# Excessive Entry and Social Inefficiencies: A Policy Experiment on Dynamic Efficiency Gains*

An-Hsiang Liu† Ralph Siebert‡ Christine Zulehner§

October, 2013

**Abstract**

This paper evaluates how different lengths of entry protection impact market structure and market performance. We formulate a dynamic oligopoly model in the tradition of Ericson and Pakes (1995) and allow entry costs to vary over time. Firms decide when to enter a market, followed by production and exit decisions. Using a detailed dataset on quarterly firm-level data on the static random access memory industry from 1974 to 2003, we find that entry costs decline by more than 90% within the first three years. We perform a policy experiment in which a social planner can control the protection length of the first entrant. Our policy experiment assesses to what extent "excessive entry" causes social inefficiencies. We especially focus on dynamic economies of scale, time variant entry costs and dynamic efficiency gains for assessing the impact on consumer and producer welfare. Our policy experiments provide evidence that the duration of entry protection has a negative impact on consumer surplus. We also find that entry protection increases social welfare if the protection duration is either sufficiently short or sufficiently long. If entry protection duration is short, the increase in monopolist's profits and entry cost saving dominate the reduction in consumer welfare, which affects total welfare positively. If the protection period is long, dynamic efficiency gains, i.e., the delay of subsequent entry and savings on entry costs impact total welfare positively.

JEL: C1, L1, L6, O3.

Keywords: Entry Costs, Dynamic Efficiency Gains, Dynamic Oligopoly, Entry Protection, Market Entry, Market Structure, Policy Experiment, Semiconductor Industry, Social Planner.

*We thank seminar participants at Purdue University, Toronto University, IUPUI and the University of Mannheim for their useful comments. We especially thank Victor Aguirregabiria, Jack Barron, Lanier Benkard, Alberto Galasso, Gautam Gowrisankaran, David Hummels, Stephen Martin, Stephen Ryan, Philipp Schmidt-Dengler, Matthew Mitchell, Carlos Serrano, Anson Soderbery, Yuya Takahashi, Justin Tobias, Mihkel Tombak and Chong Xiang for their helpful discussions and valuable suggestions on this or an earlier draft. We gratefully acknowledge financial support from the Jubiläumsfonds of Oesterreichische Nationalbank (grant 11038). All errors are our own.

†Wuhan University, Department of Economics, Economics and Management School, Address: Wuhan University Luojiashan, Wuhan, China, Email: anhliu0512@126.com.

‡Purdue University, Department of Economics, Krannert School of Management, Address: 403 West State Street, West Lafayette, IN 47907-2056, United States, Email: rsiebert@purdue.edu.

§Johannes Kepler University Linz and Austrian Institute of Economic Research (Wifo), Address: Altenberger Strasse 69, A-4040 Linz, Austria, Email: christine.zulehner@jku.at.

# 1 Introduction

Many policy makers and economists argue that market entry stimulates competition and economic growth. Prominent studies have shown that market entry increases competition and lowers prices.[1] A large number of studies also noted the possibility of finding socially inefficient levels of market entry. Unrestricted or free entry can result in an "excessive" number of firms entering a market, i.e., the number of firms entering a market is above the socially desirable number of firms.[2] This especially applies to industries characterized by high fixed or entry costs, in which incumbent firms have the opportunity to benefit from declining average costs, i.e., economies of scale. New entrants impose a negative externality and steal production from incumbents (business stealing effect) which increases average costs and causes social inefficiencies. The ultimate impact on social welfare depends on whether the resulting price reduction from increased competition dominates the reduction in average costs resulting from economies of scale when only a single incumbent serves the market.

Another source of social inefficiencies relates to the fact that new entrants invest in entry costs which have been already incurred by incumbent firms, also referred to as dynamic inefficiencies or wasteful duplication in entry costs, see e.g., Pesendorfer (2003). Dynamic inefficiencies are particularly pronounced when entry costs decline over time, which is the case in several industries as industry experts and managers frequently state. The reason is that cost of capital highly depreciates over time since production machinery becomes obsolete and more inexpensive over time.[3] Since early entry is more desirable to the entrant than it is to society, new entrants enter too early and incur an associated entry cost that is too high from a social welfare point of view. Therefore, beyond "excessive entry" in terms of the number of firms operating in a market, a further welfare concern addresses the question if firms enter "too early". A social planner is aware of multiple firms incurring high entry costs and would delay subsequent entry to avoid dynamic inefficiencies. While firms only consider their private benefits from entering and ignore the social losses resulting from business stealing and dynamic inefficiencies, a social

[1] See e.g., Carlton (1983), Dunne, Roberts and Samuelson (1988), Geroski (1989), Bresnahan and Reiss (1990, 1991), Berry (1992), Toivanen and Waterson (2000 and 2005), Klepper (2002), Mazzeo (2002), Campbell and Hopenhayn (2005), Davis (2006), Foster, Haltiwanger and Krizan (2006) and Seim (2006).

[2] See e.g., Chamberlin (1933), Spence (1976a, 1976b), Dixit and Stiglitz (1977), von Weizsacker (1980), Perry (1984), Mankiw and Whinston (1986), and Sutton (1991).

[3] For example, a new semiconductor manufacturing plant is equipped with production machinery that can grow obsolescent in 3 years. More detailed information on declining entry costs over time is provided in the next section.

planner would incorporate those externalities.

The aim of our study is to assess the social inefficiencies caused by "excessive entry" and address the following question: How severe are the social inefficiencies caused by excessive entry, and what is the impact on social welfare?

We perform a policy experiment in which a social planner can protect the first entrant for a limited number of periods from other entrants and evaluate the impact on consumer and producer welfare.[4] The social planner compares social welfare under free entry with the social welfare that is generated when a first entrant is protected for a certain number of periods to be the single operator in the market and subsequent entry is delayed.[5] We explicitly account for economies of scale, declining entry costs over time and dynamic efficiency gains generated from delaying subsequent entry.[6]

Only few empirical studies focus on excessive entry: Berry and Waldfogel (1999) assess social inefficiency in the U.S. radio broadcasting industry, and quantify the excess number of firms relative to the social optimum. They find that there are too many radio stations under free entry (on average 18.6 per market, compared with a socially optimal number of 4.8). Hsieh and Moretti (2003) show that a socially inefficient large number of real estate agents enter markets in which housing prices are high. Those studies assume time invariant entry costs. However, very little is known on how entry costs evolve over time. To the best of our knowledge we are not aware of any study that quantifies the consequences of "excessive entry" in a fully dynamic study allowing for entry costs to decline over time. One reason is that the cost of entry is rarely observed by econometricians. Also, the identification of time varying entry costs, as well as the impact on social welfare requires a rich database, which existence is limited.

Our study concentrates on the Static Random Access Memory (SRAM) industry which belongs to the family of semiconductors and are used as an input for electronic devices.[7] The SRAM industry provides a natural setting for our study for the following reasons: First, the

[4]Several studies focus on the question to what extent government intervention benefits politicians and bureaucrats without necessarily increasing consumer welfare, see McChesney (1987), De Soto (1990), Shleifer and Vishny (1993) and Djankov et al. (2002) among others. Therefore, beyond evaluating the impact on consumer surplus our policy experiment also assesses the impact on producer surplus.

[5]Gilbert and Shapiro (1990) remark that the negative impact on consumer welfare make limitations on the length of entry protection necessary.

[6]Note, that we concentrate on the social planner controlling the entry time of subsequent entrant after the first firms has entered the market. We are aware that other configurations related to an optimal timing to enter markets are feasible and would probably further maximize social welfare. However, this question is too complex based on currently available methods in the area of dynamic oligopoly models.

[7]A detailed industry description is provided in Section 2.

SRAM industry is characterized by closely related or homogeneous products, which emphasizes the presence of business stealing effects. Second, the industry is characterized by (dynamic) economies of scale or learning by doing effects, subsequent excessive entry increases average cost and reduce social welfare. Third, the SRAM industry is characterized by high entry costs which drastically decline over time due to a high depreciation rate for cost of capital and the associated investments in production machinery. Fourth, the SRAM industry is characterized by free entry. For example, firms are not able to patent the information storage capacity of those memory chips.[8] Fifth, the SRAM industry provides detailed information on firms' entry choices since it is characterized by free entry and firms enter several new product generations at different periods. Our dataset also contains detailed firm-level production data for specific product generations over time and allows us to observe frequent changes in market structures, due to entry and exit, over time. This allows us to identify firms' policy functions, discounted profit flows evaluated at different entry periods and the evolution of entry costs over time. To summarize, the SRAM industry represents a perfect base to evaluate the impact of excessive entry on social welfare.

We formulate a dynamic oligopoly model in the tradition of Ericson and Pakes (1995), in which firms make entry, exit and production decisions. Firms are assumed to be rational and forward looking agents who maximize their discounted payoffs. Using quarterly firm-level data from 1974 to 2003, we apply a two-stage estimation procedure by Bajari, Benkard and Levin (2007). The first stage estimates firms' entry, exit and production policies. Based on those estimated policy functions, we perform forward simulations and calculate the discounted continuation values. Our estimation results show reasonable estimates for the policy functions and continuation values. In a second stage, we distort the timings of entry to identify the evolution of entry costs over time. The results emphasize a considerable decline in entry costs of more than 90% within the first three years.

Based on our structural estimates, we perform counterfactual experiments and evaluate from a social planner's perspective the extent to which free entry causes excessive entry and social inefficiencies. We perform an experiment which protects the first entrant for a limited number of periods from other entrants, which allows us to assess the dynamic efficiency gains due to

[8]The patents rather protect process innovation and manufacturing processes which improve production costs, rather than new technologies that are related to the introduction of new memory chips. Section 2.1 explains the facts in more detail.

eliminating excess entry. We evaluate how the protection of a first entrant impacts market structure, performance and consumer and producer surplus. The results show that a longer-lasting entry protection of the first entrant reduces consumer welfare compared to the free entry case. Most interestingly, social welfare increases if the duration of entry protection is sufficiently short (less than 1 year) or long (more than 6 years). If entry protection is short social welfare increases, because dynamic efficiency gains in form of entry cost savings (and avoiding excessive entry) outweigh the losses in consumer surplus. As a consequence, industry profits and total surplus increases. If the length of entry protection is in the intermediate range social welfare decreases, since the increase in producer surplus due to savings on entry costs is not sufficiently large to compensate for the loss in consumer surplus. If the entry protection period is sufficiently long, social welfare increases due to increased industry benefits and dynamic efficiency gains or entry cost savings. Moreover, the profit of the protected incumbent significantly increases as the entire industry output is concentrated on one producer, such that significant cost reductions are exploited from (dynamic) economies of sale effects. Both, the entry cost savings and the protected incumbent's gain in profit, dominate the potential gains from more firms entering and higher consumer welfare. To summarize, our results suggest that excessive entry is existent due to the fact that firms enter too early and a delayed entry would achieve dynamic efficiency gains, which improves social welfare.

The remainder of the paper is organized as follows. Section 2 describes the institutional features of the semiconductor industry and present summary statistics. Section 3 introduces our dynamic oligopoly model and Section 4 describes the econometric model. In Section 5 we discuss the estimation results and the results for our counterfactual experiments. We conclude in Section 6.

## 2 Institutional and industry background

In this section, we provide information on entry regulation and describe the institutional background of the semiconductor industry. Next, we describe the SRAM industry in more detail and also provide a description of the data sources and summary statistics.

## 2.1 Entry regulations

We perform counterfactual policies which assess the social inefficiencies caused by "excessive entry" and evaluate the impact on consumer and producer welfare. In our experiment, a social planner is able to protect the first entrant for a limited number of periods from other entrants. This policy experiment is especially relevant, since governments frequently enact entry regulation policies that grant exclusive rights to specific firms to provide a good or a service for a limited period of time with the purpose of increasing social welfare. Examples are utility services (gas, electricity), airlines, railroad, and petroleum among many others. One prominent governmental-protected monopolist was the American Telephone and Telegraph (AT&T) company which provided local and long distance telephone services. Several policy debates discuss the advantages and disadvantages of entry regulation. Entry regulations serve the purpose to avoid excessive entry, especially in industries characterized by high entry or fixed costs, e.g., airlines, telecommunications, and public utility services. In this case, average costs of incumbent firms decline and new entrants steal production from incumbent firms, such that incumbents operate at higher average costs. Entry regulations also serves the purpose to avoid wasteful duplication of entry costs. Many policy makers and economists who argue against entry regulation are concerned about temporary higher prices causing harm to consumer surplus. They refer to the fact that free entry stimulates competition and reduces prices. Excellent empirical studies on entry regulation are the following: Djankov et al. (2008) show that stricter regulation of entry increases producer surplus as it is associated with more corruption and a larger shadow economy and not with less pollution. Suzuki (2013) consider how planning regulations affect entry in the lodging industry. For further information on optimal number of firms and market entry in the context of regulation, see also Schaumans and Verboven (2008) and Ferrari and Verboven (2010). Empirical studies which focus on the interaction between oligopolies and innovation are Goettler and Gordon (2011 and 2012) and Kim (2013), among many others.

## 2.2 The SRAM industry

In the following, we describe the manufacturing process for static random access memory chips (SRAMs), which are one type of semiconductors. Semiconductors consist of memory chips, microprocessors, and application-specific integrated circuits. It is frequently highlighted that the semiconductor industry has a significant impact on productivity growth, since they serve

as an input in computers, electronic devices like digital cameras or cell phones, automotive products and household appliances, among many others.

SRAMs are one type of memory chips designed for the purpose of storing and retrieving information. The chips are classified into generations according to their capacities of information storage. We would like to emphasize that firms are not able to claim a patent on the capacity of information storage on memory chips. The patents rather protect process innovation and manufacturing processes which improve production costs, rather than new technologies that are related to the introduction of new memory chips. Hence, it is important to note that firms mostly file patents for new process technologies rather than new products. Managers in the semiconductor managers reassured that almost all patents granted refer to new manufacturing processes, rather than new generations or the associated memory capacity, chip and architectural designs. We can conclude that the SRAM industry is characterized by free market entry, i.e., free entry into different SRAM generations.

The cost of entering a new SRAM generation requires a tremendous amount of capital and includes the initial purchase price and installation of new fabs with new production machineries and equipment. The new production machines are generation-specific and are developed outside the semiconductor industry. Hence, entry cost is considered exogenous as it is determined by the technological progress of production machines developed outside the SRAM industry. Interviews with experts and managers in the industry reassured the fact, that cost of capital in equipment is the main determinant of entry and entry costs primarily consist of the costs of equipment. The cost of capital is significant and amounts to roughly half the production cost of a wafer. For example, in 2004, Samsung, Hynix, Toshiba, Intel and Micron announced a capital spending between 2 and 4.5 billion U.S. dollars. Note that these numbers are rarely reported and also represent averages only.

As mentioned above, the cost of capital highly depreciates over time and entry costs decline as firms enter a market at a later point of time. Data on entry patterns in this industry confirm the argument of declining entry costs: More than 50% of the firms still entered a product generation 5 years after the first firm entered, which suggests that late entry still is profitable and supports the notion that entry costs decline over time. Hence, declining entry costs over time are a further potential source for dynamic inefficiencies and "excessive entry". Against the background of declining entry costs over time, firms face a trade off in determining the optimal

time to enter: An early market entry requires a higher entry cost, but ensures higher profit streams due to facing fewer firms and lower competition. Late entry returns lower discounted profits, but requires firms to pay only lower entry costs, see also Fudenberg and Tirole (1985), Riordan (1992), and Klette and Kortum (2004). Figure 3 shows that firms have a different optimal timing to enter which gives rise to the fact that firm heterogeneities, i.e., different production costs due to own learning within and between generations are important aspects to consider.[9]

The manufacturing process of memory chips is highly complex. Advanced photolithographic and chemical processes to etch electrical circuits onto the wafer surface are necessary to improve a chip's performance. Memory chips are cut from silicon wafers, which determines silicon as the base material entering the manufacturing process. The production yield rate, which is defined as the percentage of wafers that successfully pass all production stages, can increase from around 20% to more than 80% through learning by doing (or dynamic economies of scale). Improvements in manufacturing processes and production cost reductions, which reduces the amount of material waste, improves manufacturing errors and increases efficiencies of operations which eventually reduces production costs. Firms learn not only from their own experience, but also from their competitors' experience spilling over through labor turnover and information transfers, as well as from experience transmitted between product generations. Spence (1981) as well as Fudenberg and Tirole (1983) and Irwin and Klenow (1994) demonstrate that such learning effects can have a sizable impact on production costs. The presence of intergenerational learning effects and the associated gains in production costs across generations also explains why firms successively introduce generations and do not skip one generation to achieve an early headstart on a subsequent generation. Note, learning also impacts firms' optimal timing decisions to enter a new chip generation. If learning from own experience within a generation is strong, firms will enter earlier, see also Spence (1981), Fudenberg and Tirole (1983), Ghemawat and Spence (1985) and Dasgupta and Stiglitz (1988). Moreover, in evaluating the impact of entry regulation on total surplus, it should be noted that dynamic economies of scale or learning by doing effects have a high cost reducing impact on incumbents at the beginning of a life cycle. This suggests that entry protection is most efficient in the first few periods where learning has the highest cost

[9]For studies focusing on the timing of adopting new technologies see, e.g., Genesove (1999), Gowrisankaran and Stavins (2004), Ackerberg and Gowrisankaran (2006), Schmidt-Dengler (2006), Sweeting (2006 and 2009) and Einav (2010),.

reducing impact. This aspect highlights the relevance of establishing counterfactual experiments that consider entry protections of different lengths.

### 2.3 Data description and summary statistics

We introduce the data sources and provide some descriptive statistics. The dataset is compiled by Gartner Inc. and includes quarterly data on the SRAM industry from January 1974 until December 2003. The data set encompasses six product generations, namely the 4Kb, 16Kb, 64Kb, 256Kb, 1Mb and 4Mb generations. For every chip generation, we observe firm-level and industry-level units produced, the average selling price, and the number of firms in the market.

Figure 2 shows the average prices of the 4Kb to 4Mb SRAM generations over time. Prices are rapidly declining throughout the product life cycle. An important phenomenon that explains this price decline for SRAMs are learning effects, see the discussion above. Figure 1 shows how industry shipments of the 4Kb to 4Mb SRAM generations evolve over time. Shipments of each generation are characterized by product life cycles, which take on similar shapes across different generations. At a certain point of time, multiple SRAM generations are offered on the market.

Another explanation for the drastic price decline relates to a changing market structure over time. As time elapses, more firms enter a specific product generation which increases competition and reduces prices. Figure 3 shows the number of firms present in the 4Kb to 4Mb SRAM generations over time. The number of firms follows an inverse U-shape within every generation. Interestingly, firms continue to enter the market many years after the product generation has been launched. To illustrate market structure in more detail, Figure 4 shows the entry process for the 64Kb generation. The figure illustrates that firms enter the 64Kb generation even 20 years after the first firm entered the generation. We observe that about half of the firms enter in the first 5 years. Most firms enter during the $8^{th}$ to the $33^{rd}$ quarter while less than 10% of the firms enter in the first two years. Note, that we explicitly focus on estimating the entry cost in the 64Kb generation, since the life cycle is complete and provides a complete set of observations. However, it is important to emphasize the fact that demand and supply account for intergenerational effect. The details are explained further below.

Firms simultaneously produce multiple product generations which indicates that firms are able to benefit from learning effects across generations. In light of intergenerational learning effects, multiproduct firms might have an incentive to produce more output in previous gen-

erations to gain more experience which reduces their production costs in the new generation, resulting in higher discounted profits and an early entry into subsequent generations. It is important to mention that intergenerational learning is one major reason why (heterogeneous) firms successively enter into different generations - rather than skipping one generation - and to achieve an early headstart for production in subsequent generations. This fact is important in our policy experiment as it ensures that firms will not skip producing a generation once entry protection becomes sufficiently long. Moreover, multiproduct firms also account for substitution patterns originated from the demand side. If chips are substitutes, the production of multiple chips generations will cannibalize the sales of other generations. Consequently, multiproduct firms will account for these substitution patterns when setting their production schedules for different generations.

# 3 Dynamic oligopoly model

In this section we introduce our dynamic oligopoly model, which allows us to control for learning by doing, time varying entry costs and dynamic efficiency gains. We construct a discrete-time infinite horizon model with time indexed by $t = 0, 1, \ldots, \infty$. Firms, denoted by $i = 1, 2, \ldots, N$, maximize the sum of profits over all periods for a specific generation $k$.[10] The model is formulated as a state space game in which firms use Markov strategies. Firms' actions in a given period determine not only their own and rival firms' current profits, but also their own and rival firms' future states. We build on the fact that firms are rational and forward looking, i.e. they derive their discounted profit streams given the evolution of the state vector and their actions. At the beginning of each period, firms simultaneously decide whether to enter generation $k$ and how much to produce in generation $k$. Firms who entered in generation $k$ in the past, also decide whether to exit generation $k$ or not.[11] We focus on pure entry strategies. Entering a new product generation incurs an entry cost of $C_t^k$, where the subscript $t$ refers to the time period

[10] We simplify our entry model to one generations and believe that this is an appropriate assumption when estimating time-variant entry costs for the following reasons. First, different product generations are usually produced at different fabs. Therefore, firms' output decisions are usually made at the fab level. Second, firms that consider to enter generation k in period t only know about the existence of generation k-1, but do not know of generation k+1 yet. Moreover, firms who actively produce one generation, introduce subsequent generations because of intergenerational learning effects. Therefore, firms skipping one generation to achieve an early headstart for the next generation is not a concern in the semiconductor industry. When considering the entry of generations $k$, we treat the states from the previous generation as predetermined and given from the data. The assumption might be restrictive, however, for other industries and should be reconsidered.

[11] Exit does not involve any cost or scrap value.

when entry occurs. The entry cost is deterministic and larger than zero. Since the entry cost is primarily a function of depreciation of investments, we assume the function to be industry and generation specific, rather than firm-specific.[12]

The observable state variables to every firm at period $t$ are denoted by a vector $s_{it} = (n_t^k, x_{it}^k, x_{-it}^k, C_t^k, z_t^k, X_{it}^{k-1})$, which contains all of the payoff relevant state variables for generation $k$. The variables $n_t^k$ refer to the number of firms in period $t$ for generations $k$ and are supposed to capture the degree of competitiveness in the markets for generation $k$. We also account for firms' learning capabilities. Firms are supposed to learn from past experience lowering marginal costs. We distinguish between learning experience accumulated within a generation (intragenerational learning) and learning experience accumulated in the previous generation (intergenerational learning). With regard to intragenerational learning effects, we allow firms to learn from their own experience ($x_{it}^k$) as well as from other firms' experience via spillovers ($x_{-it}^k$). $z_t^k$ includes all the demand and cost shifters which are known to the industry. $X_{it}^{k-1} = (n_t^{k-1}, x_{it}^{k-1}, x_{-it}^{k-1})$ are state variables from the previous generation. These variables are assumed to be exogenous and given from the data. The entry of generation $k$ may be determined by the relative competitiveness in the markets of generations $k(n_t^k)$ and $k-1$ ($n_t^{k-1}$). We allow firms to learn from their own production experience in the previous generation ($x_{it}^{k-1}$). We also account for firms' intergenerational learning effects via spillovers resulting from other firms' production experience in the previous generation ($x_{-it}^{k-1}$). Finally, the time-variant entry cost $C_t^k$ enter the state vector.

## 3.1 Evolution of states

### Number of firms

We define the transition of the number of firms $n_t^k$ within generation $k$ as follows:

$$n_t^k = n_{t-1}^k + ne_t^k - nx_t^k, \tag{1}$$

where $ne_t^k$ is the number of entering firms and $nx_t^k$ the number of exiting firms.

### Learning

[12] Note that firms might have different entry times in generation $k$ due to the fact that we allow for heterogeneous firm sizes in generation $k-1$.

Firm $i$'s marginal cost of production is characterized by several learning components. The first component describes a firm's intragenerational learning from own experience. Firm's own learning within generation $k$ is given as follows:[13]

$$x_{it}^k = x_{it-1}^k + q_{it-1}^k, \tag{2}$$

where $q_{it-1}^k$ is firm $i$'s output in generation $k$ at time $t-1$ and $x_{i0}^k = 0$, assuming that there is no own experience in the beginning of the product cycle of each product generation. The other components describe the extent to which firm $i$ learns from other firms' experience via spillovers within and between generations. Intragenerational spillovers are specified as follows:

$$x_{-it}^k = x_{-it-1}^k + \sum_{j \neq i} q_{jt-1}^k, \tag{3}$$

with $x_{-i0}^k = 0$, assuming that there are no spillovers in the beginning of the product cycle of each product generation. In general, potential entrants in a specific generation may benefit from own experience in the previous generation as well as intragenerational and intergenerational spillovers.

**Productivity shocks**

In each period, each incumbent in generation $k$ receives a private productivity shock $\nu_{it}^k$. The shocks are drawn from a known normal distribution $G^k$ with zero means and constant variances $\sigma_k^2$ and are independently and identically distributed across firms and across periods.

**Marginal costs**

The marginal cost function depends on the set of payoff-relevant state variables and follows the specification in Irwin and Klenow (1994) to account for own learning and spillover learning. Firm $i$'s marginal cost $mc_{it}^k$ is specified as follows:

$$\ln(mc_{it}^k) = \rho_0^k + \rho_i + \rho_1 \ln(P_t^{sil}) + \rho_2 \ln\left(x_{it}^k + \rho_3 x_{-it}^k + \rho_4 \left[x_{it}^{k-1} + \rho_3 x_{-it}^{k-1}\right]\right) + \nu_{it}^k, \tag{4}$$

Marginal costs are described by generation-specific intercepts ($\rho_0^k$), firm fixed effects ($\rho_i$), the price of silicon $P_t^{sil}$ and several learning components, i.e., intragenerational own learning ($x_{it}^k$)

[13] Note that the laws of motion for the state variables are deterministic.

and learning from others via spillovers ($x^k_{-it}$), intergenerational own learning ($x^{k-1}_{it}$) and learning through spillovers $x^{k-1}_{-it}$. The parameters $\rho_2$ and $\rho_4$ measure the impact of the own intra- and intergenerational learning effects on the marginal costs. The parameter $\rho_3$ allows for a separate impact of own learning and learning via spillovers on marginal costs.

**Strategies**

Firm $i$'s strategy is given by $\sigma^k_i(s, \nu_i)$, where $\sigma_i$ is the set of firm $i$'s strategies that describe the production, entry and exit decisions. The set of state vectors ($s_{it}$) are summarized in $s$ and $\nu_i$ includes the generation-specific firm-level productivities. Rival firms' strategies are denoted by $\sigma_{-i}$.

## 3.2 Profit

Each firm maximizes its future discounted profits from generation $k$. If a firm is active in generation $k$, its per period profit is given by:

$$\pi^k_{it}(\sigma^k_i, \sigma^k_{-i}, s_t, \nu^k_{it}) = q^k_{it}[p^k_t - mc^k_{it}(\nu^k_{it})],$$

where the variable $p^k_t$ represents the price for generation $k$. The price is a function of the industry output $q^k_t = \sum_i q^k_{it}$ and observable demand shifters $d_t$. Hence, firm $i$'s per period profits for generation $k$ at time $t$ depend on the actions of all firms, the state vector, and firm $i$'s private shock. A firm earns zero profit if it is inactive in generation $k$.

If behavior is given by a Markov perfect equilibrium strategy profile $\sigma^k = (\sigma^k_i, \sigma^k_{-i})$, firm $i$'s value function can be written recursively:

$$V_i(s, \sigma^k; \theta) = E_\nu \left[ \pi^k_i(\sigma^k, s, \nu^k_i) - I[C^k](\sigma^k) + \beta \int V_i(s'; \sigma^k) dP(s'|\sigma^k, s) \middle| s \right], \tag{5}$$

where $\theta$ denotes a vector of parameters and I refers to an indicator function which takes on a value of 1 if a firm enters a market and 0 otherwise. $P$ describes the evolution from today's state $s$ to tomorrow's state $s'$.

### 3.3 Timing

The timing of our model can be summarized as follows. In each period $t$, events occur in the following order:

1. Firm $i$ considers the following state $s_{it} = (n_t^k, x_{it}^k, x_{-it}^k, C_t^k, z_t^k, X_{it}^{k-1})$.

2. Given the state vector, potential entrants make their entry decisions for generation $k$ and incumbents make their exit decisions for generation $k$ .

3. Incumbent firms in generation $k$ observe their productivity shocks $\nu_{it}^k$, choose outputs $q_{it}^k$ and realize profits.

4. The state moves to $(n_{t+1}^k, x_{it+1}^k, x_{-it+1}^k, C_{t+1}^k, z_{t+1}^k, X_{it+1}^{k-1})$.

## 4 Econometric model

In this section, we describe our econometric model. We apply the two-step estimation method developed by Bajari, Benkard and Levin (2007) to estimate the cost of entering the $64K$ SRAM generation. The first stage includes the estimation of the demand, the policy functions and the marginal cost functions. All policies are functions of the observed state, $s$, including the entry cost $C_t^k$ and the productivity shocks $\nu_{it}^k$ which are unobserved to the econometrician. We include a time trend in all of the policy functions to proxy for the declining entry costs over time. In the second stage, we apply forward simulations to calculate the continuation values. Finally, we recover the entry cost over time that rationalizes firms' policies.

### 4.1 First Stage

**Demand**

We assume that every generation is homogeneous in itself, but different generations represent differentiated goods. The demand function is specified log-linearly as follows:

$$\ln(q_t^k) = \alpha_0^k + \alpha_1^k \ln(P_t^k) + \alpha_2 \ln(P_t^S) + \alpha_3 \ln(GGDP_t) + \alpha_4 \ln P_t^{DRAM} + \alpha_5 Time_t + d_t^k, \quad (6)$$

where we denote the vector of coefficients with $\alpha$. We account for generation-specific demand shifters ($\alpha_0^k$). $P_t^k$ is the price for generation $k$. Note that we allow for generation-specific price

elasticities ($\alpha_1^k$). $P_t^S$ is the price for the substitutable generation.[14] $GGDP_t$ is the growth rate of GDP, $P_t^{DRAM}$ is the price of DRAMs and $Time$ refers to a time trend.

**Marginal Costs**

In line with our theoretical model, we retrieve firms' marginal costs for the $k = 64K$ generations ($mc_{it}^{64K}$) using the estimated price elasticities $\hat{\alpha}_1^{64K}$ of our demand equation (6) and the observed market shares, $ms_{it}^{64K}$, see also Siebert and Zulehner (2011):

$$mc_{it}^{64K} = P_t^{64K}\left(1 + \frac{ms_{it}^{64K}}{\hat{\alpha}_1^{64K}}\right). \tag{7}$$

**Output Policy**

The output policy depends on the set of observed state variables, the productivity shock and a time trend. Hence, the output policy accounts for factor prices, demand shifters, market structure, intra- and intergenerational own learning and learning via spillovers. The output function for the $64K$ generation is specified as:

$$\begin{aligned}\ln(q_{it}^{64K}) = \gamma_i + \gamma_1 \ln(P_t^{sil}) + \gamma_2 \ln P_t^{DRAM} + \gamma_3 GDP_t + \gamma_4 \ln(n_t^{16K}) + \gamma_5 \ln(n_t^{64K}) + \gamma_6 \ln(x_{it}^{16K}) \\ + \gamma_7 \ln(x_{it}^{64K}) + \gamma_8 \ln(x_{-it}^{16K}) + \gamma_9 \ln(x_{-it}^{64K}) + \gamma_{10} Time_t + \gamma_{11}\hat{\nu}_{it}^{64K} + \epsilon_{it}^{64K},\end{aligned} \tag{8}$$

where $\gamma_i$ is a firm level fixed effect. We assume that the error term $\epsilon_{it}^{64K}$ is i.i.d normally distributed. $\hat{\nu}_{it}^{64K}$ is the residual from the marginal cost estimation and is interpreted as the productivity shock. A positive shock increases the marginal cost of production and hence reduces a firm's output. Thus, we should expect a negative sign for $\gamma_{11}$.

**Entry and Exit Policies**

We estimate the following probit model for entering the $64K$ generation:[15]

$$Pr(e_{it}^{64K} = 1; s) = \Phi\Big(\lambda_i + \lambda_1 \ln(n_t^{64K}) + \lambda_2 \ln(x_{it}^{16K}) + \lambda_3 \ln(x_{-it}^{64K}) + \lambda_4 Time_t\Big), \tag{9}$$

where $\Phi(.)$ is the cumulative distribution function of the standard normal. The entry policy

[14] As substitute we use the average SRAM price of the adjacent generations.

[15] Note, since our theoretical model explicitly formulates entry in generation $k$, we only have to estimate the entry policy for the $64K$ generation. This is also sufficient for our policy experiments to evaluate the lengths of entry regulations in the $64K$ generation.

depends on the number of firms active in the $64K$ generation ($n_t^{64K}$), firm $i$'s experience from the previous $16K$ generation ($x_{it}^{16K}$), the spillover from other firms in the $64K$ generation ($x_{-it}^{64K}$) and a time trend.

The exit policies for the $64K$ generation is formulated as:

$$Pr(ex_{it}^{64K} = 1; s) = \Phi\Big(\psi_0 + \psi_1 \ln(n_t^{64K}) + \psi_2 \ln(x_{it}^{64K}) + \psi_3 \ln(x_{-it}^{64K}) + \psi_4 Time_t\Big), \tag{10}$$

where the probability of exit is a function of the number of firms, the own experience and other firms' experience in the current generation.

## 4.2 Second Stage

Using the estimated policy functions, we take draws for the productivity shocks for each incumbent in each period and forward simulate the sequences of optimal outputs for each firm. In each simulation $l$, we record $t_{il}^*$ (firm $i's$ optimal timing of entering the $64K$ generation in simulation $l$). Using $N \times L$ observations, where $N$ is the number of firms and $L$ is the number of simulations, we generate further potential equilibrium paths. Hence, based on the optimal timing ($t_{il}^*$), the demand equation, the marginal cost function, the drawn productivity shocks and the optimal policies, we compute a simulated sequence of outputs. Next, we compute the discounted profits before subtracting the entry costs, i.e., $\widetilde{V}_{il}(t_{il}^*)$. To summarize, we compute the value functions applying the following steps for each simulation:

1. We draw one productivity shock for each incumbent at each period i.e., $\nu_{it}^{64K}$.

2. We use the optimal production, entry and exit policies to update firms' production choices $\hat{q}_{it}^{64K}$ and state variables $\hat{s}_{it}$. Note that the draws of productivity shocks enter $\hat{q}_{it}^{64K}$ and $\hat{s}_{it}$(see equation 8).

3. For every period, we update prices $\hat{p}^{64K}(\hat{q}_{it}^{64K}, \hat{q}_{-it}^{64K})$ using the estimated demand equation.

4. We compute the vector of marginal costs $\hat{mc}_i^{64K}$ using the estimated marginal cost function, $\hat{q}_{it}^{64K}$, $\hat{s}_{it}$ and $\nu_{it}^{64K}$.

5. We calculate the value functions: $\widetilde{V}_{il}(t_{il}^*, t_{-il}^*) = \sum_{t=0}^{\infty} \beta^t (\hat{\pi}_{it}^{64K})$.

In order to identify the evolution of entry costs over time, we compare the changes in the discounted values at the optimal entry times with the values at the distorted and therefore suboptimal entry times to derive the upper and lower bounds for the entry costs.

Regarding the upper bound of the entry costs, we compare the discounted values (net of entry costs) at the optimal entry time $\widetilde{V}_{il}(t^*_{il}) - C(t^*_{il})$ with the discounted value (net of entry costs) from entering one period later, i.e., $\widetilde{V}_{il}(t^*_{il}+1) - C(t^*_{il}+1)$. Based on firms' rational expectations, we can write:

$$\widetilde{V}_{il}(t^*_{il}) - C(t^*_{il}) \geq \widetilde{V}_{il}(t^*_{il}+1) - C(t^*_{il}+1)$$

Rearranging this inequality yields:

$$\Delta C^{max} = \widetilde{V}_{il}(t^*_{il}) - \widetilde{V}_{il}(t^*_{il}+1) \geq C(t^*_{il}) - C(t^*_{il}+1), \tag{11}$$

where $\Delta C^{max}$ defines the maximum change in entry costs between period $t$ and period $t+1$ and represents the upper bound of the slope of the entry costs, see Figure 5.

Regarding the lower bound of the entry costs, we compare the discounted values (net of entry costs) when the optimal entry time is $(t_{il}+1)^*$ with the discounted value (net of entry costs) evaluated at an earlier suboptimal entry time $(t_{il})$:

$$\widetilde{V}_{il}(t_{il}+1)^* - C(t_{il}+1)^* \geq \widetilde{V}_{il}(t_{il}) - C(t_{il})$$

Again, rearranging the inequality gives:

$$\Delta C^{min} = \widetilde{V}_{il}(t_{il}) - \widetilde{V}_{il}(t_{il}+1)^* \leq C(t_{il}) - C(t_{il}+1)^*, \tag{12}$$

where $\Delta C^{min}$ represents the lower bound of the slope of entry cost and defines the minimum entry costs between period $t$ and period $t+1$, see Figure 6.

Next, we search over the entry costs that best rationalize the two conditions - inequalities (11) and (12) - using a minimum distance approach. We therefore define a penalty, when inequalities (11) and (12) are violated.[16] Then, $\Delta C$ can be recovered by searching over $c$, that minimizes

[16] A penalty is defined when the simulated discounted value (net of entry costs) is higher at a suboptimal entry time than at the optimal entry time.

the weighted sum of penalties. Denoting $A_t \equiv \{ (i,l) \mid t^*_{il} = t \}$ as the set of observations where firms optimally enter at period $t$, and $A_{t+1} \equiv \{ (i,l) \mid (t_{il}+1)^* = t+1 \}$ as the set of observations where firms optimally enter at period $t+1$:

$$\widehat{\Delta C(t)} = \arg\min_c \left\{ \frac{1}{|A_t|} \sum_{A_t} \min \left[ 0, \widetilde{V}_{il}(t^*_{il}) - (\widetilde{V}_{il}(t^*_{il}+1) + c) \right] \right.$$
$$\left. + \frac{1}{|A_{t+1}|} \sum_{A_{t+1}} \min \left[ 0, \widetilde{V}_{il}(t_{il}+1)^* - (\widetilde{V}_{il}(t_{il}) - c) \right] \right\} \quad (13)$$

where $|\cdot|$ refers to the number of elements in the set.

To summarize, we perform the following steps in our estimation algorithm:

1. We estimate the demand equation (6) to retrieve the elasticity of demand.
2. We use the elasticity of demand in order to calculate the marginal costs using the Lerner Index, see equation (7).
3. We estimate the marginal cost function to retrieve the cost parameters entering in equation (4).
4. We estimate the production policies (equation 8), the entry policy (equation 9) and the exit policy (equation 10).
5. We use the cost parameters and the optimal policies to apply forward simulations and to calculate the value functions (before entry costs), i.e., $\widetilde{V}_{il}(t^*_{il})$ and $\widetilde{V}_{il}(t_{il}+1)^*$.
6. We distort the optimal time to enter and use the policies to calculate value functions (before entry costs) i.e., $\widetilde{V}_{il}(t^*_{il}+1)$ and $\widetilde{V}_{il}(t_{il})$.
7. Finally, we use the penalty conditions and apply a minimum distance procedure to recover the slope of the entry costs (equation 13).

# 5 Estimation results and policy experiment

In this section, we describe our estimation results. We first discuss the results for the demand and the marginal cost function. Afterwards, we discuss the estimation results for the firms' production, entry and exit policies. Finally, we describe the results for the entry costs and discuss the results of our policy experiment.

## 5.1 First stage results

**Demand**

To estimate industry demand, we pool the data from all product generations and estimate equation (6) using ordinary least squares and instrumental variable estimation techniques. We instrument for potential correlation of the price with the error term using supply-side variables: price of silicon ($P_t^{sil}$) and cumulative industry output ($\sum_{t'=1}^{t-1} q_{t'}^k$) of the current generation. The estimation results are shown in Table 1. We estimate two specifications: the first specification incorporates a constant price elasticity across product generations (Columns 1 and 2). The second specification incorporates generation-specific price elasticities (Columns 3 and 4). We employ a Chow test to test for identical price elasticities across product generations. The F-statistics returns a value of 6.01 which leads us to reject the null hypothesis that prices elasticities are equal to each other.

Comparing the results in columns 3 and 4, the instrumental variable estimation returns more elastic price elasticities of demand than the ordinary least squares estimation. This result indicates an upward bias of the ordinary least squares estimates and gives rise to the fact that prices are endogenous. To test for endogeneity, we conduct a Durbin-Wu-Hausman test. The test statistics is $F_{(1,300)} = 619.99$ for the first specification and it is $F_{(4,293)} = 6.40$ for the second specification. We therefore reject the exogeneity of prices in both specifications.

In the following, we discuss the results shown in column 4. The estimated price elasticities range from $-2.75$ to $-3.19$, confirming a highly elastic market demand in the SRAM industry. In comparing the price elasticities throughout different generations, we find that the $64K$ generation is characterized by the least elastic demand, which explains the fact that the average prices in this generation stay at higher levels. Demand becomes more elastic in the subsequent $256K$ and $1M$ generations. Finally, our estimations return positive cross-price elasticities, confirming the fact that SRAM chips are substitutes across generations.

**Marginal cost** The marginal cost function, equation (4), is estimated using a non-linear least square procedure. We assume that the error in the marginal cost function is normally distributed. The results are shown in Table 2. Column 1 reports the estimation results without a time trend. All coefficients are significantly different from zero, carry the expected sign and reveal economically sensible magnitudes. In addition, the model fits remarkably well with an

adjusted $R^2$ of about 0.90. We find a positive effect of input prices on output prices. The results further return a learning parameter of $-0.34$, which corresponds to a learning rate of 21%.[17] This result confirms that the SRAM industry is characterized by strong intragenerational learning effects. Moreover, we find that there are spillovers within one generation and across generations. The spillover effect within one generation is of about 30% (or 6% overall), i.e., firms are able to benefit from about a third of other firms' output. Compared to this effect, the intergenerational spillover effect is small. About 10% (or 2%) from the output produced in the earlier generation contributes to learning.

Column 2 reports the estimation results from a second specification that includes a linear time trend. The negative significant time trend reflects the fact that marginal costs decline over time. The inclusion of the time trend slightly alleviates the magnitude of intra- and intergenerational learning. The learning effects, however, still remain economically and statistically significant.

Since the residual of the marginal cost estimation is interpreted as private productivity shocks, we use the standard deviation of the residuals, 0.3, for the distribution of productivity shocks. Therefore, we take draws of productivity shocks from a normal distribution with zero mean and 0.3 standard deviation in the forward simulation.

**Output policy**

We estimate the output policies specified in equation (8). Table 3 shows the estimation results for the $64K$ generation using an ordinary least square estimator.[18] An increase in the price of silicon increases marginal costs and hence reduces output. The coefficients of the demand shifters carry the expected signs. Firms produce more when the size of the market is larger ($GDP$) and when the price of a substitute is higher ($P^{DRAM}$). The coefficients of the number of firms within a generation is negative and significant. A more competitive market, i.e. a larger number of firms in the market, reduces firms' output. Interestingly, however, firms produce more in the $64K$ generation when the $16K$ generation becomes more competitive.

Our estimation results also reconfirm the significant intragenerational learning effects. The

[17] The learning rate is defined as $1 - 2^{\rho_2}$, i.e. the percentage reduction of marginal cost when (past) experience doubles, see also Irwin and Klenow (1994).

[18] We also estimated the output equations using an instrumental variable estimator to correct for a potential endogeneity of the lagged own cumulative output. As instruments we use the twice and three times lagged variables of own cumulative output. The differences between the estimated coefficients from the instrumental variable and the ordinary least squares estimations are negligible. In addition, the Hausman test fails to reject the consistency of the ordinary least squares estimates. Moreover, the Hausman test statistics of 97.76 reject the random effects specifications.

coefficients of the own cumulative outputs in the current generation is positive and highly significant. This illustrates the fact that firms raise production in order to gain more production experience which reduces own marginal costs via learning effects. In terms of intragenerational spillovers, other firms' cumulative output within the $64K$ generation ($x_{-i}^{64K}$) has an impact which is not significantly different from zero. This result seems to be surprising as Table 2 shows that intragenerational spillovers are prevalent. The estimation results of the output policy can be explained by the existence of two opposing effects: a direct effect having an impact on a firm's production via spillovers, and an indirect effect, as a firm produces less when other firms' marginal costs decline. From our estimation results, both effects cancel each other out.

The intergenerational own learning effects ($x_i^{16K}$) and spillover effects ($x_{-i}^{16K}$) are insignificant. This result is consistent with the marginal cost estimates. Finally, the estimate of the coefficient of $\hat{\nu}_i$ is negative and significant as expected. A positive shock which rises marginal cost will reduce output.

**Entry and exit policy** We estimate the probabilities of entry, equation (9), and exit, equation (10), using probit models. The results are shown in Table 4. The first column reports the results of the entry policy. The results show that firms are more likely to enter when more firms have already entered the $64K$ generation in the past. The significant and positive estimate of cumulative output in the $16K$ generation emphasizes the importance of experience from earlier generations through intergenerational own learning effects. The estimated positive and significant time trend indicates that at identical industry states but different time periods, firms choose to enter in a later period in order to save on entry costs. This result provides a first insight that entry costs are decreasing over time.

Column 2 in Table 4 show the results for the exit policy.[19] Firms are more likely to exit the $64K$ generation if they missed out on gaining sufficient production experience, and if many other firms entered the market. Moreover, firms are less likely to exit if other firms' cumulative outputs are larger.

[19] We do not include firm fixed effect in the exit policy since not all firms have exited the $64K$ generation at the end of data period

## 5.2 Second stage results

Using our estimates, we apply forward simulations to calculate the discounted values (before entry costs) at different states given the optimal policy functions. We compare these discounted values to those generated from distorting the entry policies as described in Section 4.2. We retrieve the slope of the entry cost for the $64K$ generation using a simulated minimum distance estimator as shown in equation (13). The resulting entry costs are illustrated in Figure 7.[20] Our results indicate that entry costs decline by 342 million U.S. dollars during the first 30 periods. More than 90% of the reduction of entry cost occurs during the first 10 periods. In the first 10 periods, entry costs rapidly decline with an average of 31.8 million U.S. dollars per period. The entry costs decline less steeply after the $11^{th}$ period, which is the time when excessive entry begins (see Figure 4). Between the $11^{th}$ and the $30^{th}$ period, entry costs decline by about 1.2 million U.S. dollars per period on average. To summarize, our entry cost estimates support the fact that early entrants incur significantly higher entry costs.

## 5.3 Counterfactual experiment

We perform a policy experiment with the purpose to assess the extent of "excessive entry" and to evaluate the impact on social inefficiencies. In our experiment, a social planner can control the protection length of the first entrant. The social planner compares social welfare under free entry with the social welfare that is generated when a first entrant is protected for a certain number of periods. Our counterfactuals explicitly account for dynamic economies of scale, declining entry costs over time and dynamic efficiency gains generated from delaying subsequent entry. We explicitly focus on the $64K$ generation because this generation is the first generation that has a complete and finalized set of observations over the life cycle. It also keeps the analysis computationally tractable. It should be emphasized at this point that evaluating the impact of entry protection on total welfare is not a straightforward exercise, as a longer protection delays entry and also reduces the entry costs for firms. A lower entry cost has a significant impact on firms' discounted values, their entry decisions, as well as their quantity choices, prices and market structure. We simulate output, prices, entry and exit depending on the different lengths of entry protection.[21]

[20] We normalized the entry costs presented in the figure by setting the entry cost of the $30^{th}$ period to 0.

[21] Note, that the social planner is only able to determine the protection period of the first entrant. We are aware that other configurations related to an optimal timing to enter markets are feasible. Moreover, we only

We formulate the policy experiment as follows: The social planner grants a monopoly right to the first firm entering the $64K$ SRAM chip generation and the right expires after a certain number of periods, which is specified by the social planner.[22] The incumbent firm receives the exclusive right to sell $64K$ chips for a specific time period. Once the entry regulation ends, potential entrants decide if they enter the chip generation, pay the corresponding entry costs and decide how much to produce.

The results are illustrated in Figures 8 and 9. Note that all changes in total welfare are evaluated relative to the free entry case. Figure 8 shows the impact of entry protection on industry consumer and producer surplus. A longer protection period increases market power for the first entrant, which causes prices to increase, output to fall, and consumer surplus to decline. The change in producer surplus is always positive comparing to the case of free entry and increases as the length of entry protection increases. The final impact on social welfare depends on the increases in producer surplus and the decreases in consumer surplus. For 3 and 4 periods of entry protection, the total surpluses are larger than that under free entry. For 5 to 14 periods of entry protection, the changes in total surplus are negative and continue to fall. This results reflect that the decreases in consumer surplus outweigh the increase in producer surplus. For 14 to 18 and 23 to 30 periods of entry protection, the changes in total surplus rise due to the strong increase in the changes in producer surplus. While the losses in consumer surplus fall steadily and level off for longer entry protection, the gains in producer surplus increase more rapidly for a longer protection period. As a result, the changes in total surplus start to rise for longer protection periods and become positive for more than 27 periods of entry protection. In sum, as consumer surplus declines steadily, the fluctuation of the change in total surplus are mainly explained by the change in producer surplus.

To better understand the change in producer surplus, we decompose the producer surplus into the monopolist's profits, other firms' profits and the savings of entry costs, see Figure 9. As the length of entry protection increases, the protected monopolist's profits drastically increase relative to the free entry case for several reasons: First, protecting the first entrant for multiple periods preserves its monopoly position and enables the protected entrant to earn

consider entry protection in the $64K$ generation. Both constraints serve the purpose to being able to solve our policy experiment. With currently existing methodologies on estimating dynamic games, those extension would not be computationally feasible and beyond the scope of the paper, but interesting topics for further research.

[22] We assume that firms anticipate neither when the planner starts the entry protection nor the duration of the protection. Therefore, other firms have no chance to outrun or leapfrog the original first entrant.

monopoly profits during these time periods. Under free entry, other firms instantly enter the market. Hence, the change in market structure lowers the first entrant's profits. Since the entry protection regime is compared to the free entry case, the monopolist's profits increase over time. Second, the monopolist produces the entire industry output and is able to further benefit from intragenerational own learning effects, which further lowers costs and increases profits. Comparing the profits gained by the monopolist to the profits lost by other firms, the monopolist's gain is lower than other firms' losses for shorter entry protection periods (less than 15 periods) but greater than other firms' losses for longer entry protection duration (more than 15 periods). With sufficiently long entry protection duration, the protected monopolist can accumulate sufficient experience in early periods and accelerate the learning paces which enables other firms to benefit from spillovers. Consequently, the industry profits become positive relative to the free entry case for more than 15 periods of entry protection.

The figure also shows that entry cost savings increase with the number of monopoly periods, which is due to the fact that longer entry protection delays successive entry. As the lengths of entry protection increase, the increasing rate of entry cost savings depends on the decreasing rate of entry cost and the distribution of entry under free entry. When the entry cost declines faster and the entries of more firms have been delayed, the savings of entry cost are larger. From Figure 9, the savings of entry cost rise more rapidly for less than 10 periods of entry protection because the entry cost decline drastically during the first 10 periods (see Figure 7). For more than 10 periods of entry protection, the savings of entry cost increases relatively slowly in general because of much lower decreasing rates of entry cost (see Figure 7). However, for 14 to 17 and 25 to 30 periods of entry protection, the savings of entry cost rise more rapidly since the entries of a larger number of firms have been delayed by the protection (see Figure 4). The entry cost savings are always larger than the change in industry profits (the combination of the monopolist's profits and other firms' profits) which makes the producer surplus to be positive relative to free entry case. For sufficiently long entry protections (more than 25 periods), both the change in the monopolist's profit and savings of entry cost increase drastically and consequently lead to more rapid increases in producer surplus.

To summarize the main results of our policy experiment: for shorter lengths of entry protection, there are losses in total welfare because of the losses in industry profit and small savings of entry costs. When the length of entry protection lasts for sufficiently long periods, the large

savings in entry costs plus the gain in monopolist profits due to learning effects rise producer surplus. Consequently, the gains in producer surplus outweigh the losses in consumer surplus and total welfare improves. Hence, under free entry firms enter too early which causes excessive entry.

## 6 Summary and concluding remarks

The purpose of this paper is to empirically investigate the impact of different lengths of entry protection on consumer and producer surplus. We estimate a dynamic structural model in which firms choose the optimal time to enter a market, and make production and exit decisions. Our estimation results show that entry costs drastically decline over time. In fact, we find that entry costs decline by more than 90% in the first 10 quarters. Our estimation results also provide evidence that firms account for learning effects when determining their optimal production. Early entrants incur higher entry costs, but are able to take advantage of learning effects early in the life cycle, which lowers their future costs and returns higher profits. Late entrants especially benefit from decreasing entry costs over time and from other technological spillovers. Our policy experiments show that declining entry costs over time play a dominant role in evaluating the impact of entry protection on total welfare. We find that consumer surplus is always declining for different lengths of entry protection due to monopoly pricing. On the other hand, entry protection increases producer surplus compared to free entry. Our results also show that different lengths of entry protection increase total surplus in the industry if the duration of the regulation is sufficiently short or long. If the protection lasts few periods, total welfare decreases relative to the free entry case as the savings in entry costs is not large enough to compensate for the decline in consumer welfare and the loss in industry profit. If the protection lasts sufficiently long, total surplus increases as the entry cost savings that result from avoiding excessive entry in the industry and the gain in the monopolist's profit compensate the losses in consumer welfare. Free entry results in excessive entry since firms under free entry only care about their own profits and not about the overall industry profits. Hence, entry protection diminishes the private incentives of entry, and optimizes social incentives by avoiding excessive entry and firms paying a socially undesirable amount on entry costs. Even though our policy experiment suggests that there is excessive entry in the industry, and entry protection would increase total surplus, we would like to remind that consumer surplus diminishes in our policy

experiments.

On a final note, it is worth emphasizing that declining entry costs over time are significant also in many other capital intensive markets, such as pharmaceuticals, the automotive and electronics industry, and many others. Therefore, it would be interesting to explore in future research if the main finding, i.e., sufficiently long entry protection has a positive impact on total welfare through avoiding excessive entry rates, also applies to other markets.

## References


Ackerberg, D. and G. Gowrisankaran, 2006, "Quantifying Equilibrium Network Externalities in the ACH Banking Industry," Rand Journal of Economics, 37, pp. 738-761.

Bajari, P., L. Benkard and J. Levin, 2007, "Estimating Dynamic Models of Imperfect Competition," Econometrica, Vol. 75, No. 5, pp. 1331-1370.

Berry, S.T., 1992, "Estimation of a Model of Entry in the Airline Industry, " Econometrica, Vol. 30, pp. 889-917.

Berry, S.T. and J. Waldfogel, 1999, "Free Entry and Social Inefficiency in Radio Broadcasting" The Rand Journal of Economics, Vol. 30, pp. 397-420.

Berry, S.T. and J. Waldfogel, 2001, "Do Mergers Increase Product Variety: Evidence from Radio Broadcasting, Quarterly Journal of Economics, Vol. 116, pp. 969-1007.

Bresnahan, T.F. and P.C. Reiss, 1990, "Entry in Monopoly Markets," Review of Economic Studies, Vol. 57, No. 4, pp. 531-553.

Bresnahan, T.F. and P.C. Reiss, 1991, "Entry and Competition in Concentrated Markets," Journal of Political Economy, Vol. 99, No. 5, pp. 977-1009.

Campbell, J.R., and H.A. Hopenhayn, 2005, "Market Size Matters, " The Journal of Industrial Economics, Vol. 53, No. 1, pp. 1-25.

Carlton, D.W., 1983, "The Location and Employment Choices of New Firms: An Econometric Model with Discrete and Continuous Endogenous Variables, " The Review of Economics and Statistics, Vol. 65, No. 3, pp. 440-449.

Chamberlin, E., 1933, "The Theory of Monopolistic Competition," Cambridge, Mass.: Harvard University Press.

Dasgupta, P. and J. Stiglitz, 1988, "Learning-by-Doing, Market Structure and Industrial and Trade Policies, " Oxford Economic Papers, 40(2), pp. 246-268.

Davis, P., 2006, "Spatial Competition in Retail Markets: Movie Theaters, " The Rand Journal of Economics, Vol. 37, No. 4, pp. 964-982.

De Soto, H., 1990, "The Other Path, " New York: Harper and Row.

Dixit, A.K., and J.E. Stiglitz, 1977, "Monopolistic competition and Optimal Product Diversity, " The American Economic Review, Vol. 67, pp. 297-308.

Djankov, S. and R. La Porta and F. Lopez-De-Silanes and A. Shleifer, 2002 "The Regulation Of Entry, " The Quarterly Journal of Economics, MIT Press, Vol. 117(1), pp. 1-37.

Dunne, T., M.J. Roberts and L. Samuelson, 1988, "Patterns of Firm Entry and Exit in U.S. Manufacturing Industries, " The Rand Journal of Economics, Vol. 19, pp. 495-515.

Einav, L., 2010, "Not All Rivals Look Alike: Estimating an Equilibrium Model of The Release Date Timing Game," Economic Inquiry, 48(2), pp. 369-390.

Ericson, R. and A. Pakes, 1995, "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," Review of Economic Studies, 62, pp. 53-82.

Ferrari,S. and F. Verboven, 2010, "Empirical analysis of markets with free and restricted entry," International Journal of Industrial Organization, 28, pp. 403-406.

Foster, L., J. Haltiwanger and C. J. Krizan, 2006, "Market Selection, Reallocation, and Restructuring in the U.S. Retail Trade Sector in the 1990s," Review of Economics and Statistics, 88(4), pp. 748-758.

Fudenberg, D. and J. Tirole, 1983, "Learning by Doing and Market Performance," Bell Journal of Economics, Vol. 14, pp. 522-530.

Genesove, D. 1999, "The Adoption of Offset Presses in the Daily Newspaper Industry in the United States," NBER Working Paper No. 7076.

Geroski, P. 1989, "Entry, Innovation and Productivity Growth," The Review of Economics and Statistics, Vol. 71, pp. 572-578.

Ghemawat, P. and A. M. Spence, 1985, "Learning Curve Spillovers and Market Performance," The Quarterly Journal of Economics, 100, pp. 839-852.

Gilbert, R.J. and C. Shapiro, 1990, "Optimal Patent Length and Breath," Rand Journal of Economics, 21, pp. 106-112.

Goettler, R.L. and B.R. Gordon, 2011, "Does AMD spur Intel to innovate more?," Journal of Political Economy, 119 (6), pp. 1141-1200.

Goettler, R.L. and B.R. Gordon, 2012, "Competition and product innovation in dynamic oligopoly?," Technical Report, University of Chicago, working paper.

Gowrisankaran, G. and J. Stavins, 2004, "Network Externalities and Technology Adoption: Lessons from Electronic Payments," Rand Journal of Economics, 35, pp. 260-276.

Hsieh, C.-T. and E. Moretti, 2003, "Can Free Entry Be Inefficient? Fixed Commissions And Social Waste In The Real Estate Industry, " Journal of Political Economy, v111(5,Oct), pp. 1076-1122.

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

Kim, M. 2013, "Strategic responses to used-goods markets: Airbus and Boeing since 1997," mimieo.

Klepper, S., 2002, "Firm Survival and the Evolution of Oligopoly," Rand Journal of Economics, Vol. 33, pp. 37-61.

Klette T.J. and S. Kortum, 2004, "Innovating Firms and Aggregate Innovation, " Journal of Political Economy, Vol. 112, No. 5, pp. 986-1018.

Mankiw, N.G. and Whinston, M.D.,1986, "Free Entry and Social Inefficiency," Rand Journal of Economics, Vol. 17, pp. 48-58.

Mazzeo, M., 2002, "Product Choice and Oligopoly Market Structure," The Rand Journal of Economics, Vol. 33, No. 2, pp. 221-242.

McChesney, F.S., 1987, "Rent Extraction and Rent Creation in the Economic Theory of Regulation," The Journal of Legal Studies 16(1), pp. 101-118.

Perry, M.K., 1984, "Scale Economies, Imperfect Competition, and Public Policy," Journal of Industrial Economics, Vol. 32, pp. 313-330.

Pesendorfer, M., 2003, "Horizontal Mergers in the Paper Industry," RAND Journal of Economics, Vol. 34(3), pp. 495-515.

Riordan, M.H., 1992, "Regulation and Preemptive Technology Adoption," Rand Journal of Economics, Vol. 23, No. 3, pp. 334-349

Schaumans, C. and F. Verboven, 2008, "Entry and regulation: evidence from health care professions, " The RAND Journal of Economics, 39(4), pp. 949-972.

Schmidt-Dengler, 2006, "The Timing of New Technology Adoption: The Case of MRI," manuscript.

Seim, K., 2006, "An Empirical Model of Firm Entry with Endogenous Product-Type Choices, " The RAND Journal of Economics, Vol. 37, No. 3, pp. 619-640

Shleifer, A. and R. W. Vishny, 1993, "Corruption " The Quarterly Journal of Economics, 108(3), pp. 599-617.

Siebert, R.B. and C. Zulehner, 2013, "The Impact of Innovation and Market Demand on Market Structure," manuscript.

Spence, A.M., 1976a, "Product Selection, Fixed Costs, and Monopolistic Competition," Review of Economic Studies, Vol. 43, pp. 217-236.

Spence, A.M., 1976b, "Product Differentiation and Welfare, " American Economic Review, Vol. 66, pp. 407-414.

Spence, A.M., 1981, "The Learning Curve and Competition," The Bell Journal of Economics, 12(1), pp. 49-70.

Sutton, J., 1991, "Sunk Costs and Market Structure," MIT Press.

Suzuki, J., 2013, "Land Use Regulation as a Barrier to Entry: Evidence from the Texas Lodging Industry," International Economic Review 54(2): pp. 495-523.

Sweeting, A., 2006, "Coordination, Differentiation and the Timing of Radio Commercials," Journal of Economics and Management Strategy, Vol. 15, No. 4, pp. 909-942.

Sweeting, A., 2009, "The Strategic Timing of Radio Commercials: An Empirical Analysis Using Multiple Equilibria," Rand Journal of Economics, Vol. 40, No. 4, pp.710-742.

Toivanen, O. and M. Waterson, 2000, "Empirical Research on Discrete Choice Game Theory Models of Entry: An Illustration," European Economic Review, Vol. 44(4-6), pp.985-992.

Toivanen, O. and M. Waterson, 2005, "Market Structure and Entry:Where's the Beef?," Rand Journal of Economics, Vol. 36, No.3, pp.680-699.

von Weizsacker, C.C., 1980, "A Welfare Analysis of Barriers to Entry," Bell Journal of Economics, Vol. 11, pp. 399-420.

# A Appendix: Tables

Table 1: Demand Estimation Results

| Variable | OLS | IV | OLS | IV |
|---|---|---|---|---|
| Constant | 16.706*** | 18.728*** | 16.903*** | 17.406*** |
| | (19.88) | (21.77) | (20.77) | (24.04) |
| $\ln(P)$ | -2.745*** | -3.187*** | | |
| | (-25.77) | (-22.15) | | |
| $\ln(P^{16K})$ | | | -2.717*** | -3.067*** |
| | | | (-13.47) | (-18.24) |
| $\ln(P^{64K})$ | | | -2.579*** | -2.751*** |
| | | | (-28.95) | (-29.31) |
| $\ln(P^{256K})$ | | | -2.804*** | -2.927*** |
| | | | (-20.57) | (-23.62) |
| $\ln(P^{1M})$ | | | -3.006*** | -3.185*** |
| | | | (-21.53) | (-22.71) |
| Dummy $64K$ | 2.041*** | 2.194*** | 1.846*** | 1.661*** |
| | (16.20) | (17.32) | (8.45) | (8.44) |
| Dummy $256K$ | 3.768*** | 4.149*** | 3.951*** | 3.742*** |
| | (20.15) | (22.52) | (14.65) | (15.44) |
| Dummy $1M$ | 5.201*** | 5.942*** | 5.889*** | 5.803*** |
| | (16.15) | (18.37) | (13.43) | (13.80) |
| $\ln(P^S)$ | 0.345* | 0.336** | 0.185 | 0.436** |
| | (1.74) | (2.02) | (1.01) | (2.43) |
| $\ln(GGDP)$ | 10.095 | 7.984 | 11.546 | 8.274 |
| | (0.79) | (0.60) | (0.95) | (0.71) |
| $\ln(P^{Dram})$ | 1.132*** | 1.223*** | 1.267*** | 1.217*** |
| | (10.48) | (10.86) | (10.31) | (9.92) |
| Time | 0.013* | -0.00003 | 0.023** | 0.014 |
| | (1.74) | (-0.00) | (2.54) | (1.53) |
| Number of observations | 314 | 310 | 314 | 310 |
| Adjusted R-squared | 0.858 | 0.817 | 0.868 | 0.844 |

The dependent variable is the logarithm of industry output ($\ln(q_t^k)$). As instruments, we use the price of silicon, cumulative total output in the current generation, and cumulative total output from the previous generation. The t-statistics are shown in parentheses below the parameter estimates, and *** (**, *) denotes a 99% (95%, 90%) level of significance.

Table 2: Marginal Cost Estimation Results

| Variable | (1) | (2) |
|---|---|---|
| $\ln(P^{sil})$ ($\rho_1$) | 1.008*** | 0.729*** |
| | (67.83) | (36.28) |
| Intragenerational own learning ($\rho_2$) | -0.344*** | -0.260*** |
| | (-61.00) | (-51.76) |
| Spillover learning ($\rho_3$) | 0.317*** | 0.308*** |
| | (6.14) | (5.19) |
| Intergenerational own learning ($\rho_4$) | 0.010*** | 0.0009** |
| | (4.63) | (2.46) |
| Dummy $64K$ | 0.239*** | 0.293*** |
| | (21.02) | (25.96) |
| Dummy $256K$ | 0.639*** | 0.731*** |
| | (52.10) | (57.62) |
| Dummy $1M$ | 1.008*** | 1.204*** |
| | (67.23) | (69.54) |
| Time | | -0.013*** |
| | | (-20.46) |
| Firm fixed effect | Yes | Yes |
| Number of observations | 5,835 | 5,835 |
| Adjusted R-squared | 0.907 | 0.913 |

Table 2 shows nonlinear least square results of the marginal cost function, as shown in equation (4). The dependent variable is the logarithm of the marginal costs. The t-statistics are shown in parentheses below the parameter estimates, and *** (**, *) denotes a 99% (95%, 90%) level of significance.

Table 3: Output Policy Estimation Results

| Variable | $64K$ |
|---|---|
| Constant | -5.971 |
| | (-1.04) |
| $GDP$ | 0.00000038*** |
| | (3.25) |
| $\ln(P^{DRAM})$ | 0.200** |
| | (1.98) |
| $\ln(P^{sil})$ | -0.391 |
| | (-1.48) |
| $\ln(n^{16K})$ | 0.430** |
| | (2.12) |
| $\ln(n^{64K})$ | -0.296 |
| | (-1.30) |
| $\ln(x_i^{16K})$ | -0.044 |
| | (-1.17) |
| $\ln(x_i^{64K})$ | 0.492*** |
| | (11.34) |
| $\ln(x_{-i}^{16K})$ | 0.643 |
| | (1.14) |
| $\ln(x_{-i}^{64K})$ | 0.013 |
| | (0.07) |
| $\hat{\nu}_i$ | -0.319** |
| | (-2.57) |
| Time | -0.094*** |
| | (-6.64) |
| Firm fixed effect | Yes |
| Number of observations | 1,693 |
| Adjusted R-squared | 0.782 |

The dependent variable is the logarithms of firm level output in the $64K$ generation. The t-statistics are shown in parentheses, and *** (**, *) denotes a 99% (95%, 90%) level of significance.

Table 4: Entry and Exit Policy Estimation Results

| Variable | Entry $64K$ | Exit $64K$ |
|---|---|---|
| Constant | -4.680$^{***}$ | -2.658$^{***}$ |
| | (-5.58) | (-3.62) |
| $\ln(n^{64K})$ | 0.170$^{***}$ | 1.104$^{*}$ |
| | (2.71) | (1.91) |
| $\ln(x_i^{16K})$ | 0.046$^{**}$ | |
| | (2.31) | |
| $\ln(x_i^{64K})$ | | -0.139$^{***}$ |
| | | (-4.43) |
| $\ln(x_{-i}^{64K})$ | 0.051 | -0.395$^{*}$ |
| | (1.63) | (-1.71) |
| Time | 0.054$^{***}$ | 0.051$^{***}$ |
| | (7.10) | (2.70) |
| Firm fixed effect | Yes | No |
| Number of observations | 3,195 | 1,701 |
| Pseudo R-squared | 0.592 | 0.058 |

Table 4 shows the results for the entry and exit policies, as shown in equations (9) and (10). The entry and exit policies are estimated using probit models. The dependent variable in the entry model takes on a value of one when a firm chose to enter the $64K$ generation and zeros before entry occurred. In the exit model, the dependent variable takes on a value of one if a firm exits the generation and zero between the firm's entry and exit decisions. Firm fixed effects are not included in the exit models since not all firms have exited in the data. The z-statistics are shown in parentheses, and $^{***}$ ($^{**}$, $^{*}$) denotes a 99% (95%, 90%) level of significance.

# B Appendix: Figures

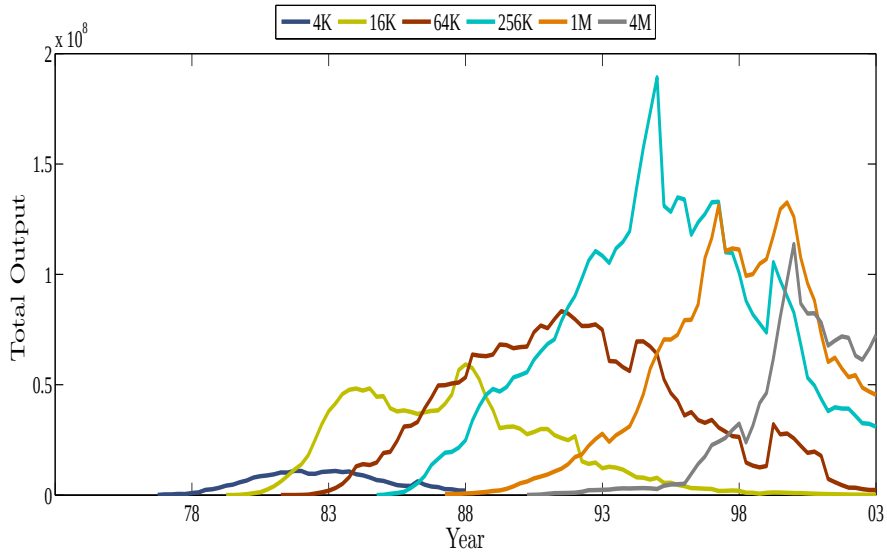


Figure 1: Industry units shipped, 1974-2003
Source: Gartner Inc.

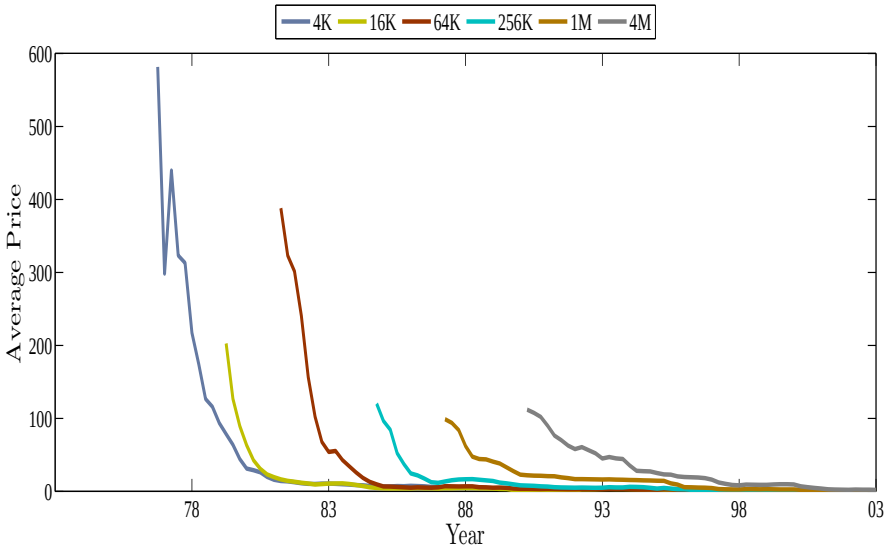


Figure 2: Average SRAM prices, 1974-2003
Source: Gartner Inc.

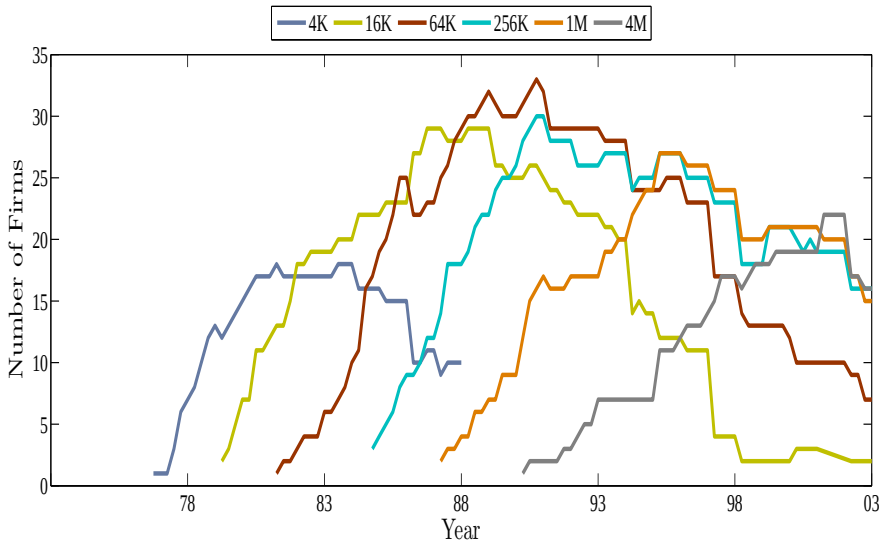


Figure 3: Number of Firms
Source: Gartner Inc.

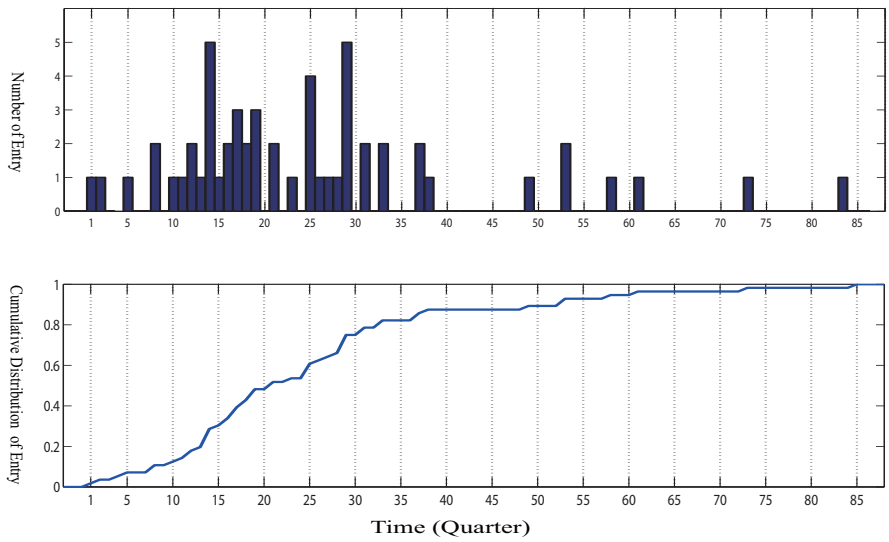


Figure 4: Distribution of Entry 64K
Source: Gartner Inc. The first entry occurs in the first quarter of 1982.

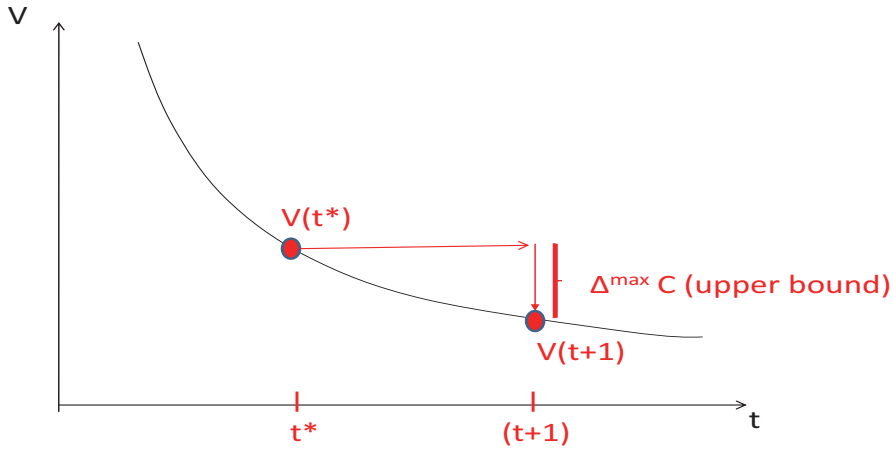


Figure 5: Identification of Entry Cost (Upper Bound)

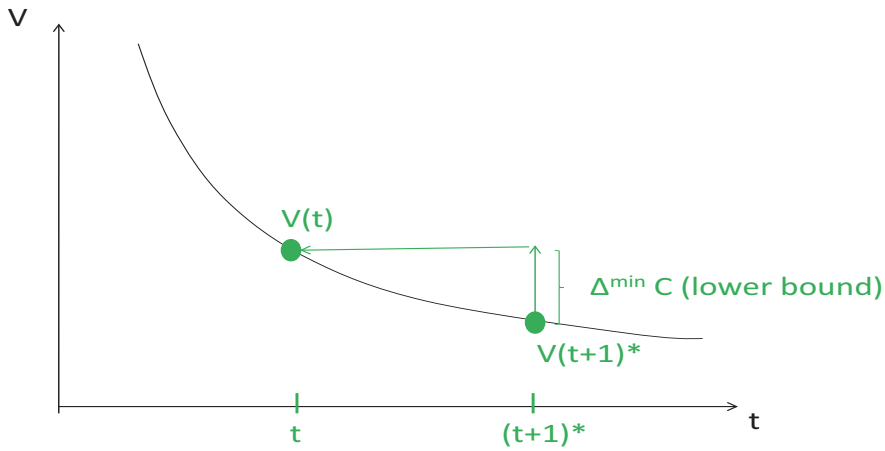


Figure 6: Identification of Entry Cost (Lower Bound)

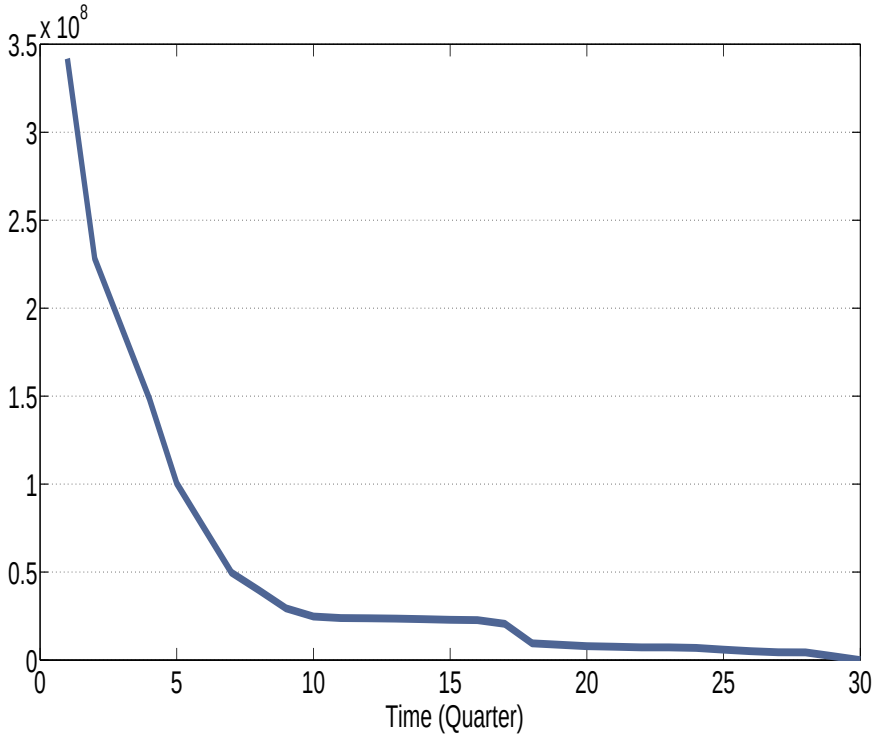


Figure 7: Slope of Entry Costs for the $64K$ generation

All values are the discounted values at the timing of the first entry in US dollars.

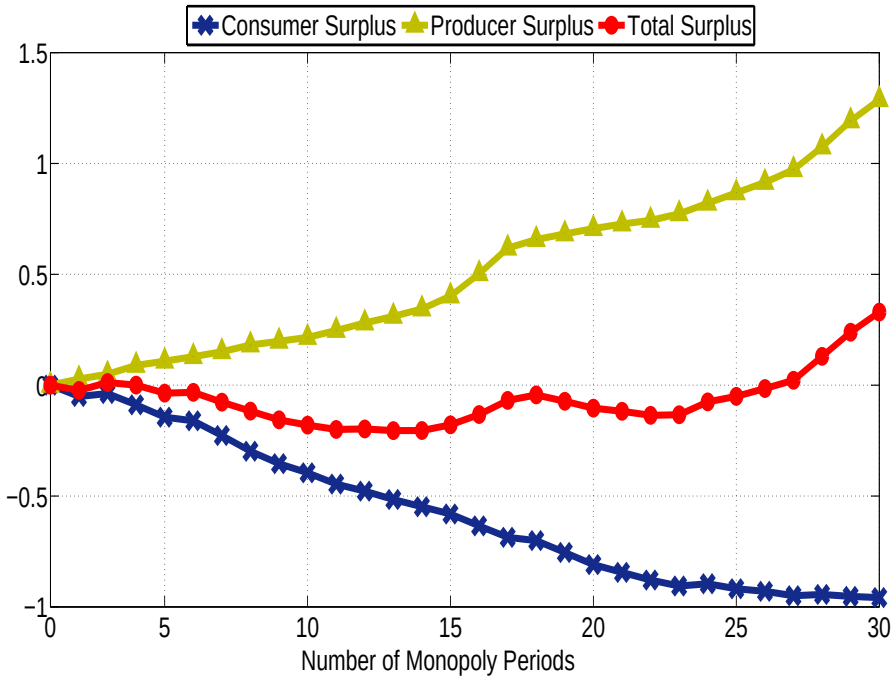


Figure 8: Change in Welfare

All values are the discounted values at the timing of the first entry in US dollars.

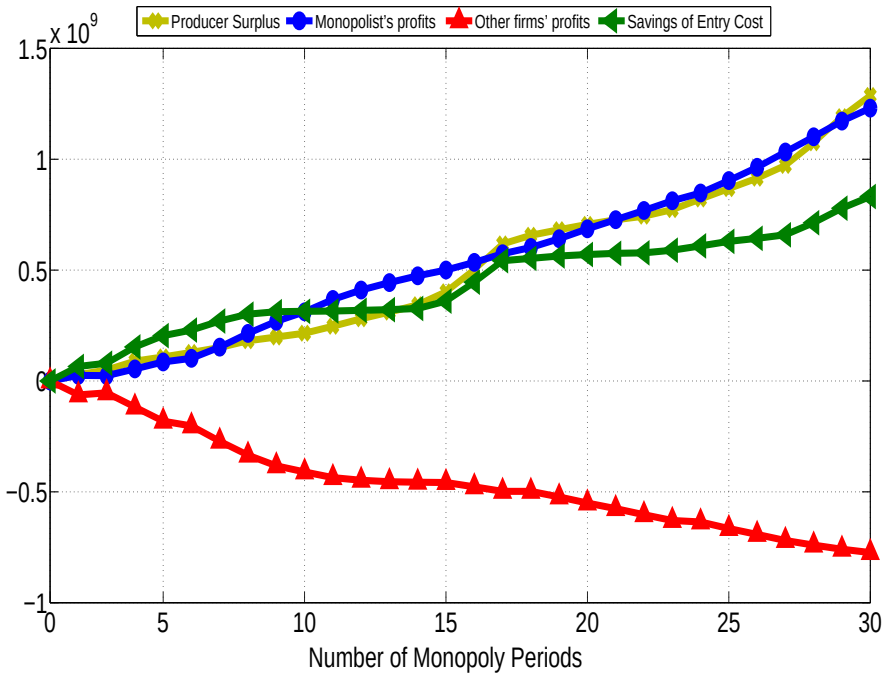


Figure 9: Change in Producer Surplus

All values are the discounted values at the timing of the first entry in US dollars.