Scand. J. of Economics 107(4), 603–630, 2005
DOI: 10.1111/j.1467-9442.2005.00429.x

Moore's Law and the Semiconductor Industry: A Vintage Model*

Ana Aizcorbe

Bureau of Economic Analysis, Washington, DC 20230, USA
ana.aizcorbe@bea.gov

Samuel Kortum

University of Minnesota and Federal Reserve Bank, Minneapolis, MN 55455, USA
kortum@econ.umn.edu

Abstract

In this paper we develop a vintage model to gain a better understanding of the semiconductor industry and its role in recent U.S. productivity gains. Unlike previous work, in our model the observed price declines of individual chips are driven by the introduction of better vintages rather than by learning economies. Dominated chips, nonetheless, continue to be produced, for a time, due to sunk investments in chip-specific production equipment. The model lends partial support to Jorgenson's hypothesis that an exogenous increase in Moore's Law could have generated the more rapid price declines, and faster productivity growth, seen after 1995.

Keywords: Equipment; innovation; microprocessor; productivity

JEL classification: *L*63; *O*31; *O*33; *O*47

I. Introduction

Dale Jorgenson's Presidential address to the American Economic Association (2001) makes a strong case that accelerated technological change in the production of semiconductors, microprocessor units (MPUs) in particular, has driven the recent increased productivity growth in the U.S. economy.[1] But,

* This paper was written for the conference in memory of Tor Jakob Klette: Technology and Change, Oslo, August 2004. A year before his death, Tor Jakob sent one of us (Kortum) a series of articles on Gordon Moore, proposing to pursue a project on semiconductors. We have greatly missed our other coauthor. We have benefited from comments by Fernando Alvarez, George Hall, Tom Holmes, Erzo Luttmer, Kalle Moene, Ed Prescott, Rob Shimer, participants at the conference in Oslo, the Philadelphic workshop on Monetary and Macroeconomics, a seminar at the University of Chicago, and the graduate IO course at the University of Minnesota. Soma Dey provided helpful comments and excellent research assistance. Any errors are our own. Any views expressed here are our own and not necessarily those of the Bureau of Economic Analysis the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

[1] See also Oliner and Sichel (2000, 2002), Jorgenson and Stiroh (2000), McKinsey Global Institute (2001), Jorgenson, Ho, and Stiroh (2002) and Congressional Budget Office (2002).

while semiconductors now figure prominently in accounts of economic growth, Jorgenson points out that there is no fully satisfactory economic model of the industry that produces them.[2] This paper is our attempt to rise to Jorgenson's challenge.

We develop a model of the MPU sector, a key segment of the semiconductor industry. Our intention is to produce a model that: (i) fits the basic facts about this sector, (ii) explains the link between technological improvements and price declines of individual chips, see Jorgenson (2001), and (iii) clarifies how competition in the industry influences these price contours, see Aizcorbe (2004). We put the theoretical model to work to examine how MPU pricing responds to quickening technological change. Consistent with Jorgenson's hypothesis, we find, however, that faster technological change leads to faster declines in chip prices. We also find that the lives of individual chips will be shorter and their introductory prices higher in an environment with faster technological change. These results hold across various assumptions on industry market structure.

Many of the basic facts surrounding the MPU sector are familiar. We observe a relationship called Moore's Law: an amazingly rapid exponential increase in the performance of chips over time (see Figure 1).[3] At the same time, we observe sharp declines in the prices of individual MPU chips over their product life (see the top panel of Figure 2). An intuitive way to tie these two facts together is that the price declines of existing products are necessitated by their need to stay competitive with newly introduced chips whose better performance traces out Moore's Law. We can capture this intuition by treating chips as homogeneous except that newer chips provide more of whatever older chips provide. Doing so obviates the need for us to entertain more sophisticated models of product differentiation, of strategic interaction among producers, or of learning-by-doing in semiconductors.[4]

[2] There are several models of the memory chip (DRAM) sector, including Baldwin and Krugman (1988) and Flamm (1996). Other work has sought to formally model and estimate the demand for these devices, as in Song (2004), as well as the cost structure facing chip makers, as in Irwin and Klenow (1994) and Siebert (2003).

[3] Moore's Law states that the number of electrical components on a chip will double over a specified time period, currently taken to be about 24 months; see Moore (1965) for the initial statement of the law.

[4] The most common explanation for these downward-sloping price contours is that learning-by-doing reduces manufacturing costs. Although the importance of learning in generating price declines for memory chips is well documented in the empirical literature, evidence for MPUs suggests that learning plays a lesser role in generating price declines there; see Aizcorbe (2005). Hatch and Mowery (1998) discuss the sources of learning curves in semiconductor production and review empirical studies devoted to estimating these curves.

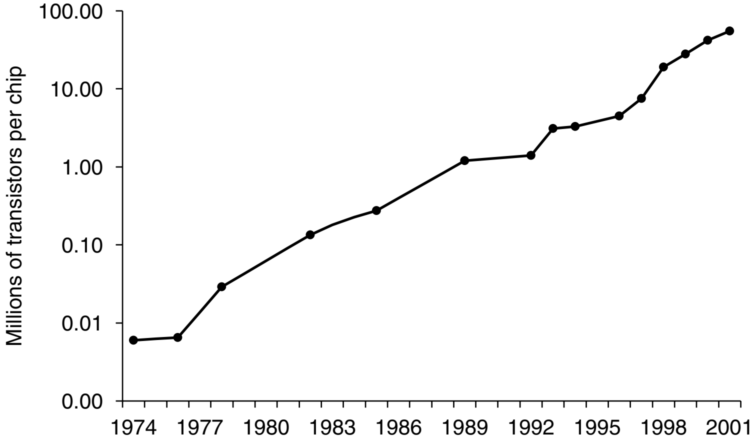

Fig. 1. Moore's Law
Source: Data on the number of transistors on Intel's chips were obtained from www.intel.com/pressroom/kits/quickreffam.htm. Dots show dates of introduction for leading-edge chips and lines are interpolations between these dates.

A remaining puzzle is the fact that a number of different chips, from the one just introduced to products that have been around one, two, or even three years, coexist in the market at the same time. Why doesn't the best product drive the others out of business? Why does the industry continue to produce an inferior chip? Our first stab at a model of the industry attempts to maintain the simplicity of treating MPUs as a commodity while coming to grips with this observation about the availability of a hierarchy of products, all on the market at any given date.

We start with a model of a competitive industry producing microprocessors; later on, we expand it to allow for different degrees of market power. Growth in the industry is driven by Moore's Law which, in the model, reflects improvements in the technology of chip-producing equipment.[5] Since technological improvements are embodied in the equipment for production, we are led to the model of vintage capital, developed by Salter (1960).[6] Investing in any vintage of chip-producing equipment is a sunk cost.

[5] For example, the sophistication of the lithography equipment determines the size of features on chips (transistors). The number of features on each chip in turn governs the chip's performance and functionality (what we refer to as chip quality in our model). The development of equipment capable of etching finer circuitry is often referred to as "process innovation" and the length of time between introductions of new equipment as the "product cycle".
[6] The vintage capital model is laid out more formally by Solow, Tobin, von Weizsacker and Yaari (1966). Moene and Wallerstein (1997) and Mitchell (2002) have recently applied the vintage capital model to labor market inequality and to the scale of production, respectively.

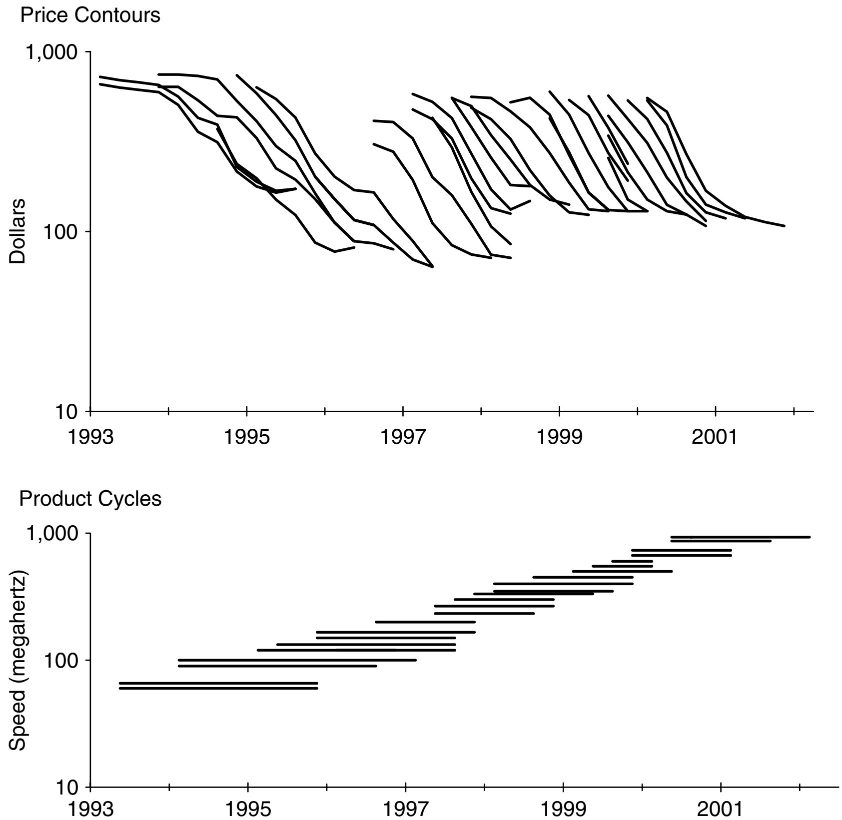

Fig. 2. Price contours, product cycles and quality improvement for Intel's desktop microprocessor chips, 1993–2002
Source: MicroDesign Resources, Inc. Contours are for Pentium desktop chips that were introduced during or after 1993 and that exited the market by 2002.

Thus, a given product continues to be sold even when it is no longer the best performing chip on the market. It drops out of the market only when the competitive price drops below the marginal cost of production.

This straightforward augmentation of the simple competitive model takes us a long way. It delivers a convincing producer-side explanation for the rapid declines in the prices of individual microprocessors over their life on the market. In the model, producers must invest in chip-producing equipment for each new chip they introduce. This investment is specific to a particular chip and is irreversible. The cost of such investments can only be recouped if the introductory price of chips is far above the unit variable cost of producing them. Producers anticipate that this price markup will rapidly deteriorate, however, as new and better chips are introduced. We can interpret the price declines as either declines in the markup of price over unit variable cost (given that the cost of the equipment is sunk) or as declines in costs themselves, if we add to unit variable cost the high but rapidly declining implicit rental price of the equipment. In either interpretation, this explanation stands in sharp contrast to the two most popular explanations for the observed price declines: (i) costs fall due to learning by doing or (ii) markups fall either due to a loss of market power as chips age and become a less differentiated commodity, as in Song (2004) and Hobijn (2002) or due to intertemporal price discrimination, as in Aizcorbe (2004). While these other explanations may have a role in a complete quantitative model, our view is that the simple explanation coming out of the vintage model should be the starting point.

Of course, one may question the relevance of a competitive model when applied to a sector dominated by a single firm, in this case Intel. We are able to show, however, that the predictions of the competitive model extend, in large part, to the analysis of a dominant firm with some degree of market power. We calibrate this extension of the model and evaluate its ability to describe basic facts about the prices of microprocessors. We also use the model to assess the consequences of the speed-up in Moore's Law that is thought to have occurred in the mid-1990s.[7]

As noted above, our model shares many of the features of vintage capital models. The industry equilibrium setting is borrowed from Lucas and Prescott (1971). The model most like ours is Jovanovic and Lach (1989). The novelty of our paper is in the application.

II. A Competitive Baseline

Our baseline model comprises a competitive industry with many chip producers. Time is continuous. Producers discount future profits at rate r.

Demand

We model the demand for chips, of varying quality, in the simplest possible way. Consumers are indifferent between having one chip of quality A or

[7] Beginning in 1995, the industry moved from a three-year to a two-year product cycle and, thus, opened up possibilities for an increase in the rate of product innovation. For additional details about this shift to a shorter technology cycle, see Semiconductor Industry Association (2003 and updates).

A chips of quality 1.[8] The output of the industry can therefore be summarized by the quantity of quality-weighted chips, X. The associated price per quality unit of chips is P.

We posit a simple demand function giving the price P at which a quantity X would be demanded:

$$P = D(X; t).$$

Indexing the function by the date, t, allows for possible shifts in demand. At any date, $D(X; t)$ is downward sloping in quantity, $D'(X; t) < 0$. For some purposes it is useful to denote the quantity demanded X^D as

$$X^D(P; t) = D^{-1}(P; t). \tag{1}$$

Supply

The supply side is more intricate. Suppose a producer enters the industry at date t by building a fabrication plant of physical capacity I (chips per year). The total sunk cost of building such a plant is qI, where q is the cost per unit of capacity. When the plant is running, the unit variable cost is a constant c per chip.[9] Thus, if a plant of size I produces chips at full capacity for l years: (i) total production over the life of the plant is Il, (ii) total variable cost is Ilc, and (iii) total cost, including the sunk cost of building the plant, is $(q + lc)I$.

We assume that the cost of building the plant (or retrofitting it) is the cost of the chip-producing equipment. A plant built at date t embodies equipment of vintage $v = t$. Such a plant produces vintage v chips, whose quality is $A(v)$. The notation v captures the fact that once the plant is built, the

[8] This assumption is admittedly problematic for MPUs for at least two reasons. First, MPUs are actually multidimensional devices with many attributes; see Grimm (1998) for a hedonic analysis of MPUs. Nonetheless, the most important feature of an MPU is speed, an attribute directly related to the number of transistors on the chip, as captured by Moore's Law. Second, even in a unidimensional world, with speed as the only relevant attribute, our assumption on demand is still quite restrictive. Computers are typically configured with only one MPU and it is difficult to argue that consumers are indifferent between owning two slow machines and one fast one. We make these extreme assumptions about demand in order to simplify the analysis so as to highlight the mechanisms, on the supply side, that drive price declines.

[9] Unit costs actually do vary both across chips of different vintages and over each chip's lifespan. For example, yield rates (the percentage of usable chips) typically increase sharply soon after the production of a new chip begins. In our view, treating costs as constant provides a good first approximation to the problem.

producer has locked in a particular vintage of technology and the plant itself will experience no technological change.[10]

Over time, however, new and better vintages of equipment become available. We assume $A(v)$ is a continuously differentiable function $A'(v) > 0$. Looking back in time, it is convenient to assume that $\lim_{v \to -\infty} A(v) = 0$. We assume that producers anticipate the evolution of $A(v)$. We take the path of $A(v)$ as exogenous to these producers.[11]

We denote cumulative investment in the industry, up to and including date t, by $K(t)$. For $v \leq t$, $K(v)$ is the total capacity of all equipment of vintage v or earlier, and $K(t) - K(v)$ is the total capacity of equipment strictly more advanced than vintage v. The constant returns to scale technology that we have assumed allows us to ignore investments in individual plants and to simply keep track of the total. Since investments in the industry are sunk, $K(t)$ is non-decreasing.[12] We want to accommodate occasional spikes of investment at particular dates, hence $K(v)$ may have discontinuities. If there is a mass $I(v)$ of investment in vintage v it equals $K(v) - \lim_{s \nearrow v} K(s)$.

The output of the industry is the total quality units of chips produced. We can compute this aggregate by integrating across all vintages of equipment in operation, weighting the physical output of vintage v chips by $A(v)$. At any date t, we only need to keep track of a single price, $P(t)$, the price of a quality unit.

At date t only fabrication plants of vintage $v \in [t - \tau(t), t]$ will be operated. Given the current price $P(t)$ we can determine $\tau(t)$ from the *competitive shutdown condition*:

$$P(t)A(t - \tau(t)) = c. \tag{2}$$

[10] To simplify the exposition, the model treats the introduction of new chips and the building of new plants as one and the same event. More generally, the model captures the introduction of new chips and any attendant setup costs, whether they be the cost of a new plant and equipment, the cost of adapting old equipment (as occurs with die shrinks), or changes in the production process that allow for the reduction of defects and for chips to test at faster speeds. We gloss over the fact that the investment required for these different types of introductions can be very different.

[11] Jovanovic and Lach (1989) allow A to evolve endogenously. They assume that A is a time-invariant function of cumulative industry investment. We do not find it plausible, however, that learning by doing is the main force behind the slope of Moore's Law for microprocessors. Our assumption is, admittedly, also rather extreme. We attribute the quality improvements in chips to improvements in the equipment that chip producers (such as Intel) purchase from suppliers. Hence, taken literally, in our model equipment producers are the ones driving technological change. We leave it for future work to explain what actually drives Moore's Law.

[12] We can accommodate a secondary market in equipment. But, we do assume that only firms in the semiconductor industry can make use of the specialized equipment used to produce chips.

All older vintages $v < t - \tau(t)$ are left idle. Otherwise they would operate at a loss, as their revenue per physical unit $P(t)A(v)$ would not cover the unit variable cost of production c.

A second condition is that entrants take advantage of all profit opportunities for investing in the industry. Let $V(t, v)$ be the value as of date t of a unit of vintage $v \leq t$ equipment:

$$V(t,v) = \int_t^{\infty} e^{-r(s-t)} \max\{[A(v)P(s) - c], 0\} ds. \tag{3}$$

The value of new equipment at date t is $V(t, t)$ while its cost is q. The absence of profit opportunities delivers the *competitive investment condition*:

$$q \geq \int_t^{\infty} e^{-r(s-t)} \max\{[A(t)P(s) - c], 0\} ds, \tag{4}$$

which holds with equality if there is any investment in vintage $v = t$ equipment.

Integrating over past vintages, the total flow of quality units that can be produced at date t, using vintages $v \in (t - \tau(t), t]$, is $\int_{t-\tau(t)}^{t} A(v)dK(v)$. Given a price P, the lower endpoint of integration is $t - \tau = A^{-1}(c/P)$, which is decreasing in P. A lower bound on the industry supply curve is thus

$$X^S(P;t) = \int_{A^{-1}(c/P)}^{t} A(v)dK(v), \tag{5}$$

which is weakly increasing in P. If there is a mass of investment $I(v)$ in vintage, $v = A^{-1}(c/P)$, then total industry supply will be not only $X^S(P; t)$ but also some fraction of the quality units that could be produced, $A(v)I(v)$, using this marginal vintage of equipment. The market-clearing condition will determine that fraction.

Equilibrium

At any date t, taking account of any new investment at that point in time, the market-clearing price per quality unit $P(t)$ must induce a total supply of quality units of chips equal to demand (1) at that price. Taking account of the possibility that the vintage $v = A^{-1}(c/P(t))$ may be partially utilized, we get the *competitive market-clearing condition*:

$$P(t) = \max\{P : X^S(P;t) \leq X^D(P;t)\}. \tag{6}$$

We consider the behavior of the industry starting from date t_0. In addition to specifying the exogenous time path of technology $A(t)$ and the demand curve $D(X; t)$ for all $t \geq t_0$, we must also specify industry investment prior to that initial date, i.e., $K(v)$ for all $v < t_0$.

A *competitive industry equilibrium* consists of time paths, over all dates $s \geq t_0$, of prices $P^c(s)$ and total capacity $K^c(s)$ (non-decreasing in s) satisfying the competitive investment condition (4) and the competitive market-clearing condition (6) for all $t \geq t_0$.[13] Given the equilibrium values of $P^c(t)$ and $K^c(t)$, the equilibrium quantity $X^c(t)$ can be read off the demand curve (1) and the equilibrium age of the oldest active vintage $\tau^c(t)$ can be read off the competitive shutdown condition (2).

A striking feature of the competitive equilibrium is that a chip's price typically exceeds the unit variable cost of producing it. Consider chips produced with vintage v equipment. As of date t, the price for such a chip is $A(v)P(t)$ and the cost of producing it is c. Suppose $v > t - \tau(t)$ so that vintage v equipment is not about to be shut down. It follows from (2) that $A(v)P(t) > c$, i.e., price exceeds unit variable cost. Since vintage v equipment is operating at full capacity, producers cannot further exploit the situation.

The way to account for the difference between price and unit variable cost is by means of the implicit rental price of equipment, or unit capital cost. The implicit rental price $q(t, v)$ must cover the loss in value of vintage v equipment as of date t along with a normal return on the equipment asset value

$$q(t, v) = -\partial V(t, v)/\partial t + rV(t, v).$$

Differentiating equation (3) yields $q(t, v) = A(v)P(t) - c$. Thus the *total unit cost* of production, the sum of the unit variable cost c and the implicit rental price of the equipment, is exactly equal to the price of the chip.

III. Market Power

We want to extend the analysis above to incorporate a firm with market power. The firm with market power is assumed to be able to commit to a price path. To simplify this extension, we imagine that equipment is rented, at competitive rental prices. The implicit rental market is a modeling device to make explicit how a profit-maximizing producer would account for the opportunity cost of the equipment it uses.

A Rental Market for Equipment

Imagine a large set of firms that rent equipment, of various vintages, to producers. Due to free entry, the future revenue from renting a piece of equipment just covers the cost of purchasing it in the first place, q.

[13] The superscript c indicates that these equilibrium magnitudes are associated with a *competitive* market structure. This notation facilitates comparisons with other market structures considered below.

Consider the problem as of date t_0 with $K(v) = 0$ for $v < t_0$.[14] We begin by solving for rental prices consistent with some $\tau(t)$ specifying that only vintages of equipment $v \in [t - \tau(t), t]$ are operated at date $t \geq t_0$. We can set the rental price for vintages that are not used to zero, $q(t, v) = 0$ for $v < t - \tau(t)$. But by the continuity of $A(v)$, it follows that the marginal vintage must also have a zero rental price, $q(t, t - \tau(t)) = 0$. For any vintage $v > t - \tau(t)$, a unit of equipment produces the same number of quality units of chips as $A(v)/A(t - \tau(t))$ units of the oldest vintage still in use. The net cost saving of using vintage $v > t - \tau(t)$ relative to the marginal vintage determines its equilibrium rental price,

$$q(t, v) = cA(v)/A(t - \tau(t)) - c. \tag{7}$$

Adding together the unit variable cost and the rental price of equipment, a producer faces a total unit cost of $cA(v)/A(t - \tau(t))$ when using vintage v equipment. The total unit cost per quality unit $C(t)$ for a chip producer at date t is

$$C(t) = c/A(t - \tau(t)), \tag{8}$$

the same across all vintages $v \in [t - \tau(t), t]$ that might be used to produce chips at that date.

Given a path $\tau(s)$ for $s \geq t$, we can integrate rental prices forward to obtain the value of a unit of vintage v equipment as of date $t \geq v$:

$$V(t, v) = \int_t^\infty e^{-r(s-t)} \max\{[cA(v)/A(s - \tau(s)) - c], 0\} ds.$$

The value of new equipment installed at date t is $V(t, t)$. Given that new equipment costs q, the absence of profit opportunities implies the *efficient investment condition*:

$$q \geq \int_t^\infty e^{-r(s-t)} \max\{[cA(t)/A(s - \tau(s)) - c], 0\} ds, \tag{9}$$

which holds with equality if there is any investment in vintage $v = t$ equipment.[15]

[14] This technical assumption allows us to ignore the issue of the producer with market power trying to take advantage of the rental firms' sunk investments.

[15] Note that one more unit of capacity in vintage t equipment allows for shutting down $A(t)/A(t - \tau(t))$ units of the oldest vintage still in use at date t, without altering total production. The cost savings is $cA(t)/A(t - \tau(t)) - c$. Similarly, by date s (with $s > t > s - \tau(s)$) the cost savings of moving production out of the oldest vintage, and into vintage t equipment, is $cA(t)/A(s - \tau(s)) - c$. The RHS of (9) is simply the present value of all such future cost savings made possible by an extra unit of investment in vintage t equipment.

We can derive the competitive investment condition (4) from the efficient investment condition (9) by substituting in the competitive shutdown condition (2). The advantage of (9), however, is that it also holds for other market structures.

A Monopolist

Suppose there is only one producer. This monopolist chooses a path of output $X(t)$, for $t \geq t_0$, so as to maximize the discounted value of industry revenue less the cost of production (including rental costs). The industry revenue (sales) function is

$$S(X;t) = XD(X;t). \tag{10}$$

The marginal revenue of an extra quality unit is

$$R(X;t) = S'(X;t) = XD'(X;t) + D(X;t).$$

To guarantee that the monopolist's problem is bounded and well behaved, we assume that the marginal revenue function is positive and strictly decreasing in X.[16] In what follows, it is convenient to denote marginal revenue as a function only of time:

$$M(t) = R(X(t);t).$$

The level of demand implied by any given level of marginal revenue is

$$X^{D_M}(M;t) = R^{-1}(M;t). \tag{11}$$

Rather than choosing output, we can, equivalently, think of the monopolist as choosing a path of marginal revenue per quality unit $M(t)$ for $t \geq t_0$. At each date, $M(t)$ must equal total unit cost per quality unit $C(t)$ from (8):

$$M(t) = C(t) = c/A(t - \tau(t)). \tag{12}$$

Given $M(t) = M$, this entails producing with vintages $v \in [A^{-1}(c/M), t]$. The maximum quantity X that the monopolist can supply using vintages $v \in (A^{-1}(c/M), t]$ is

$$X^{S_M}(M;t) = \int_{A^{-1}(c/M)}^{t} A(v)dK(v), \tag{13}$$

[16] A sufficient condition for a decreasing marginal revenue function is that $D(X, t)$ is weakly concave in X. Some convex demand curves, for example a constant elasticity demand curve with an elasticity of demand strictly greater than one, also yield marginal revenue decreasing in X.

where total supply adds in any production using the marginal vintage $v = A^{-1}(c/M)$. Consistency between supply and demand requires

$$M(t) = \max\{M : X^{S_M}(M;t) \leq X^{D_M}(M;t)\}. \tag{14}$$

The solution to the monopolist's problem is time paths, for all $s \geq t_0$, of marginal revenue $M^m(s)$, total capacity $K^m(s)$, and age of the oldest operating vintage $\tau^m(s)$, satisfying (9), (12) and (14), at all dates $t \geq t_0$. Given the values of $M^m(t)$ and $K^m(t)$, the equilibrium output $X^m(t)$ can be read off (11). The industry price is $P^m(t) = \mu^m(t)M^m(t)$, where $\mu^m(t) = \epsilon^m(t)/(\epsilon^m(t) - 1)$ is the monopoly markup ratio and $\epsilon^m(t) = P^m(t)/[D'(X^m(t);\, t)X^m(t)]$ is the elasticity of demand.[17]

A Dominant Firm

Now consider a dominant firm (a technological leader) facing a competitive fringe of technological laggards. The laggards make their output decisions, taking as given the output path chosen by the dominant firm. The dominant firm can commit to an entire output path, taking account of how the laggards will react. We can describe the equilibrium as if the dominant firm sets a price, since it effectively does so via its output choice.[18]

Suppose the laggards' technology (or productivity) is a fraction $0 < \rho \leq 1$ of the leader's. Vintage v equipment used by a laggard will produce chips of quality $A_L(v) = \rho A(v)$, while, using the same vintage of equipment, the dominant firm's chips will be of quality $A(v)$. In all other respects the dominant firm and the laggards are the same.

---

[17] The monopoly problem is parallel to the competitive equilibrium, with marginal revenue per quality unit $M(t)$, rather than price per quality unit $P(t)$, guiding production and investment decisions. Replacing $P(t)$ with $M(t)$ in the competitive shutdown condition yields

$$M(t)A(t - \tau(t)) = c.$$

This condition is equivalent to (8) after setting marginal revenue per quality unit $M(t)$ equal to total marginal cost per quality unit $C(t)$. Replacing $P(s)$ with $M(s)$ in the competitive investment condition (4) yields

$$q \geq \int_t^\infty e^{-r(s-t)} \max\{[A(t)M(s) - c], 0\}ds,$$

which holds with equality if there is any investment in vintage $v = t$ equipment. This condition is equivalent to (9) after substituting in (8) evaluated at $s = t$ and $C(t) = M(t)$.

[18] More sophisticated dynamic dominant firm models are analyzed in Kydland (1979) and Gowrisankaran and Holmes (2004). In the present application, the rental market in equipment makes the dynamic problem static from the point of view of individual firms. The outcome is identical to Bertrand competition in prices.

At date t the laggards can rent equipment of vintage v at the competitive rental price $q(t, v) = cA(v)/A(t - \tau(t)) - c$. This equipment produces chips of quality $\rho A(v)$ for them. Including the cost of running the equipment c, their total unit cost of production per quality unit at date t is $C_L(t) = c/[\rho A(t - \tau(t))]$, the same for any vintage $v \in [t - \tau(t), t]$. In order to dissuade entry, the dominant firm must choose $X(t)$ so that

$$D(X(t); t) \leq C_L(t).$$

Thus, we can think of the dominant firm as choosing a price path subject to the restriction that $P(t) \leq C_L(t)$.

The solution to the dominant firm's problem is time paths, for all $s \geq t_0$, of price $P^d(s)$, total capacity $K^d(s)$ and age of the oldest operating vintage $\tau^d(s)$, satisfying the efficient investment condition (9) as well as

$$\int_{t-\tau^d(t)}^{t} A(v)dK^d(v) \leq X^D(P^d(t); t) \leq \lim_{s \nearrow [t-\tau^d(t)]} \int_{s}^{t} A(v)dK^d(v)$$

and

$$P^d(t) = \min\left\{\frac{c}{\rho A(t - \tau^d(t))}, \frac{\varepsilon^d(t)c}{(\varepsilon^d(t) - 1)A(t - \tau^d(t))}\right\}, \quad \varepsilon^d(t) > 1,$$

or $P^d(t) = c/[\rho A(t - \tau^d(t))]$ if $\epsilon^d(t) \leq 1$, where $\epsilon^d(t) = P^d(t)/[D'(X^d(t); t)X^d(t)]$ and $X^d(t) = X^D(P^d(t); t)$, at all dates $t \geq t_0$.

By setting $\rho = 1$ we get back to a competitive industry. If demand is elastic, then for a small enough value of ρ, at each date $1/\rho > \varepsilon^d(t)/(\varepsilon^d(t) - 1)$, i.e., we get back to monopoly. In general we define the markup implied by the dominant firm model by $\mu^d(t) = \min\{1/\rho, \varepsilon^d(t)/(\varepsilon^d(t) - 1)\}$.

IV. Implications of Moore's Law

The essence of Moore's Law is that technology grows exponentially

$$A(t) = e^{gt},$$

where $g > 0$ is the rate of technological change. In particular, Moore's Law suggests that chip quality doubles every two years, i.e., $g = 0.35$. Given exponential technology growth combined with strictly positive investment, the solution exhibits some striking properties, common to all the market-structure scenarios. As we show below, strictly positive investment is guaranteed if we limit downward shifts in the demand curve. Other than that, our results carry through quite generally to different functional forms for demand.

Life of a Vintage

With positive investment we know that the efficient investment condition (9) must hold with equality. Plugging the exponential path of technology into (9) and imposing equality:

$$q = \int_t^\infty e^{-r(s-t)} \max\{[ce^{g(t-[s-\tau(s)])} - c], 0\} ds,$$

for all $t \geq t_0$. This equation has a very simple solution $\tau(t) = \tau$ for $t \geq t_0$. Plugging in this solution and applying the change of variable $x = s - t$ gives

$$q = \int_0^\tau e^{-rx} c[e^{g(\tau - x)} - 1] dx. \tag{15}$$

Solving this integral, τ is the solution to

$$\frac{q}{c} = \frac{e^{g\tau} - e^{-r\tau}}{r+g} - \frac{1 - e^{-r\tau}}{r}. \tag{16}$$

The value of τ is not only constant over time but is also independent of market structure. The life of equipment is increasing in the cost of financing equipment purchases r and in the up-front purchase price relative to the unit variable cost of operating it, q/c. Equipment life declines in the rate of technological change, g; it is optimal to replace equipment sooner if better equipment becomes available more rapidly.[19]

Prices

We can summarize our results about the price per quality unit under different market structures by

$$P(t) = \mu(t)C(t),$$

where $C(t)$ is the total unit cost of production per quality unit, given by equation (8). The markup is given by either: $\mu(t) = 1$ under perfect competition, $\mu(t) = \mu^m(t)$ under monopoly, or $\mu(t) = \mu^d(t)$ under the dominant firm model.

[19] It is easiest to see these results from equation (15) after dividing both sides by c. Denote

$$f(z; r, g) = \int_0^z e^{-rx}[e^{g(z-x)} - 1]dx$$

so that $f(z; r, g) = q/c$ at $z = \tau$. By inspection, f is a monotonically increasing function of z, with $f(0; r, g) = 0$ and rising without bound as z gets arbitrarily large. Thus, given r, g and q/c, there is a unique solution $\tau > 0$, increasing in q/c. Since, for given z, the function $f(z; r, g)$ shifts down with an increase in r and shifts up with an increase in g, the solution τ is increasing in r and decreasing in g.

Applying exponential technological change and our result on the constancy of τ, the marginal cost equation (8) gives

$$C(t) = be^{-gt}, \tag{17}$$

where b is a constant given by

$$b = ce^{g\tau}. \tag{18}$$

Thus, for a given markup, prices fall at rate g, with b determining the location of the price path. To get more intuition for b, from (7) note that the rental price of equipment of vintage v at date t is

$$q(t, v) = be^{-g(t-v)} - c,$$

for $v \geq t - \tau$. The rental price of new equipment is $b - c$, but it declines as the equipment ages. The rental price hits zero for equipment of age τ.

A central question in our analysis is how prices are influenced by the speed of Moore's Law, g. We cannot say much in general about the markup $\mu(t)$ since it depends on the market structure and properties of the demand curve. We can, however, say something about how the total unit cost of production is affected by a change in g. Of course, a higher value of g means prices fall faster. But, we can also show that a higher g will be associated with a higher b.[20] As we showed above, with faster technological change, each vintage of equipment will have a shorter life, τ. To recoup the sunk cost of purchasing new equipment over a shorter horizon, the rental price of equipment must start at a higher level.

In summary, the implication for the price path is

$$P(t) = ce^{g\tau}\mu(t)e^{-gt}. \tag{19}$$

To verify our proposed solution we need to show that there is a strictly increasing path of capacity $K(t)$ that equates supply and demand at these prices,

$$\int_{t-\tau}^{t} e^{gv} dK(v) \leq X^D(\mu(t)be^{-gt}; t) \leq \lim_{s \nearrow [t-\tau]} \int_{s}^{t} e^{gv} dK(v).$$

[20] To prove this result, rewrite (15), using (18), as

$$f(z; \tau, g) = \int_0^{\tau} e^{-rx}[ze^{-gx} - c]dx,$$

so that $f(z;\ \tau,\ g) = q$ at $z = b$. By inspection, f is a monotonically increasing function of z. Suppose the rate of technological change rises of $g' > g$. We showed above that the resulting life of equipment is $\tau' < \tau$. With faster technological change and shorter equipment lives we get $f(\tau',\ b,\ g') < q$. Thus $f(\tau',\ z,\ g') = q$ at $z = b' > b$.

A sufficient condition is that the resulting path of output $X(t)$ is non-decreasing.[21]

Investment and Output

Thus far, we have said little about the implication of the model for investment in the industry, other than to require that investment be positive. In general, the implications for investment can be quite intricate. For example, a positive demand shock may lead to a spike in investment. But, when that spike is retired τ years later, there will be another spike to replace it (with the latter spike diminished in size due to technological change). Thus, in principle, the path of investment reflects current shocks to demand along with echoes of past shocks.

Here, we analyze the behavior of investment and output in a much more restrictive setting. In particular, we now assume that the demand curve features a constant price elasticity $\varepsilon > 1$ and a secular trend λ that may be positive or negative (but not too negative as we show below):

$$P(t) = D(X(t); t) = e^{\lambda t}[X(t)/a]^{-1/\varepsilon},$$

or, equivalently,

$$X^D(P(t); t) = ae^{\lambda\varepsilon t}P(t)^{-\varepsilon}, \tag{20}$$

where a determines the scale of the industry. This assumption on demand implies a constant markup μ: (i) $\mu = 1$ under perfect competition, (ii) $\mu = \mu^m = \varepsilon/(\varepsilon - 1)$ under monopoly, and (iii) $\mu = \mu^d = \min\{1/\rho, \varepsilon/(\varepsilon - 1)\}$ under the dominant firm model. We simply treat μ as a parameter reflecting market structure.

Furthermore, we set the initial conditions on capacity so as to remove any initial spike in investment. We conjecture, and then verify, that in this setting the path of new investment is $\dot{K}(t) = ke^{ht}$, where $k > 0$ and h are constants to be determined. Given the conjectured investment path, the quantity of output supplied is

$$X^S = \int_{t-\tau}^{t} e^{gv}\dot{K}(v)dv = \int_{t-\tau}^{t} e^{gv}ke^{hv}dv = \frac{k}{g+h}e^{(g+h)t}(1 - e^{-(g+h)\tau}).$$

[21] In terms of the demand curve, that outcome is guaranteed if, for all P, $X^D(P; t') \geq X^D(P; t)$ for all $t' \geq t$. The content of the restriction is to rule out large negative shocks to demand. Our proposed equilibrium would be broken by demand shocks so negative that vintages less than τ years old are taken out of production.

Setting $P(t) = \mu b e^{-gt}$, equating supply and demand, and rearranging:

$$\mu b e^{-gt} = e^{\lambda t} e^{\frac{-(g+h)t}{\varepsilon}} \left[\frac{k/a}{g+h} (1 - e^{-(g+h)\tau}) \right]^{-\frac{1}{\varepsilon}}.$$

Equating the growth rates on both sides of the equation above we get $h = \varepsilon(\lambda + g) - g$. Equating the multiplicative factors on both sides,

$$k = a\varepsilon(\lambda + g)(\mu c)^{-\varepsilon} e^{-\varepsilon g \tau} / (1 - e^{-\varepsilon(\lambda+g)\tau}).$$

For the investment path to be positive we need $k > 0$ and hence $\varepsilon(\lambda + g) > 0$. Our conjecture is thus verified under the restriction that demand does not contract too fast, i.e., $\lambda > -g$. Note, however, that this restriction does not rule out the possibility of a declining investment path.

For $t \geq t_0$, the equilibrium path of investment is

$$\dot{K}(t) = \frac{a\varepsilon(\lambda + g)(\mu c e^{g\tau})^{-\varepsilon}}{1 - e^{-\varepsilon(\lambda+g)\tau}} e^{[\varepsilon(\lambda+g)-g]t},$$

and the path of industry output is

$$X(t) = a(\mu c e^{g\tau})^{-\varepsilon} e^{\varepsilon(\lambda+g)t}.$$

The initial conditions on capacity, to guarantee no initial spike in investment, are

$$\dot{K}(v) = \frac{a\varepsilon(\lambda + g)(\mu c e^{g\tau})^{-\varepsilon}}{1 - e^{-\varepsilon(\lambda+g)\tau}} e^{[\varepsilon(\lambda+g)-g]v},$$

for $v \in [t_0 - \tau, t_0)$.[22]

While output and investment are trending over time, the ratio of investment expenditure to industry revenue is constant:

$$\frac{q\dot{K}(t)}{P(t)X(t)} = \frac{q\varepsilon(\lambda + g)}{\mu c e^{g\tau}(1 - e^{-\varepsilon(\lambda+g)\tau})}. \tag{21}$$

V. Microprocessor Prices

The sharpest and most robust implications of the model relate to the price contours of individual products, shown in Figure 2. Here we look at how the model fares in explaining these patterns. We then consider what the model has to say about changes that would result from a speed-up in Moore's Law. Except for such a one-time unexpected change, we continue to impose the restriction of Moore's Law.

[22] The level of capacity $K(v)$ in vintages prior to $v = t_0 - \tau$ is irrelevant.

Price Contours

We have data on the prices at various dates t of particular MPUs, defined by their introduction dates. Using the introduction date, we associate a chip with some vintage $v \leq t$. In the model and in the data, we denote by $p(t, v)$ the price at date t of the chip introduced at date v.

We denote the associated total unit cost of production by $c(t, v) = C(t)A(v) = q(t, v) + c$. Holding fixed any vintage v, $c(t, v)$ declines at rate g with t. Vintage v drops out of the market at date $t = v + \tau$ at which point $c(v + \tau, v) = c$. The total unit cost of producing any vintage of chip when it first enters the market is $c(v, v) = b$. To summarize, the total unit cost of production declines at rate g from the level b to the level c during the life of vintage v, which runs from date v to $v + \tau$.[23]

We can take this prediction about the total unit cost of production to derive implications for prices of individual chips, $p(t, v) = \mu(t)c(t, v)$. Thus:

$$p(t, v) = \mu(t)be^{-g(t-v)},$$

for $0 \leq t - v \leq \tau$. On a logarithmic scale, price contours will decline in parallel, with typical slope g, although the slope may vary if the markup changes over time due to shifts in demand or movements along a demand curve that is not constant elasticity. At any given date, the more recent vintages (larger v) sell for more. But holding fixed the age of the vintage, $t - v$, the price varies only due to variation in $\mu(t)$. For expositional purposes, in what follows we consider the case of a fixed markup $\mu(t) = \mu$.

What would look different if the rate of technological change were $g' > g$? First, price contours would be steeper. Second, and more surprisingly, the price of a chip when it is introduced would be higher. Thus, more rapid technological change leads to faster price declines, but from a higher level. Products drop out of the market sooner as well.

Transition Dynamics

What happens if there is a permanent unexpected increase in the rate of technological change from g to g' at date t_0? Suppose that the price per quality unit at t_0 is that determined in the stationary configuration. To simplify the discussion we focus on the case of a fixed μ and no shocks to

[23] It is typically assumed that price declines of individual chips reflect falling production costs, say, due to learning by doing. Here, unit production costs do not fall; they are fixed at the value c. For a chip producer renting his equipment, the rental price for his equipment falls over time so that the total unit cost of production falls. For a chip producer owning his equipment, markups over unit variable cost c fall over the life of a product. The markup of price over c (in excess of μ) goes to finance the up-front cost of the equipment. This markup fades away over time until it hits the value μ when the product is dropped.

demand, $D(X; t) = D(X)$. Thus, the price level is $P(t_0) = \mu b e^{-gt_0}$, with $b = e^{g\tau}$, at the date when Moore's Law speeds up.

From our earlier discussion, we know that if the industry were to continue investing in new equipment, the price level would have to jump to $\mu b' e^{-gt_0}$ with $b' = e^{g'\tau} > b$. Such a jump in price is inconsistent with market clearing. Market clearing demands that there be no change in price for a period of time until date t_1, where $b'/b = e^{g'(t_1 - t_0)}$. Starting at date t_1 the industry falls back into the Moore's-Law solution with prices now falling at rate g'. Thus $P(t) = \mu e^{g(\tau - t_0)}$ for $t \in [t_0, t_1]$ and $P(t) = \mu e^{g(\tau - t_0)} e^{-g(t - t_1)}$ for $t \geq t_1$.

In the interval of time from date t_0 to date t_1 industry investment falls to zero. The reason is that during this period, the price of chips is not high enough to compensate investors for the higher rate of depreciation they will experience with technological change occurring at rate g'. Since there is no investment during this time interval, each chip on the market at date t_0 remains on the market at least through date t_1 and the price of each chip remains constant. After date t_1 all chip prices begin to decline at rate g' and older vintages begin to drop out of the market again. In effect, producers wait to invest until the technology of the new equipment improves enough to allow them to recoup their investments.

Numerical Illustrations

We now turn to a quantitative assessment of the model's ability to match the price contours shown in Figure 2 and to shed light on the consequences of a speed-up in Moore's Law.

Table 1 quantifies the basic facts about these price contours, shown as averages across the products within given chip families (such as the Pentium I family). Chips sold over the 1990s typically entered the market at a price of \$500–600, with the price falling to the \$100–150 range by the time each chip exited the market. Over the decade, the amount of time that chips were sold on the market fell from about three years for chips introduced in the early 1990s to about one year for chips introduced at the end of the decade. The rate of price decline over chips' life spans was rapid early in the decade (almost 80 percent for the Pentium I chips) and even more rapid by the end of the decade (over 100 percent for the Pentium III chips). Growth in the number of transistors per chip was around 20–25 percent early in the decade and doubled with the introduction of the Pentium I chip in 1997; see Table 2. Similarly, growth in chip speed (another indicator of the growth of chip quality) averaged between 20 and 30 percent and doubled with the introduction of the Pentium III chip. That these indicators of quality growth are typically far slower than the staggering price declines, presents a challenge to our model, as discussed below.

Our base-case calibration uses various sources of information, independent of the price data discussed above, to pin down values for the parameters: r, q, c and g. With these four parameters, we can make predictions

Table 1. *Intel's desktop chip families, 1985–1999*

	Desktop chip families				
	386	486	Pentium I	Pentium II	Pentium III
1. Introduction date	1985Q4	1989Q2	1993Q1	1997Q2	1999Q1
2. Chips in family	10	15	13	7	15
	averages over all chips in chip family				
Prices (dollars)					
3. Introduction	268	656	516	587	506
4. Exit	90	102	115	146	135
5. Lifespan (years)	3.0	2.8	1.9	1.4	1.2
	annual rates				
6. Price change over lifespan (percent)	−36	−66	−79	−99	−110
Growth between family introductions (percent):					
7. Number of transistors	42	25	21	53	45
8. Speed of chip	21	18	36	29	63

Notes: Introduction dates and the number of chips for each chip family (lines 1–2) are reported on Intel's website at www.intel.com/pressroom/kits/quickreffam.htm. Average prices and lifespans for chips in each family (lines 3–5) were calculated using chip-level data from Dataquest, Inc. (1985–1993) and from MicroDesign Resources (1993–1999). Price change over chips' lifespans (line 6) was calculated using $\ln\{P(t+s)/P(t)\}/s$, where the P's denote the prices given in lines 3 and 4 and the s denotes chips' lifespans given in line 5. Growth rates in the quality variables between family introductions (lines 7–8) measure the change in the quality of the cutting-edge chips in the quarter when the chip families were introduced, where the "cutting-edge chip" is the chip that either contains the most transistors (for line 7) or is the fastest (for line 8) of all the chips on the market in the quarter the chip family made its debut. For example, the growth in the speed of the Pentium I chips was calculated as the log difference in the speed of the first Pentium I and Pentium II chips divided by the number of quarters between the introduction of the two chips and expressed as an annual rate (multiplied by 4).

Table 2. *Selected attributes for Intel's desktop microprocessors and price indexes for semiconductor devices, 1985–1999 (growth rates at annual rates)*

	1985–1994	1995–1999
Intel's microprocessors		
1. Number of transistors	28	53
2. Speed (megahertz)	20	43
Price indexes		
3. Semiconductor devices	−15	−49
4. Microprocessors	−31	−96

Sources: Authors' calculations using data on the characteristics of Intel's chips from www.intel.com/pressroom/kits/quickreffam.htm, price indexes for semiconductor devices underlying Oliner and Sichel (2000), and price indexes for microprocessors from Grimm (1998)—for 1985–1994—and Aizcorbe (2002)—for 1995–1999.

about cost contours $c(t, v)$, i.e., the total unit cost of producing a chip at date t that first entered the market at date v. We do not need to take a stand on the parameters of demand, although we must assume that negative shocks to demand never lead to a stall in investment.

The parameter values for our base case are summarized in the top line of the left side of Table 3. We set the real interest rate to $r = 0.07$, a standard figure for the real return on equity. We base our value of the setup cost, q, on estimates of investment in plant and equipment and on annual unit capacity, obtained from MicroDesign Resources Inc. (MDR). These estimates imply a setup cost of somewhere between \$50 and \$90 per unit of annual capacity; using the midpoint of that range, we set $q = 70$.[24] For the unit cost c (which in our model includes the cost of running the equipment but not renting it or purchasing it), we used MDR data on the average cost per chip (ranging from \$50 to \$100 over 1993–1997) with an adjustment to exclude the depreciation for plant and equipment that is included in their estimates. Using the midpoint cost of this range and subtracting \$30, our estimate of the value of amortized plant and equipment investment spread over all chips, yields our base-case parameter for unit production cost: $c = 45$.[25] We also need a number for the rate of technological change. Moore's Law is typically approximated as a doubling in quality every two years, implying $g = 0.35$.

[24] We used data and anecdotal evidence on production of chips using the 0.25 micron process to obtain this estimate. There were four fabs that, we assume, incurred about \$1–2 billion per fab for the retooling to run this process. Our estimate of q divides this setup cost by MDR's estimate of a combined annual capacity of 85 million chips at the four fabs.

[25] To adjust the reported cost data, we divide the \$4–8 billion setup costs for the four fabs by the total number of chips they produced using this process (189 million) and subtract the resulting \$30 per chip from the MDR estimate. An alternative method gave a similar estimate: using data from Intel's financial statements to mimic the four-year, straight-line depreciation that MDR uses to amortize capital costs and dividing those costs by the total number of chips that Intel produced yields an estimate of a little over \$30 depreciation per chip.

Table 3. *Sensitivity of model to underlying parameter values*

Parameters				Implications	
r	q	c	g	τ	b
			Baseline case		
0.07	70	45	0.35	2.6	112
			Discount rate		
0.02	70	45	0.35	2.5	108
0.07	70	45	0.35	2.6	112
0.12	70	45	0.35	2.7	116
			Setup cost		
0.07	50	45	0.35	2.3	101
0.07	70	45	0.35	2.6	112
0.07	90	45	0.35	2.9	124
			Unit variable cost		
0.07	70	25	0.35	3.3	79
0.07	70	45	0.35	2.6	112
0.07	70	75	0.35	2.1	156
			Moore's Law		
0.07	70	45	0.25	3.2	100
0.07	70	45	0.35	2.6	112
0.07	70	45	0.70	1.8	159

Note: The four parameters—the Moore's Law parameter (g), the discount rate (r), setup costs (q) and unit variable cost (c)—imply chips' lifespans (τ) and the total unit cost of production at introduction (b).

Given these values for r, q, c and g, (16) delivers the product life of a chip and, given τ, (18) delivers the initial total unit cost of producing a chip. Table 3 displays the results under various scenarios. For the base case we obtain $\tau = 2.6$ years and $b = 112$ dollars. Our implied value for the life of a vintage is close to the data, at least for the early 1990s. The value for b means that the total unit cost of production is \$112 when a chip is introduced and falls to \$45 when it is dropped 2.6 years later. The remaining rows in Table 3 explore the sensitivity of the model to changes in the underlying parameters. The most notable perturbation (the last row) is a doubling in the coefficient of Moore's Law, which reduces the lifetime of a chip by nearly a year, while raising the initial total unit cost of production by close to \$50.

To obtain predictions about the price contours $p(t, v)$ we need a value of μ. Based on the results for cost contours in Table 3, it is clear that this markup must be substantially greater than one, the value in the competitive baseline, if the model is to come close in its price predictions. Only the monopoly and dominant-firm versions of the model stand a chance to justify the observed price levels. To obtain a simple formula for the markup, we assume a constant-elasticity demand curve (as in Section IV). We take the demand elasticity to be $\varepsilon = 1.5$, as reported in Flamm (1997). Although this estimate applies to the entire semiconductor industry, we use it for the MPU sector since it is the only one available. Under monopoly, the implied markup is $\mu = \varepsilon/(\varepsilon - 1) = 3$. For the dominant-firm model we need to quantify Intel's technological lead over Advanced Micro Devices Inc. (AMD). This lead was thought to be about two years as of 1995, which, with $g = 0.35$, means $A_L(v) = A(v)/2$, or $\rho = \frac{1}{2}$.[26] The implied markup is $\mu = \min\{1/\rho, \varepsilon/(\varepsilon - 1)\} = 2$. Thus we consider markups of 2 and 3.

Table 4 shows the resulting implications for price contours under various assumptions on the other parameters. The introduction prices implied by the

Table 4. *Markups and sensitivity of implied prices to changes in underlying parameters*

				Implied prices	
r	q	c	g	Introduction	Exit
			Markup: $\mu = 2$		
			Setup costs		
0.07	50	45	0.35	202	90
0.07	70	45	0.35	224	90
0.07	90	45	0.35	248	90
			Unit variable cost		
0.07	70	25	0.35	158	50
0.07	70	45	0.35	224	90
0.07	70	75	0.35	312	150
			Moore's Law		
0.07	70	45	0.25	200	90
0.07	70	45	0.35	224	90
0.07	70	45	0.70	318	90
			Markup: $\mu = 3$		
			Setup cost		
0.07	50	45	0.35	303	135
0.07	70	45	0.35	336	135
0.07	90	45	0.35	372	135
			Unit variable cost		
0.07	70	25	0.35	237	75
0.07	70	45	0.35	336	135
0.07	70	75	0.35	468	225
			Moore's Law		
0.07	70	45	0.25	300	135
0.07	70	45	0.35	336	135
0.07	70	45	0.70	477	135

[26] McKinsey Global Institute (2001) came up with the two-year figure by comparing introduction dates for chips that were roughly comparable and that were produced by both Intel and AMD. They also reported that Intel's lead eroded over the late 1990s. With technological change itself quickening after 1995, it is not clear that AMD was catching up in terms of "technological distance", the relevant metric for ρ.

model are typically lower than the average introduction prices seen in the data, but the exit prices fit the data fairly well. Given that the implied life spans in the model are fairly close to those in the data, the problem lies in the fact that the model implies slower price declines than those seen in the data. Although a higher markup helps the model match introduction prices, it also raises the exit price and, thus, does not lead to a better match with the data. The model fares best in the last row of Table 4, based on the higher markup of $\mu = 3$ along with faster technological change of $g = 0.70$. It is difficult, however, to justify such rapid technological change except in the late 1990s.[27]

Suppose the coefficient of Moore's Law doubled during the 1990s to $g' = 0.70$. As seen by comparing the third and fourth columns of Table 5, such a speed-up implies higher introduction prices (477 vs. 336 dollars), shorter life spans (1.8 vs. 2.6 years), and faster price declines over the life of each chip (70 vs. 35 percent). The latter two predictions are consistent with what is seen in the data following 1995; contours became steeper and chips' market lives became shorter. The prediction that introduction prices increase is not borne out in the data, as these prices actually edged down beginning in 1995.

In the model, prices would not immediately begin falling faster following a speed-up in Moore's Law. Figure 3 illustrates the results of a permanent but unexpected change from $g = 0.35$ to $g' = 0.70$ at date t_0. With the price level at date t_0 continuing to equate supply following the speed-up,

Table 5. *Illustration of change in Moore's Law* ($\mu = 3$)

	Pentium chips		Simulations	
	I 1993	III 1999	Base case $g = 0.35$	Speed-up $g = 0.70$
Prices (dollars)				
Introduction	516	506	336	477
Exit	115	135	135	135
Lifespan (years)	1.9	1.2	2.6	1.8
Price change over lifespan (percent)	−79	−110	−35	−70

Notes: Prices and lifespan information are as reported in Table 2 (for the Pentium chips) and Table 4 (for the simulations). The price change over each chip's lifespan was calculated by applying the formula $\ln\{P(t + s)/P(t)\}/s$, where the P's and s are the prices and chip's lifespan reported in the table.

[27] One possibility is that equating quality improvement with Moore's Law is too simplistic. From a consumer's point of view, a doubling of speed might make an MPU more than twice as valuable, an idea suggested by Luttmer (2004). In that case we might be able to justify a larger value of g.

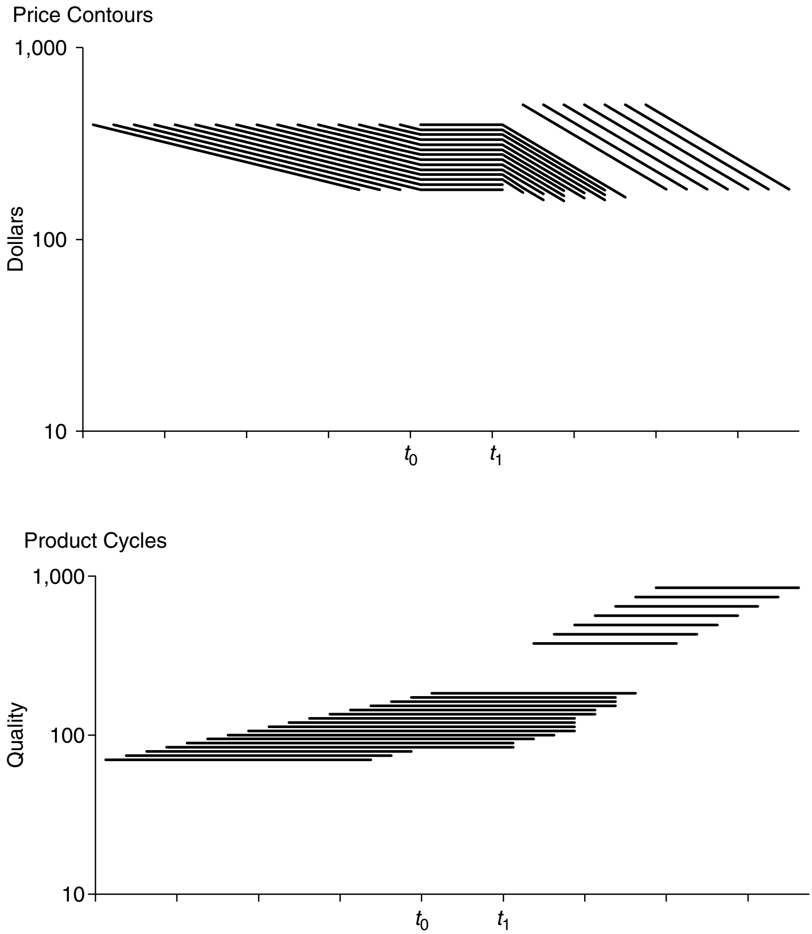

Fig. 3. Price contours, product cycles and quality improvement with a speed-up in Moore's Law

investment stalls and no new chips are introduced for a while. Absent entry, each existing chip's price remains at its t_0 level until date t_1. At t_1 the quality of new equipment is sufficiently better, whereby equipment purchases are again justified and investment resumes. Beyond that date, prices for all chips fall at the faster rate, $g' = 0.70$. Taking t_0 to be 1995, the actual price contours, graphed in Figure 2, do not follow such a stark pattern.

We conclude this section with an evaluation of the model's implications for investment in the MPU sector. To do so requires that we impose the constant-elasticity demand curve (20) which implies industry sales growth of $\dot{S}/S = \lambda\varepsilon + g(\varepsilon - 1)$. Based on MDR estimates, we calculate that Intel's revenue from chips grew 20 percent per year over the period 1993–2000. Given $\varepsilon = 1.5$ and $g = 0.35$, we infer $\lambda = 0.017$. Given all these parameter values (including the baseline parameters in Table 3), the implied investment share from (20) is 15 percent if $\mu = 3$ and 23 percent if $\mu = 2$. Thus, we straddle the 19–22 percent range reported in Intel's financial statements on the value of additions to machine and structures relative to total revenues.

VI. Conclusion

In this paper we develop a theoretical model to better understand the behavior of the microprocessor sector, an important segment within the semiconductor industry. Despite its simplicity, the model captures two basic features of the MPU market.

First, the vintage-capital feature of the model accommodates the fact that MPUs of different qualities coexist in the market. Producers make product-specific irreversible investments in equipment each time they introduce a new chip. Once these sunk costs have been borne, it makes sense to keep producing a chip, even after better versions have become available, as long as the price exceeds the variable cost of production.

Second, the model predicts that prices fall over the life of each chip, reflecting declines in markups (of price over unit variable costs) for existing chips when new and superior products are introduced. Downward-sloping price contours have elsewhere been explained either by falling costs arising from learning economies or falling markups arising from market-power considerations. In our model, there is no learning by doing and, although we can accommodate market power, the prediction of declining prices holds even under perfect competition. In that case the wedge between price and the unit variable cost of production reflects the implicit rental price of vintage-specific equipment, which starts out high but eventually falls to zero when the vintage of the chip it is used to produce drops out of the market.

The most basic shortcoming of the present model is its prediction that the rate of price decline over the life of a chip should equal the coefficient of Moore's Law. Prices of individual chips consistently fall faster than the rate at which technology advances as measured by growth of transistors per chip or chip speed. What accounts for this deviation between the model and the data? Our current thinking is that the problem stems from our extreme simplification of the demand side. Luttmer (2004) shows how we might introduce different types of consumers buying chips at different ends of the quality spectrum. In such a model it is possible that high-end chips would command higher markups, thus leading to an additional force for rapid declines in chip prices as chips age. Even if such an extension is required before the model's quantitative implications can be taken too seriously, we think the present model serves as a good theoretical benchmark.

References

Aizcorbe, A. (2005), Why Are Semiconductor Prices Falling So Fast? A Decomposition, Economic Inquiry, forthcoming.

Aizcorbe, A. (2004), Product Introductions and Productivity at Intel in the 1990s, mimeo, Bureau of Economic Analysis, Washington, DC.

Baldwin, R. E. and Krugman, P. R. (1988), Market Access and International Competition: A Simulation Study of 16K Random Access Memories, in R. Feenstra (ed.), *Empirical Methods for International Trade*, MIT Press, Cambridge, MA.

Congressional Budget Office (2002), The Role of Computer Technology in the Growth of Productivity, CBO paper, Washington, DC.

Flamm, K. (1996), *Mismanaged Trade? Strategic Policy and the Semiconductor Industry*, Brookings Institution Press, Washington, DC.

Flamm, K. (1997), More for Less: The Economic Impact of Semiconductors, Semiconduction Industry Associaton.

Gowrisankaran, G. and Holmes, T. (2004), Mergers and the Evolution of Industry Concentration: Results from the Dominant-firm Model, *RAND Journal of Economics 35* (3), 561–582.

Grimm, B. T. (1998), Price Indexes for Selected Semiconductors, 1974–96, *Survey of Current Business 78*, 8–24.

Hatch, N. and Mowery, D. C. (1998), Process Innovation and Learning by Doing in Semiconductor Manufacturing, *Management Science 44* (11), 1461–1477.

Hobijn, B. (2002), On Both Sides of the Quality Bias in Price Indexes, *Staff Reports* no. 157, Federal Reserve Bank of New York.

Irwin, D. A. and Klenow, P. J. (1994), Learning by Doing Spillovers in the Semiconductor Industry, *Journal of Political Economy 102* (6), 1200–1227.

Jorgenson, D. W. (2001), Information Technology and the U.S. Economy, *American Economic Review 91* (1), 1–32.

Jorgenson, D. W. and Stiroh, K. J. (2000), U.S. Economic Growth in the New Millennium, *Brookings Papers on Economic Activity 1*, 125–211.

Jorgenson, D. W., Ho, M. S. and Stiroh, K. J. (2002), Projecting Productivity Growth: Lessons from the U.S. Growth Resurgence, *Economic Review*, Federal Reserve Bank of Atlanta, 1–13.

Jovanovic, B. and Lach, S. (1989), Entry, Exit, and Diffusion with Learning by Doing, *American Economic Review 79* (4), 690–699.

Kydland, F. (1979), A Dynamic Dominant Firm Model of Industry Structure, *Scandinavian Journal of Economics 81* (3), 355–366.

Lucas, R. E. and Prescott, E. C. (1971), Investment under Uncertainty, *Econometrica 39* (5), 659–681.

Luttmer, E. G. J. (2004), Measuring Multi-sector Balanced Growth, mimeo, University of Minnesota, Minneapolis, MN.

McKinsey Global Institute (2001), *US Productivity Growth 1995–2000: Understanding the Contribution of Information Technology Relative to Other Factors*, McKinsey & Company, Washington, DC.

MicroDesign Resources, *Intel Microprocessor Forecast*, various years, MicroDesign Resources, Inc., Sebastopol, CA.

Mitchell, M. F. (2002), Technological Change and the Scale of Production, *Review of Economic Dynamics 5*, 477–488.

Moene, K. O. and Wallerstein, M. (1997), Pay Inequality, *Journal of Labor Economics 15* (3), 403–430.

Moore, G. E. (1965), Cramming More Components onto Integrated Circuits, *Electronics 38* (8).

Oliner, S. D. and Sichel, D. E. (2000), The Resurgence of Growth in the Late 1990s: Is Information Technology the Story?, *Journal of Economic Perspectives 14*, 3–22.

Oliner, S. D. and Sichel, D. E. (2002), Information Technology and Productivity: Where Are We Now and Where Are We Going?, *Economic Review*, Federal Reserve Bank of Atlanta, 15–44.

Salter, W. E. G. (1960), *Productivity and Technical Change*, Cambridge University Press, Cambridge.

Semiconductor Industry Association (2003), *International Technology Roadmap for Semiconductors*, San Jose, CA (update 2005), http://www.public.itrs.net/.

Siebert, R. (2003), Learning by Doing and Multiproduction Effects Over the Life Cycle: Evidence from the Semiconductor Industry, CEPR Discussion Paper no. 3734.

Solow, R. M., Tobin, J., von Weizsacker, C. C. and Yaari, M. (1966), Neoclassical Growth with Fixed Factor Proportions, *Review of Economic Studies 33* (2), 79–115.

Song, Minjae (2004), Measuring Consumer Welfare in the CPU Market: An Application of the Pure Characteristics Demand Model, mimeo, Georgia Institutes of Technology, Atlanta, GA.