left(\frac{\alpha C}{(1-\alpha) w}\right)^{\alpha} \tilde{\theta}^{\frac{\alpha}{\alpha-1}} A M \int_{\tilde{\theta}}^{\infty} (\theta^{\frac{1}{1-\alpha}}) dS(\theta). \tag{4}$$

this assumption is not at all essential. Allowing C to be heterogenous, e.g. $C \in [\underline{C}, \overline{C}]$ does not change the analysis. A detailed proof is available from the author.

[7] Assuming $0 < \alpha < 1$ implies that firms face diseconomy of scale in the short run when technology and market condition are given. See Lucas (1978).

[8] This is also a simplifying assumption but not essential. Our main conclusions remain unchanged if the entry incurs a sunk cost. The proof is available from the author.

The corresponding number of firms is

$$N = M \int_{\tilde{\theta}}^{\infty} dS(\theta). \tag{5}$$

At equilibrium, the market supply equals the market demand. Equation 1 and 4 then imply the relative industry GDP, defined as PY/C, to be

$$\frac{PY}{C} = \frac{1}{1-\alpha} \tilde{\theta}^{\frac{1}{\alpha-1}} M \int_{\tilde{\theta}}^{\infty} (\theta^{\frac{1}{1-\alpha}}) dS(\theta). \tag{6}$$

If C, the foregone earning of firm human capital, grows with the mean income μ of the economy (e.g. $C = \phi\mu$) and the parameters α and $S(\theta)$ are time-invariant, there exists an important time-series relationship between the firm numbers N and the relative industry GDP, PY/μ.

Proposition 1 *In a competitive market, the number of firms is positively related with the relative industry GDP, i.e.* $\partial N/\partial(PY/\mu) > 0$.

Proof. It follows Equation 5 and 6. ■

2.2 Examples and Remarks

There should be no surprise to find out a comovement between the number of firms N and the relative industry GDP, PY/μ. In fact, the profit of the marginal firm increases with PY (the total industry GDP) and the opportunity cost of participation is determined by μ (the mean income of the economy). It follows that the ratio PY/μ keeps track with the viable number of firms.[9]

This result is fairly robust for a homogeneous-product industry under free entry equilibrium. Supporting empirical evidence can be found from many industries including TV (see Figure 3, 4 and section 7.3).[10] To derive this result, the production

[9] Under some addtional assumptions, such as θ follows a Pareto distribution or a degenerate distribution, we can further show that the number of firms is proportional to the relative industry GDP, i.e. $N \propto PY/\mu$. However, those assumptions are not necessary for the purpose of this paper.

[10] Data source: Gort and Klepper (1982).

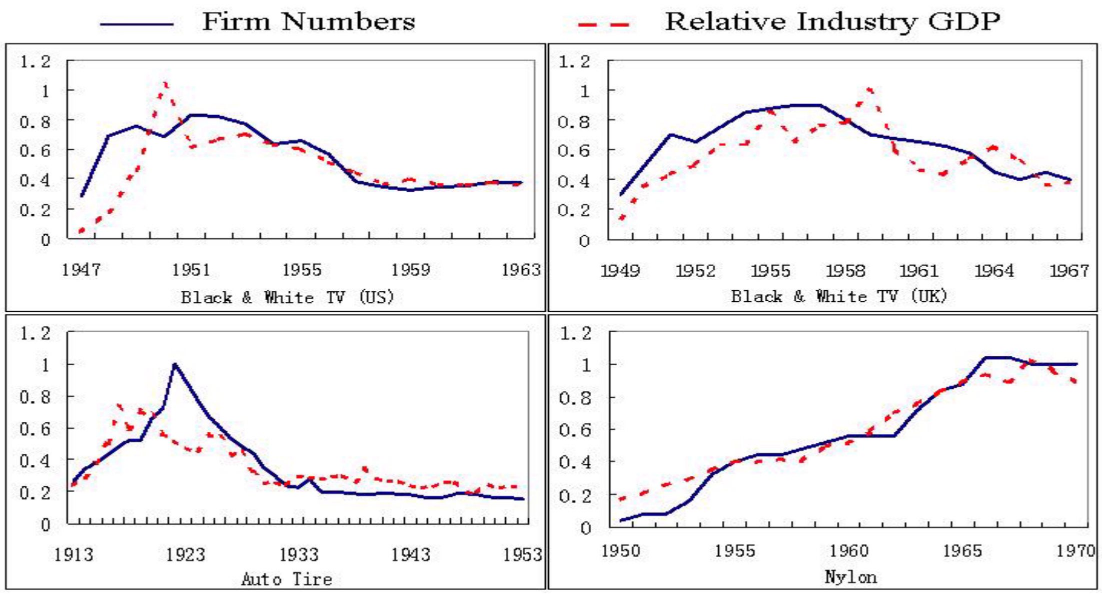

Figure 3: Firm Numbers and Relative Industry GDP: Evidence I

function that we assumed is not crucial but will help make our following dynamic analysis more tractable. However, assuming α and $S(\theta)$ to be time-invariant is necessary.[11] In fact, by not allowing α and $S(\theta)$ to change, we constrain the role that the inter-firm heterogeneity can play in the industry dynamics, which is a major distinction between our theory and the existing supply-side literature. In Appendix A, we relax the assumptions to consider more supply-side effects, and show the essence of our analysis still carry over.

3 The Demand Structure

Now let us turn to the demand side. To explain the industry evolution, it is very crucial to understand the dynamics of new product demand. In this section, we propose a novel analysis on that.

[11] Throughout the paper, we do not require θ to be fixed over time for each individual firm, but rather the overall distribution $S(\theta)$ is time-invariant. See section 5.3. for more discussions.

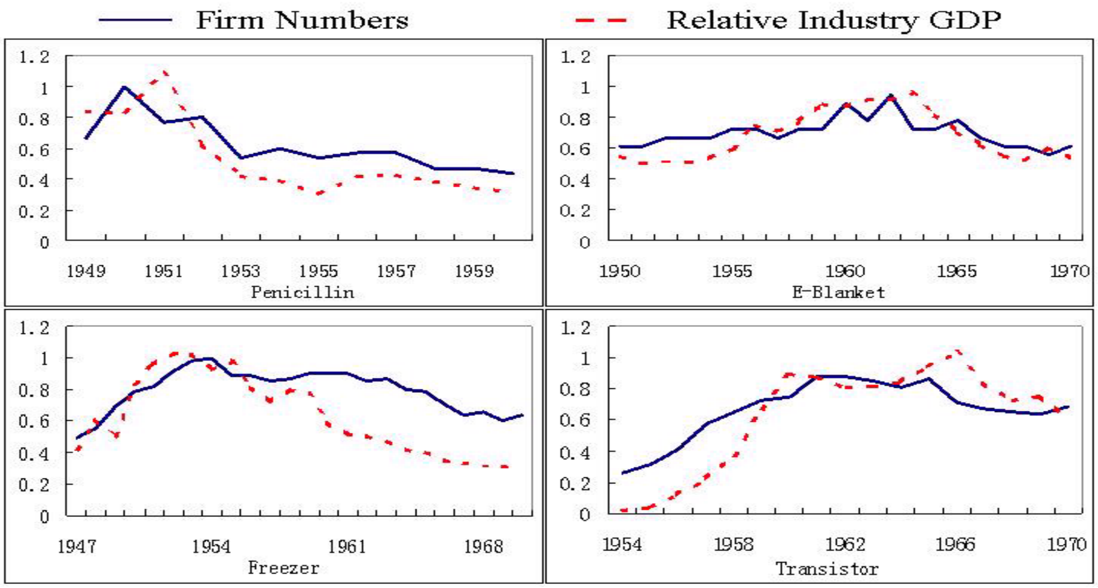

Figure 4: Firm Numbers and Relative Industry GDP: Evidence II

3.1 Questions on Traditional Wisdom

It has long been a challenge to explain the demand diffusion of new products. In the economics and marketing literature, the most popular theory relies on the behavioral assumption of social contagion, i.e. consumers imitate early adopters. This explanation has been formalized by introducing the logistic model and its variants since 1950s (Griliches 1957, Mansfield 1961 and Bass 1969).

The logistic model assumes that the hazard rate of adoption rises with cumulative adoption.

$$\frac{\dot{F}_t}{1-F_t} = vF_t \Longrightarrow F_t = \frac{1}{[1+(\frac{1}{F_0}-1)e^{-vt}]} \tag{7}$$

where F_t is the fraction of consumers who have adopted the product at time t, and v is a constant contagion parameter.

Though the logistic model has traditionally fit data very well, some important issues remain unclear. In particular, assuming homogenous adopters and exogenous diffusion parameters does not explain why diffusion rates are historically so different

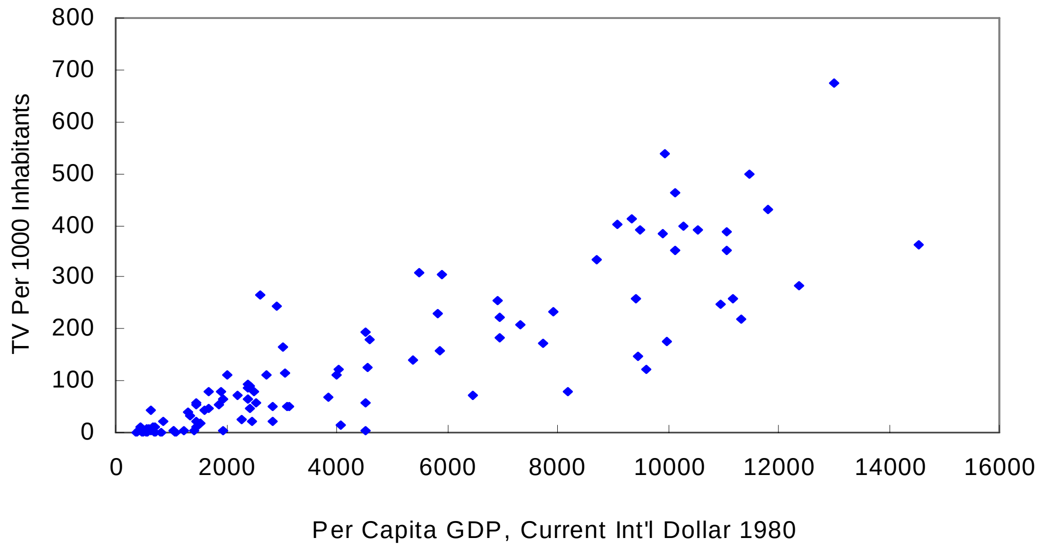

Figure 5: Per Capita GDP and TV Adoption in 104 Countries, 1980

across countries (regions), consumer groups and products. To compromise those issues, many following studies added rather ad hoc assumptions to the contagion framework, e.g. assuming the diffusion parameters to be function of traits like region, consumer type and product.

However, the key question is – how much does the contagion effect really matter? Figure 5 plots the TV adoption rate verse per capita GDP for 104 countries in 1980.[12] At that time, TV was no longer a new product so that there really leaves little room for contagious spread of information to explain the diffusion.[13] However, it is evident that adoption is so strongly related with income across countries.

[12]Data source: *UN Common Database.*

[13]Sometimes, the social contagion model is interpreted more broadly to include network effects of consumption. However, as Bowden & Offer (1994) shows, the US enjoyed a uniformly faster adoption than the UK in consumer appliances including TV, Cloth Washer, Cloth Dryer, Dishwasher, Electric Blanket, Freezer, Radio, Refrigerator, Vacuum Cleaner and etc. Many of them certainly have litter, if any, network effects of consumption among consumers.

3.2 Alternative Approach

This paper is hence motivated to take a different approach. Without assuming social contagion framework, we derive the logistic diffusion curve as a demand function generated from heterogeneity of consumers.[14] As a result, it becomes more clear how the diffusion process is shaped by economic forces like price and income.

The model is as follows. Assume a new nondurable product sells for price P in the market.[15] An individual consumer adopts it only if her disposable income I_d on that product allows her to do so, i.e.

$$y = \begin{cases} 1 & \text{if} \quad I_d \geqslant P, \\ 0 & \text{if} \quad I_d < P. \end{cases}$$

Consumers are heterogenous in their disposable income I_d on this new product. This heterogeneity comes from their different income and preference. We assume for an individual consumer the disposable income I_d is the product of her total income I ($I \geq 0$) and propensity of spending c ($1 \geq c \geq 0$), i.e. $I_d = cI$. I and c are independently distributed over the population. A direct implication of this assumption is that new products diffuse faster in higher-income groups.[16]Figure 6 presents supporting evidence from the Black & White TV in the US and the UK.[17]

Notice that the disposable income I_d is generally not directly observable, but it can be nonetheless inferred from the observables. In particular, the way we model it suggests that with a given distribution of preference c, a higher mean (inequality) of total income I implies a higher mean (inequality) of disposable income I_d.

[14] This is not a totally new idea (e.g. Cramer (1969) and Bonus (1973) derive normal diffusion from individual consumers' Quasi-Engel curves), but the way this paper derives logistic diffusion and connects it to demand models is new.

[15] In chapter 6 and 7 we will see that introducing durability does not change the analysis.

[16] Denote $H(c)$ the cdf function of c. For a given income group the fraction of adopter is $\Pr(I_d \geqslant P \mid I) = 1 - H(P/I)$ and it rises with I.

[17] Data source: US from Bogart (1972), UK from Emmett (1956).

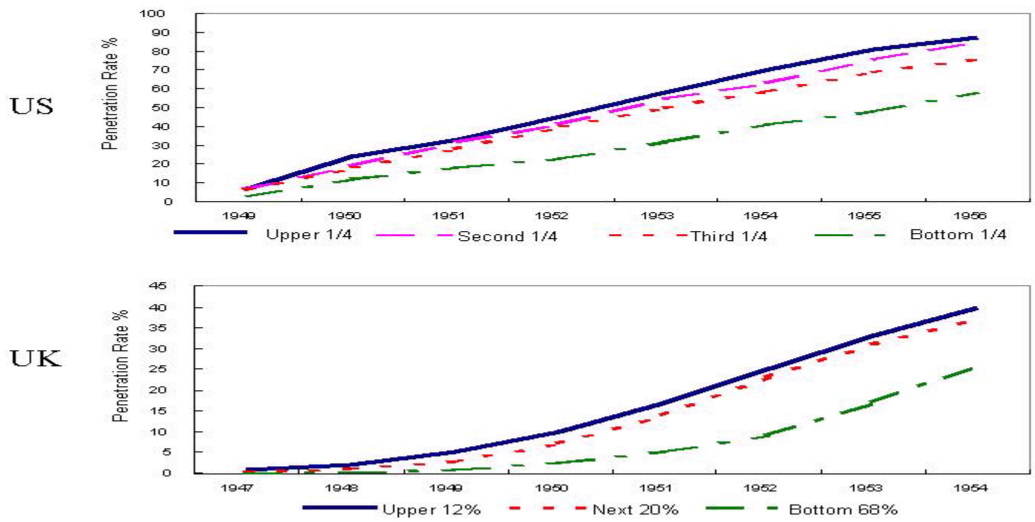

Figure 6: TV Penetration Rates by Income Class: US and UK

3.3 An Explicit Formulation: Log-logistic Distribution

To take our analysis a step further, we have to model the disposable income more explicitly. By the way that it is constructed, we know that I_d is distributed over the positive domain $[0, \infty)$, so its distribution tends to be positively skewed. The possible candidates for this group of distribution are far from unique, so we choose to pick a reasonable one.

In the following discussion, we introduce the log-logistic distribution as our specific example.[18] The reason for us to pick the log-logistic distribution is not only because it serves as an easily tractable example, but also because it connects our study to the typically observed logistic diffusion curves as we will show next.

The log-logistic distribution is defined as the distribution of a variate whose logarithm is logistically distributed. Assuming that the disposable income I_d follows the

[18] The application of log-logistic distribution in economics has a long history, traced back to the study of Lomax (1954) on business failure rates, and to Fisk (1961) on the size distribution of income. Figure 7 gives an example of using the Fisk (log-logistic) distribution to fit US family income in year 1970 (Data source: *Statistical Abstract of the United States*).

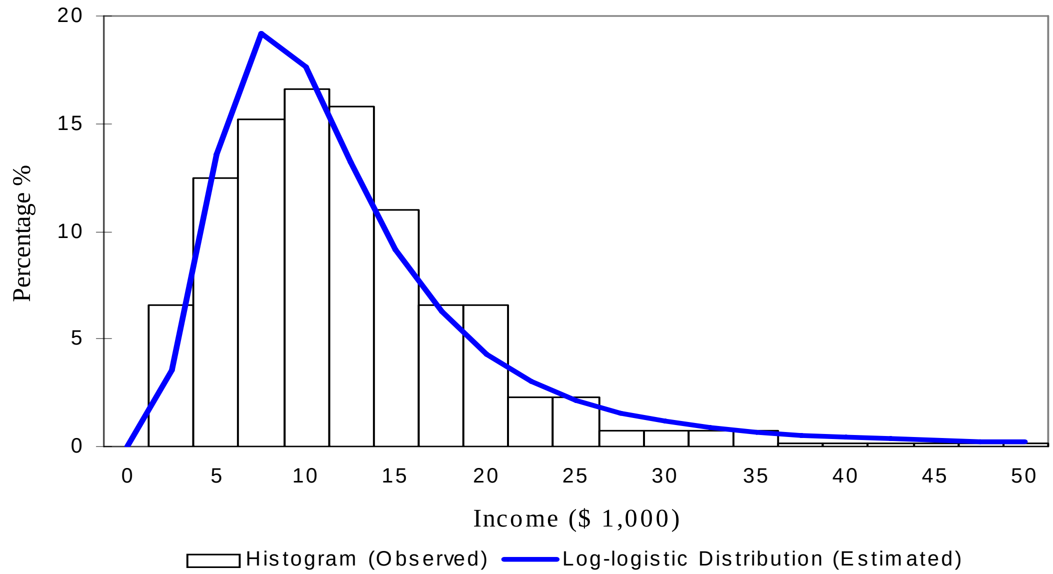

Figure 7: US Family Income Distribution 1970

log-logistic distribution, the cdf function is given as

$$G_{I_d}(x) = 1 - \frac{1}{1 + a_1 x^{a_2}}$$

with the mean $E(I_d)$ and Gini coefficient $g(I_d)$ given as

$$E(I_d) = a_1^{-1/a_2}\Gamma(1 + \frac{1}{a_2})\Gamma(1 - \frac{1}{a_2}); \qquad g(I_d) = \frac{1}{a_2}$$

where Γ denotes the gamma function $\Gamma(\mu) \equiv \int_0^\infty i^{\mu-1}\exp(-i)di$.

Hence, we may rewrite the cdf function into a more meaningful form:

$$G_{I_d}(x) = 1 - \frac{1}{1 + (\frac{\Gamma(1+g)\Gamma(1-g)}{E(I_d)}x)^{1/g}}$$

where $g = g(I_d)$.

Recall $E(I_d) = E(c)\mu$. We can now derive the adoption rate F to be a function of price, mean income and other parameters:

$$F = 1 - G_{I_d}(P) = \frac{1}{1 + \eta(P/\mu)^{1/g}} \tag{8}$$

where $\eta = (\Gamma(1+g)\Gamma(1-g)/E(c))^{1/g}$.

3.4 Endogenous Diffusion vs. Exogenous Diffusion

The appealing feature of introducing log-logistic distribution is to endogenize the logistic diffusion curves. To see that, let us assume that the price declines at a constant rate $P_t = P_0 e^{-\rho t}$, and mean income grows at a constant rate $\mu_t = \mu_0 e^{zt}$. Then we can rewrite Equation 8 as follows

$$F_t = \frac{1}{1 + \eta[P_0/\mu_0]^{1/g} e^{-(\rho+z)t/g}}. \tag{9}$$

Comparing Equation 9 with Equation 7, we realize that our formula is equivalent to the logistic model under very reasonable assumptions. In particular, the diffusion parameters traditionally treated exogenous now have clear economic meanings – the contagion parameter v is determined by the growth rates of price and income, and the initial condition F_0 is the fraction of adopters who can afford the new product at the initial price and mean income:

$$v = (\rho + z)/g; \qquad F_0 = \frac{1}{1 + \eta[P_0/\mu_0]^{1/g}}.$$

This result is also empirically plausible. Sultan et al. (1990) analyzed the parameter estimates of 213 published application of the logistic model and its extension. They report the average value of $v = 0.38$. Jeuland (1993, 1994) finds that the value of v is rarely greater than 0.5 and rarely less than 0.3. Assigning reasonable values of Gini coefficient and growth rates of income and price, our model can easily generate the value of v within that range.

4 The Industry Equilibrium

4.1 The Momentary Equilibrium

Combining the supply and demand analyses, we are now ready to derive the industry equilibrium. At a point of time, the industry equilibrium implies (1) individual firms take the market price as given and maximize profit; (2) individual consumers take the market price as given and maximize utility; (3) industry price, output and firm numbers are uniquely determined to clear the market. All the findings are summarized as follows:

$$P = \frac{C^{1-\alpha}w^{\alpha}}{(1-\alpha)^{1-\alpha}(\alpha)^{\alpha}A\tilde{\theta}}; \tag{10}$$

$$Y = (\frac{\alpha C}{(1-\alpha)w})^{\alpha}\tilde{\theta}^{\frac{\alpha}{\alpha-1}}AM\int_{\tilde{\theta}}^{\infty}(\theta^{\frac{1}{1-\alpha}})dS(\theta); \tag{11}$$

$$Y = mF = \frac{m}{1+\eta(P/\mu)^{1/g}}; \tag{12}$$

$$N = M\int_{\tilde{\theta}}^{\infty}dS(\theta). \tag{13}$$

Notice that m denotes the total number of potential consumers. It is reasonable to assume that the normalized ratio M/m is a constant, and does not vary with the population size.

This is a system of four equations with four unknowns. The solution suggests that the equilibrium values of P, Y, $\tilde{\theta}$ and N are endogenously determined by four important parameters: technology A, mean income μ, foregone earning C and input price w. With reasonable assumptions on the law of motion for those parameters, we will then be able to characterize the time path of industrial evolution.

4.2 Law of Motion Equations

4.2.1 Learning-by-Doing Technology Progress

Technology progress is commonly observed over the industry life cycle. Many theoretical and empirical studies (e.g. Arrow 1962, Boston Consulting Group 1972) have identified learning by doing as one of the most important sources. Therefore, we assume that the technology A is determined by the cumulative industry output:

$$A_t = A_0(Q_t)^\gamma \tag{14}$$

in which $Q_t = \int_0^t Y(s)ds + Q_0$ and γ is the learning rate.

Equation 14 implies that only aggregate cumulative output matters to every producer's productivity. In fact, a firm's own contribution to Q may matter more, especially at high frequencies (Irwin & Klenow 1994, Thompson & Thornto 2001). However, at lower frequencies, the distinction between own and outside experience should fade given a wide range of channels by which information diffusion can occur.[19]

4.2.2 Income Growth and Its Effects

As an economywide variable, it is reasonable to assume the mean income μ grows at an exogenous rate z. The foregone earning C of firm human capital, as we discussed in section 2.1, may grow with the mean income μ:

$$\begin{aligned} \mu_t &= \mu_0 e^{zt} \qquad \text{with} \quad \mu_0 > 0, \quad z > 0; \\ C_t &= \phi \mu_t \qquad \text{with} \quad \phi > 0. \end{aligned}$$

The law of motion for the input price w is a little complicated. Since we assume there is only one input in our model, w is actually a composite price index for both labor and non-labor inputs. Though the price of labor inputs may grow with the mean income, the price of non-labor inputs such as capital and materials does not. Therefore, it is reasonable to assume that $d(w/\mu)/d\mu < 0$. A simple formulation is

$$w_t = \sigma \mu_t^\psi \qquad \text{with} \quad \sigma > 0, \quad \psi < 1.$$

4.3 The Dynamic Equilibrium

Given the law of motion equations, it becomes clear that there are two driving forces of industry dynamics – technological learning (due to cumulative production) and

[19] Lieberman (1987) lists many of these channels: employees may be hired away; products can be reverse-engineered; patents can be invented around or infringed and etc.

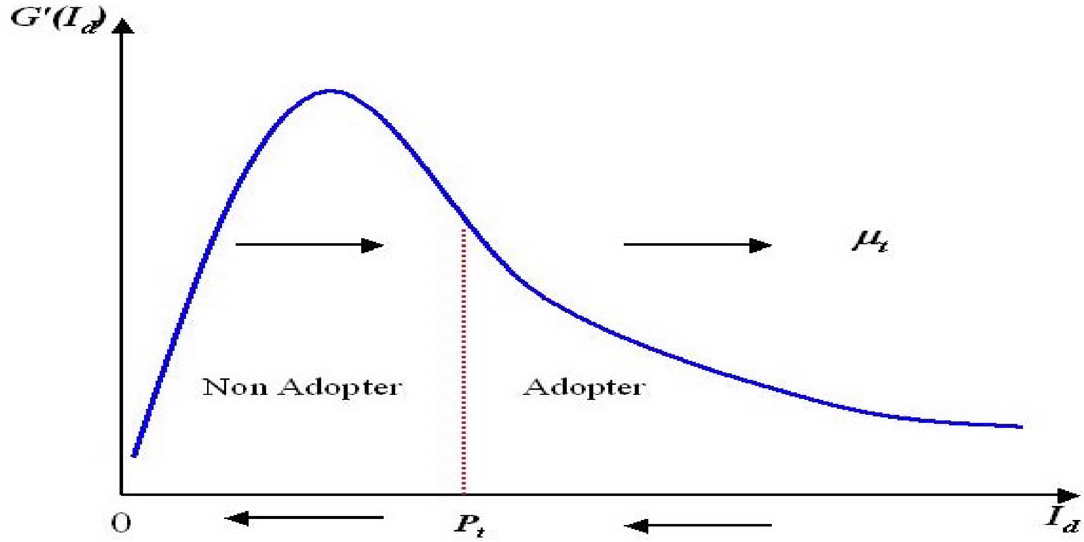

Figure 8: Product Diffuses as Technology and Income Change

income growth. As the initial adoption of a new product starts, these two forces interact to generate further technology progress and demand diffusion, and keep this process going (See Figure 8).

The dynamic industry equilibrium can be summarized as follows:

$$\frac{P_t}{\mu_t} = \frac{(\phi)^{1-\alpha}(\sigma\mu_t^{\psi-1})^{\alpha}}{(1-\alpha)^{1-\alpha}(\alpha)^{\alpha}A_t\tilde{\theta}_t}; \tag{15}$$

$$Y_t = \left(\frac{\alpha\phi\mu_t^{1-\psi}}{(1-\alpha)\sigma}\right)^{\alpha}\tilde{\theta}_t^{\frac{\alpha}{\alpha-1}}A_t M\int_{\tilde{\theta}_t}^{\infty}(\theta^{\frac{1}{1-\alpha}})dS(\theta); \tag{16}$$

$$Y_t = mF_t = \frac{m}{1+\eta(P_t/\mu_t)^{1/g}}; \tag{17}$$

$$N_t = M\int_{\tilde{\theta}_t}^{\infty} dS(\theta); \tag{18}$$

$$A_t = A_0(Q_t)^{\gamma} \quad \text{where} \quad Q_t = \int_0^t Y(s)ds + Q_0; \tag{19}$$

$$\mu_t = \mu_0 e^{zt} \tag{20}$$

from which we are ready to characterize the time path of industry evolution.

5 The Industry Evolution

5.1 Industry Dynamics: Characterization

With the assumption of learning by doing, the market demand equation 17 implies a first-order differential equation

$$\dot{Q}_t = f(Q_t, t) = \frac{m}{1 + \eta(\frac{P}{\mu}(Q_t, t))^{1/g}} \tag{21}$$

where the relatively price P_t/μ_t is a function of (Q_t, t), and the function is determined by the equilibrium conditions 15 - 17 as follows

$$\frac{m}{1 + \eta(P_t/\mu_t)^{1/g}} = (\frac{\alpha\mu_t^{1-\psi}}{\sigma})^{\frac{\alpha}{1-\alpha}} (P_t/\mu_t)^{\frac{\alpha}{1-\alpha}} A_t^{\frac{1}{1-\alpha}} M \int_{\frac{(\phi)^{1-\alpha}(\sigma\mu_t^{\psi-1})^{\alpha}}{(1-\alpha)^{1-\alpha}(\alpha)^{\alpha} A_t P_t/\mu_t}}^{\infty} (\theta^{\frac{1}{1-\alpha}}) dS(\theta), \tag{22}$$

$$\text{where} \qquad A_t = A_0(Q_t)^{\gamma}; \qquad \mu_t = \mu_0 e^{zt}.$$

Characterizing the solution to the differential equation, we have following findings:

Theorem 1 (Unique Solution): *There exists a unique dynamic path $Q(t)$ of $\dot{Q}_t = f(Q_t, t)$ for $t \geq 0$ which satisfies $Q(0) = Q_0$.*

Proof. Given f is continuously differentiable, it satisfies the Lipschitz condition:

$$| f(x,t) - f(y,t) | \leq L | x - y | \quad where \quad L = \sup | \partial f/\partial Q | .$$

Theorem 1 then follows Theorem 5 of p23 in Birkhoff & Rota (1968). ■

Theorem 2 (Unique Shakeout): *Given a log-logistic distribution of disposable income, there exists a unique shakeout.*

Proof. With a log-logistic distribution, the relative industry GDP is given by

$$\frac{P_t Y_t}{\mu_t} = (P_t/\mu_t)\frac{m}{1+\eta(P_t/\mu_t)^{1/g}}.$$

Therefore

$$\partial N_t/\partial t = \frac{\partial N_t}{\partial(\frac{P_t Y_t}{\mu_t})}\frac{\partial(\frac{P_t Y_t}{\mu_t})}{\partial(P_t/\mu_t)}\frac{\partial(P_t/\mu_t)}{\partial t}.$$

As shown in Proposition 1, $\partial N_t/\partial(\frac{P_t Y_t}{\mu_t}) > 0$. Furthermore, Equation 22 implies that P_t/μ_t decreases with A_t and μ_t. Since $A_t = A_0(Q_t)^\gamma$ and $\mu_t = \mu_0 e^{zt}$ are strictly increasing with time t, we have $\partial(P_t/\mu_t)/\partial t < 0$. In addition, we have

$$\frac{\partial(\frac{P_t Y_t}{\mu_t})}{\partial(P_t/\mu_t)} \lesseqqgtr 0 \quad \text{for} \quad P_t/\mu_t \gtreqqless [\frac{g}{\eta(1-g)}]^g.$$

Hence, the unique shakeout occurs at $(P_t/\mu_t)^* = [\frac{g}{\eta(1-g)}]^g$ and the corresponding adoption rate is $F^* = 1-g$. If $P_0/\mu_0 > [\frac{g}{\eta(1-g)}]^g$, the firm numbers initially rise and later fall. If $P_0/\mu_0 < [\frac{g}{\eta(1-g)}]^g$, the firm numbers decline from the beginning. ■

Theorem 3 (Comparison Theorem): *Anything else being equal, the relative price P_t/μ_t falls more quickly, the diffusion F_t proceeds faster and the timing of shakeout t^* comes earlier if (1) technology is better (higher Q_0, higher A_0 or higher γ); (2) mean income is higher (higher μ_0 or higher z); (3) market size is larger (higher m); (4) input price is lower (lower ϕ, lower σ or lower ψ).*

Proof. Let us take γ as an example, and the other proofs are similar. The proof takes the following three steps:

(1) Equation 21 and 22 define $\dot{Q}_t = f(Q_t, t)$, where f satisfies the Lipschitz condition. Since $\partial f/\partial\gamma > 0$ for any given (Q_t, t), a higher γ leads to a higher Q_t (hence higher A_t) at any time t. This result follows Theorem 8 and Corollary 2 of p25-26 in Birkhoff & Rota (1968).

(2) With $\partial A_t/\partial\gamma > 0$ at any time t, Equation 22 and 17 imply that a higher γ leads to a lower relative price and faster diffusion:

$$\frac{\partial(P_t/\mu_t)}{\partial\gamma} = \frac{\partial(P_t/\mu_t)}{\partial A_t}\frac{\partial A_t}{\partial\gamma} < 0; \qquad \frac{\partial F_t}{\partial\gamma} = \frac{\partial F_t}{\partial(P_t/\mu_t)}\frac{\partial(P_t/\mu_t)}{\partial\gamma} > 0.$$

(3) At the time of shakeout t^*, $\partial(P_tY_t/\mu_t)/\partial t = 0$ and $\partial^2(P_tY_t/\mu_t)/\partial t^2 < 0$,

$$\frac{\partial(P_tY_t/\mu_t)}{\partial t} = 0 \Longrightarrow \frac{\partial(P_t/\mu_t)}{\partial t}\frac{m}{1+\eta(P_t/\mu_t)^{1/g}}\{1 - \frac{\eta/g}{\eta+[P_t/\mu_t]^{-1/g}}\} = 0.$$

Since we have proved in Theorem 2 that $\frac{\partial(P_t/\mu_t)}{\partial t} < 0$, we have

$$J = \frac{m}{1+\eta(P_t/\mu_t)^{1/g}}\{1 - \frac{\eta/g}{\eta+[P_t/\mu_t]^{-1/g}}\} = 0 \quad \text{and} \quad \frac{\partial J}{\partial(P_t/\mu_t)} < 0.$$

Therefore,

$$\frac{\partial t^*}{\partial\gamma} = -\frac{\partial^2(P_tY_t/\mu_t)/\partial t\partial\gamma}{\partial^2(P_tY_t/\mu_t)/\partial t^2} = -\frac{\frac{\partial(P_t/\mu_t)}{\partial t}\frac{\partial J}{\partial(P_t/\mu_t)}\frac{\partial(P_t/\mu_t)}{\partial\gamma}}{\partial^2(P_tY_t/\mu_t)/\partial t^2} < 0.$$

Hence a higher γ leads to an earlier shakeout. ■

5.2 Industry Dynamics: An Intuitive Illustration

More intuitively, the industry dynamics can be illustrated as follows. First, Equation 17 implies that there is a downward-sloping demand curve on $(P/\mu, F)$. Notice that only the normalized demand $F = Y/m$ matters for our discussion:

$$F = \frac{1}{1+\eta(P/\mu)^{1/g}} \tag{23}$$

and the inverse demand function is convex ($\partial^2(P/\mu)/\partial F^2 > 0$) for $F \in (0, \frac{1+g}{2})$, but concave ($\partial^2(P/\mu)/\partial F^2 < 0$) for $F \in (\frac{1+g}{2}, 1)$.

Second, Equation 15 and 16 suggest that the supply curve is upward sloping on $(P/\mu, F)$, and shifts to the right if the technology A or mean income μ is higher. The normalized supply F is given as

$$F = (\frac{\alpha\mu^{1-\psi}}{\sigma})^{\frac{\alpha}{1-\alpha}}(P/\mu)^{\frac{\alpha}{1-\alpha}}A^{\frac{1}{1-\alpha}}\frac{M}{m}\int_{\frac{(\phi)^{1-\alpha}(\sigma\mu^{\psi-1})^{\alpha}}{(1-\alpha)^{1-\alpha}(\alpha)^{\alpha}A(P/\mu)}}^{\infty}(\theta^{\frac{1}{1-\alpha}})dS(\theta). \tag{24}$$

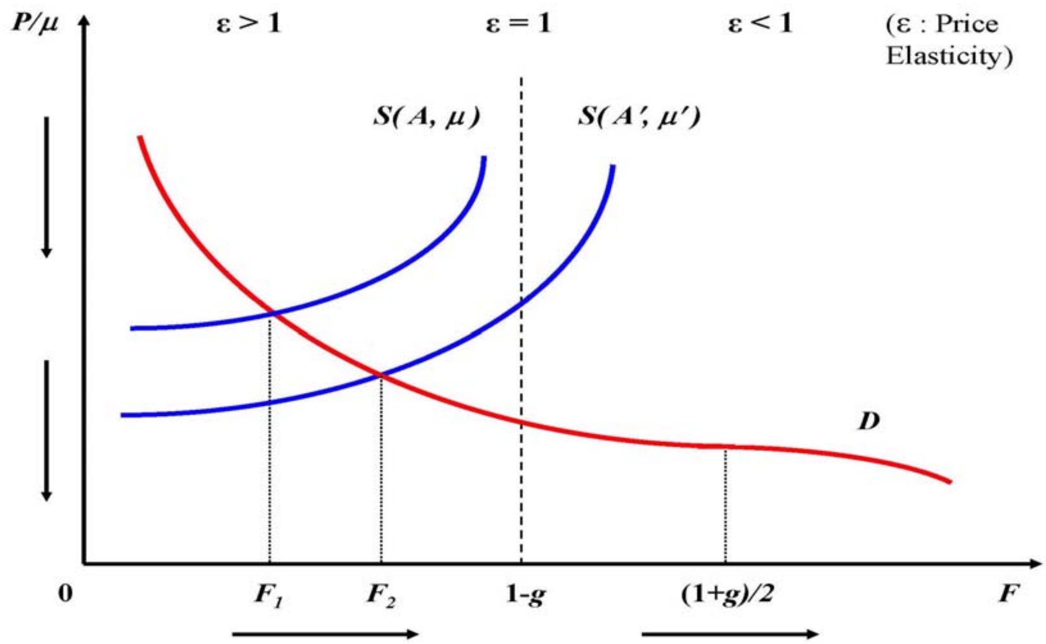

Figure 9: Industry Life Cycle: An Illustration

Plotting the demand and supply curves on the graph of $(P/\mu, F)$ for a given technology A and mean income μ, we can then pin down the solution for the momentary equilibrium with Figure 9.

For the dynamic analysis, we need notice that there is an important property of the demand function 23 for the relative price elasticity,

$$\varepsilon = \left| \frac{\partial F/F}{\partial(P/\mu)/(P/\mu)} \right| = \frac{(1-F)}{g},$$

which decreases with F and achieves unit at $F^* = 1 - g$. It suggests something crucial for the time path of relative industry GDP, PF/μ, as well as the number of firm N: if an industry starts from an initial condition that $F_0 < 1 - g$, the supply curve will keep shifting to the right as the technology and mean income improve and the industry achieves the unique shakeout at $F^* = 1 - g$.[20]

[20]It is clear that our core results hold in more generality. The only key assumption we need is the decreasing price elasticity of demand. Introducing log-logistic distribution enables us to justify

With Figure 9, we can clearly identify the two driving forces of industrial evolution and discuss their effects as follows.

- In the presence of technology progress ($\gamma > 0$) but no income growth ($z = 0$), the supply curve shifts to the right due to cumulative production. As the result, the industry relative price P/μ as well as the absolute price P keeps falling, and the product penetrates into lower-income groups. Eventually, the demand growth is overtaken by the technology progress so less firms are needed. Hence we observe the aggregate demand turns inelastic and the shakeout starts.

- In the presence of income growth ($z > 0$) but no technology progress ($\gamma = 0$), the supply curve shifts to the right due to income growth. Though the absolute price P may not fall (e.g. $\partial P/\partial t > 0$ if $0 \leq \psi < 1$), the relative price P/μ keeps falling and induces more adoption.[21] Eventually, as the market demand turns inelastic to the relative price, the growth of industry profit is outstripped by the growth of foregone earnings of human capital so the shakeout comes in.

- In the presence of both technology progress ($\gamma > 0$) and income growth ($z > 0$), the supply curve shifts to the right due to both cumulative production and income growth, and the two driving forces interact to result in demand diffusion and industry shakeout.

With Figure 9, it is also easy to understand the comparative dynamics proved in Theorem 3: if the technology is better (higher Q_0, higher A_0 or higher γ), the mean

---

this assumption with typically observed diffusion curves and uncover the links with the underlying income distribution.

[21] In the literature of international economics, it is termed as the "Balassa-Samuelson Effect" that a non-tradable good typically has a higher price in a richer country due to the higher foregone earning for producing it. However, in spite of the higher price, the consumers in a richer country typically consume more of the good given their higher income. It is consistent with our model that it is the relative price P/μ rather than the absolute price P that drives the demand diffusion.

income is higher (higher μ_0 or higher z), the market size is larger (higher m) or the input prices is lower (lower ϕ, lower σ or lower ψ), it contributes to the cumulative production and/or income growth so the supply curve shifts faster to the right. As the result, the industry achieves faster demand diffusion and earlier shakeout.

The above discussion provides a meaningful explanation for the variation of industrial evolution across countries. For the US and the UK, the income inequality and consumer preference are similar, so their threshold adoption level for shakeout $F^* = 1 - g$ should be close. Therefore, the US, being relatively richer and larger, tends to enjoy faster demand diffusion and earlier industry shakeout.

5.3 Industry Life Cycle: Further Implications

Our model also delivers rich implications on other aspects of the industry life cycle.

As suggested, firms are different in size and profit due to their heterogenous efficiency (i.e. $\partial y^*_\theta / \partial \theta > 0$ and $\partial \pi_\theta / \partial \theta > 0$). Assume each individual firm's efficiency θ_t is fixed over time. Since the minimum ability requirement $\tilde{\theta}_t$ falls before the shakeout and rises afterwards, we then observe that high-efficiency firms enter the industry earlier and survive longer. It is usually termed as "first mover advantage" though it indeed is firms who have advantage that move first.

This result can be easily generalized. In fact, our analysis does not require a fixed θ_t for each individual firm over time. As a simple example, we may assume

$$\ln \theta_t = \ln \theta + \varepsilon_t \tag{25}$$

where θ is an individual firm's fixed efficiency, and ε_t are i.i.d. random shocks. It implies that the overall distribution of θ_t is time-invariant, but individual firms have idiosyncratic shocks.[22] As a result, we may observe both entry and exit before and after the shakeout, and the net number of firms evolves as our model describes.

[22] For example, if $\ln \theta$ and ε_t are normally distributed, θ_t then has a time-invariant log-normal distribution.

Moreover, since firm efficiency is time persistent, we also observe early entrants tend to be larger and have higher survival rate. This result is consistent with well-established findings of industry studies on firm age and size effects (Evans 1987a, 1987b, Dunne, Roberts and Samuelson 1988, 1989, Audretsch 1991). Simons (2002) shows it is particularly true for the US and UK TV industries.

As technology improves, each individual firm's output tends to grow. Given the assumption 25, our model suggests that surviving firms tend to have the same proportionate growth, usually quoted as the "Gibrat's Law".[23] To see that, notice Equation 2 implies

$$\ln(y^*_{\theta,t}) = \alpha \ln \frac{\alpha\phi}{(1-\alpha)\sigma} + \alpha(1-\psi)\ln \mu_t + \ln(A_t) - \frac{\alpha}{1-\alpha}\ln \tilde{\theta}_t + \frac{1}{1-\alpha}(\ln\theta + \varepsilon_t),$$

which suggests

$$\ln(y^*_{\theta,t}) - \ln(y^*_{\theta,t-1}) = \alpha(1-\psi)\ln \frac{\mu_t}{\mu_{t-1}} + \ln \frac{A_t}{A_{t-1}} - \frac{\alpha}{1-\alpha}\ln \frac{\tilde{\theta}_t}{\tilde{\theta}_{t-1}} + \frac{1}{1-\alpha}(\varepsilon_t - \varepsilon_{t-1}).$$

Since $\partial Y/\partial t > 0$, Equation 16 implies

$$\alpha(1-\psi)\ln \frac{\mu_t}{\mu_{t-1}} + \ln \frac{A_t}{A_{t-1}} - \frac{\alpha}{1-\alpha}\ln \frac{\tilde{\theta}_t}{\tilde{\theta}_{t-1}} > 0.$$

Therefore, given $\frac{1}{1-\alpha}(\varepsilon_t - \varepsilon_{t-1})$ are i.i.d. random shocks, we tend to observe that surviving firms grow at the same positive rate.

Furthermore, as the industry evolves, the market concentration ratio displays the "U" shape over time, and the industry profitability goes to the reverse direction. To see that, denote λ_q to be the market share for the top q firms that survive the period of interest. We have

$$\lambda_q = \frac{M\int_{S^{-1}(1-q/M)}^{\infty} y^*_\theta dS(\theta)}{M\int_{\tilde{\theta}}^{\infty} y^*_\theta dS(\theta)} => \frac{d\lambda_q}{dN} = \frac{d\lambda_q}{d\tilde{\theta}}\frac{d\tilde{\theta}}{dN} < 0.$$

[23] Alternatively, if we assume $\ln\theta_t = \lambda \ln\theta_{t-1} + \varepsilon_t$ with $0 < \lambda < 1$ and ε_t i.i.d. normally distributed, θ_t also follows a time-invariant log-normal distribution but the growth of a surviving firm is decreasing in size. All the other results remain unchanged. It is consistent with studies of Evans and of Dunne, Roberts and Samuelson.

Also, the industry profitability π/μ rises before the shakeout and falls afterwards:

$$\pi/\mu = M\int_{\tilde{\theta}}^{\infty}[(\theta/\tilde{\theta})^{\frac{1}{1-\alpha}}-1]\phi dS(\theta) \Longrightarrow \frac{d(\pi/\mu)}{dN} = \frac{d(\pi/\mu)}{d\tilde{\theta}}\frac{d\tilde{\theta}}{dN} > 0.$$

6 Extension to Durable Goods

So far our theoretical work is built on consumer nondurable goods. However, the analysis can be readily extended to durables. This extension will not only be of its own theoretical interest, but also help our empirical work.

It is the issue of durability that complicates the analysis of durable goods. For a durable good, consumers actually pay rental price for the service from the stock of the good, and the producers are paid the output price to supply the increment of stock to meet the demand. Therefore, some modifications are needed for our original model.[24]

First, we have to derive the rental price from the output price. At equilibrium, the output price is the discounted sum of expected future rents, i.e.

$$P_t = E_t\sum_{\tau=t}^{\infty}\frac{(1-\delta)^{\tau-t}}{(1+r)^{\tau-t}}R_\tau,$$

where r is the market interest rate and δ is the depreciation rate. It implies that the rental price R_t can be written into the following form

$$R_t = [1-\frac{1-\delta}{1+r}E_t(\frac{P_{t+1}}{P_t})]P_t \tag{26}$$

and the consumers make their adoption decisions based on this rental price

$$F_t = \frac{1}{1+\eta(R_t/\mu_t)^{1/g}}. \tag{27}$$

Second, the output Y_t for durable goods is made of two parts. One is the demand growth of the stock $m(F_t - F_{t-1})$. The other is the replacement demand $\delta m F_{t-1}$.

[24]The model is now set in discrete time so that it can be directly brought to the empirical study in the next section.

Therefore, the total output is

$$Y_t = m(F_t - F_{t-1} + \delta F_{t-1}). \tag{28}$$

For a durable good, the rest equilibrium conditions stay the same as the non-durable goods. Notice that if $\delta = 1$, the full depreciation, we actually get back to the case of nondurable goods.

Assuming learning-by-doing technology progress and constant income growth, we can then characterize the industry dynamics. As we go through the empirical study in the next section, we will see that with some minor reinterpretation most our theoretical analyses for nondurable goods remain unchanged for durable goods, and are supported by the data.

7 Model Estimation

In this section we mainly use the data of TV, a durable good, to estimate our model. Given the complex dynamic system and limited data, it is generally difficult to directly test against competing theories with standard statistical tests. However, the estimation results show that our model fits the data as well as, if not better than, alternative theories.

7.1 Data

The origin of TV industry can be traced back to 1930s, when innovation and first production of B&W TV started in the US and the UK. However, WWII resulted in the curtailment of TV production in both countries and it was not until after the war that the TV market got off the ground. In our study, we focus on the TV industry evolution from late 1940s to late 1960s, namely the B&W TV age. During that period, the US and the UK were the two major countries that pioneered the TV adoption and production, and these two markets were segmented.

To estimate our model, yearly data of 7 variables for the B&W TV industry in both the US and the UK are collected. The dataset includes the number of TV producers, TV output, value of TV output, household numbers, TV adoption rate, nominal GDP per capita and CPI. For the UK, we also collect the data of TV licence fee.

Two additional data are also used to extend our empirical work. One is the Gort & Klepper (1982) dataset, which covers the firm numbers, price and output for many US industries from their beginning years until 1970. The other is a panel dataset of B&W TV adoption across 49 US continental states in the 1950s and 1960s.

A detailed description of our datasets is provided in Appendix B.

7.2 Estimation Strategy

Our model requires estimating a four-equation system for market structure, pricing, adoption and output. For a consumer nondurable goods, it is the system of equations 15 - 20 that we should estimate. For a consumer durable good like TV, it is the system that we discussed in section 6. Since both of them involve a system of equations, we then have to deal with the issue of simultaneity. There are two approaches that we may consider as follows.

- OLS (Ordinary Least Squares): If the heterogeneity of firms is negligible, OLS yields consistent parameter estimates. Indeed under the assumption of identical firms, the system of equations for a nondurable good can be simplified as follows:

$$\frac{P_t}{\mu_t} = \frac{(\phi)^{1-\alpha}(\sigma\mu_t^{\psi-1})^{\alpha}}{(1-\alpha)^{1-\alpha}(\alpha)^{\alpha}A_t} \Longrightarrow \ln(P_t/\mu_t) = \kappa + \alpha(\psi-1)\ln(\mu_t) - \gamma\ln(Q_{t-1}) + \varepsilon_t; \tag{29}$$

$$F_t = \frac{1}{1+\eta(P_t/\mu_t)^{1/g}} \Longrightarrow \ln(\frac{1}{F_t}-1) = \beta + \frac{1}{g}\ln(P_t/\mu_t) + \epsilon_t \tag{30}$$

where κ and β are constants, ε_t and ϵ_t are random errors. For a durable good, the price equation 29 stays the same, and the adoption equation 30 still holds if the consumer expects the price to decline at an approximately constant rate which typically fits well with data. Notice that the system of equations 29 and 30 is *recursive* since each of the endogenous variables can be determined sequentially and the errors from each equation are independent of each other. In a system of this sort, the OLS is the appropriate estimation procedure.

- 2SLS (Two-Stage Least Squares): If the heterogeneity of firms is not negligible, we have to confront the problem of simultaneity. Notice that in the case of nondurable goods, equation 29 actually is a reduced-form linear equation for price. Hence it can be estimated with OLS in the first-stage regression and the fitted values of the dependent variable $\ln(P_t/\mu_t)$ can then be used in the second-stage regression for the adoption equation 30. However, for durable goods, equation 29 may be less robust since the lagged adoption rate is also involved.

In the following sections, we report the estimation results of OLS since the 2SLS results are very similar.

7.3 Market Structure Estimation

Proposition 1 predicts comovement of the firm numbers and the relative industry GDP, i.e. $\partial N/\partial(PY/\mu) > 0$. To check it with data, we apply the following regression to the US and UK B&W TV industries and ten other US industries in the Gort & Klepper (1982) dataset:

$$\ln N_t = a + b\ln(PY/\mu)_t + \nu_t$$

where ν_t is assumed to be a Gaussian white noise process.[25]

The regression results are reported in the following Table 1:

[25] The results are robust to alternative assumptions such as the error has time trend or serial correlation.

Table 1: Testing Proposition 1

Product	Data Range	b $(S.E.)$	$b \leq 0$	$adj.R^2$	$Corr$ $(N, PY/\mu)$
Black & White TV (US)	1947-1963	0.31 (0.12)	R	0.28	0.61
Black & White TV (UK)	1949-1967	0.57 (0.11)	R	0.57	0.72
Blanket, Electrical	1950-1970	0.43 (0.08)	R	0.55	0.75
Computer	1955-1970	0.10 (0.02)	R	0.59	0.63
Freezer	1947-1970	0.31 (0.07)	R	0.42	0.69
Nylon	1950-1970	1.65 (0.13)	R	0.89	0.97
Penicillin	1949-1960	0.49 (0.10)	R	0.67	0.78
Pens, Ball Point	1951-1970	1.01 (0.25)	R	0.45	0.81
Styrene	1943-1970	1.04 (0.29)	R	0.30	0.42
Tapes, Recording	1958-1970	0.73 (0.08)	R	0.90	0.92
Tires, Automobile	1913-1953	1.30 (0.13)	R	0.72	0.78
Transistors	1954-1970	0.29 (0.02)	R	0.92	0.88

R: Reject at 5% significance level

For all 12 products included in the test, we reject the null hypothesis $b \leq 0$ at 5% significance level, which suggests that Proposition 1 works well for this sample group of products. Given the evidence presented in Figure 3 and 4, the results are not surprising at all. As we mentioned before, it is still possible to find counterexamples but that does not necessarily invalidate our analysis (See Appendix A).

7.4 Price Estimation

We estimate here three alternative models on price. Model (P-2), derived from our theory (Equation 29), estimates the relative price using real per capita income and

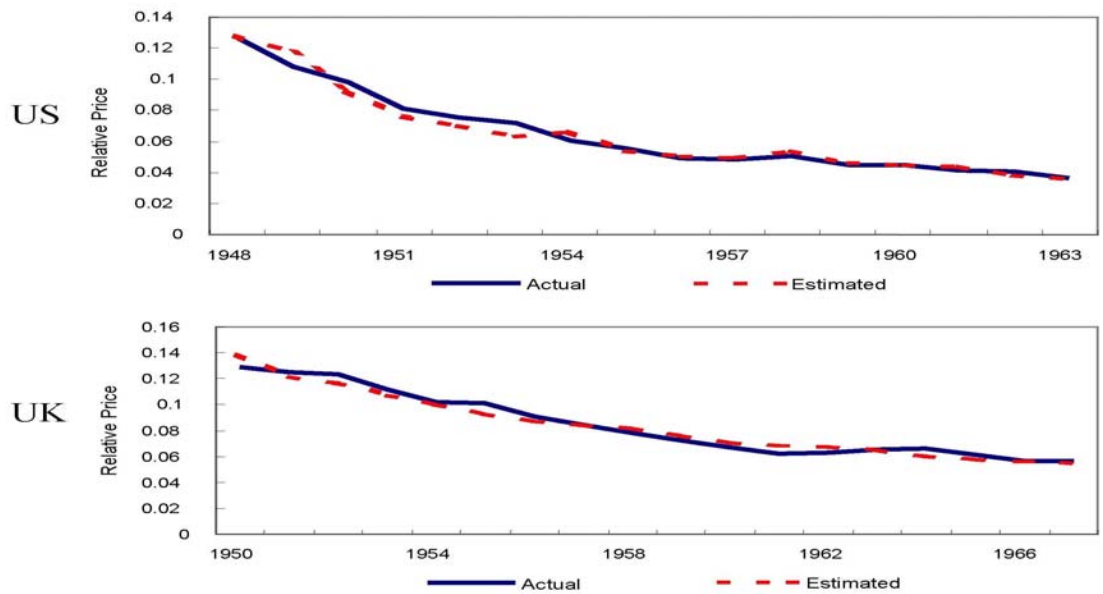

Figure 10: TV Relative Price Estimation: US and UK

cumulative industry output. The parameters to be estimated are changing rate of relative input price $\alpha(\psi-1)$ and technological learning rate γ. For comparison, we also estimate two additional models. In Model (P-1), the relative price is estimated with a time trend only, which provides the average annual price change. In Model (P-3), the cumulative output of the US production is included in the estimation of the UK TV relative price so we can estimate how much the UK producers may have benefited from the technology spillover from the US.

$$\ln(P_t/\mu_t) = \kappa + \omega t + \varepsilon_t; \quad \text{(P-1)}$$

$$\ln(P_t/\mu_t) = \kappa + \alpha(\psi-1)\ln(\mu_t) - \gamma\ln(Q_{t-1}) + \varepsilon_t; \quad \text{(P-2)}$$

$$\ln(\frac{P_{uk,t}}{\mu_{uk,t}}) = \kappa + \alpha(\psi-1)\ln(\mu_{uk,t}) - \gamma\ln(Q_{uk,t-1} + hQ_{us,t-1}) + \varepsilon_t \quad \text{(P-3)}$$

where μ is real per capita GDP in 1953 dollar (pound) and Q is cumulative output.

The results are in Table 2. All the parameter estimates have the expected signs and most are statistically significant.[26] The estimation of Model (P-1) shows the

[26]Notice that the regressions may involve nonstationary time series so that the t test becomes

US enjoyed a much faster annual declining rate of the relative TV price than the UK. By estimating the learning by doing equation, Model (P-2) suggests that the US advantage was due to a faster declining rate of input price in addition to the larger cumulative output. Figure 10 provides the data fitting using Model (P-2).

Since the TV industry was developed earlier and faster in the US, we may wonder whether and how much the UK producers could have learned from the US experience. Model (P-3) tests this hypothesis by including the US cumulative output into UK relative price equation. The regression is conducted using nonlinear least squares. Comparing the estimation results with Model (P-2), we find that introducing the US experience does not improve fitting the UK price data, and the parameter estimates are almost unchanged.[27] Moreover, the magnitude of h, the coefficient of US experience, is very small if it exists at all. This finding is consistent with our assumption that these two markets were technologically segmented during that period.

Table 2: TV Price Estimation

Data	Model	ω (S.E.)	$\alpha(\psi - 1)$ (S.E.)	$-\gamma$ (S.E.)	h (S.E.)	$adj.R^2$
US	(P-1)	-0.08^* (0.005)				0.93
(1948-1963)	(P-2)		-2.61^* (0.38)	-0.06^* (0.02)		0.97
UK	(P-1)	-0.05^* (0.003)				0.94
(1949-1967)	(P-2)		-1.26^* (0.23)	-0.08^* (0.02)		0.96
	(P-3)		-1.26 (0.71)	-0.08 (0.25)	0.003 (0.77)	0.96

* Statistically significant at 5% level.

less meaningful. However, OLS still consistently estimates the parameters as long as the regression equations are correctly specified. Cointegration tests are conducted for each model specification, but the results are not conclusive because of the small sample size.

[27] It is not surprising that the estimates of Model (P-3) are not statistically significant given the limited data we use to estimate a nonlinear model.

7.5 Adoption Estimation

We estimate here four models for the B&W TV adoption. First, if the diffusion is a social contagion process, we have the logistic model:

$$\ln(\frac{1}{F_t} - 1) = \beta + wt + \epsilon_t \tag{A-1}$$

where $\beta = \ln(\frac{1}{F_0} - 1)$, $w = -v$.

For a consumer durable good like TV, if the decision makers predict a constant price declining rate, i.e. $E_t(\frac{P_{t+1}}{P_t}) = \rho$, our adoption equations 26 and 27 suggest

$$\ln(\frac{1}{F_t} - 1) = \beta + \frac{1}{g}\ln(P_t/\mu_t) + \epsilon_t \tag{A-2}$$

where $\beta = \ln\eta + \frac{1}{g}\ln(1 - \frac{1-\delta}{1+r}\rho)$.

To check the robustness of our theory, we also conduct a panel estimation using TV diffusion data across 49 US continental states in year 1950, 1955, 1959 and 1963.

$$\ln(\frac{1}{F_{i,t}} - 1) = \beta + \frac{1}{g}\ln(P_t/\mu_{i,t}) + u_i + \epsilon_t \tag{A-3}$$

where $F_{i,t}$ is the TV adoption rate of state i at time t, $\mu_{i,t}$ is the per capita income of state i at time t, and u_i is the fixed effect of state i.

One thing special about the UK TV market is that the UK government imposes a TV licence fee, which is in fact a tax on TV ownership. As our theory predicts, it would further delay the TV adoption and shakeout. To include the TV licence fee into our regression, Equations 26 and 27 suggest

$$\ln(\frac{1}{F_t} - 1) = \beta + \frac{1}{g}\ln(R\frac{P_t}{\mu_t} + \frac{L_t}{\mu_t}) + \epsilon_t \tag{A-4}$$

where $R = [1 - \frac{1-\delta}{1+r}\rho]$ is the TV rental rate, L_t is the TV licence fee at year t.

The estimation results are in Table 3. In term of R^2 value, we find the endogenous diffusion models (A-2) and (A-4) fit data better than Model (A-1). Of course, this result by itself is not sufficient to reject the contagion model, but at least suggests our theory, built on consumer heterogeneity, provides a very competitive explanation.

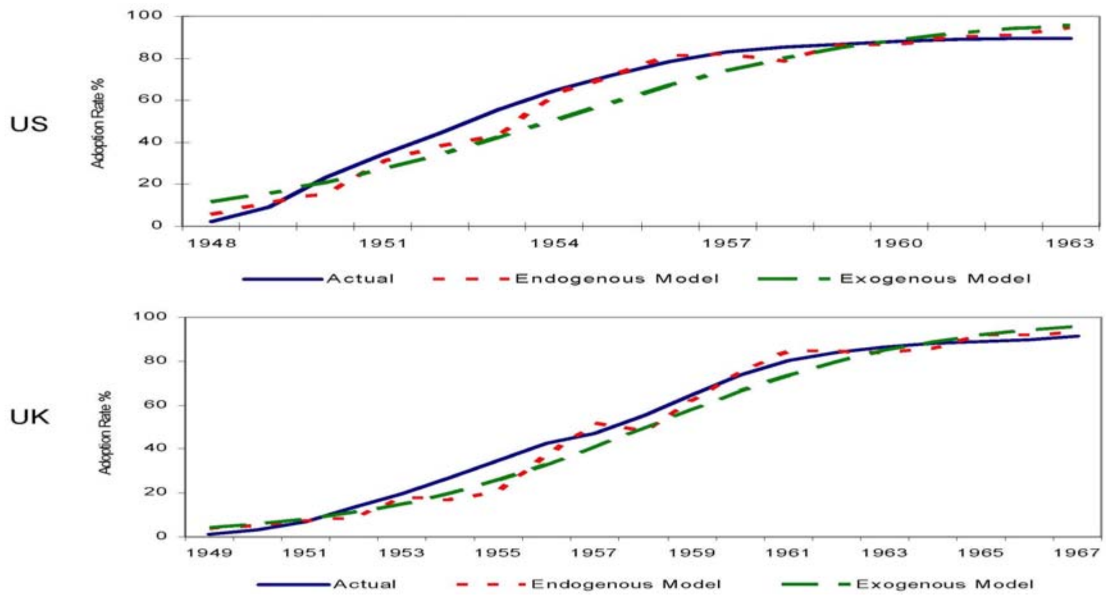

Figure 11: TV Adoption Estimation: US and UK

Model (A-3) reports the panel-data estimates using the fixed-effect model (random-effect model is rejected by the Hausman specification test). Impressively, the results are very close to what we estimate using Model (A-2). It hence gives us more confidence on our theory. See Figure 11 for the data fitting.

Table 3: TV Adoption Estimation

Data	Model	w $(S.E.)$	R $(S.E.)$	$1/g$ $(S.E.)$	$adj.R^2$
United States	(A-1)	-0.34^* (0.06)			0.84
(1948-1963)	(A-2)			4.52^* (0.44)	0.95
	(A-3)			5.38^* (0.30)	0.55
United Kingdom	(A-1)	-0.35^* (0.04)			0.94
(1949-1967)	(A-2)			6.46^* (0.61)	0.94
	(A-4)		0.25^* (0.11)	8.56^* (0.97)	0.95

* Statistically significant at 5% level.

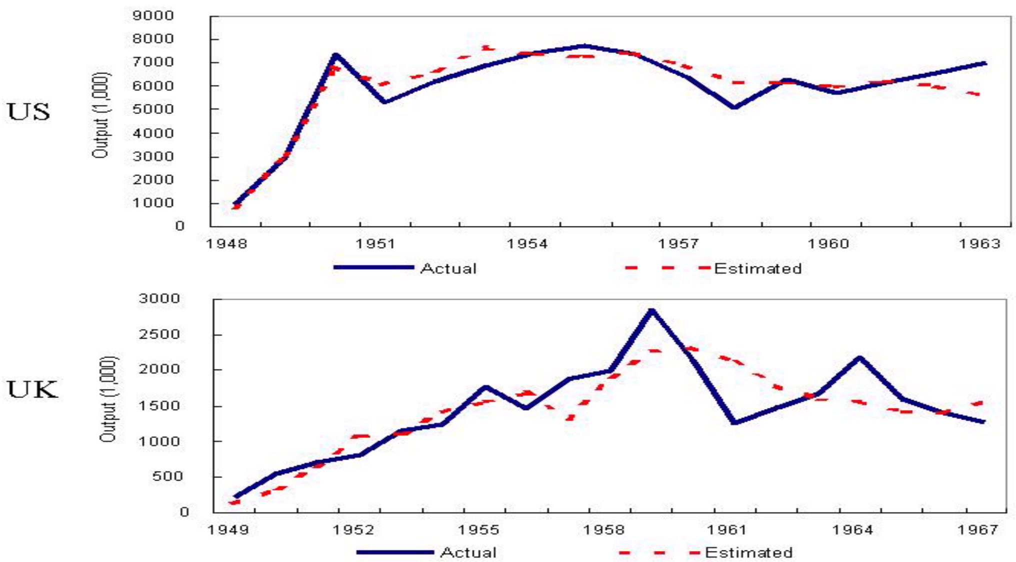

Figure 12: TV Output Estimation: US and UK

7.6 Output Estimation

Having estimated the adoption rates of B&W TV, the output estimation becomes straightforward. The demand function takes the following simple form as Equation 28:

$$Y_t = m_t(F_t - F_{t-1} + \delta F_{t-1}) + \upsilon_t$$

where m_t is the number of households at time t, δ is the annual depreciation rate. It implies that the annual output consists of two parts: the first purchase from the new adopters and the replacement purchase from the existing adopters.

Table 4: TV Output Estimation

Data	$\underset{(S.E.)}{\delta}$	$adj.R^2$
US (1948-1963)	$\underset{(0.005)}{0.10^*}$	0.84
UK (1949-1967)	$\underset{(0.009)}{0.08^*}$	0.53

* Statistically significant at 5% level.

Given the data of Y_t, m_t and F_t, the only parameter to estimate is the depreciation rate δ. Though a constant δ seems to be a relatively strong assumption, the estimation results in Table 4 suggest that it does explain the data pretty well. See Figure 12 for the data fitting.

7.7 Demand Factors, Diffusion and Shakeout

The regression results show that our model fits the time paths of industry variables very well. We are now ready to explore the specific roles that demand factors played in shaping the TV industry evolution in the US and the UK.

The facts are for the UK in that period, the per capita GDP was 70-80% of the US and the number of households was one third of the US. As our model suggests, the lower mean income together with the smaller market size would lead to a sustainedly higher TV price-income ratio in the UK than the US.[28] In addition, the TV licence fee in the UK added further cost of adoption. Therefore, TV diffusion in the UK was delayed. The TV price-income ratio is plotted in Figure 13 (a).

How did this delayed diffusion affect the timing of shakeout for a durable good like TV? Recall the early demand of durable goods is mainly the first purchase:

$$Y_t \approx m(F_t - F_{t-1}) \tag{31}$$

where m is assumed to be a constant market size. Given the sustainedly higher P_t/μ_t ratio, the TV market expansion was postponed in the UK so that the peak of first purchase came in much later. The data of $(F_t - F_{t-1})$ is plotted in Figure 13 (b).

Comparing Figure 12 with Figure 13, we notice that the actual TV output declined less severely in the US than what Equation 31 predicts, which was due to the fast population growth during that period.[29] However, the industry output did level off

[28] The absolute price of B&W TV in the UK was not necessarily higher than the US as we have discussed in section 5.1. Using both official exchange rate and PPP, we found the price was actually lower in the UK.

[29] During 1948-1963, the household numbers grew at 2% annually in the US and 0.5% in the UK.

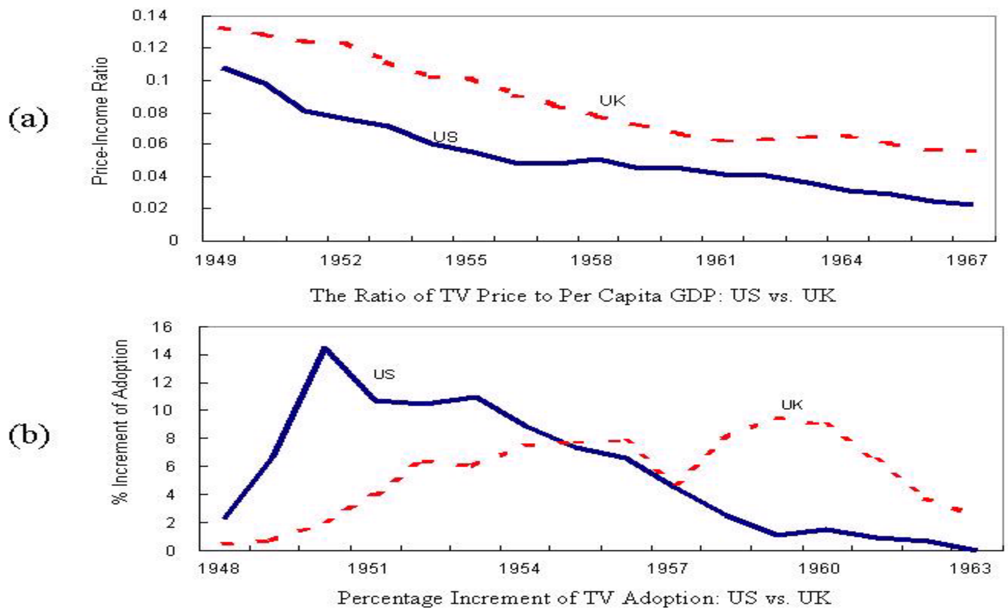

Figure 13: TV Relative Price and Adoption Increment: US vs. UK

in early 1950s, and the relative industry GDP as well as the firm numbers started declining at that time. In the case of UK, where the population grew much slower, the time path of actual output is very close to what we calculate with Equation 31, which kept increasing until 1959. Consequently, the relative industry GDP as well as the firm numbers had not declined until then.

8 Concluding Remarks

Understanding the industry life cycle is an important frontier of economics. One limitation of the existing literature is its brief treatment of the demand side. Major theories, like Jovanovic and MacDonald 1994 or Klepper 1996, emphasize the effects of inter-firm technological differences but did little to explain the impacts of demand characteristics. This paper therefore complements existing knowledge by exploring a previously largely unexplored demand side. First, it shows that demand changes alone could drive a shakeout. Second, it connects characteristics of demand, including income distribution and market size, to industry life cycle dynamics including timing of shakeout and observed adoption rates. In addition, this paper provides a new explanation for logistic demand diffusion curves other than the traditional social contagion theories.

To simplify the analysis, our model assumes a stylized portrayal of firms. Firms are assumed to face decreasing return to scale in the short run, achieve optimal production scale immediately upon entry, and change optimal production without cost. Moreover, the span of control α and efficiency distribution $S(\theta)$ are assumed time-invariant. Although these assumptions are not unusual and work reasonably well in the paper, we need be aware that not all of them always hold, and some may work better than others in a case-by-case base. In particular, when the comovement between firm numbers and relative industry GDP breaks down,[30] it signals that additional explanations may be required outside our theory (e.g. a supply-side explanation). However, it is possible to extend our theory to consider additional supply-side effects. As an example, Appendix B modifies the model to allow an increasing α. As a result, the shakeout starts earlier than the decline of relative industry GDP, but nonetheless our major findings on diffusion and shakeout remains valid.[31]

Two final comments. First, the competitive market does not internalize spillover of learning, so the equilibrium derived in the model is not Pareto optimal. A social planner would prefer a faster diffusion and earlier shakeout. Second, the close-country framework can be generalized. In an open world, a country may specialize in certain industries to explore the comparative advantage. Then it should be the world income distribution and world market size that shape the industry life cycle.

[30] For example, the shakeout may start earlier than the decline in relative GDP (e.g. Auto and Color TV), or may not occur at all (e.g. Shampoo). I thank Steven Klepper for pointing this out.

[31] This extension may explain the Auto shakeout in the 1910s. For Color TV, the shakeout was probably also related to the change of $S(\theta)$ considering the entry of Japanese firms at the time (see footnote 4).

Appendix A: Increasing Span of Control

In the paper, we have shown that the relative industry GDP is a good indicator for the number of firms. However, it is possible to find counterexamples. A dramatic case is the automobile, for which the firm shakeout came in around 1910 but the relative industry GDP kept rising up to the Great Depression. Does that mean our theory is not consistent with the fact?

Not necessarily. Recall that Proposition 1 is derived on the assumption that the span of control parameter α is constant over time. In the case of automobile, it is less likely to be valid. In fact, as the assembly line was introduced in 1910s the auto industry became more and more capital intensive, so the technology progress may have also worked through the increasing span of control. If we take that into account, it is possible for the number of firms to deviate from the relative industry GDP so the shakeout comes in earlier. However, this extension does not necessarily invalidate our previous analysis. In the following, we construct an example with endogenously increasing span of control to clarify this point.

For simplicity, we assume the industry has identical firms. The production function is $y = \phi_\alpha A x^\alpha$ where $\phi_\alpha = (1-\alpha)^{-(1-\alpha)}\alpha^{-\alpha}$. Assume that there is only labor input, and the entrepreneur's forgone earning equals labor's wage, i.e. $C = w$, and mean income of the economy is fixed over time. We then have the momentary industry equilibrium conditions as follows:

$$P = \frac{C}{A}; \tag{32}$$

$$Y = mF = \frac{m}{1+\eta(P/\mu)^{1/g}}; \tag{33}$$

$$N = \frac{1-\alpha}{C} PY. \tag{34}$$

Assume that there are two channels of learning by doing:

$$A_t = A_0 Q_t^\gamma; \qquad \alpha_t = 1 - \alpha_0 Q_t^{-\lambda}$$

where $Q_t = \int_0^t Y(s)ds + Q_0$.

Characterizing the dynamics of this system, we can show most our analyses stay unchanged except that the shakeout of firms now starts earlier than that of GDP.

Lemma 1 *The number of firms starts declining when price elasticity falls to* $1+\lambda/\gamma$.

Proof. Given $P_t = \frac{C}{A_0}Q_t^{-\gamma}$ and $1-\alpha_t = \alpha_0 Q_t^{-\lambda}$, we have

$$N_t = \frac{1-\alpha_t}{C}P_tY_t = \alpha_0 A_0^{\lambda/\gamma}(P_t/C)^{1+\lambda/\gamma}Y_t.$$

Hence, the peak of firm numbers is determined as follows:

$$\frac{\partial N_t}{\partial t} = 0 => (1+\lambda/\gamma)\frac{\dot{P}_t}{P_t} + \frac{\dot{Y}_t}{Y_t} = 0 => \varepsilon^* = 1+\lambda/\gamma.$$

■

Recall the industry GDP starts declining at $\varepsilon^* = 1$. Since the demand function 33 has a declining price elasticity, the shakeout of firm numbers hence starts earlier at a lower adoption level, $1 - \frac{g(\lambda+\gamma)}{\gamma}$, instead of $1-g$ for the industry GDP.

Using Figure 14, we can also check the effects of parameter changes. In particular, the supply curve shifts down faster if the technology is better (higher Q_0, higher A_0 or higher γ), the mean income is higher (higher μ), the market size is larger (higher m) or the input price is lower (lower C (w)). As a result, it leads to a faster demand diffusion and earlier industry shakeout.

So far, we have not linked the foregone earning C and input price w to the mean income μ. What if we do? For the model to be analytically solvable, we no longer distinguish the entrepreneur's foregone earning from the input price, i.e. $C = w$. It is a little different from what we assumed in the paper: $C = \phi\mu$ and $w = \sigma\mu^{\psi}$ $(\psi < 1)$. Therefore, we have to discuss two special cases. In one case, $C = w = \phi\mu$, a higher mean income will be fully transferred into a higher production cost so that it does not result in a lower relative price P/μ for a given technology. Hence it does not lead to a faster diffusion and earlier shakeout. In the other case, if $C = w = \sigma\mu^{\psi}$ $(\psi < 1)$,

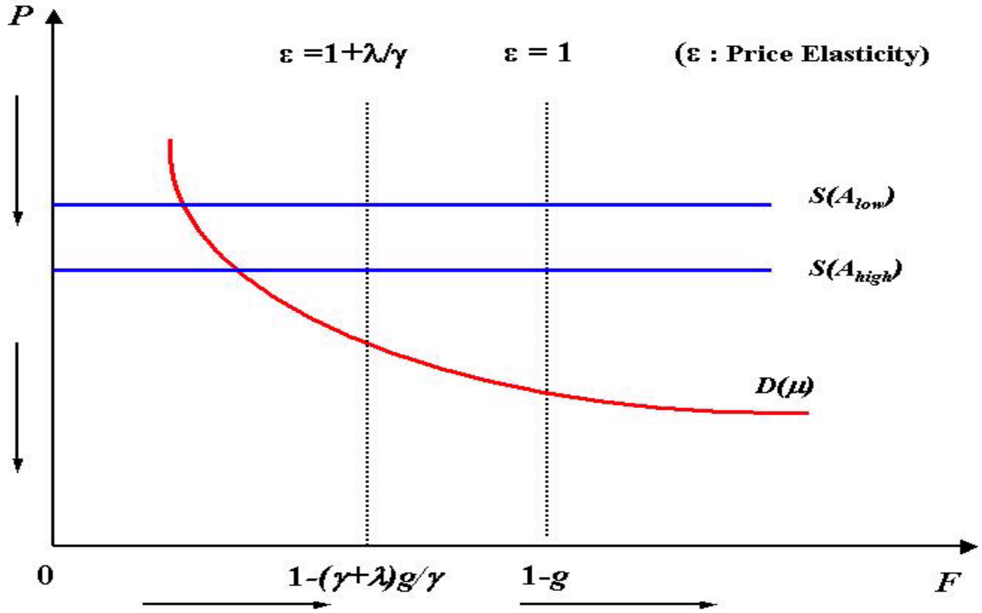

Figure 14: Industry Life Cycle with Increasing Span of Control

a higher mean income will only be partly transferred into the production cost so that the relative price P/μ is lower for a given technology. As the result, it does lead to a faster diffusion and earlier shakeout.

Appendix B: Data Details

- *US-UK TV Dataset* – The US data starts as early as 1946 when the B&W TV was initially introduced, and ends at 1963 when the sale of color TV took off. Most of the data (the number of TV producers, TV output, value of TV output, household numbers and TV adoption rate) are drawn from periodic editions of *Television Factbook.* The nominal GDP per capita is from Johnston & Williamson (2003) and CPI is from *International Historical Statistics: the Americas*, 1750-1993. The UK data also starts from 1946 but ends a little later than the US at 1967 when the color TV was introduced. The number of TV producers is cited from Simons (2002). The TV output and value are from *Monthly Digest of Statistics (1946-1968).* The household numbers are calculated by population (from *UN Common Database*) divided by average household size (from *UN Demographic Yearbook*). The TV adoption rate is from Table AI of Bowden & Offer (1994), and TV licence fee is from the BBC press office. The nominal GDP per capita is from Officer (2003) and the CPI is from *International Historical Statistics: Europe, 1750-1993.*

- *Gort-Klepper Dataset* – It covers time-series data of firm numbers (46 industries), price (23 industries) and output (25 industries) from the beginning of each industry up to 1970. 10 industries are selected for our empirical study considering their long enough coverage and continuos observations.

- *US TV Panel Dataset* – It covers the TV adoption rate and personal income across 49 continental states of the US at year 1950, 1955, 1959 and 1963. The TV adoption rate is drawn from the *Television Factbook*, and the personal income is drawn from the *Statistical Abstract of the United States.*

References

[1] Arnold, Erik, (1985). *Competition and Technological Change in the Television Industry.* Macmillan, London.

[2] Arrow, Kenneth, (1962). "The Economic Implications of Learning by Doing," *Review of Economic Studies*, 29(3), 155-173.

[3] Audretsch, David B., (1991). "New-Firm Survival and the Technological Regime," *Review of Economics and Statistics*, 73:3, 441-450.

[4] Bass, F. M., (1969). "A New Product Growth Model for Consumer Durables," *Management Science*, 15 (Jan.), 215-27.

[5] Bogart, Leo, (1972). *The Age of Television.* F. Ungar Pub. Co., New York.

[6] B.C.G., (1972). "Perspectives on Experience," *Technical Report*, Boston.

[7] Bowden, Sue and Avner Offer, (1994). "Household Appliances and the Use of Time: The United States and Britain Since the 1920s," *Economic History Review*, New Series, 47 (Nov.), 725-748.

[8] Comin, Diego and B. Hobijn, (2004). "Cross-Country Technological Adoption: New Evidence and Questions," *Journal of Monetary Economics*, (Jan.).

[9] David, Paul, (1969). "A Contribution to the Theory of Diffusion," mimeo.

[10] Dunne, T., M. Roberts, and L. Samuelson, (1988). "Patterns of Firm Entry and Exit in U.S. Manufacturing Industries," *RAND Journal,* 19:4, 495-515.

[11] Dunne, T., M. Roberts, and L. Samuelson, (1989). "The Growth and Failure of U.S. Manufacturing Plants," *Quarterly Journal of Economics*, (Nov.), 104(4), 671-98.

[12] Emmett, B.P., (1956). "The Television Audience in the United Kingdom," *Journal of the Royal Statistical Society*, Series A, 119, 3, 284-311.

[13] Evans, David S., (1987a). "The Relationship between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries," *Journal of Industrial Economics*, (Jun.), 35(4), 567-81.

[14] Evans, David S., (1987b). "Tests of Alternative Theories of Firm Growth," *Journal of Political Economy*, (Aug), 95(4), 657-74.

[15] Fisk, Peter R., (1961). "The Graduation of Income Distributions," *Econometrica*, 29, 2, 171-185.

[16] Gort, Michael and Steven Klepper, (1982). "Time Paths in the Diffusion of Product Innovations," *The Economic Journal*, 92, (Sept.), 630-653.

[17] Griliches, Zvi, (1957). "Hybrid Corn: An Exploration in the Economics of Technological Change," *Econometrica*, 25 (Oct.), 501-522.

[18] Grossman, Gene M. and Elhanan Helpman, (1991). "Endogenous Product Cycles," *The Economic Journal*, 101, 1212-1229.

[19] Hopenhayn, H., (1993). "The Shakeout," mimeo, Universitat Pompeu Fabra.

[20] Horsky, Dan, (1990). "A Diffusion Model Incorporating Product Benefits, Price, Income and Information," *Marketing Science*, 9 (Autumn), 342-365.

[21] Horvath, Michael, Fabiano Schivardi and Michael Woywode, (2001). "On Industry Life-Cycles: Delay, Entry, and Shakeout in Beer Brewing," *International Journal of Industrial Organization*, 19, 1023-1052.

[22] Jovanovic, Boyan and G. M. MacDonald, (1994). "The Life Cycle of a Competitive Industry," *Journal of Political Economy*, 102 (Apr.), 322-347.

[23] Jovanovic, Boyan and Peter L. Rousseau, (2002). "Moore's Law and Learning-By-Doing," *Review of Economic Dynamics*, 5 (2), 346-375.

[24] Klepper, Steven and E. Graddy, (1990). "The Evolution of New Industries and the Determinants of Market Structure," *Rand Journal of Economics*, 21, 27-44.

[25] Klepper, Steven, (1996). "Entry, Exit, Growth, and Innovation over the Product Life Cycle," *American Economic Review*, 86 (June), 562-583.

[26] Klepper, Steven and K. Simons, (2000). "The Making of an Oligopoly: Firm Survival and Technological Change in the Evolution of the U.S. Tire Industry," *Journal of Political Economy*, 108, 4, 728-758.

[27] Levy, Jonathan D., (1981). *Diffusion of Technology and Patterns of International Trade: The Case of Television Receivers.* Ph.D. Dissertation, Yale University.

[28] Lucas, Robert E., (1978). "On the Size Distribution of Business Firms," *The Bell Journal of Economics*, 9, 2, 508-523.

[29] Lucas, Robert E., (1993). "Making a Miracle," *Econometrica*, 61, 251-272.

[30] Mansfield, Edwin, (1961). "Technical Change and the Rate of Imitation," *Econometrica*, 29 (Oct.), 741-766.

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

[32] Rob, Rafael, (1991). "Learning and Capacity Expansion under Demand Uncertainty," *Review of Economic Studies*, 58 (June), 655-675.

[33] Simons, Kenneth, (2002). "Predictable Cross-Industry Heterogeneity in Industry Dynamics," working paper, Department of Economics, University of London.

[34] Sutton, John, (1997). "Gibrat's Legacy," *Journal of Economics Literature*, Vol. 35, No.1, (Mar.), 40-59.

[35] Utterback, James and Fernando Suárez, (1993). "Innovation, Competition, and Industry Structure," *Research Policy*, 22, 1-21.

[36] Vernon, R., (1966). "International Investment and International Trade in the Product Cycle," *Quarterly Journal of Economics*, 80, 190-207.

[37] Wang, Zhu. (2005) "Technology Innovation and Market Turbulence: A Dotcom Example," *Working Paper*, Payments System Research Department, Federal Reserve Bank of Kansas City.

[38] Zeira, Joseph, (1999). "Information Overshooting, Booms and Crashes," *Journal of Monetary Economics*, 43, 237-257.