Mergers, Innovation, and Entry-Exit Dynamics: Consolidation of the Hard Disk Drive Industry, 1996–2015*

Mitsuru Igami† Kosuke Uetake‡

July 28, 2016

Abstract

How far should an industry be allowed to consolidate when competition and innovation are endogenous? We extend Rust's (1987) framework to incorporate a stochastically alternating-move game of dynamic oligopoly, and estimate it using data from the hard disk drive industry, in which a dozen global players consolidated into only three in the last 20 years. We find plateau-shaped equilibrium relationships between competition and innovation, with systematic heterogeneity across time and productivity. Our counterfactual simulations suggest the optimal policy should stop mergers when five or fewer firms exist, highlighting a dynamic welfare tradeoff between ex-post pro-competitive effects and ex-ante value-destruction side effects.

Keywords: Antitrust, Competition and innovation, Dynamic oligopoly, Dynamic welfare tradeoff, Entry and exit, Horizontal mergers, Industry consolidation.

JEL classifications: L13, L41, L63, O31.

*For comments, we thank John Rust, Paul Ellickson, April Franco, Joshua Gans, and Allan Collard-Wexler, who discussed earlier versions of the paper, as well as participants at various seminars and conferences. For inside stories and insights, we thank Jeff Burke, Tu Chen, Finis Conner, MyungChan Jeong, Peter Knight, Currie Munce, Reggie Murray, Orie Shelef, and Lawrence Wu, as well as Mark Geenen and his team at *TRENDFOCUS*, including John Chen, Don Jeanette, John Kim, and Tim Luehmann. Financial support from the Yale Center for Customer Insights is gratefully acknowledged.

†Yale Department of Economics. E-mail: mitsuru.igami@yale.edu.

‡Yale School of Management. Email: kosuke.uetake@yale.edu.

1 Introduction

How far should an industry be allowed to consolidate? This question has been foundational for antitrust policy since its inception in 1890 as a countermeasure to merger waves (c.f., Lamoreaux 1985). Conventional merger analysis takes a proposed merger as given and focuses on its immediate effects on competition, which is expected to decrease after a target firm exits, and efficiency, which might increase if sufficient "synergies" materialize.[1] Such a static analysis would be appropriate if mergers were completely random events in isolation from competition and innovation, and if market structure and firms' productivity evolved exogenously over time. However, Demsetz (1973) cautioned that monopolies are often endogenous outcomes of competition and innovation. Berry and Pakes (1993) conjectured such dynamic factors could dominate static factors. Indeed, in 100% of high-tech merger cases, the antitrust authority has tried to assess potential impacts on innovation but found little guidance in the economics literature.[2] This paper proposes a tractable dynamic oligopoly model in which mergers, innovation, and entry-exit are endogenous, estimates it using data from the process of industry consolidation among the manufacturers of hard disk drives (HDDs) between 1996 and 2015, and quantifies a dynamic welfare tradeoff by simulating hypothetical merger policies.

Mergers in innovative industries represent an opportunity to kill competition and acquire talents, which make them strategic and forward-looking choices of firms.[3] Besides the static tradeoff between market power and efficiency, merger policy needs to consider both ex-post and ex-ante impact. Ex post, a merger reduces the number of competitors and alters their productivity profile, which will change the remaining firms' incentives for subsequent mergers and innovation. Theory predicts mergers are strategic complements (e.g., Qiu and Zhou 2007); hence, a given merger increases the likelihood of subsequent mergers. Its impact on subsequent innovation is more complicated because the competition-innovation relationship crucially hinges on the parameters of demand, supply, and investment (e.g., Sutton 1998). These ex-post changes in competition and innovation will have ex-ante impacts as well, because a tougher antitrust regime will lower firms' expected profits and option values of staying in the market, which will in turn reduce their ex-ante investments in productivity, survival,

[1] See Williamson (1968), Werden and Froeb (1994), and Nevo (2000), for example.

[2] See survey by Gilbert and Greene (2015).

[3] According to Reggie Murray, the founder of Ministor, "Most mergers were to kill competitors, because it's cheaper to buy them than to compete with them. Maxtor's Mike Kennan said, 'We'd rather buy them than have them take us out,' referring to Maxtor's acquisition of Quantum in 2001" (January 22, 2015, in Sunnyvale, CA). See Appendix A for a full list of interviews with industry veterans.

and market entry. Thus, merger policy faces a tradeoff between the ex-post pro-competitive effects and the ex-ante value-destruction side effects. Their exact balance depends on the parameters of demand, cost, and investment functions; hence, the quest for optimal merger policy is a theoretical as well as empirical endeavor.

Three challenges haunt the empirical analysis of merger dynamics in the high-tech context. First, mergers in a concentrated industry are rare events by definition, and the nature of the subject precludes the use of experimental methods; hence, a model has to complement sparse data. Second, an innovative industry operates in a nonstationary environment and tends to feature a globally concentrated market structure,[4] which creates a methodological problem for the application of two-step estimation approaches, because (at most) only one data point exists in each state of the world, which is too few for nonparametric estimation of conditional choice probabilities (CCPs). Third, workhorse models of dynamic oligopoly games such as Ericson and Pakes (1995) entail multiple equilibria, which preclude the application of full-solution estimation methods such as Rust (1987), because parameter estimates will be inconsistent when a single vector of parameter values predicts multiple strategies and outcomes. We solve these problems by developing a tractable model with unique equilibrium, incorporating the nonstationary environment of the HDD industry, and extending Rust's framework to a dynamic game with stochastically alternating moves.

The paper is organized as follows. In section 2, we introduce a simple model of a dynamic oligopoly with endogenous mergers, innovation, and entry-exit. We depart from the simultaneous-move tradition of the literature and adopt sequential or alternating moves. An unsatisfactory feature of a sequential-move game is that the assumption on the order of moves will generate an artificial early-mover advantage if the order is deterministic (e.g., Gowrisankaran 1995, 1999; Igami 2015, 2016). Instead, we propose a random-mover dynamic game in which the turn-to-move arrives stochastically. Dynamic games with stochastically alternating moves have been used as a theoretical tool since Baron and Ferejohn (1989) and Okada (1996). Iskhakov, Rust, and Schjerning (2014, 2016) used it to numerically analyze competition and innovation. We find it useful as an empirical model as well. We combine this random-mover modeling with the HDD market's fundamental feature that the industry is now mature and declining: a finite horizon. With a finite horizon and stochastically alternating moves, we can solve the game for a unique equilibrium by backward induction from the final period, in which profits and values become zero. At most only one firm moves within a period and makes a discrete choice between exit, investment in productivity, or

[4] Sutton (1998) explains this feature by low transport costs (per value of product) and high sunk costs.

merger proposal to one of the rivals. Thus, the dynamic game becomes a finite repetition of an effectively single-agent discrete-choice problem. We estimate the sunk costs associated with these discrete alternatives by using Rust's (1987) maximum-likelihood method with the nested fixed-point (NFXP) algorithm.

In section 3, we describe key features of the HDD industry and the outline of data. This high-tech industry has experienced massive waves of entry, shakeout, and consolidation, providing a suitable context for studying the dynamics of mergers and innovation. We explain several product characteristics and institutional backgrounds that inform our subsequent analysis, such as fierce competition among undifferentiated "brands" and an industry-wide technological trend called Kryder's Law (i.e., exogenous technological improvements in areal density).[5] Our dataset consists of three elements. Panel A contains aggregate HDD shipments, HDD price, disk price, and PC shipments, which we use to estimate demand in section 4.1. Panel B is firm-level market shares, which we use to estimate variable costs and period profits in section 4.2. Panel C records firms' dynamic choices between merger, innovation, and entry-exit, which we use to estimate sunk costs in section 4.3.

In section 4, we take three steps to estimate (i) demand, (ii) variable costs, and (iii) sunk costs, respectively, each of which pairs a model element and a data element as follows. In section 4.1, we estimate a log-linear demand model from the aggregate sales data in Panel A, treating each gigabyte (GB) as a unit of homogeneous data-storage services. We use two cost shifters as instruments for prices: the price of disks (key components of HDDs) and a time trend, both of which reflect Kryder's Law.[6] To control for demand-side dynamics that could arise from the repurchasing cycle of personal computers (PCs), we also include PC shipments as a demand shifter.

In section 4.2, we infer the implied marginal cost of each firm in each period from the observed market shares in Panel B, based on the demand estimates in section 4.1 and a Cournot model (with heterogeneous costs across firms) as a mode of spot-market competition. The firm's first-order condition (FOC) provides a one-to-one mapping from its observed market share to its marginal cost (productivity). Our preferred interpretation of Cournot competition is Kreps and Scheinkman's (1983) model of quantity pre-commitment followed by price competition, given all firms' cost functions (i.e., productivity levels). Effective production capacities are highly "perishable" in our high-tech context, because Kryder's

[5] Kryder's Law is an engineering regularity that says the recording density (and therefore storage capacity) of HDDs doubles approximately every 12 months, just like Moore's Law, which says the circuit density (and therefore processing speeds) of semiconductor chips doubles every 18-24 months

[6] The modeling of Kryder's Law is beyond the scope of this paper, and we regard this industry-wide trend as an exogenous technological process that progresses deterministically. See also section 6.

Law makes old manufacturing equipment obsolete within a few quarters. Hence, our notion of "quantity pre-commitment" is the amount of re-tooling efforts each firm makes in each quarter, which determines its effective output capacity for that period. Likewise, the real-world counterpart to our notion of cost (productivity) is intangible assets, such as the state of tacit knowledge embodied by teams of engineers, rather than durable physical capacities. Our profit-margin estimates strongly correlate with accounting profit margins in the firms' income statements.

In section 4.3, we estimate the sunk costs of merger, innovation, and entry, based on the observed choice patterns in Panel C and the benefits of these actions (i.e., streams of period profits) from section 4.2. Our dynamic discrete-choice model in section 2 provides a clear mapping from the observed choices and their associated benefits to the implied costs of these choices, which is analogous to the way Cournot FOC mapped output data and demand elasticity into implied costs. For example, if we observe many mergers despite small incremental profits, the model will reconcile these observations by inferring a low cost of merger: revealed preference.[7] Our firm-value estimates match closely with the actual acquisition prices in the historical merger deals.

In section 4.4, we investigate the equilibrium relationships between innovation, merger, and market structure, based on our estimates of optimal strategies (i.e., CCPs of innovation and merger) from section 4.3. Three patterns emerge. First, the incentive to innovate increases steeply as the number of firms increases from 1 to 3, reflecting the dynamic preemption motives as in Gilbert and Newbery (1982). This pattern is robust across years and productivity levels. Second, this competition-innovation relationship becomes heterogeneous and nonmonotonic with more than three firms: (i) the innovation rate increases at a decreasing rate monotonically at high-productivity firms and in early years; (ii) it is flat at mid-level firms and in middle years; and (iii) it often decreases at low-level firms and in late years. Thus, our structural competition-innovation curve exhibits a "plateau" shape instead of the famous "inverted U." Moreover, this systematic heterogeneity suggests (high) continuation values are a key factor in sustaining the (positive) competition-innovation relationship. Third, mergers become more attractive as the industry matures, and all kinds of pairs can merge. But high types tend to acquire more often, and low types are more popular as targets, because high types gain more from increased concentration, and low types are a

[7] Computationally, the calculation of the likelihood function is the heaviest part because, for each candidate vector of parameter values, we use backward induction to solve a *nonstationary* dynamic game with 8 different types of firms and 77,520 industry states in each of the 360 periods. We perform this subroutine in C++, and the estimation procedure takes less than a week.

cheaper means for stochastic productivity gains (i.e., synergies).

In section 5, we conduct counterfactual policy simulations to answer our main question: How far should the industry should be allowed to consolidate? In section 5.1, we find the optimal static (or "commitment") policy is to block mergers if five or fewer firms exist. In section 5.2, we clarify the underlying mechanism behind this finding by decomposing the dynamic welfare tradeoff between the ex-post pro-competitive benefits of blocking mergers and the ex-ante value-destruction side effects, which reduce both competition and innovation in early years. In section 5.3, we find this optimal policy threshold ($\underline{N} = 5$) can be relaxed slightly if the industry is declining quickly and firms are failing anyway. In section 5.4, we find the optimal dynamic (or ex-post "surprise") policy is to initially promise no merger enforcement at all ($\underline{N}^{ante} = 1$) and then block mergers once the industry reaches three firms ($\underline{N}^{post} = 3$). This policy, however, relies entirely on the authority's ability to surprise (and the naïveness of firms); hence, its feasibility and desirability are dubious in the long run.

We conclude in section 6 by discussing other policy implications and limitations. The current de-facto policy of $\underline{N} = 3$ is somewhat stricter than our optimal threshold of $\underline{N} = 5$ for the HDD industry, but the welfare outcomes under these two policies are not drastically different. By contrast, our results suggest allowing mergers to duopoly or monopoly ($\underline{N} = 2$ or 1) will have a negative welfare impact that is orders of magnitude larger. Thus, our main message is "2 are few and 6 are many."[8]

1.1 Literature Context

Dynamic welfare tradeoff is a classical theme in the literature on market structure and innovation (c.f., Scotchmer 2004). Tirole (1988, p. 390) summarizes Schumpeter's (1942) basic argument that "if one wants to induce firms to undertake R&D one must accept the creation of monopolies as a necessary evil." He then proceeds to discuss this "dilemma of the patent system" but concludes that "the welfare analysis is relatively complex, and more work is necessary before clear and applicable conclusions will be within reach" (p. 399), which is exactly the purpose of this paper.

Traditional oligopoly theory suggests the main purpose of mergers is to kill competition and increase market power. Stigler (1950) added a twist to this thesis by conjecturing that, because a merger increases concentration at the industry level and non-merging parties can free-ride on merging parties' efforts, no firms would want to take initiatives to merge. Salant, Switzer, and Reynolds (1983) proved this idea in a symmetric Cournot model, although Perry

[8]We do not model collusion, but our finding resonates with Selten's (1973) "4 are few and 6 are many."

and Porter (1985) and Deneckere and Davidson (1985) revealed the fragility of the free-riding result, which crucially relied on symmetry across firms. Farrell and Shapiro (1990) used a Cournot model with cost heterogeneity across firms, and formalized the notion of "synergy" as an improvement in the marginal cost of merging firms (above and beyond the convergence of the two parties' pre-merger productivity levels). We follow their modeling approach and definition of synergy. The latest reincarnations of this strand is Mermelstein, Nocke, Satterthwaite, and Whinston's (2014, henceforth MNSW) numerical theory of duopoly with mergers and investments, which Marshall and Parra (2015) extend to more general market structures. We provide a structural empirical companion to this literature.

Rust (1987) pioneered the empirical methods for dynamic structural models by combining dynamic programming and discrete-choice modeling, and proposed a full-solution estimation approach. Much of the empirical dynamic games literature has evolved within Ericson and Pakes's (1995) framework, and two-step methods have been developed to estimate this class of models.[9] However, typical empirical contexts of innovative industries (i.e., nonstationarity and global concentration) pose practical challenges to these methods, which led us to propose the pairing of a random-mover dynamic game (in a nonstationary environment and a finite horizon, as in Pakes 1986) with Rust's estimation approach.[10]

Applications of dynamic games to mergers include Gowrisankaran's (1995, 1999) pioneering computational work, Stahl (2011), and Jeziorski (2014). Applications to innovation include Benkard (2004), Goettler and Gordon (2011), Kim (2015), and Igami (2015, 2016).[11] Applications to entry and exit are the largest literature, including Ryan (2012), Collard-Wexler (2013), Takahashi (2015), Arcidiacono, Bayer, Blevins, and Ellickson (2015), and Igami and Yang (2016). We have not found any empirical application of stochastically alternating-move games, but Iskhakov, Rust, and Schjerning (2014, 2016) numerically study Bertrand duopoly with "leap-frogging" process innovations.

2 Model

This section describes our empirical model. Our goal is to incorporate a dynamic oligopoly game of mergers and innovation within Rust's (1987) dynamic discrete-choice model.

[9] Aguirregabiria and Mira (2007); Bajari, Benkard, and Levin (2007); Pakes, Ostrovsky, and Berry (2007); Pesendorfer and Schmidt-Dengler (2008).

[10] Egesdal, Lai, and Su (2015) propose MPEC algorithm as an alternative to NFXP, which is conceptually feasible but currently impractical for nonstationary, sequential-move games, due to extensive use of memory. See Iskhakov, Lee, Rust, Schjerning, and Seo (2016) for a recent tune-up to NFXP.

[11] Ozcan (2015) and Entezarkheir and Moshiri (2015) analyze panel data on patents and mergers.

2.1 Setup

Time is discrete with a finite horizon, $t = 0, 1, 2, \ldots, T$, where the final period T is the time at which the demand for HDDs becomes zero. Each of the finite number of incumbent firms, $i = 1, 2, \ldots n_t$, has its own productivity on a discretized grid with unit interval, $\omega_{it} \in \{\omega^1, \omega^2, \ldots \omega^{\max}\}$, which represents the level of tacit knowledge embodied by its team of R&D engineers and manufacturing engineers. Given the productivity profile, $\omega_t \equiv \{\omega_{it}\}_{i=1}^{n_t}$, these incumbents participate in the HDD spot market and earn period profits, $\pi_{it}(\omega_t)$. Thus, ω_t constitutes the payoff-relevant state variable along with the time period t, which subsumes both the time-varying demand situation and the industry-wide technological trend (i.e., Kryder's Law). We specify and estimate $\pi_{it}(\omega_t)$ in section 4.

We assume a potential entrant (denoted by $i = 0$ and state ω^0) exists in every period and chooses whether to enter or wait when its turn-to-move arrives.[12] Upon entry, it becomes active at the lowest productivity level, $\omega_{i,t+1} = \omega^1$. If it stays out, $\omega_{i,t+1} = \omega^0$. Each of the two actions entails a sunk cost and an idiosyncratic cost shock, κ^{a0} and $\varepsilon(a_{it}^0)$, where $a^0 \in A^0 = \{enter, out\}$. An incumbent chooses between exit, innovation, merger, and staying alone without taking any major action (which we call "idling"), when its turn arrives. Each of these dynamic actions, $a \in A = \left\{exit, innovate, \{\text{propose merger to rival } j\}_{j \neq i}, idling\right\}$, entails a sunk cost, κ^a, and an idiosyncratic cost shock, $\varepsilon(a_{it})$. We follow Rust (1987) to assume $\varepsilon(a_{it}^0)$ and $\varepsilon(a_{it})$ are independently and identically distributed (i.i.d.) type-1 extreme value.

The three actions by incumbents induce the following transitions of ω_{it}. First, all exits are final and imply liquidation, after which the exiter reaches an absorbing state, $\omega_{i,t+1} = \omega^{00}$ ("dead"). Second, innovation in the HDD context involves the costly implementation of retooling or upgrading of manufacturing equipment to improve quality-adjusted productivity,[13] $\omega_{i,t+1} = \omega_{it} + 1$. Third, an incumbent may propose merger to one of the other incumbents and enter a bilateral bargaining. We consider two bargaining protocols: (i) Nash bargaining with equal bargaining powers between the acquirer and the target (henceforth "NB"), and (ii) take-it-or-leave-it offer by the acquirer to the target ("TIOLI").

Horizontal mergers and synergies in the HDD context are not so much about the realloca-

[12] In our data, entry had all but ceased by January 1996 (i.e., the beginning of our sample period) and our main focus is on the process of consolidation, but we incorporate entry to keep our model sufficiently general, so that it can be applied to the entire life cycle of an industry in principle. Another reason is that at least one episode of entry actually existed. Finis Conner founded Conner Technology in the late 1990s.

[13] We say "quality-adjusted" productivity because the industry-wide technological trend is always improving product quality (in terms of areal density) at a deterministic rate according to Kryder's Law; hence, the ω_{it}s here should be understood as the de-trended version of raw productivity.

tion of tangible assets (e.g., physical production capacities), which are "perishable" and tend to become obsolete within a few quarters anyway, as about combining teams of engineers who embody tacit knowledge.[14] Thus, a natural way to model the evolution of post-merger productivity is to follow Farrell and Shapiro (1990) and specify $\omega_{i,t+1} = \max\{\omega_{it}, \omega_{jt}\} + \Delta_{i,t+1}$, where i and j are the identities of the acquirer and the target, respectively, and $\Delta_{i,t+1}$ is the realization of stochastic improvement in productivity. The first term on the right-hand side reflects the convergence of the merging parties' productivity levels, which Farrell and Shapiro called "rationalization," and the second term represents what they called "synergies." Given the discrete grid of ω_{it}'s (and the fact that mergers in a concentrated industry are rare events by definition), a simple discrete probability distribution is desirable; hence, we specify $\Delta_{i,t+1} \sim Poisson(\lambda)$ i.i.d., where λ is the expected value of synergy.

2.2 Timing

Standard empirical models of strategic industry dynamics such as Ericson and Pakes (1995) assume simultaneous moves in each period. However, if any of the n firms can propose merger to any other firm in the same period, every proposal becomes a function of the other $n(n-1)-1$ proposals, which will lead to multiple equilibria. Instead, we consider an alternating-move game in which the time interval is relatively short and only (up to) one firm has an opportunity to make a dynamic discrete choice within a period. Gowrisankaran (1995, 1999) and Igami (2015, 2016) are examples of such formulation with deterministic orders of moves, but researchers usually do not have theoretical or empirical reason to favor one specific order over the others. A deterministic order is particularly undesirable for analyzing endogenous mergers, because early-mover advantages will translate into stronger bargaining powers, tilting the playing field and equilibrium outcomes in favor of certain firms.

For these reasons, we use stochastically alternating moves and model the timeline within each period as follows.

1. Nature chooses at most one firm (say i) with "recognition" probability, ρ_i, at the beginning of each period.

2. Mover i observes the current industry state, ω_t, forms rational expectations about its future evolution, $\{\omega_\tau\}_{\tau=t+1}^{T}$, and draws i.i.d. shocks, $\varepsilon(a_{it})$, which represent random

[14] According to Currie Munce of HGST, a big rationale for consolidation is that "As further improvement becomes technically more challenging, the industry has to pool people and talents, which would lead to further break-through" (February 27, 2015).

private costs associated with the dynamic actions. If i is an incumbent, $\varepsilon(a_{it})$ includes ε_{it}^{x}, ε_{it}^{c}, ε_{it}^{i}, and $\left\{\varepsilon_{ijt}^{m}\right\}_{j}$, for exit, idling, investment, and merger proposal to rival incumbent j, respectively. These target-specific ε_{ijt}^{m}'s represent transient and idiosyncratic factors, and do not enter merger negotiation.[15]

3. Based on these pieces of information and their implications, mover i makes the discrete choice $a_{it} \in A_{it}$, immediately incurring the associated sunk cost, κ^{a}, and the idiosyncratic cost shock, $\varepsilon(a_{it})$. If i is an incumbent and chooses to negotiate a potential merger with incumbent j, the two parties bargain over the acquisition price, p_{ij}, which is a dollar amount to be transferred from i to j upon agreement. Our baseline specification of the bargaining protocol is NB, but we also consider an alternative specification, TIOLI.[16] If the negotiation breaks down, no transfer takes place, i's turn ends without any other action or other merger negotiation, and j will remain independent.

4. All incumbent firms (regardless of the stochastic turn to move) participate in the spot-market competition, earn period profits, $\pi_{it}(\omega_t)$, and pay the fixed cost of operation, κ^{c}, which includes the costs of continual efforts to keep up with the industry-wide technological progress (i.e., Kryder's Law).

5. Mover i implements its dynamic action, and its state evolves accordingly. If i is merging, it draws stochastic synergy, $\Delta_{i,t+1}$, which determines the merged entity's productivity in the next period, $\omega_{i,t+1}$.

These steps are repeated T times until the industry comes to an end.

2.3 Dynamic Optimization and Equilibrium

Whenever its turn to move arrives, a firm makes a discrete choice to maximize its expected net present value. Its strategy, σ_i, consists of a mapping from its effective state (a vector of the productivity profile ω_t, time t, and the draws of $\varepsilon_{it} = \{\varepsilon(a_{it})\}_{a \in A}$) to a choice $a_{it} \in A_{it}$—a complete set of such mappings across all t, to be precise. We may integrate out ε_{it}

[15] For example, consider senior manager M, who goes to one of the numerous Irish pubs in Silicon Valley, bumps into a rival firm's manager, has a good time, and comes up with an idea of merger, after which he goes back to the headquarters and recommends the idea. The board agrees and sends out another manager, K, as their delegate. Manager K bargains with his counterpart, but neither of them knows or cares about Manager M's happy-hour experience that triggered the negotiation, that is, ε_{ijt}^{m}. We thank Allan Collard-Wexler for suggesting this interpretation over dinner.

[16] No systematic record exists on the actual merger negotiations, and the details are likely to be highly idiosyncratic. In the absence of solid evidence, we prefer keeping the specification as neutral as possible.

and consider σ_i as a collection of the ex-ante optimal choice probabilities conditional on $(\omega_{it}, \omega_{-it}, t)$.

The following Bellman equations characterize an incumbent firm's dynamic optimization problem.[17] Mover i's value *after* drawing ε_{it} is

$$V_{it}\left(\omega_t, \varepsilon_{it}\right) = \pi_i\left(\omega_t\right) + \max\left\{\bar{V}_{it}^x\left(\omega_t, \varepsilon_{it}^x\right), \bar{V}_{it}^c\left(\omega_t, \varepsilon_{it}^c\right), \bar{V}_i^i\left(\omega_t, \varepsilon_{it}^i\right), \left\{\bar{V}_{ijt}^m\left(\omega_t, \varepsilon_{ijt}^m\right)\right\}_j\right\}, \quad (1)$$

where $\bar{V}_{it}^a$s represent conditional (or "alternative-specific") values of exiting, idling, innovating, and proposing merger to rival j, respectively,

$$\begin{aligned}
\bar{V}_i^x\left(\omega_t, \varepsilon_{it}^x\right) &= -\kappa^x + \varepsilon_{it}^x + \beta E\left[\Lambda_{i,t+1}\left(\omega_{t+1}\right)|\omega_t, a_{it} = exit\right], && (2)\\
\bar{V}_i^c\left(\omega_t, \varepsilon_{it}^c\right) &= -\kappa^c + \varepsilon_{it}^c + \beta E\left[\Lambda_{i,t+1}\left(\omega_{t+1}\right)|\omega_t, a_{it} = idle\right], && (3)\\
\bar{V}_i^i\left(\omega_t, \varepsilon_{it}^i\right) &= -\kappa^c - \kappa^i + \varepsilon_{it}^i + \beta E\left[\Lambda_{i,t+1}\left(\omega_{t+1}\right)|\omega_t, a_{it} = invest\right], \text{ and} && (4)\\
\bar{V}_{ij}^m\left(\omega_t, \varepsilon_{ijt}^m\right) &= -\kappa^c - \kappa^m + \varepsilon_{ijt}^m - p_{ij}\left(\omega_t\right) + \beta E\left[\Lambda_{i,t+1}\left(\omega_{t+1}\right)|\omega_t, a_{it} = merge\ j\right]. && (5)
\end{aligned}$$

Mover i's value *before* drawing ε_{it} is

$$\begin{aligned}
EV_{it}\left(s_t\right) &= E_\varepsilon\left[V_{it}\left(s_t, \varepsilon_{it}\right)\right] && (6)\\
&= \pi_i\left(s_t\right) + \gamma + \ln\left[\exp\left(\tilde{V}_{it}^x\right) + \exp\left(\tilde{V}_{it}^c\right) + \exp\left(\tilde{V}_{it}^i\right) + \sum_{j \neq i}\exp\left(\tilde{V}_{ijt}^m\right)\right],
\end{aligned}$$

where γ is Euler's constant and $\tilde{V}_{it}^a$ is the deterministic part of $\bar{V}_i^a\left(\omega_t, \varepsilon_{it}^a\right)$, that is, $\tilde{V}_{it}^a \equiv \bar{V}_i^a\left(\omega_t, \varepsilon_{it}^a\right) - \varepsilon_{it}^a$. In equations 2 through 5, $\Lambda_{i,t+1}$ represents i's expected value at $t+1$ *before* nature picks a mover at $t+1$,

$$\Lambda_{i,t+1}\left(\omega_{t+1}\right) = \rho_i\left(\omega_{t+1}\right) EV_{i,t+1}\left(\omega_{t+1}\right) + \sum_{j \neq i}\rho_j\left(\omega_{t+1}\right) W_{i,t+1}^j\left(\omega_{t+1}\right). \quad (7)$$

This "umbrella" value is a recognition probability-weighted average of mover's value (EV_{it}) and non-mover's value $\left(W_{it}^j\right)$. Nobody knows exactly who will become the mover before nature picks one. When nature picks $j \neq i$, non-mover i's value (before j draws ε_{jt} and takes

[17] Appendix B features the corresponding expressions for the potential entrant.

an action) is

$$
\begin{aligned}
W_{it}^{j}\left(\omega_t\right) = \; & \pi_i\left(\omega_t\right) - \kappa^c + E_{it}\left[\Pr\left(a_{jt} = exit\right)\right] \beta E\left[\Lambda_{i,t+1}\left(\omega_{t+1}\right)|\omega_t, a_{jt} = exit\right] \\
& + E_{it}\left[\Pr\left(a_{jt} = idle\right)\right] \beta E\left[\Lambda_{i,t+1}\left(\omega_{t+1}\right)|\omega_t, a_{jt} = idle\right] \\
& + E_{it}\left[\Pr\left(a_{jt} = invest\right)\right] \beta E\left[\Lambda_{i,t+1}\left(\omega_{t+1}\right)|\omega_t, a_{jt} = invest\right] \\
& + E_{it}\left[\Pr\left(a_{jt} = merge\ i\right)\right] p_{ji}\left(\omega_t\right) \\
& + \sum_{k \neq i,j} E_{it}\left[\Pr\left(a_{jt} = merge\ k\right)\right] \beta E\left[\Lambda_{i,t+1}\left(\omega_{t+1}\right)|\omega_t, a_{jt} = merge\ k\right],
\end{aligned} \tag{8}
$$

where $E_{it}\left[\Pr\left(a_{jt} = action\right)\right]$ is non-mover i's belief over mover j's choice. These value functions entail the following ex-ante optimal choice probabilities:

$$
\Pr\left(a_{it} = action\right) = \frac{\exp\left(\tilde{V}_{it}^{action}\right)}{\exp\left(\tilde{V}_{it}^{x}\right) + \exp\left(\tilde{V}_{it}^{c}\right) + \exp\left(\tilde{V}_{it}^{i}\right) + \sum_{j \neq i} \exp\left(\tilde{V}_{ijt}^{m}\right)}. \tag{9}
$$

In equilibrium, these probabilities constitute the non-movers' beliefs over the mover's choice. We use these optimal choice probabilities to construct a likelihood function for estimation in section 4.3.

The bargaining protocol determines the equilibrium acquisition price, p_{ij}. Under NB, the two parties jointly maximize the following expression:

$$
\begin{aligned}
& \left\{\beta E\left[\Lambda_{i,t+1}\left(\omega_{t+1}\right)|\omega_t, merge\ j\right] - p_{ij} - \beta \Lambda_{i,t+1}\left(\omega_{t+1} = \omega_t\right)\right\}^{\zeta} \\
& \times \left\{p_{ij} - \beta \Lambda_{j,t+1}\left(\omega_{t+1} = \omega_t\right)\right\}^{1-\zeta},
\end{aligned} \tag{10}
$$

where $\zeta \in [0, 1]$ represents the bargaining power of the acquirer (i here), which equals .5 under NB (with 50-50 split) and 1 under TIOLI. The last term in each bracket is the disagreement payoff.[18]

We solve this dynamic game for a unique sequential equilibrium in pure strategies that are type-symmetric. Note that ε_{it}'s are i.i.d. shocks whose realizations do not affect anyone's future payoff except through the actual choice a_{it}; hence, we may solve this game by backward induction from the final period, T. At T, all firms' profits and continuation values are zero, so no decision problem exists. At $T-1$, a single mover (denoted by $i = T-1$) draws

[18] Only up to one deal (between i and j here) can be negotiated within a period.This setting is not as restrictive as it might seem at a first glance, because the time interval is relatively short and all other potential deals in the future are embedded in the disagreement payoff (i.e., each firm's stand-alone continuation value). This specification shares the flavor of Crawford and Yurukoglu (2012) and Ho (2009).

ε_{T-1} and takes whichever action a_{T-1} maximizes its expected net present value. At $T-2$, another mover ($i = T-2$) draws ε_{T-2} and makes its discrete choice, in anticipation of (i) the evolution of ω_t from $T-2$ to $T-1$, (ii) the recognition probabilities and other common factors, and (iii) the optimal CCPs of all types of potential movers at $T-1$, which imply the transition probabilities of ω_t from $T-1$ to T. This iterative process repeats itself until the initial period $t = 0$.

An equilibrium exists and is unique. First, each of the (at most) T discrete-choice problems has a unique solution given the i.i.d. draws from a continuous distribution. Second, in each period t, only (up to) one firm solves this problem in our alternating-move formulation. Third, mover t's choice completely determines the transition probability of ω_t to ω_{t+1}, but it cannot affect future movers' optimal CCPs at $t+1$ and beyond in any other way. In other words, this game is effectively a sequence of T single-agent problems. By the principle of optimality, we can solve it by backward induction for a unique equilibrium.

2.4 Other Modeling Considerations

To clarify our modeling choices, we discuss five alternative modeling possibilities that we have considered: (i) an infinite horizon, (ii) continuous time, (iii) heterogeneous recognition probabilities, (iv) alternative bargaining protocols, and (v) private information on synergies.

First, we have chosen a finite horizon over an infinite one primarily because we study the process of industry consolidation in an innovative, nonstationary industry. Another reason is multiple equilibria. Iskhakov, Rust, and Schjerning (2016) find numerous equilibria in a stochastically alternating-move duopoly game of innovation with an infinite horizon. Multiple equilibria would preclude the use of full-solution estimation methods and counterfactual analysis.

Second, continuous time modeling is an attractive alternative, but Arcidiacono, Bayer, Blevins, and Ellickson (2015) acknowledge that the feasibility of its application to a nonstationary environment is currently unknown. Another problem with a shorter or infinitesimal time interval in our context is its potential conflict with the i.i.d. idiosyncratic shocks and timing assumptions. For major and infrequent decisions such as mergers, the actual decision making and implementation take at least a month or a quarter. Shorter intervals would imply firms draw i.i.d. random shocks every day or week. Incorporating a persistent unobserved state could alleviate this problem but create another technical challenge.

Third, some firms might be more active in M&A than others, and recognition probabilities can accommodate such heterogeneity. For example, making ρ_i depend on ω_{it} would be

conceptually straightforward, albeit computationally costly. One problem with this idea is that we have no theory. Another problem is identification. Because we have no theoretical or empirical foundation for a priori specification of asymmetric ρ_i's, we prefer keeping it symmetric and instead focus on the extent of heterogeneity *in the equilibrium CCP estimates.* Indeed, section 4.4 shows that high-type firms are more likely to acquire low-type firms.

Fourth, regarding the specifications NB and TIOLI, we may leave the bargaining powers, ζ, as free parameter and try to estimate them. However, mergers in a concentrated industry are rare events by definition, which leads to a data environment with only a handful of actual acquisition deals to estimate ζ. Thus, we pre-specify NB and TIOLI as alternative models, and implement both as a sensitivity analysis.

Fifth, regarding the nature of synergies, we may consider a more complicated model of synergies with private information (e.g., some firms might privately know their merger would yield particularly high $\Delta_{i,t+1}$), but we have chosen not to add more structures, for three reasons. First, such non-trivial private information will constitute unobserved state variables and generate a selection problem, which is an interesting problem but beyond the scope of this paper. Second, no systematic record exists on firms' subjective assessments of "chemistry;" hence, the identification of such factors would appear hopeless without strong additional assumptions anyway. Third, our simple model of $\Delta_{i,t+1}$ as a completely random draw actually seems the most consistent with our personal interview with Finis Conner, the co-founder of Seagate Technology, the founder of Conner Peripherals, and the founder of Conner Technology. Having founded two Fortune-500 companies in the HDD industry and engaged in some of the historical mergers, he is an embodiment of the industry's highest-quality private information. Nevertheless, he stated, "You have to dive into the water to see where the skeletons are," which means even an industry veteran would not know the internal functioning of the other firms sufficiently to predict the synergy realizations with much precision, until after the actual mergers take place.[19] Thus, ours is an empirical model of Finis Conner. We keep our synergy function simple, and conduct sensitivity analysis with respect to λ in section 4.3.

For these reasons, we believe our specification strikes the right balance amid many conceptual and practical challenges.

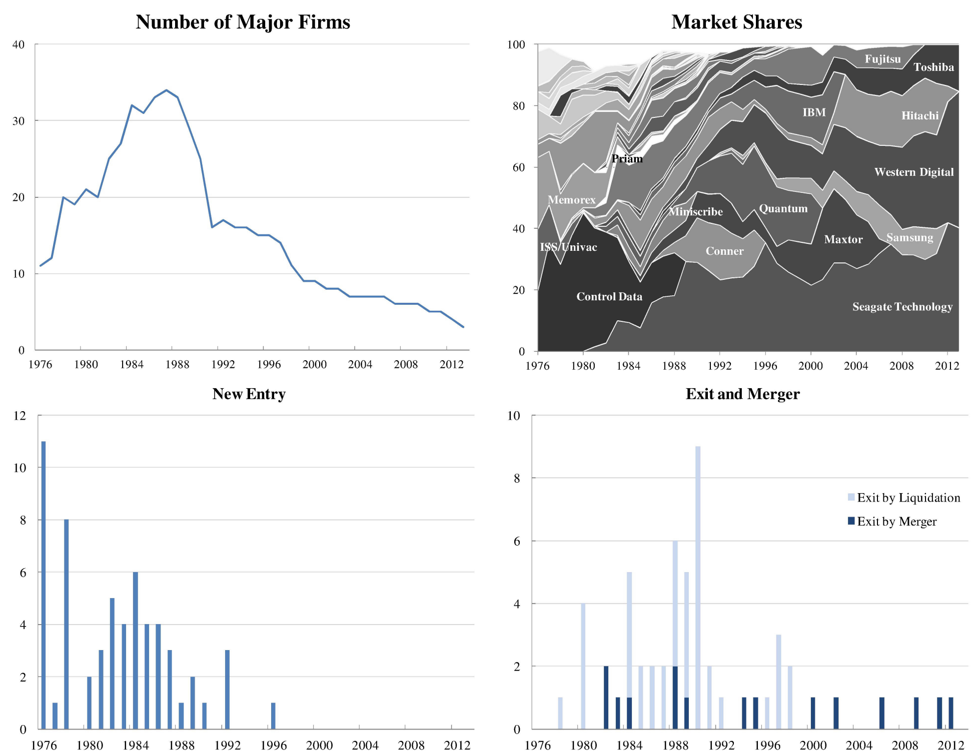
[19] From author's personal interview on April 20, 2015, in Corona del Mar, CA.

3 Data

3.1 Institutional Background and Product Characteristics

Computers are archetypical high-tech goods that store, process, and transmit data. HDDs, semiconductor chips, and network equipment perform these tasks, respectively. HDDs offer the most relevant empirical context to study mergers and innovation in the process of industry consolidation. The industry has experienced massive waves of entry and exit, followed by mergers among a dozen survivors (Figure 1).

Figure 1: Evolution of the World's HDD Industry



Note: The number of firms counts only the major firms with market shares exceeding 1% at some point of time. See Igami (2015, 2016) on product and process innovations during the 1980s and 1990s.

The manufacturing of HDDs requires engineering virtuosity in assembling heads, disks, and motors into an air-tight black box, managing volume production in a reliable and economical manner, and keeping up with the technological trend that constantly improves quality and efficiency (Kryder's Law).

Despite such complexity, HDDs are also one of the simplest products in terms of eco-

Figure 2: Product Characteristics of HDDs

Note: Left panel shows 3.5-inch HDDs of Hitachi GST, Western Digital, and Seagate Technology. Right panel shows evidence of successful marketing efforts by Microsoft and Intel (and lack thereof by HDD makers).

nomics because they are "completely undifferentiated product" according to Peter Knight, former vice president of Conner Peripherals and Seagate Technology, and former president of Conner Technology.[20] Consumers typically do not observe or distinguish "brands" (Figure 2, left). Moreover, HDDs are physically durable but do not drive the repurchasing cycle of PCs. Microsoft and Intel ("Wintel") do, as is evident from the fact that PC users tend to be aware of the technological generations of operating systems (OS) and central processing units (CPU) but not HDDs (Figure 2, right), which means the demand for HDDs can be usefully modeled within a static framework as long as we control for the PC shipments as a demand shifter. These product characteristics inform our demand analysis in section 4.1.

Two institutional features inform our analysis of the supply side in section 4.2. First, unlike car manufacturing in Japan (say), the manufacturers of PCs and HDDs do not engage in long-term contracts or relationships in a strict sense. The architecture of a PC is highly modular, and standardized interfaces connect its components, which makes different "brands" of HDDs technologically substitutable. Furthermore, "second sourcing" has long been a standard practice in the computer industry, by which a downstream firm keeps close contact with multiple suppliers of a key component so that a backup supplier or two will always exist in cases of accidental supply shortage at the primary one. According to Peter Knight, "Compaq, HP, nobody cared who makes their disk drives. They bought the lowest-price product that had reasonable quality. There was no reason for single-sourcing." Second, PC makers might appear to have consolidated as much as HDD makers, but the actual market structure of the global PC industry is more fragmented. The average combined market share of the top four vendors (i.e., CR4) between 2006 and 2015 is 52.5%, which is considered between "low" and "medium" concentration. By contrast, the HDD industry's

[20]From author's personal interview on June 30, 2015, in Cupertino, CA. See also section 4.1.

average CR4 is 91.6% during the same period.[21]

Finally, our data include some kind of solid-state drives (SSDs), but we do not explicitly model them, because (i) pure SSDs comprised less than 10% of industry sales even in the last five years of our sample period, (ii) they are made of NAND flash memory (a type of semiconductor devices), whose underlying technology is totally different from HDD's magnetic recording technology, and (iii) NAND flash memories are supplied by a different set of firms (i.e., semiconductor chip makers specialized in flash memories). Modeling SSDs means modeling the semiconductor industry. However, most SSDs for desktop PCs are actually hybrid HDDs which combine a small NAND part with HDDs. These hybrids are part of our HDD data, and their increasing presence is captured as a secular trend of quality improvement in our data analysis.[22] Thus, when we control for the industry-wide technological trend, we are incorporating Kryder's Law for HDDs as well as Moore's Law for semiconductor devices in a reduced-form manner.

Table 1: Summary Statistics

Variable	Unit of measurement	Number of observations	Mean	Standard deviation	Minimum	Maximum
Panel A						
HDD shipments, Q_t	Exabytes*	78	15.882	17.368	0.021	53.196
HDD price, P_t	\$/Gigabytes*	78	14.991	37.305	0.032	178.617
Disk price, Z_t	\$/Gigabytes*	78	1.952	5.252	0.005	23.508
PC shipments, X_t	Million units	78	29.286	7.188	14.468	40.312
Panel B						
Market share, ms_{it}	%	590	13.2	11.0	0.0	45.7
Panel C						
Indicator$\{a_{it} = merge\}$	0 or 1	1,766	0.0034	0.0582	0	1
Indicator$\{a_{it} = invest\}$	0 or 1	1,766	0.0142	0.1181	0	1
Indicator$\{a_{it} = exit\}$	0 or 1	1,766	0.0028	0.0531	0	1
Indicator$\{a_{it} = enter\}$	0 or 1	233	0.0043	0.0654	0	1
Variable profit, π_{it}	Million \$	see note	42.70	91.19	0.00	9725.22

Note: 1 exabytes (EB) = 1 billion gigabytes (GB) = 1 billion bytes. Panel A is recorded in quarterly frequency at the aggregate level, Panel B is quarterly at the firm level, and Panel C is monthly at the firm level. π_{it} is our period-profit estimate and contains 42,325,920 values across 7 productivity levels, 78 quarters, and 77520 industry states. See sections 4.1 and 4.2.

Source: *TRENDFOCUS* Reports (1996–2015).

[21] Modeling the entire supply chain of PCs and HDDs as bilateral oligopoly would be an interesting exercise, but it is beyond the scope of this paper, whose main focus is horizontal mergers and long-run dynamics.

[22] Pure SSDs have become common for note PCs, but we focus on HDDs (including hybrids) for desktop PCs, which is still the mainstream market for HDDs.

3.2 Three Data Elements

Our empirical analysis will focus on the period between 1996 and 2015 for three reasons. First, most of the exits prior to the mid-1990s were shakeouts of fringe firms that occurred through plain liquidation, whereas our main interest concerns mergers in the final phase of industry consolidation. Second, the de-facto standardization of both product design and manufacturing processes had mostly finished by 1996. Specifically, the 3.5-inch form factor had come to dominate the desktop market (see Igami 2015), and manufacturing operations in Southeast Asia had achieved the most competitive cost-quality balance (see Igami 2016). Third, our main data source, *TRENDFOCUS*, an industry publication series, started most of its systematic data collection at the quarterly frequency in 1996.[23]

Table 1 summarizes our main dataset, which consists of three elements corresponding to three steps of our empirical analysis in the next section. Panel A is the aggregate quarterly data on HDD shipments, HDD price, disk price, and PC shipments,[24] which we use to estimate HDD demand in section 4.1. Panel B is the firm-level market shares at the quarterly frequency, a graphic version of which is displayed in Figure 1 (top right). We use demand estimates and Panel B to infer the variable cost of each firm in each period in section 4.2. Panel C is a systematic record of firms' dynamic choices between merger, R&D investment, and entry/exit, at the monthly frequency. Panel C includes some elements that are derived from the other two panels, such as the indicator of R&D investment and the equilibrium variable profits.[25] We use these dynamic choice data and stage-game payoffs to estimate the implied sunk costs associated with these actions in section 4.3.

4 Empirical Analysis

We flesh out our model (section 2) with the actual data (section 3), which contained three elements: (A) aggregate sales, (B) firm-level market shares, and (C) dynamic discrete choice. Each of these data elements is paired with a model element and an empirical method to estimate demand, variable costs, and sunk costs. Table 2 provides an overview of such model-data-method pairing as well as section 4's roadmap.

[23] By contrast, Igami (2015, 2016) used *Disk/Trend* Reports (1977–1999), an annual publication series. Other studies of the HDD industry, such as Christensen (1993) and Gans (2016), also focus on this period.

[24] Appendix C features more details on Panel A, including visual plots of these variables.

[25] Appendix D.2 explains the details of this data construction.

Table 2: Overview of Empirical Analysis

Section	Step	Model	Data	Method
4.1	Demand	Log-linear demand	Panel A	IV regression
4.2	Variable cost	Cournot competition	Panel B	First-order condition
4.3	Sunk cost	Dynamic discrete choice	Panel C	Maximum likelihood

Note: See section 2 for the dynamic game model, and section 3 for the three data elements.

4.1 Demand Estimation

We follow Peter Knight's characterization of HDDs as "completely undifferentiated products" (see section 3.1). To be precise, HDDs come in a few different data-storage capacities (e.g., 1 terabytes per drive), but all firms are selling these products with "the same capacities, the same speed, and similar reliability" at any given moment, so that cost becomes the only dimension of competition.[26] Most consumers, including the authors, do not even know which "brand" of HDDs are installed inside their desktop PCs, and PC manufacturers typically do not let consumers choose a brand. Thus, homogeneous-good demand and Cournot competition are useful characterizations of the spot-market transactions.

To ensure our data format is consistent with our notion of product homogeneity, we consider units of data storage (measured in bytes) as undifferentiated goods. We specify a log-linear demand for raw data-storage functionality of HDDs,

$$\log Q_t = \alpha_0 + \alpha_1 \log P_t + \alpha_2 \log X_t + \varepsilon_t, \tag{11}$$

where Q_t is the world's total HDD shipments in exabytes (EB = 1 billion GB), P_t is the average HDD price per gigabytes (\$/GB), X_t is the PC shipments (in million units) as a demand shifter, and ε_t represents unobserved demand shocks.

Because the equilibrium prices in the data may correlate with ε_t, we instrument P_t by Z_t, the average disk price per gigabyte (\$/GB). Disks are one of the main components of HDDs, and hence their price is an important cost shifter for HDDs. Disks are made from substrates of either aluminum or glass. The manufacturers of these key inputs are primarily in the business of processing materials, and only a small fraction of their revenues come from the HDD-related products. Thus, we regard Z_t as exogenous to the developments within the HDD market. In Table 3, column 1 shows OLS estimates, and column 2 shows the IV estimates with disk prices as Z_t. The estimates for price elasticity, α_1, are within the standard

[26] From author's personal interview on June 30, 2015, in Cupertino, CA.

errors of each other. This finding of inelastic demand (i.e., $|\alpha_1| < 1$) is rationalizable under oligopoly but creates a conceptual problem under monopoly (i.e., its profit-maximizing price would be arbitrarily high). Consequently, we will ignore the top 5% of consumers with the highest willingness to pay to keep monopoly price finite.

Table 3: Demand Estimates

Dependent variable: log total EB shipped	(1) OLS	(2) IV-1	(3) IV-2	(4) IV-2 rolling (mean)
log price per GB (α_1)	$-.8549^{***}$	$-.8244^{***}$	$-.8446^{***}$	$-.8420$
	(.0188)	(.0225)	(.0259)	(−)
log PC shipment (α_2)	$.8430^{***}$	1.0687^{***}	$.9198^{***}$	.7836
	(.1488)	(.1817)	(.2180)	(−)
Constant (α_0)	-1.6452^{***}	-2.4039^{***}	-1.9033^{***}	-1.3196
	(.4994)	(.6084)	(.7320)	(−)
Number of observations	78	78	78	−
Adjusted R^2	.9971	.9971	.9972	−
First-stage regression				
IV for HDD price	−	*Disk price*	*Time trend*	*Time trend*
F-value	−	3009.80	742.14	−
Adjusted R^2	−	.9889	.9469	−

Note: Heteroskedasticity- and autocorrelation-consistent standard errors are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. See Appendix D.1 for column 4.

Although we believe disk prices represent an exogenous cost shifter, one might still suspect the existence of some unknown endogeneity problems because, after all, disks and HDDs are close neighbors in the supply chain of computers. To address this concern, we use a logarithm of time trend as an alternative IV in column 3. This IV relies on a classical notion of technological progress as a time trend, which is particularly natural in the HDD context. Kryder's Law dictates a secular trend in the improvement of areal density (i.e., bytes per square inch), which mechanically translates into the reduction of materials cost per byte, because the same number of disks can store more information.

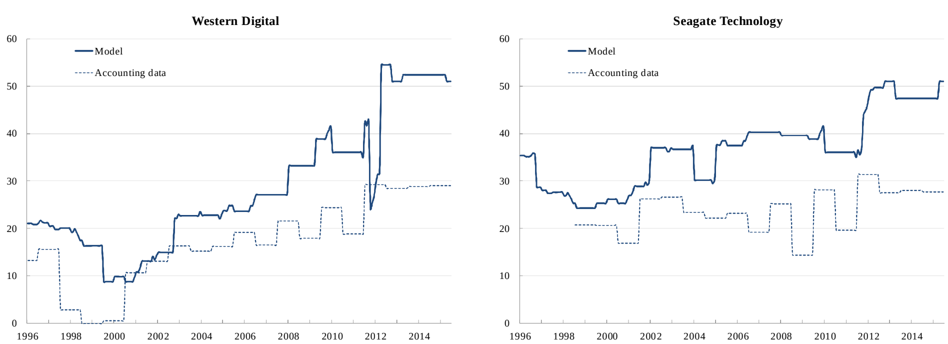
Yet another concern is that consumers' preferences might have changed over two decades. Casual empiricism suggests people have dramatically increased the amount of data usage in everyday life, which could alter the demand parameters over time. To investigate this matter, we use rolling estimation in which we roll through the sample of 78 quarters with a 12-quarter window, using 12 observations for estimation at a time. Detailed results do not fit the table format; hence, we plot the estimated coefficients against time in Appendix D.1, and report only their time averages in column 4 of Table 3. We use this last specification for the subsequent analysis.

Other concerns and modeling considerations include (i) demand-side dynamics, such as durability of HDDs and the repurchasing cycle of PCs, (ii) supply-side dynamics, such as long-term contracts with PC makers, and (iii) non-HDD technological dynamics, including SSDs and the semiconductor industry. Our summary views are as follows: (i) the physical durability of HDDs does not determine the dynamics of PC demand; (ii) the actual interaction between HDD makers and PC makers is more adequately described as spot-market transactions rather than a long-term relationship; and (iii) our analysis incorporates the non-HDD technological trend and the growing presence of hybrid HDDs as part of Kryder's Law. Section 3.1 provides further details.

4.2 Variable Costs and Spot-Market Competition

The second data element is the panel of firm-level market shares (Figure 1, top right), which we will interpret through the analytical lens of Cournot competition, for two reasons. Despite selling undifferentiated high-tech commodities, HDD makers' financial statements report positive profit margins (see dotted lines in Figure 3), which suggests the Cournot model as a reasonable metaphor for analyzing their spot-market interactions. Another appeal is that the classical oligopoly theory of mergers has mostly focused on the Cournot model (see section 1.1), which brings conceptual clarity and preserves economic intuition.

Figure 3: Comparison of Profit Margins (%) in the Model and Financial Statements



Note: The model predicts economic variable profits, whereas the financial statements report accounting profits (gross profits), and hence they are conceptually not comparable. The correlation coefficient between the model and the accounting data is .8398 for Western Digital, and .5407 for Seagate Technology. With a management buy-out in 2000, Seagate Technology was a private company until 2002, when it re-entered the public market. These events caused discontinuity in the financial record.

Each of the n_t firms observes the profile of marginal costs $\{mc_{it}\}_{i=1}^{n_t}$ as well as the concurrent HDD demand, and chooses the amount of re-tooling efforts to maintain effective output

level, q_{it}, to maximize its variable profit,

$$\pi_{it} = (P_t - mc_{it})\, q_{it}, \tag{12}$$

where P_t is the price per GB of a representative HDD at t and mc_{it} is the marginal cost, which is predetermined at $t-1$ and constant with respect to q_{it}.[27] Firm i's first-order condition is

$$P_t + \frac{\partial P}{\partial Q} q_{it} = mc_{it}, \tag{13}$$

which provides one-to-one mapping between q_{it} (observed) and mc_{it} (implied) given P_t in the data and $\partial P/\partial Q$ from the demand estimates. Intuitively, the higher the firm's observed market share, the lower its implied marginal cost.

The interpretation of mc_{it} requires special attention in the high-tech context. As we discussed in section 2 regarding synergies, "productivity" in HDD manufacturing is not so much about tangible assets as about tacit knowledge embodied by teams of engineers. Thus, our preferred interpretation of Cournot spot-market competition follows Kreps and Scheinkman's (1983) model of quantity pre-commitment followed by price competition, given the cost profile (i.e., all active firms' productivity levels).[28]

Appendix D.2 shows the details of our marginal-cost estimates and how we convert them into productivity levels, ω_{it}, for the subsequent dynamic analysis. Meanwhile, we focus our main-text exposition on an external validity check of our static model. Figure 3 compares the model's predictions with accounting data, in terms of profit margins at Western Digital (left) and Seagate Technology (right), respectively. Our model takes as inputs the demand estimates and the marginal-cost estimates, and predicts equilibrium outputs, prices, and hence each firm's variable profit margin in each year,

$$m_{it}(\omega_t) = \frac{P_t(\omega_t) - mc_{it}}{P_t(\omega_t)}, \tag{14}$$

[27] In principle, we may replace this constant marginal-cost specification with other functional forms. In the high-tech context, however, marginal costs are falling every period across the industry, and the only geographical market is Earth. Thus, one cannot rely on either inter-temporal or cross-sectional variation in data to identify marginal-cost curves nonparametrically.

[28] One might wonder whether such "pre-committed quantities" are hard-wired to physical production capacities. In the context of high-tech manufacturing, *effective* physical capacities are highly "perishable" because of the constant improvement in the industry's basic technology (i.e., Kryder's Law), which makes previously installed manufacturing equipment obsolete. Thus, we prefer a rather abstract phrase "quantity pre-commitment," to "capacity" because the latter could mislead the reader to imagine "durable" physical facilities.

under any industry state, ω_t (i.e., the number of firms and their productivity levels). The solid lines represent such predictions of economic profit margins along the actual history, whereas the dotted lines represent gross profit margins (i.e., revenue minus cost of revenues) in the firms' financial statements.

Economic profits and accounting profits are different concepts, which explains the existence of systematic gaps in their levels. On average, (economic) variable profit margins are higher than (accounting) gross profit margins by 11.4 and 13.8 percentage points at these firms, respectively, because the former excludes fixed costs of operation and sunk costs of investment, whereas the latter includes some elements of fixed and sunk costs.[29] Thus, correlation is more important than levels, which is .8398 for Western Digital, and .5407 for Seagate Technology. If we accept accountants as conveyors of truth, this comparison should confirm the relevance of our spot-market model.

These static analyses are interesting by themselves, but merger policy will affect not only firms' spot-market behaviors but also their incentives for mergers and investments, and hence the entire history of competition and innovation. Thus, a complete welfare analysis of industry consolidation requires endogenous mergers, innovation, and entry-exit dynamics, which are the focus of the subsequent sections.

4.3 Sunk Costs and Dynamic Discrete Choice

The third data element is the panel of firms' discrete choices between mergers, innovation, entry, and exit, which we will interpret through the dynamic model. We have already estimated profit function, that is, period profits of all types of firms, in each period, in each industry state, $\pi_{it}(\omega_t)$. In other words, we observe the actual choices and the "benefit" side of the equation; hence, the "cost" side of the equation is the only unknown now.

Configuration

Table 4 lists all the parameters and key specifications of our model. Before engaging in the MLE of the core parameters, $(\kappa^i, \kappa^m, \kappa^e)$, we determine the values of the other parameters either as by-products of the previous two steps or directly from auxiliary data.

[29] For example, manufacturing operations in East Asia accounted for 41,304, or 80.8%, of Seagate's 50,988 employees on average between 2003 and 2015, whose wage bills constitute the labor component of the "cost of revenues" in terms of accounting. However, some of these employees spent time and effort on technological improvements, such as the re-tooling of manufacturing equipment for new products (i.e., product innovation), as well as the diagnosis and solution of a multitude of engineering challenges to improve the cost effectiveness of manufacturing processes (i.e., process innovation), which are sunk costs of investment in terms of economics.

Table 4: List of Parameters and Key Specifications

Parameter	Notation	Empirical approach
1. Static estimates		
Demand	$\alpha_0, \alpha_1, \alpha_2$	See section 4.1
Variable costs	mc_{it}	See section 4.2
Period profits	$\pi_{it}(\omega_t)$	See section 4.2
2. Dynamics (sunk costs)		
Innovation, mergers, and entry	$\kappa^i, \kappa^m, \kappa^e$	MLE (main task of section 4.3)
Fixed cost of operation	$\kappa_t^c(\omega_{it})$	Accounting data (see Appendix D.3)
Liquidation value	$\kappa^x = 0$	Industry background
3. Dynamics (transitions)		
Discount factor (annual)	$\beta = .9$	Calibrated to the literature's standard
Prob. stochastic depreciation	$\delta = .0190$	Implied by mc_{it}
Average synergy	$\lambda = 1$	Implied by mc_{it} (sensitivity analysis with 0 & 2)
4. Other key specifications		
Terminal period	$T = Dec$-2025	Sensitivity analysis with Dec-2020 & Dec-2015
Bargaining power	NB: $\zeta = .5$	Sensitivity analysis with TIOLI: $\zeta = 1$

First, we pin down the other two κ's as follows. The fixed cost of operations and keeping up with Kryder's Law, $\kappa_t^c(\omega_{it})$, comes directly from the accounting data on sales, general, and administrative (SGA) expenses, and are allowed to vary over time and across a firm's productivity level.[30] We set liquidation value, κ^x, to zero because tangible assets quickly become obsolete and have no productive use outside the HDD industry. The variance of $\varepsilon(a_{it})$ is also estimable in principle. Our plain logit specification implicitly assumes $Var(\varepsilon) = \pi^2/6$, where π is the mathematical constant.[31]

Second, three parameters govern transitions. The discount factor is calibrated to $\beta =$ .9 at an annualized rate, a standard level in the literature. We introduce the possibility of exogenous and stochastic depreciation of ω_{it} at the end of every period, because our estimates of mc_{it} (or equivalently, ω_{it}) exhibit occasional deterioration with probability $\delta =$ .0190. Likewise, our mc_{it} estimates suggest the extent of synergy. The average post-merger improvement is approximately \$1 (measured in terms of the discretized bin, to be precise),[32] which constitutes our "estimates" of the Poisson synergy parameter,

$$\hat{\lambda}_{MLE} = \frac{1}{\#_m} \sum_{m=1}^{\#_m} \Delta_m, \tag{15}$$

where $\#_m$ is the number of mergers in the data, and Δ_m is the productivity improvement

[30] See Appendix D.3 for details.

[31] Igami (2016) estimates $Var(\varepsilon)$ in the same industry, and finds it statistically indistinguishable from $\pi^2/6$. This paper builds on this finding to alleviate the computational burden for estimation.

[32] See Appendix D.2 for details.

from merger m. Mergers in a concentrated industry are rare events ($\#_m = 6$ in our main sample), and most IO economists feel skeptical about merging parties' claim about synergy. Consequently, we consider $\lambda = 1$ as our baseline "calibration" and conduct sensitivity analysis with $\lambda = 0$ (no synergy) and $\lambda = 2$ (strong synergy) instead of arguing over what its "right" value should be.

Third, two aspects of our dynamic model require fine-tuning. The first such aspect is the terminal condition. Our sample period ends in 2015Q2, but the HDD industry does not; hence, we need to assume something about the post-sample end game. Our baseline specification is relatively optimistic to assume the HDD demand continues to exist until the end of year 2025, with linear interpolation between June 2015 and December 2025. Our sensitivity analyses employ more pessimistic scenarios, with $T = Dec$-2015 and Dec-2020. The second aspect is bargaining protocols. Our baseline specification is Nash bargaining with equal bargaining powers between acquirer and target, $\zeta = .5$, but we also estimate the TIOLI version, $\zeta = 1$.

Extending Rust (1987) to Random-Mover Dynamic Games

Having determined the baseline configuration, we proceed to estimate $(\kappa^m, \kappa^i, \kappa^e)$. The outline of our MLE procedure follows Rust (1987), who constructed the likelihood of bus-engine replacement as a function of Harold Zurcher's decisions regarding whether to replace old bus engines (choice data), their mileages (observed state variable), and the sunk cost of replacement (parameter to be estimated). Just as his identification of the replacement cost relied on the variation in the mileage of bus engines (i.e., observed differences in payoff-relevant state across time), our identification of $(\kappa^i, \kappa^m, \kappa^e)$ relies on variation in period profits and their dynamic counterparts (i.e., expected net present values associated with discrete alternatives). Similarly, just as his NFXP approach nested the solution of Harold Zurcher's optimal choice problem (the "inner loop") within the calculation of the likelihood function (to be maximized in the "outer loop"), our likelihood function nests the HDD makers' optimal choice problem. Thus, our overall scheme closely follows Rust's.

We differ from Rust (1987) in three respects: (i) the HDD makers' optimal choice problem takes place within a dynamic game, rather than being a single-agent problem; (ii) their turns-to-move arrive stochastically rather than deterministically; and (iii) the underlying payoffs change over time and eventually disappear. Feature (i) fundamentally complicates the estimation problem because games generally entail multiple equilibria, which would make estimates inconsistent because one cannot use model-generated CCPs to pin down parameter

values if a single parameter value predicts multiple CCPs. Our solution is three-fold. First, we use an alternating-move formulation to streamline the decision problems, so that only (up to) one player makes a choice in each period. Second, we avoid tilting the playing field (i.e., assuming a deterministic sequence would embed early-mover advantage *a priori*) by making the turn-to-move stochastic, which led us to feature (ii) in the above. Third, we exploit the high-tech context of feature (iii) to set a finite time horizon, which enables us to solve the game for a unique equilibrium by backward induction. In other words, we address methodological challenges stemming from feature (i) by crafting (ii) and exploiting (iii), so that the overall scheme of estimation can proceed within the NFXP framework. Thus, we regard our approach as an extension to Rust (1987) as well as an illustration of a particular kind of dynamic game that is amenable to NFXP.

The optimal choice probabilities of entry, exit, innovation, and mergers in equation 9 constitute the likelihood function. Firm i's contribution at t is

$$l_{it}\left(a_{it}|s_t;\kappa\right)=\rho_i(s_t)\prod_{action\in A_{it}(s_t)}\Pr\left(a_{it}=action\right)^{1\{a_{it}=action\}}, \tag{16}$$

where $1\{\cdot\}$ is an indicator function. The MLE is

$$\hat{\kappa}_{MLE}=\arg\max_{(\kappa^i,\kappa^m,\kappa^e)}\frac{1}{T}\frac{1}{I}\sum_t\sum_i\ln\left[l_{it}\left(a_{it}|\omega_t;\kappa^i,\kappa^m,\kappa^e\right)\right], \tag{17}$$

where T is the number of sample periods and I is the number of firms.

The realizations of turns-to-move are not always evident in the data; hence, the implementation of MLE needs to distinguish "active" periods in which some firm took an action (such as exit, merger, or entry) and altered ω_t, and "quiet" periods in which no firm made any such proactive moves. Specifically, we incorporate the random turns to move by setting

$$\hat{\rho}_i\left(s_t\right)=\begin{cases}1 & \text{if } a_{it}\in\{exit,merge,enter\}\text{, and}\\ \frac{1}{n_{\max}}\Pr\left(a_{it}=idle,out\right) & \text{if } a_{it}\in\{idle,out\}\ \ \forall i.\end{cases} \tag{18}$$

That is, when exit, merge, or entry is recorded in the data, we may assign probability 1 to the turn-to-move of the firm that took the action, whereas in a "quiet" period, nature may have picked any one of the firms that subsequently decided to idle (or stay out) and did not alter ω_t.

Results

Table 5, column 1 shows our baseline estimates with (i) Nash bargaining with equal bargaining powers, $\zeta = .5$, (ii) mean synergy from the data, $\lambda = 1$, and (iii) optimistic terminal condition, $T = Dec$-2025. As a sensitivity analysis, column 2 alters ζ, columns 3 and 4 alter λ, and columns 5 and 6 alter T. All the specifications lead to similar estimates that are within the 95% confidence interval of each other.

Table 5: MLE of Dynamic Parameters and Sensitivity Analysis

Specification	(1)	(2)	(3)	(4)	(5)	(6)
Bargaining (ζ):	.5 (NB)	1 (TIOLI)	.5	.5	.5	.5
Synergy (λ):	1	1	0	2	1	1
Terminal period (T):	2025	2025	2025	2025	2020	2015
κ^i	3.0365	3.0335	3.0411	3.0326	3.0353	3.0283
	$[2.65, 3.48]$	$[2.64, 3.47]$	$[2.65, 3.48]$	$[2.64, 3.47]$	$[2.65, 3.48]$	$[2.64, 3.47]$
κ^m	5.2043	5.9239	4.9135	5.4411	5.2040	5.2049
	$[4.47, 6.14]$	$[5.19, 6.86]$	$[4.18, 5.85]$	$[4.47, 6.14]$	$[4.47, 6.14]$	$[4.47, 6.14]$
κ^e	5.2069	4.8092	5.3330	5.0667	5.3416	5.4529
	$[-]$	$[-]$	$[-]$	$[-]$	$[-]$	$[-]$
Log likelihood	-320.6502	-321.0507	-320.3957	-320.8940	-320.8981	-323.8451

Note: The 95% confidence intervals are constructed from the likelihood-ratio tests.

The directions of these differences are logical and provide an intuitive understanding of identification. Compare κ^m in columns 1 ($\zeta = .5$) and 2 ($\zeta = 1$). The TIOLI assumption in column 2 gives greater bargaining power to acquirers, lowers acquisition prices, p_{ij}, and increases values of mergers, $\bar{V}_{ij}^m$. *Ceteris paribus*, the TIOLI model predicts higher CCPs of merger, $\tilde{P}_{ij}^m$, but the *actual* CCPs in the data, $\bar{P}_{ij}^m$, do not change, which decouples these two objects (i.e., $\tilde{P}_{ij}^m > \bar{P}_{ij}^m$). Consequently, the only way for the model to reconcile them is to increase κ^m, so that $\tilde{P}_{ij}^m$ comes down again (i.e., $\tilde{P}_{ij}^m \approx \bar{P}_{ij}^m$). The same mechanism applies to the sensitivity of κ^m in columns 3 ($\lambda = 0$) and 4 ($\lambda = 2$), where the expected synergy level plays the same role as ζ in column 2. By contrast, columns 5 ($T = Jan$-2020) and 6 ($T = Jan$-2015) suggest the assumption on a time horizon hardly affects any estimates, because terminal values are relatively small and the data variation in the sample period remains unchanged.

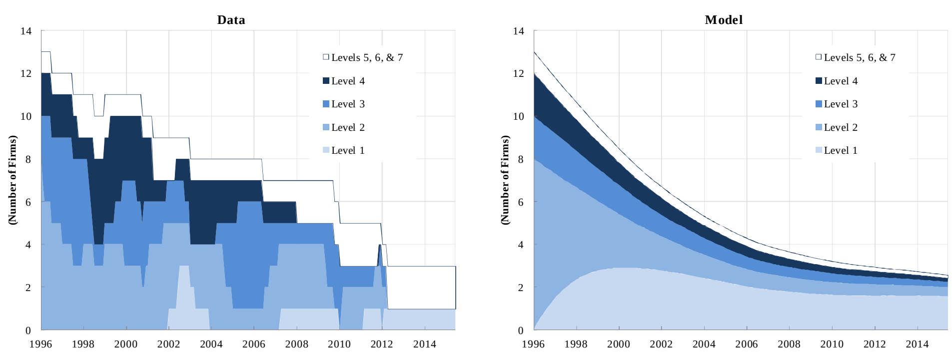
The innovation cost, κ^i, is in the ballpark of HDD makers' R&D spending. The entry cost, κ^e, does not carry meaningful confidence intervals, because almost any high value of κ^e could rationalize the data that contain only one entry; hence, our estimate is a lower bound. Nevertheless, one entry is more informative than zero entry. "Almost" any high κ^e can rationalize the data, but it cannot be arbitrarily high, because it must permit some

possibility that a potential entrant with a lucky draw of ε_{it}^{e} (such as Finis Conner, who founded Conner Technologies in the late 1990s) would choose to enter.

Besides these sunk costs, the NFXP estimation provides the equilibrium value and policy functions as by-products. Hence, as an external validity check, we may compare these model-generated enterprise values with the actual acquisition prices in the six merger cases.[33] The comparison reveals that at least three out of the six historical transaction values closely match the target firms' predicted values. See Appendix D.3 for further details.

Another way of assessing fit is to compare the actual and predicted trajectories of market structure (Figure 4). The estimated model generates a smooth version of the industry consolidation process in the data, with approximately three firms remaining at the end of the sample period. The model also replicates some aspects of their productivity composition (e.g., the survival of a few low-level firms). We believe the estimated model provides a reasonable benchmark with which we can compare welfare performances of hypothetical antitrust policies in section 5.

Figure 4: Fit of the Estimated Model (Number of Firms)



Note: The model outcome is the average of 10,000 simulations based on the estimated model. The productivity types are defined on a discretized grid of levels 1 through 7, each step of which corresponds to an approximately \$1 reduction in marginal cost.

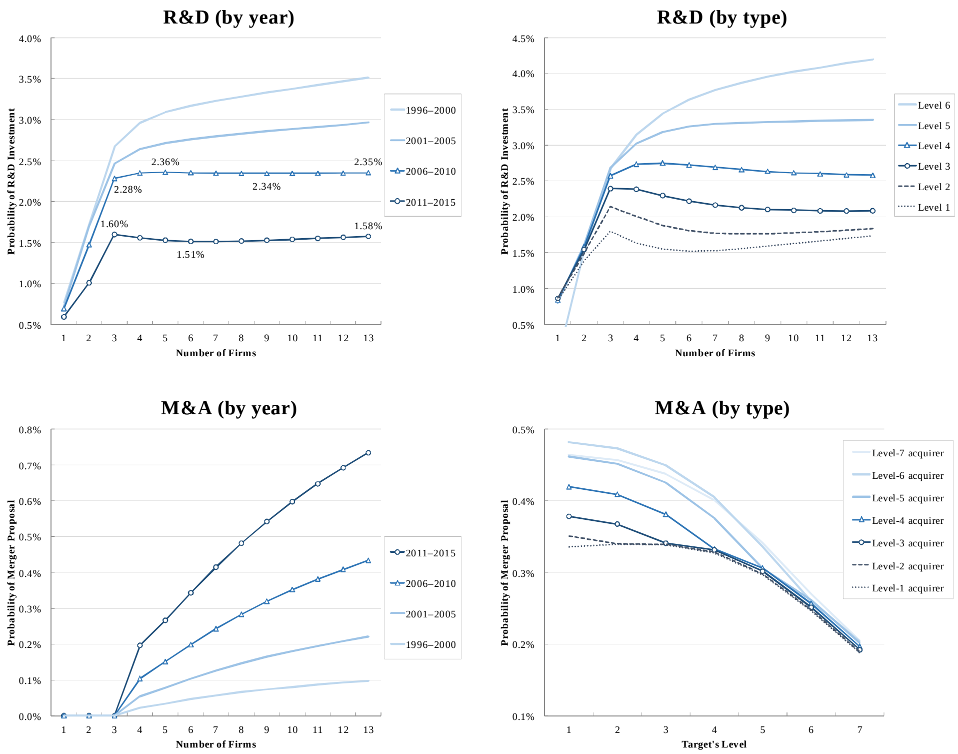
4.4 Competition, Innovation, and Merger

Whereas the value-function estimates and the simulations of industry dynamics were useful for assessing fit, the policy functions are interesting by themselves because they represent

[33]In principle, we may use these six observed acquisition prices to "estimate" the bargaining parameter, ζ. However, we prefer calibrating ζ because six cases are too few for precise estimation.

structural relationships between competition, innovation, and merger. Figure 5 shows the equilibrium R&D and M&A strategies by year, type, and market structure.

Figure 5: Plateaus and Cascades in Equilibrium Strategies



Note: Each graph summarizes the equilibrium strategies for R&D and M&A, by averaging the structural CCP estimates across ω_{it}, s_t, or t.

The top panels feature a plateau-shaped relationship between the optimal R&D investment (vertical axis) and the number of firms (horizontal axis). Regardless of how we slice the equilibrium strategy, the incentive to innovate sharply increases between one, two, and three firms, because a monopolist has little reason to replace itself (Arrow 1962), whereas duopolists and triopolists have to race and preempt rivals (Gilbert and Newbery 1982). Explained in this way, the upward slopes may look obvious in hindsight, but this pattern is actually not so obvious. A static Cournot model (or other standard models of imperfect competition) predicts the opposite, because $\pi_i(n)$ decreases with n, and the incremental profit from innovation, $\Delta\pi_i \equiv \pi_i\left(\omega^{high}\right) - \pi_i\left(\omega^{low}\right)$, also decreases with n.[34] Thus, the fact that positive slopes came out of our Cournot-based model highlights the importance of dynamic incentives. Dynamics make innovation increase with competition.

[34] See Igami (2016) for a stylized model, and Dasgupta and Stiglitz's (1980) for rent dissipation.

After three or four firms, the plateaus exhibit ample heterogeneity both across time (top-left panel) and productivity (top-right panel). Innovation rates are high and increasing with n in early years and at high-productivity firms because continuation values (and hence the incremental value of investment) are high. By contrast, the incentives are low and often decreases with n in later years and at low-productivity firms because the possibility of exit becomes more realistic in such cases, and increased competition make them give up. Thus, "heterogeneous plateaus" are the structural-empirical cousin of the celebrated "inverted-U" curve (e.g., Scherer 1965, Aghion et al. 2005), and option values and the probability of death govern the plateaus' heterogeneity.

The incentive for merger is equally intriguing. The bottom-left panel of Figure 5 plots the optimal M&A strategy as a function of time and competition. Two patterns emerge. First, mergers become more popular in later years because killing rivals becomes more attractive when the fixed cost of operations has grown and new entry stops. Second, more mergers occur when more rivals exist, because they represent potential merger targets. Once we divide the CCP by n_t, the merger-competition slope (per active firm) becomes virtually flat.

Who merges with whom? The bottom-right panel of Figure 5 plots the CCP of merger (sliced by the *acquiring* firm's level) against the *target* firm's level. Three patterns emerge. First, all combinations are possible, as is the case in our data.[35] Second, high types acquire more than low types, because the former expect higher values from reduced competition and increased productivity. Third, low types are more attractive targets than high types, which seems intuitive, but the mechanism is subtle. On the one hand, eliminating a low type does not soften competition by much; hence, the benefits are limited. On the other hand, low types' reservation values are low, so they represent a cheaper means to obtain synergy draws. Our results incorporate all of these economic factors in equilibrium. We do not model every single detail of M&A, because this paper focuses on long-run industry dynamics. Nevertheless, the fact that rich nuances come out of our relatively simple model highlights the fruitfulness of analyzing mergers as dynamic and endogenous choices.

5 Optimal Policy and Dynamic Welfare Tradeoff

How far should an industry be allowed to consolidate? We are now ready to simulate welfare outcomes under hypothetical merger policies, and our short answer is five. The following subsections clarify exactly how we reach this finding (section 5.1), the underlying mechanism

[35] We observe two high-low, one mid-mid, two mid-low, and one low-mid mergers between 1996 and 2015.

that led to this finding (section 5.2), how the outcomes may change if the industry is declining quickly (section 5.3), and the possibility of "smarter" policies (section 5.4).

5.1 Commitment Policy

Table 6 shows the highest present value of social welfare is achieved under a static (or "commitment") policy in which antitrust authorities block mergers if n_t is five or less (i.e., $\underline{N} = 5$). Each of the nine columns reports the discounted sums of consumer surplus (CS), producer surplus (PS), and social welfare (SW) under a hypothetical regime with $\underline{N} \in \{1, 2, ..., 9\}$ relative to the baseline model ($\underline{N} = 3$).

We set $\underline{N} = 3$ in the baseline (estimated) model based on the following evidence. The FTC reports that in merger enforcement concerning high-tech markets between 1996 and 2011, no merger was blocked until the number of "significant competitors" reached three. Specifically, (i) none of the 5-to-4 mergers were blocked; (ii) 33% of the 4-to-3 merger proposals were blocked; and (iii) 100% of the 3-to-2 and 2-to-1 proposals were blocked.[36] Thus, $\underline{N} = 3$ is an accurate description of the actual policy during our sample period. This *de facto* rule of the game is a shared perception among antitrust practitioners and firms in Silicon Valley, according to our conversations with former chief economists at the FTC and the Antitrust Division of the DOJ, antitrust economic consultants, as well as senior managers at the HDD manufacturers.

Table 6: Welfare Performance of Commitment Policies

Threshold number of firms ($\underline{N}$)	1	2	3	4	5	6	7	8	9
Discounted Jan-1996 value									
Consumer surplus (%)	▲24.12	▲7.40	±0	+0.45	+0.65	+0.73	+0.77	+0.80	+0.82
Producer surplus (%)	+118.46	+32.68	±0	▲3.11	▲5.26	▲5.55	▲8.12	▲9.19	▲10.39
Social welfare (%)	▲14.64	▲4.74	±0	+0.21	+0.26	+0.24	+0.18	+0.13	+0.07
Undiscounted sum									
Consumer surplus (%)	▲79.61	▲20.74	±0	+1.14	+1.62	+1.84	+1.97	+2.06	+2.11
Producer surplus (%)	+278.70	+64.64	±0	▲4.87	▲7.62	▲9.21	▲10.67	▲11.68	▲12.64
Social welfare (%)	▲48.16	▲13.24	±0	+0.62	+0.81	+0.87	+0.86	+0.85	+0.81

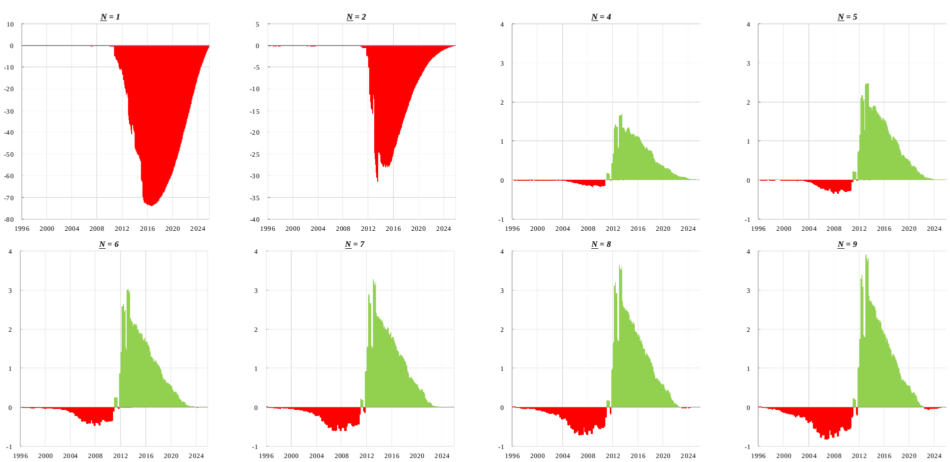
Note: All welfare numbers are expressed in terms of percentage change from the baseline outcome under $\underline{N} = 3$. These changes might appear small because most of the counterfactual policies' impacts are concentrated in later periods, at which point the market size is shrinking, and discounting attenuates their values as of January 1996.

Computational implementation is straightforward. We estimated the baseline model by

[36] See Federal Trade Commission (2013), Table 4.7 entitled "Number of Significant Competitors in Electronically-Controlled Devices and Systems Markets." Our model can incorporate similar policy regimes based on HHI thresholds instead of $\underline{N}$, but we found no clear HHI threshold in the report.

searching over the parameter space of $(\kappa^i, \kappa^m, \kappa^e)$ to maximize the likelihood of observing the actual choice patterns in the data (in the outer loop), and by solving the dynamic game by backward induction to calculate the predicted choice patterns based on the model (in the inner loop) in which the sunk cost of merger is κ^m when $n_t > 3$ but ∞ when $n_t \leqslant 3$. Simulating welfare outcomes under an alternative regime is simpler than estimation. First, solve the counterfactual game with the same parameter estimates $\left(\hat{\kappa}^i, \hat{\kappa}^m, \hat{\kappa}^e\right)$ but in a different policy environment ($\underline{N} \neq 3$) just once, and obtain the optimal choice probabilities in the counterfactual equilibrium. Second, use these CCPs to simulate 10,000 counterfactual industry histories, $\{s_t\}_{t=0}^T$. Third, calculate $\{(CS_t, PS_t, SW_t)\}_{t=0}^T$ along each simulated history, take their average across the 10,000 simulations, and summarize their (undiscounted) time profiles in terms of time-0 discounted present values.

Figure 6: Time Profile of Undiscounted Welfare (% change)



Note: Each panel shows the average of 10,000 simulations under an alternative policy regime.

The threshold policy $\underline{N} = 5$ optimally trades off the ex-post pro-competitive effect of blocking mergers with the negative ex-ante value-destruction side effects. Figure 6 visualizes the dynamic welfare tradeoff by showing the time profiles of undiscounted social welfare, relative to the baseline performance under $\underline{N} = 3$. The most staggering patterns emerge from the more permissive regimes ($\underline{N} = 1$ and 2), in which the deadweight losses under monopoly and duopoly during the second half of the sample period dominates any positive changes during the first half. Thus, an obvious finding from our welfare analysis is that allowing mergers to monopoly or duopoly is a bad idea, even if we account for potentially positive side effects on ex-ante incentives to enter and innovate.

By contrast, stricter policies ($\underline{N} = 4$ through 9) generate more nuanced and qualitatively interesting welfare tradeoffs between positive ex-post performances and negative ex-ante side effects. The positive changes reflect the direct impact of blocking mergers, but this pro-competitive effect is partially offset by preceding negative changes, which seem to suggest some side effects on entry and R&D investments. Both the positive and the negative effects become larger as the threshold rises, but their rates of change are not always in balance. That is, the negative side effects grow faster than the positive main effect, to the extent that the net improvement peaks at $\underline{N} = 5$ and then declines (Table 6). Thus, our main finding is the optimality of $\underline{N} = 5$ and the inter-temporal tradeoff it epitomizes. Tougher merger policy is not unambiguously better, and this subtlety would have been totally absent had we employed a static framework. The next subsection dissects its underlying mechanisms.

5.2 Decomposing the Dynamic Welfare Tradeoff

Figure 7 illustrates this dynamic welfare tradeoff, first by decomposing the changes in consumer surplus into the changes in competition and innovation, and then by further decomposing (i) the changes in competition into the changes in mergers and entry/exit, and (ii) the changes in innovation into the changes in mergers and in-house investments.

The summary of welfare performances under alternative policy regimes in the previous subsection indicated the presence of dynamic tradeoff. This subsection clarifies its underlying mechanisms by focusing on changes in CS, whose good "summary statistic"[37] is price in the current empirical context with homogeneous goods.

The top panel of Figure 7 shows the CS performance of the optimal policy ($\underline{N} = 5$) relative to the baseline ($\underline{N} = 3$). The line graph represents the change of price (Δp_t), and the two bar graphs show its two determinants, the average markup (Δm_t), and the average marginal cost (Δmc_t) across active firms.[38] Their accounting relationship permits our first decomposition,

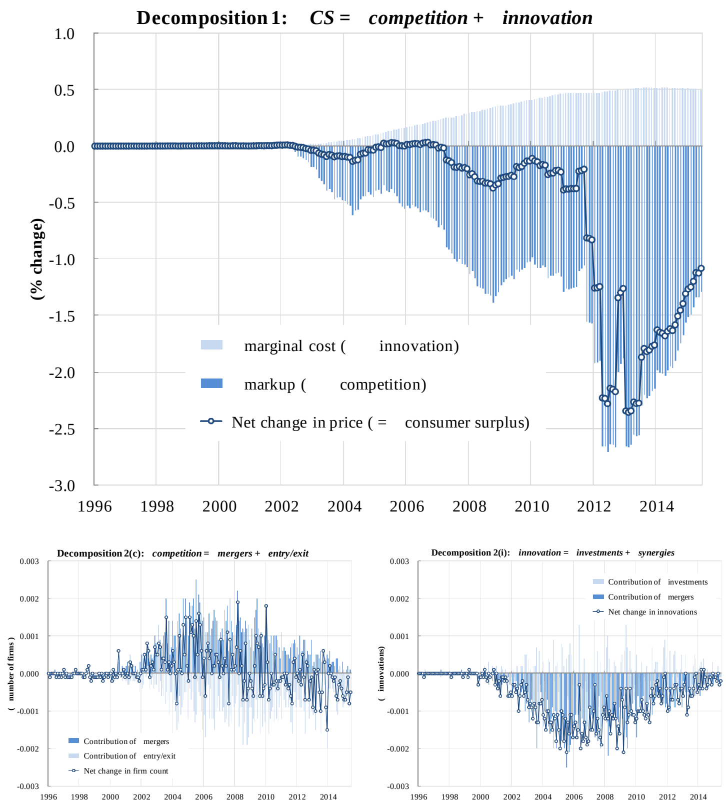
$$\Delta CS_t \propto \Delta p_t = \Delta m_t + \Delta mc_t. \tag{19}$$

In words, CS increases when the price decreases, which can be the result of either reduced

[37] We do not intend to claim Δp completely determines ΔCS. It does not, because we allow both the covariate (i.e., X_t) and the parameters (i.e., $\hat{\alpha}_t$s) to change over time. Here we mean "summary statistic" for illustration purposes only.

[38] The average markup and cost do not reflect all of the changes in individual firms' performances, but these measures are "sufficient" for our current purpose of illustrating the key determinants of welfare under different policy regimes.

Figure 7: Decomposition of Dynamic Welfare Tradeoff



Note: We use the (un-weighted) average marginal cost across firms. Alternative summary statistics such as the market share-weighted average do not qualitatively alter the decomposition patterns.

markups (i.e., increased "competition"), reduced costs (i.e., increased productivity or "innovation"), or both. Thus, we are decomposing ΔCS into changes in "competition" and "innovation." The graph suggests Δm_t and Δmc_t tend to move in opposite directions and cancel each other out.

The two bottom panels of Figure 7 further decompose the "competition" and the "innovation" effects of the optimal policy regime. On the competition side, Δm_t reflects changes in the number of firms, Δn_t, which in turn reflects the changes in *net entry* $(= \textit{entry} - \textit{exit})$, ΔNE_t, and M&As, ΔMA_t,

$$\Delta m_t \propto \Delta n_t = \Delta NE_t + \Delta MA_t. \tag{20}$$

In the bottom-left panel, the line graph shows Δn_t, and the two bar graphs show ΔNE_t and ΔMA_t, respectively. The contribution of the policy to competition (Δn_t) is mostly positive

from the merger channel (ΔMA_t) because more mergers are blocked under $\underline{N} = 5$ than under $\underline{N} = 3$, which represents the classical pro-competitive effect of blocking mergers.

However, this gain is partially offset by the negative contribution from the entry/exit channel (ΔNE_t). The mechanism behind this countervailing effect is partial destruction of firms' continuation values. The reduced opportunities for mergers mean reduced profit margins due to more competition in the future, reduced opportunities for productivity growth by stochastic synergy draws, and reduced opportunities for profitable exit by being acquired, all of which reduce the option value of entry and survival. Entry and survival require lump-sum and continual investments (i.e., κ^e and κ^c) to catch up and keep up with the industry-wide technological trend, respectively. Dimmer prospects for future profit opportunities reduce expected benefits but not expected costs, which is the reason for less entry and more exits; hence, $\Delta NE_t < 0$.

The microeconomic mechanism is similar but more complicated and interesting on the innovation side. Δmc_t primarily reflects the productivity levels of firms, $\Delta \omega_{it}$, which can change either as a result of in-house R&D investments, ΔRD_t, or stochastic synergy draws upon successful mergers, ΔMA_t,

$$\Delta mc_t \propto \Delta \omega_{it} = \Delta RD_t + \Delta MA_t. \tag{21}$$

The bottom-right panel of Figure 7 features a line graph representing the counts of $\Delta \omega_{it}$, and two bar graphs representing ΔRD_t and ΔMA_t, respectively. The $\underline{N} = 5$ policy's contribution to "innovation" is negative on the synergy side because it blocks more mergers than under the $\underline{N} = 3$ regime.

However, the in-house R&D channel does not seem to sufficiently offset these forgone synergies, for two reasons. First, the policy promotes more competition ex post, but this increased competition does not necessarily translate into increased R&D. As Figure 5 illustrates, the equilibrium R&D investment CCPs all but stop increasing after three or four firms, and occasionally decrease. Second, the aforementioned destruction of firms' continuation values re-surfaces here and discourages in-house investments. The side effects of value destruction are not limited to entry and survival; they affect all kinds of investments including in-house R&D, which is an investment in productivity growth and hence incremental profits and values. The overall level of continuation values decreases, and so does the incremental values from making such investments. Thus, the impact of value destruction manifests itself through multiple side effects, which is why the "optimal" policy ($\underline{N} = 5$) does not substantially outperform its neighboring thresholds such as $\underline{N} = 3$ and $\underline{N} = 4$.

5.3 Failing Firms and Declining Industries

In a mature industry such as HDDs, regulators often have to deal with "failing firms," that is, firms that (i) are in imminent danger of failure (in a more severe condition than insolvency and close to ceasing operations), (ii) cannot be reorganized in Chapter 11 bankruptcy, and (iii) cannot find an alternative purchaser (or other less anti-competitive uses) of their assets.[39] To our knowledge, no formal economic analysis exists on this subject, because a systematic evaluation of failing firms requires a framework like ours. Exits (through liquidation) in our model meet all of the three criteria for "failing firms;" hence, our model can handle such cases, in principle. However, the equilibrium CCPs of exit are less than 10% in almost all states and periods in our baseline estimate. Consequently, we have chosen not to study failing firms per se but to ask a broader question regarding the optimal policy toward declining industries, in which exits become more likely.[40]

Table 7: Commitment Policies in Fast-Declining Industry

Threshold number of firms ($\underline{N}$)	1	2	3	4	5	6	7	8	9
$T = Dec$-2020									
Consumer surplus (%)	▲17.37	▲6.18	±0	+0.42	+0.77	+0.84	+0.89	+0.89	+0.93
Producer surplus (%)	+90.89	+28.83	±0	▲3.08	▲5.93	▲7.33	▲8.89	▲10.13	▲11.63
Social welfare (%)	▲10.28	▲3.89	±0	+0.19	+0.33	+0.31	+0.25	+0.17	+0.10
$T = Dec$-2015									
Consumer surplus (%)	▲10.58	▲4.52	±0	+0.34	+0.48	+0.49	+0.54	+0.54	+0.53
Producer surplus (%)	+64.42	+23.86	±0	▲2.77	▲4.66	▲5.93	▲7.58	▲8.87	▲10.35
Social welfare (%)	▲5.73	▲2.69	±0	+0.14	+0.15	+0.08	+0.01	▲0.07	▲0.17

Note: All welfare measures net present values as of January 1996, expressed in terms of percentage change from the baseline outcome under $\underline{N} = 3$.

Should the authority relax its merger policy in declining industries? We will answer this question as follows. We capture the notion of "declining industry" (and hence higher exit rates in equilibrium or "failing firms") by hypothetically eliminating much of the HDD demand in the post-sample period (i.e., after June 2015). Our baseline model assumes the demand will linearly decline to zero between June 2015 and December 2025, reflecting what we presume to be a consensus forecast among industry participants. By contrast, this subsection simulates alternative industry dynamics in which the demand converges to zero in December 2020 and December 2015 (i.e., $T = Dec$-2020 and Dec-2015), five and 10

[39] See McFarland and Nelson (2008) for legal details.

[40] "Declining industries" do not constitute a valid defense in the US legal context (except under a brief period during the Great Depression), and the permission of "recession cartels" in Japan was repealed in 1999. We are not using this phrase in a strictly legal sense.

years earlier than our baseline scenario. We solve these new games for equilibrium CCPs, simulate 10,000 histories, and calculate their average welfare performances. We maintain the baseline policy regime ($\underline{N} = 3$) throughout these procedures. Finally, within each of these hypothetical demand scenarios (i.e., $T = Dec$-2020 and Dec-2015), we compute welfare outcomes under alternative policy regimes (i.e., $\underline{N} \neq 3$) so that we can determine the optimal policy under each end-game scenario.

Table 7 demonstrates how the dynamic welfare tradeoff alters its balance, albeit slightly, when the industry is declining more quickly. The optimal static threshold continues to be five, but permitting mergers to quadropoly, triopoly, or even duopoly would not be as harmful as in Table 6, because not many consumers will be harmed when the demand is disappearing precipitously. The fact that $\underline{N} = 5$ continues to be optimal might appear surprising, but the contribution of the last five or 10 years of the industry's life cannot be too large in terms of present value as of January 1996. Hence, witnessing *any* qualitative change, such as the negative SW performance of $\underline{N} = 8$ and 9 under $T = Dec$-2015 scenario, is actually surprising. Thus, the optimal policy is likely to feature more relaxed thresholds, but the difference is small.

5.4 Opportunistic Policy

Thus far, we have considered only a static (or time-invariant) policy design that commits the authority to a particular merger threshold. We have intentionally kept our discussions within such static thresholds because of their simplicity and direct connection to the practitioners' rule of thumb. Detailed analysis of dynamic welfare tradeoff is quite complicated even under such a simple policy design. Nevertheless, a sophisticated reader would be wondering if the authority can craft a smarter policy than simply committing to $\underline{N} = 5$. Our short answer is "yes" in the short run and "no" in the long run.

Table 8 considers "smart" policies in which the authority acts opportunistically and alters the merger threshold ex post.[41] The optimal surprise policy is to initially promise no antitrust scrutiny at all (i.e., declare $\underline{N}^{pre} = 1$). An elusive quest for monopoly profits will attract massive entry and innovation early on (i.e., no value-destruction side effects). However, when the industry reaches $n_t = 3$, the planner should start blocking mergers, so

[41] We refrain from simulating more complicated policies (and their possible strategic interactions with the firms) because intuitive understanding of the results will become increasingly more difficult, the actual policy implementation will become impractical, and we could not find anecdotal or quantitative evidence. Nevertheless, these ideas do stimulate theoretical curiosity, and we refer the reader to MNSW (2014) and Jeziorski (2014) for such investigations.

that firms have to compete to death (i.e., $\underline{N}^{post} = 3$). This surprise ban on mergers at triopoly will ensure sufficient pro-competitive outcome ex post.[42]

Table 8: Performance of Opportunistic Policies

Promised threshold ($\underline{N}^{pre}$)	1	1	1	1	1	1	1	1	1
True threshold ($\underline{N}^{post}$)	1	2	3	4	5	6	7	8	9
Consumer surplus (%)	▲24.12	▲7.86	+1.04	+0.69	+0.75	+0.80	+0.86	+0.91	+0.91
Producer surplus (%)	+118.46	+34.50	▲5.24	▲4.52	▲5.95	▲7.11	▲8.71	▲9.96	▲11.10
Social welfare (%)	▲14.64	▲5.04	+0.62	+0.34	+0.30	+0.27	+0.23	+0.19	+0.11

Note: All welfare measures net present values as of January 1996, expressed in terms of percentage change from the baseline outcome under $\underline{N} = 3$ (both promised and true).

To some readers, this simulation experiment might appear too complicated and unrealistic at a first glance, but negative surprises are facts of life. In the American political context, for example, consider a long spell of the Republican "pro-business" regime, followed by stronger regulatory oversight under the Democratic regime. Another example is the inception of the Chinese antitrust policy in 2008. Its Ministry of Commerce (MOFCOM) almost stopped the latest HDD merger between Western Digital and HGST in 2012, which the authorities in the United States, Japan, South Korea, and Europe had already cleared. Thus, we believe the academic literature should clarify the pros and cons of surprise changes, so that policy makers can at least understand the true meaning of such actions.

In the long run, such a "smart" policy is not going to be wise, because governments cannot fool financial markets forever. One industry might be tricked, but the subsequent cohorts of high-tech industries may not. The authority can surprise only once.

6 Conclusion

This paper proposed an empirical model of mergers and innovation to study the process of industry consolidation, with HDDs as a working example. We used quantitative methods to clarify the dynamic welfare tradeoff inherent in antitrust policy, and determined the welfare-maximizing merger threshold to be five in the HDD context. That is, "4 are few and 6 are many" (Selten 1973).

[42] Computationally, we implement these opportunistic policies as follows. First, we start simulating the industry's history by using the equilibrium CCPs under $\underline{N} = \underline{N}^{pre} = 1$, which corresponds to the $\underline{N} = 1$ counterfactual in section 5.1. Second, whenever the simulated n_t reaches the true (unannounced) threshold, $\underline{N}^{post} \geqslant 1$, our algorithm switches to the equilibrium CCPs under $\underline{N} = \underline{N}^{post}$ and keeps simulating the history until $t = T$. Third, collect 10,000 simulated histories and calculate their average welfare performance. This average is the outcome we attribute to each pair $(\underline{N}^{pre}, \underline{N}^{post})$ that represents a particular ex-post policy.

This finding is specific to the parameters of consumers' preferences, production technology, and investment technology in our data; hence, each high-tech industry requires careful modeling and measurement, just like the actual enforcement of antitrust policy proceeds case by case. The fact that the authority has permitted mergers to triopoly in the HDD market (i.e., beyond our optimal threshold of five) does not appear particularly troubling, because the magnitude of welfare differences is small as long as the threshold is three or higher. Thus, the danger of type II errors (i.e., not rejecting what needs to be rejected) is not overwhelming. By contrast, permitting mergers to duopoly or monopoly would lead to negative welfare impacts that are larger by an order or two of magnitude.

Our model focuses on the direct or "unilateral" effect of mergers on prices through market structure and productivity, and does not incorporate the "coordinated" effect with respect to collusive conducts, such as those studied by Selten (1973) or Miller and Weinberg (2015). Hence, the negative effect on consumer surplus in our study represents a lower bound, and the actual harm of monopoly and duopoly could be greater in practice.

Moore's Law (or its HDD-equivalent, Kryder's Law) is another important subject beyond the scope of this paper. The coincidence of merger waves in the semiconductor industry and the slowdown of Moore's Law has prompted popular press to causally interpret this correlation: the structure-conduct-performance (SCP) paradigm. However, our structural analysis suggests a slower demand growth and exploding costs of innovation are the primary suspects for causing *both* more mergers and less innovation. Our framework can incorporate such secular technological trends on a larger scale without any conceptual difficulty, because the only change will be to re-define a larger state space that can span a wider range of productivity levels. A larger state space, however, creates a computational problem. Our current implementation already uses approximately 48 gigabytes of physical memory (DRAM). A drastic expansion of state space requires more data-storage capacity with faster access speed. Consequently, our structural estimation of Moore's Law has to wait until Moore's Law makes such a venture possible.[43]

[43] We would also need new IVs for demand estimation if we endogenize Moore's Law.

Appendices: Table of Contents

Appendix A lists our interviews with industry veterans. Appendices B, C, and D supplement the details of sections 2 (Model), 3 (Data), and 4 (Empirical Analysis), respectively.

Appendix A List of Interviews

For confidentiality reasons, we do not quote from our personal interviews with the industry sources. The only exceptions are historical overviews and remarks on events in the distant past (by the standard of Silicon Valley). Nevertheless, almost every modeling choice, parameterization, and estimation result has tight connections to the actual data-generating process, which we learned through these interviews.

Table 9: Interviews with Industry Sources

#	Date	Location	Name	Affiliation (position)
1	Various	TRENDFOCUS office (Cupertino, CA)	Mark Geenen John Kim John Chen Don Jeanette	TRENDFOCUS (president & VPs) Microscience International Komag Toshiba, Fujitsu
2	1/22/2015	Fibbar MaGees Irish pub (Sunnyvale, CA)	Reggie Murray	Ministor (founder) Maxtor (thin-film head) Memorex
3	2/27/2015	HGST/IBM office (San Jose, CA)	Currie Munce	HGST/IBM (SSD)
4	3/5/2015	SIEPR (Stanford, CA)	Lawrence Wu	NERA Consulting (president)
5	3/11/2015	SIEPR (Stanford, CA)	Orie Shelef	Former merger consultant
6	3/23/2015	Residence (Monte Sereno, CA)	Tu Chen	Komag (founder)
7	4/17/2015	Seagate headquarters (Cupertino, CA)	Jeff Burke	Seagate (VP of strategic marketing & research)
8	4/20/2015	Residence (Corona del Mar, CA)	Finis Conner	Conner Technology (founder) Conner Peripherals (founder) Seagate (co-founder) International Memories Inc. Shugart Associates (co-founder)
9	6/30/2015	BJ's restaurant & brewery (Cupertino, CA)	Peter Knight	Conner Technology (president) Conner Peripherals (senior VP) IBM
10	7/1/2015	Gaboja restaurant (Santa Clara, CA)	MyungChan Jeong	HGST/IBM (R&D engineer) Seagate, Maxtor Samsung Electronics

Note: Affiliations are listed from new to old. VP stands for vice president. SIEPR stands for the Stanford Institute for Economic Policy Research, where Igami spent his 2014–2015 sabbatical.

Appendix B Supplementary Materials for Section 2

Potential Entrant's Problem

Section 2.3 focused on the exposition of incumbent firms' problem. This section explains the detail of potential entrant's problem.

If nature picks a potential entrant i as a proposer, i draws $\varepsilon_{it}^0 = (\varepsilon_{it}^e, \varepsilon_{it}^o)$ and chooses to enter or stay out, which entail the following alternative-specific values:

$$\bar{V}_i^e(s_t, \varepsilon_{it}^e) = -\kappa^e + \varepsilon_{it}^e + \beta E[\Lambda_{i,t+1}(s_{t+1})|s_t, a_{it} = enter], \text{ and} \tag{22}$$

$$\bar{V}_i^o(s_t, \varepsilon_{it}^o) = \varepsilon_{it}^o + \beta E[\Lambda_{i,t+1}(s_{t+1})|s_t, a_{it} = out], \tag{23}$$

respectively. Thus, the potential entrant's value *after* drawing ε_{it}^0 is

$$V_{it}^0(s_t, \varepsilon_{it}^0) = \max\left\{\bar{V}_i^e(s_t, \varepsilon_{it}^e), \bar{V}_i^o(s_t, \varepsilon_{it}^o)\right\}, \tag{24}$$

and its expected value *before* drawing ε_{it}^0 is

$$EV_{it}^0(s_t) = E_\varepsilon\left[V_{it}^0(s_t, \varepsilon_{it}^0)\right] = \gamma + \ln\left[\exp\left(\tilde{V}_{it}^e\right) + \exp\left(\tilde{V}_{it}^o\right)\right]. \tag{25}$$

These expressions correspond to equations 1 through 6 in the main text.

When the non-mover is a potential entrant, its non-mover expected value is simpler than the incumbent's version in equation 8,

$$\begin{aligned} W_{it}^{0j}(s_t) = \; & \sigma_{it}(a_{jt} = exit)\beta E[\Lambda_{i,t+1}(s_{t+1})|s_t, a_{jt} = exit] \\ & + \sigma_{it}(a_{jt} = stay)\beta E[\Lambda_{i,t+1}(s_{t+1})|s_t, a_{jt} = idle] \\ & + \sigma_{it}(a_{jt} = invest)\beta E[\Lambda_{i,t+1}(s_{t+1})|s_t, a_{jt} = invest] \\ & + \sum_{k \neq i,j} \sigma_{it}(a_{jt} = merge\ k)\beta E[\Lambda_{i,t+1}(s_{t+1})|s_t, a_{jt} = merge\ k], \end{aligned} \tag{26}$$

because it does not earn a profit, pay a fixed cost, or become a merger target.

When nature picks a potential entrant j as a mover, equations 8 and 26 become

$$\begin{aligned} W_{it}^{j}(s_t) &= \pi_i(s_t) - \kappa^c \\ &+\sigma_{it}\left(a_{jt}^0 = enter\right)\beta E\left[\Lambda_{i,t+1}(s_{t+1})\,|s_t, a_{jt}^0 = enter\right] \\ &+\sigma_{it}\left(a_{jt}^0 = out\right)\beta E\left[\Lambda_{i,t+1}(s_{t+1})\,|s_t, a_{jt}^0 = out\right], \text{ and} \end{aligned} \tag{27}$$

$$\begin{aligned} W_{it}^{0j}(s_t) &= \sigma_{it}\left(a_{jt}^0 = enter\right)\beta E\left[\Lambda_{i,t+1}(s_{t+1})\,|s_t, a_{jt}^0 = enter\right] \\ &+\sigma_{it}\left(a_{jt}^0 = out\right)\beta E\left[\Lambda_{i,t+1}(s_{t+1})\,|s_t, a_{jt}^0 = out\right] \end{aligned} \tag{28}$$

for an incumbent non-mover and a potential entrant non-mover, respectively.

These value functions entail the following optimal choice probabilities before potential-entrant mover i draws ε_{it}^0,

$$\Pr\left(a_{it}^0 = action\right) = \frac{\exp\left(\tilde{V}_{it}^{action}\right)}{\exp\left(\tilde{V}_{it}^{e}\right) + \exp\left(\tilde{V}_{it}^{o}\right)}, \tag{29}$$

which corresponds to equation 9.

Appendix C Supplementary Materials for Section 3

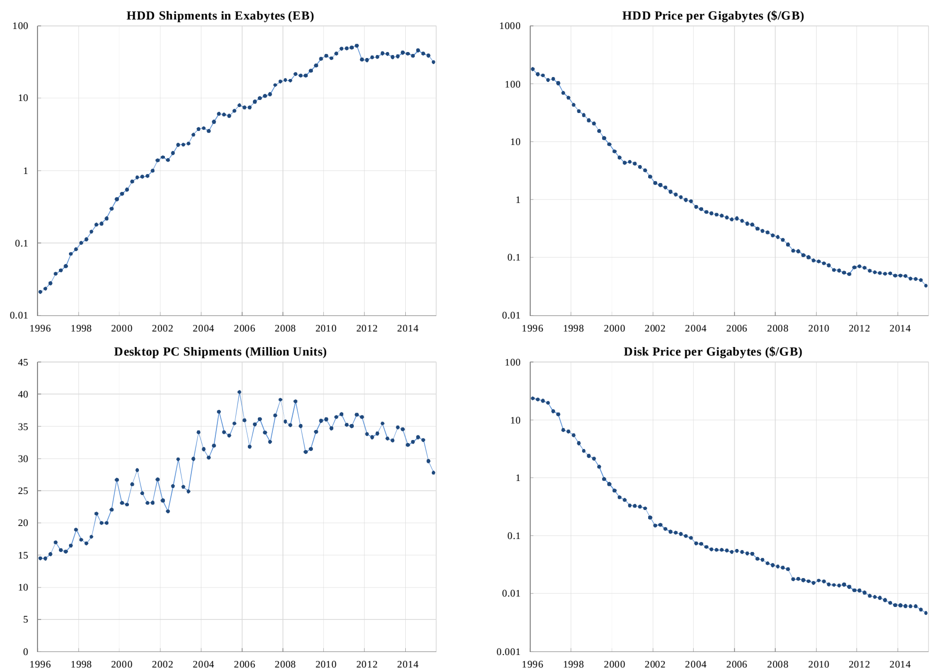
Data Patterns Underlying Demand Estimation (Panel A)

Figure 8 summarizes data patterns of Panel A, that is, the four variables for demand estimation (Q_t, P_t, X_t, Z_t). The HDD shipment volume in EB (Q_t) has grown steadily on the back of PC shipments (X_t) as the upper- and lower-left panels show. The HDD price per GB (P_t) has been decreasing as a result of Kryder's Law. With this secular trend in storage density, the disk price per GB (Z_t) has fallen dramatically, because more data can be stored on the disk surface of the same size. The upper- and lower-right panels capture these trends. Thus, the downward trends in P_t and Z_t reflect both process innovation (i.e., lower marginal costs) and product innovation (i.e., higher "quality" or data-storage capacity per HDD unit) in this industry.

Market Shares before and after Mergers (Panel B)

In section 3, we visualized and summarized the data patterns of firm-level market shares (Panel B) in Figure 1 and Table 1. In this section, we supplement these exhibits with the list of 14 merger cases and the Herfindahl-Hirschman Index (HHI).

Figure 8: Data for Demand Estimation at the Level of Gigabytes (GB)



Note: See Sections 3.2 and 4.1 for summary statistics and demand estimation, respectively.

Table 10 shows the combined market share of the acquiring firm and the target firm declined after merger in each of the 14 cases, which suggests the theoretical prediction of free-riding by the non-merging parties could be a real phenomenon. At the same time, the acquiring firms managed to achieve expansions relative to their individual pre-merger market shares, which is consistent with our interviews with the industry participants, in which they explained gaining market shares as the primary motivation for mergers. Finally, a larger firm acquires a smaller firm in most of the cases, which seems intuitive.

Figure 9 overlays the historical HHI on the number of firms, n_t. The HHI correlates negatively with n_t by construction. It started at around $2,000$ in the late 1970s, decreased to $1,000$ in the mid 1980s due to massive entry, and was mostly unaffected by the shakeouts because fringe firms' liquidation-exit did not really change the surviving firms' market shares. Once n_t reached 10 around year 2000, the consolidation process through mergers increased the HHI from $1,500$ to $2,500$ during the first decade of the 21st century, and then to almost $4,000$ on the back of the 5-to-4 and 4-to-3 mergers.

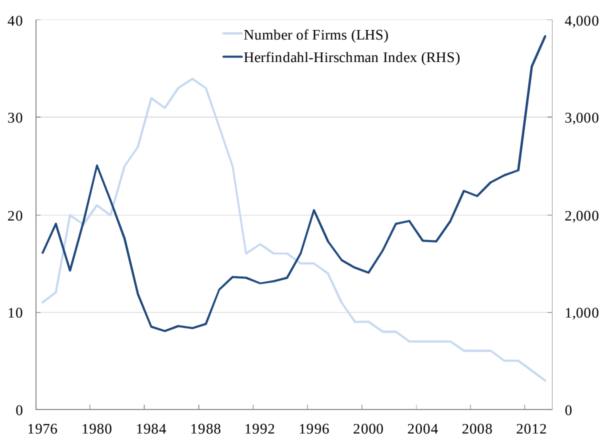
Table 10: Market Shares before and after Mergers (%)

Year	Target name	Acquiror name	ms^T Before	ms^A Before	$ms^T + ms^A$ Before	$ms^T + ms^A$ After
1982	Memorex	Burroughs	7.83	1.85	9.68	2.73
1983	ISS/Univac/Unisys	Control Data	0.75	27.08	27.83	19.85
1984	Vertex	Priam	0.93	2.52	3.45	2.78
1988	Plus Dev.	Quantum	0.89	1.41	2.30	4.64
1988	Imprimis	Seagate	13.92	18.16	32.08	29.23
1989	MiniScribe	Maxtor	5.68	4.99	10.68	8.53
1994	DEC	Quantum	1.65	18.60	20.25	20.68
1995	Conner	Seagate	11.94	27.65	39.58	35.41
2001	Quantum	Maxtor	13.87	13.87	27.73	26.84
2002	IBM	Hitachi	13.86	3.64	17.50	17.37
2006	Maxtor	Seagate	8.19	29.49	37.67	35.27
2009	Fujitsu	Toshiba	4.41	10.32	14.72	11.26
2011	Samsung	Seagate	6.89	39.00	45.89	42.82
2012	Hitachi	Western Digital	20.32	24.14	44.46	44.27

Note: ms^T and ms^A denote the target and the acquiring firms' market shares, respectively. For each merger case, "before" refers to the last calendar quarter in which ms^T was recorded separately from ms^A, and "after" is four quarters after "before." Alternative time windows including 1, 8, and 12 quarters lead to similar patterns.

Source: *DISK/TREND* Reports (1977–99), *TRENDFOCUS* Reports (1996–2014), and interviews.

Figure 9: Herfindahl-Hirschman Index (HHI) of the Global HDD Market



Note: The HHI is the sum of the squares of the firm's market shares.

Appendix D Supplementary Materials for Section 4

D.1 Supplementary Materials for Section 4.1

Demand Estimates by Subsample

Our initial demand estimates (Table 3, columns 1, 2, and 3) used the entire sample period, implicitly assuming the demand function remained constant over time. However, changing uses of digital technology could have altered the consumers' willingness to pay for the same amount of data storage. To investigate this possibility, we estimate our demand model using two subsamples (i.e., the first and the second halves). Table 11 shows the first-half and the second-half estimates for the main parameter, the price coefficient (α_1), are within the 95% confidence intervals of each other, across all three specifications. Thus, consumers' valuation for gigabytes of data storage does not exhibit a time trend in a statistically significant manner.

Table 11: Demand Estimates by Subsample

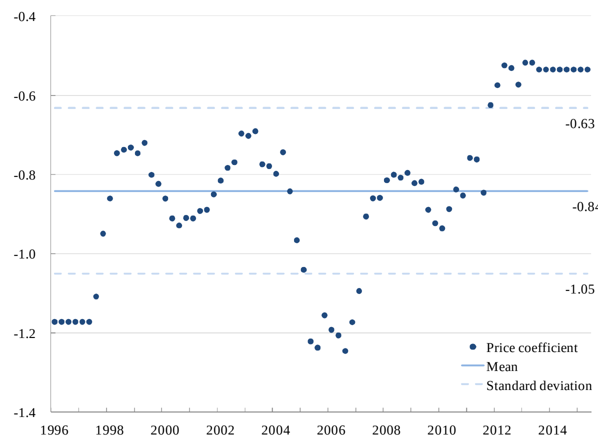
Dependent variable:	(1)		(2)		(3)	
log total EB shipped	OLS		IV-1		IV-2	
Subsample period:	First half	Second half	First half	Second half	First half	Second half
log price per GB (α_1)	$-.8165^{***}$	$-.8594^{***}$	$-.8188^{***}$	$-.8504^{***}$	$-.8959^{***}$	$-.8624^{***}$
	(.0246)	(.0264)	(.0172)	(.0233)	(.0484)	(.0249)
log PC shipment (α_2)	$.8053^{***}$	1.6302^{***}	$.7896^{***}$	1.6191^{***}	.2773	1.6340^{***}
	(.1728)	(.2422)	(.1222)	(.3809)	(.3191)	(.3797)
Constant (α_0)	-1.6405^{***}	-4.3901^{***}	-1.5868^{***}	-4.3337^{***}	.1649	-4.4094^{***}
	(.5863)	(.8718)	(.4102)	(1.3306)	(1.0895)	(1.3275)
Number of observations	39	39	39	39	39	39
Adjusted R^2	.9972	.9746	.9973	.9765	.9966	.9766
First-stage regression						
IV for HDD price	–	–	*Disk price*	*Disk price*	*Time trend*	*Time trend*
F-value	–	–	2973.32	536.17	350.51	1056.23
Adjusted R^2	–	–	.9944	.9638	.9346	.9824

Note: Standard errors are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Figure 10 shows the price-elasticity estimates from a refined version of the estimation by subsample, in which we roll through the full sample with a 12-quarter window. The use of wider windows (i.e., 16 and 20 quarters) produced similar patterns, whereas a narrower window with eight quarters led to highly volatile estimates.

We see no obvious trend. The exceptions are several low values (i.e., *more* elastic demand) at the beginning, and higher values (i.e., *less* elastic demand) at the end, but the

Figure 10: Rolling Estimates of Price Elasticity



Note: Each dot represents a 12-quarter rolling estimate of the price coefficient, α_1. This plot visualizes Table 3 (column 4).

first six and the last six quarters are carbon copies of the adjacent estimates because one can run fewer than 78 regressions on a sample of 78 observations. The cyclicality has no obvious explanations either. The IT boom around 2000 coincides with marginally more elastic demand, but no such event accompanied another streak of elastic demand around 2006. Thus, we do not see a time trend or systematic fluctuations in our rolling estimates of people's willingness to pay. Nevertheless, we believe it is natural for the demand parameters for high-tech products to exhibit some variation across time, and have chosen to use this rolling-estimation version as the baseline for subsequent analysis.

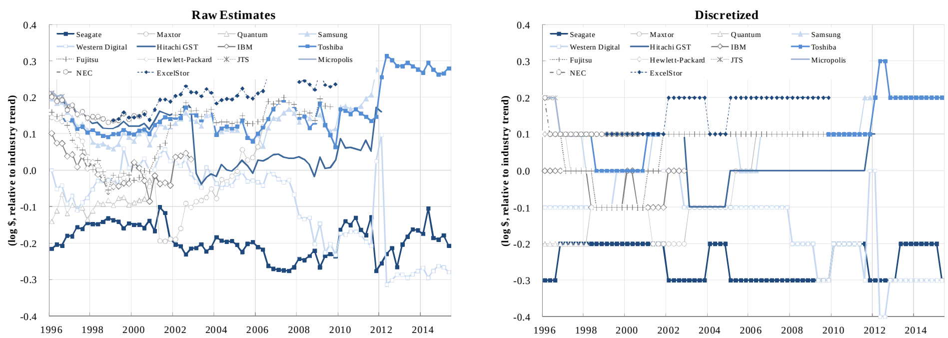
D.2 Supplementary Materials for Section 4.2

Discretization of Productivity Levels

We define the empirical state space by discretizing the levels of firm-specific productivity based on the marginal cost estimates in section 4.2. Figure 11 (left) plots the trajectories of marginal costs at the firm level between 1996 and 2015. Because the entire industry has experienced a secular trend of cost reduction, we de-trend these estimates and express them relative to the trajectory of Kryder's Law, in the natural logarithm of dollars.

To parameterize the dynamic oligopoly game parsimoniously and keep it computationally tractable, we discretize this relative marginal-cost space as shown in Figure 11 (right). This discretization scheme eliminates small wiggles of productivity evolution but preserves the overall patterns of these firms' relative performances, including their major shifts as well

Figure 11: Marginal Cost Estimates and Their Discretization



Note: The left panel plots our marginal cost estimates. The right panel displays its discretized version.

as leader-follower differences (at least most of the persistent ones). Finer grids resulted in too many zig-zag patterns, frequently amplifying small wiggles that happened to cross the discretization thresholds. Coarser grids tended to eliminate such noises, but the transitions between levels became too infrequent and each of these productivity changes became too impactful in terms of its profit implications via Cournot competition. After experimenting with these alternative grids, we have come to prefer the 0.1 log-dollar grid because it appears to strike the right balance between noise reduction and smooth transitions.

These discretized marginal cost estimates (say, $\overline{mc}_{it}$s) span the state space of firm-specific productivity levels, which will be denoted by $\omega_{it} \in \{\bar{\omega}_1, \bar{\omega}_2, ..., \bar{\omega}_M\}$, where $M = 7$ with our preferred grid. Note the ranking convention reverses as we redefine marginal costs as productivity levels. That is, a lower marginal cost will be referred to as a high-productivity level.

D.3 Supplementary Materials for Section 4.3

Fixed Costs and Accounting Data

We determine the fixed cost of operations and technological catch-up, κ^c, directly from accounting data rather than estimating it along with the three sunk-cost parameters $(\kappa^i, \kappa^m, \kappa^e)$ in section 4.3, for the following reasons. Our previous experience with the estimation of dynamic games (i.e., Igami 2015, 2016; Igami and Yang 2016) suggests the fixed cost of operations is an order of magnitude smaller than the sunk costs of entry and other major investments (e.g., product and process innovations). Moreover, the fixed-cost estimates tend

to be statistically indistinguishable from zero when sparse data are used, and play a relatively minor role in the overall performance of the dynamic models. Thus, rather than adding κ^c as another parameter to the main estimation procedure, we prefer pinning it down separately from auxiliary data, such as the firms' financial statements.

Accounting data are not always conceptually equivalent to the objects in economic models, as our discussion of profits in section 4.2 clarifies. But they are nevertheless useful for some purposes, such as fixing the values of a relatively unimportant parameter that cannot be precisely estimated anyway. Our notion of κ^c is something stable over time, and the accounting data on SGA and R&D expenses share this property.

Table 12: Summary Statistics of Accounting Data on Fixed Costs

Variable	Unit of measurement	Number of observations	Mean	Standard deviation	Minimum	Maximum
Fixed cost, κ^c	Million $	35	$1,078$	686.7	230.9	$2,422$
Year, t	Fiscal year	35	$2,007$	5.419	$1,996$	$2,015$
Productivity level, ω_{it}	Levels 1–7	35	4.943	0.996	3	6
Indicator$\{i = Seagate\}$	0 or 1	35	0.428	0.502	0	1
Indicator$\{(i,t) \in Special\}$	0 or 1	35	0.114	0.323	0	1

We estimate κ^c from the financial statements of Seagate Technology and Western Digital between 1996 and 2015. We rely on these firms simply because they are the only publicly traded companies for which systematic records exist. Moreover, they specialize in the manufacturing of HDDs, whereas other survivors such as Hitachi and Toshiba are conglomerates and disclose limited information on HDD-specific activities. The two firms clearly represent a highly selective sample but not a terrible source of information when our only purpose is to capture a ballpark trend in operating costs over two decades.

Table 12 shows summary statistics. Sample size is smaller than 40 (i.e., two firms times 20 years) because Seagate became privately owned for financial restructuring in 2000 and its financial statements lost consistency after it went public again. Our main variable is *fixed cost*, which is the sum of SGA and R&D expenses. The right-hand-side variables include *year*, *productivity level* (based on the discretized version of our marginal-cost estimates), *Seagate dummy* (the omitted category is Western Digital), and a *special-occasion dummy* (to distinguish abnormal periods for Western Digital when its facilities were hit by a natural disaster).

Table 13 shows the results of OLS regressions. The time trend is positive and statistically significant, whereas the productivity level (i.e., control for concurrent firm sizes) is positive

Table 13: Fixed-Cost Estimates from Accounting Data

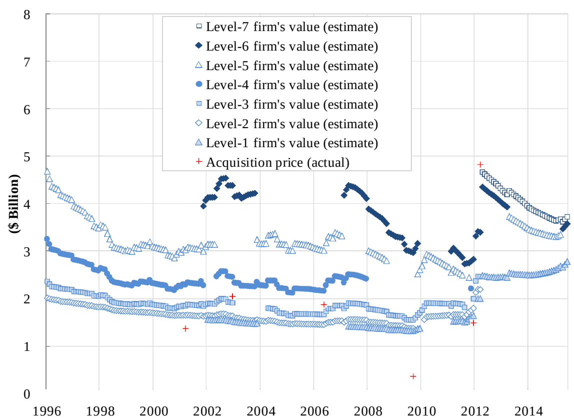
Dependent variable: Fixed cost, κ^c	(1) OLS	(2) OLS	(3) OLS	(4) OLS
Year (t)	89.80***	–	61.73***	48.13***
	(11.16)	(–)	(13.95)	(9.81)
Productivity level (ω_{it})	–	540.09***	332.30***	25.96
	(–)	(74.50)	(75.86)	(72.76)
$I\{(i,t) \in Special\}$	–	–	–	728.61***
	(–)	(–)	(–)	(132.69)
$I\{i = Seagate\}$	–	–	–	1,182***
	(–)	(–)	(–)	(187.4)
Number of observations	35	35	35	35
Adjusted R^2	0.633	0.603	0.746	.888

but imprecisely estimated presumably because of multi-collinearity. Historically, Seagate spent more than Western Digital, but the latter had to spend large sums to recover from a flood in Thailand in October 2011. We use predicted fixed costs based on the last (full) specification as $\kappa_t^c(\omega_{it})$ in our main estimation task in section 4.3.

Implied vs. Actual Acquisition Prices

We conduct a sanity check of fit by comparing predicted enterprise values and the actual acquisition prices. Figure 12 plots our firm-value estimates along the historical path of market structure in the data, and overlays the actual transaction prices in the six merger cases from *Thomson*'s financial data (marked by red crosses). Because target firms' stand-alone values underpin their equilibrium acquisition prices in our model, comparison of the estimated values and the actual acquisition prices provides a ballpark assessment of the fit in terms of dollar values. In at least half the cases, each of the acquisition prices is located close to the estimated value of firms with the corresponding productivity level (1, 2, 3, or 4) and stays within the range of the focal level and its adjacent level. Thus, we regard the estimated model as a reasonable benchmark with which we may compare our counterfactual simulation to assess the impacts of a hypothetical merger policy.

Figure 12: Firm-Value Estimates and Actual Acquisition Prices



Note: Red crosses represent the actual acquisition prices in the six merger cases from *Thomson* database. The other seven markers represent our estimates of equilibrium firm values along the historical path of market structure in the data.

References

[1] Aghion, Philippe, Nick Bloom, Richard Blundell, Rachel Griffith and Peter Howitt (2005) "Competition and innovation: An inverted-U relationship," *Quarterly Journal of Economics*, 120 (2): 701–728.

[2] Aguirregabiria, Victor, and Pedro Mira (2007) "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75 (1): 1–53.

[3] Arcidiacono, Peter, Patrick Bayer, Jason R. Blevins, and Paul B. Ellickson (2015) "Estimation of Dynamic Discrete Choice Models in Continuous Time with an Application to Retail Competition," *Review of Economic Studies*, forthcoming.

[4] Arrow, Kenneth J. (1962) "Economic Welfare and the Allocation of Resources to Invention," in R.R. Nelson (Ed.), *The Rate and Direction of Economic Activity.* N.Y., Princeton University Press.

[5] Bajari, Patrick, C. Lanier Benkard, and Jonathan Levin (2007) "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75 (5): 1331–70

[6] Baron, David P., and John A. Ferejohn (1989) "Bargaining in Legislatures," *American Political Science Review*, 83 (4): 1181–1206.

[7] Benkard, C. Lanier (2004) "A Dynamic Analysis of the Market for Wide-Bodied Commercial Aircraft." *Review of Economic Studies*, 71: 581–611.

[8] Berry, Steven T., and Ariel Pakes (1993) "Some Applications and Limitations of Recent Advances in Empirical Industrial Organization: Merger Analysis," *American Economic Review Papers and Proceedings*, 83(2): 247–252.

[9] Christensen, Clayton M.. 1993. "The Rigid Disk Drive Industry: A History of Commercial and Technological Turbulence." *Business History Review*, 67: 531–88.

[10] Collard-Wexler, Allan (2013) "Demand Fluctuations in the Ready-Mix Concrete Industry," *Econometrica*, 81 (3): 1003–1037.

[11] Crawford, Gregory S., and Ali Yurukoglu (2012) "The Welfare Effects of Bundling in Multichannel Television Markets," *American Economic Review*, Vol. 102, pp. 643-85.

[12] Dasgupta, Partha and Joseph Stiglitz (1980) "Industrial Structure and the Nature of Innovative Activity," *Economic Journal*, Vol. 90, No. 358, pp. 266-293.

[13] Demsetz, Harold (1973) "Industry structure, market rivalry, and public policy," *Journal of Law and Economics*, April, 1–90.

[14] Deneckere, Raymond and Carl Davidson (1985) "Incentives to Form Coalitions with Bertrand Competition," *RAND Journal of Economics*, Vol. 16, pp. 473-86.

[15] Egesdal, Michael, Zhenyu Lai, and Che-Lin Su (2015) "Estimating dynamic discrete-choice games of incomplete information," *Quantitative Economics*, 6: 567–597.

[16] Entezarkheir, Mahdiyeh, and Saeed Moshiriz (2015) "Merger Induced Changes of Innovation: Evidence from a Panel of U.S. Firms," mimeo.

[17] Ericson, Richard, and Ariel Pakes (1995) "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economic Studies*, 62 (1): 53–82.

[18] Farrell, Joseph and Carl Shapiro (1990) "Horizontal Mergers: An Equilibrium Analysis," *American Economic Review*, Vol. 80, pp. 107-26.

[19] Federal Trade Commission (2013) "Horizontal Merger Investigation Data: Fiscal Years 1996–2011."

[20] Gans, Joshua (2016) *The Disruption Dilemma*, The MIT Press, MA.

[21] Gilbert, Richard J., and Hillary Greene (2015) "Merging Innovation into Antitrust Agency Enforcement of the Clayton Act," *George Washington Law Review*, 83: 1919–1947.

[22] ——— and David Newbery (1982) "Preemptive Patenting and the Persistence of Monopoly," *American Economic Review*, 72(2): 514–26.

[23] Goettler, Ronald, and Brett Gordon (2011) "Does AMD spur Intel to innovate more?" *Journal of Political Economy*, 119(6): 1141–200.

[24] Gowrisankaran, Gautam (1995) "A Dynamic Analysis of Mergers," Ph.D. dissertation, Yale University.

[25] ——— (1999) "A Dynamic Model of Endogenous Horizontal Mergers," *RAND Journal of Economics*, Vol. 30, pp. 56-83.

[26] Ho, Kate (2009) "Insurer-Provider Networks in the Medical Care Market," *American Economic Review*, Vol. 99, pp. 393-430.

[27] Igami, Mitsuru (2015) "Estimating the Innovator's Dilemma: Structural Analysis of Creative Destruction in the Hard Disk Drive Industry, 1981–1998," *Journal of Political Economy*, forthcoming.

[28] ——— (2016) "Industry Dynamics of Offshoring: The Case of Hard Disk Drives," mimeo.

[29] ——— and Nathan Yang (2016) "Unobserved Heterogeneity in Dynamic Games: Cannibalization and Preemptive Entry of Hamburger Chains in Canada," *Quantitative Economics*, 7 (2): 483–521.

[30] Iskhakov, Fedor, John Rust, and Bertel Schjerning (2014) "The Dynamics of Bertrand Price Competition with Cost-Reducing Investments," manuscript, Georgetown University.

[31] ——— (2016) "Recursive Lexicographical Search: Finding All Markov Perfect Equilibria of Finite State Directional Dynamic Games." *Review of Economic Studies*, 83 (2): 658-703.

[32] Iskhakov, Fedor, Jinhyuk Lee, John Rust, Bertel Schjerning, and Kyoungwon Seo (2016) "Comment on 'Constrained Optimization Approaches to Estimation of Structural Models'," *Econometrica*, 84 (1): 365–370.

[33] Jeziorski, Przemyslaw (2014) "Empirical Model of Dynamic Merger Enforcement: Choosing Ownership Caps in U.S. Radio," mimeo.

[34] Kim, Myongjin (2015) "Strategic Responses to Used Goods Markets: Airbus and Boeing." Manuscript, University of Oklahoma.

[35] Kreps, David M. and Jose A. Scheinkman (1983) "Quantity precommitment and Bertrand competition yield Cournot outcomes," *The Bell Journal of Economics*, Vol. 14, No. 2, pp. 326–37.

[36] Lamoreaux, Naomi R.. 1985. *The Great Merger Movement in American Business, 1895-1904* (New York: Cambridge University Press). Reissued in paperback, 1988.

[37] Marshall, Guillermo, and Álvaro Parra (2015) "Mergers in Innovative Industries," mimeo.

[38] McFarland, Henry, and Philip Nelson (2008) "Failing Firms and Declining Industries," *Issues in Competition Law and Policy*, 3: 1691–1716.

[39] Mermelstein, Ben, Volker Nocke, Mark A. Satterthwaite, and Michael D. Whinston (2014) "Internal Versus External Growth in Industries with Scale Economies: A Computational Model of Optimal Oligopoly Policy," mimeo.

[40] Miller, Nathan H., and Matthew C. Weinberg (2015) "Mergers Facilitate Tacit Collusion: Empirical Evidence from the U.S. Brewing Industry," Manuscript, Georgetown University.

[41] Nevo, Aviv (2000) "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry," *RAND Journal of Economics*, Vol. 31, No. 3, pp. 395-421.

[42] Okada, Akira (1996) "A Noncooperative Coalitional Bargaining Game with Random Proposers," *Games and Economic Behavior*, 16: 97–108.

[43] Ozcan, Yasin (2015) "Innovation and Acquisition: Two-Sided Matching in M&A Markets," mimeo.

[44] Pakes, Ariel (1986) "Patents as options: Some estimates of the value of holding European patent stocks," *Econometrica*, 54 (4): 755–784.

[45] ———, Michael Ostrovsky, and Steven Berry (2007) "Simple estimators for the parameters of discrete dynamic games (with entry/exit examples)," *RAND Journal of Economics*, 38 (2): 373–399.

[46] Perry, Martin K. and Robert H. Porter (1985) "Oligopoly and the Incentive for Horizontal Merger," *American Economic Review*, Vol. 75, pp. 219-27.

[47] Pesendorfer, Martin, and Philipp Schmidt-Dengler (2008) "Asymptotic Least Squares Estimators for Dynamic Games," *Review of Economic Studies*, 75 (3): 901–928.

[48] Qiu, Larry and Wen Zhou (2007) "Merger Waves: a Model of Endogenous Mergers," *RAND Journal of Economics*, Vol. 38, pp. 214-226.

[49] Rust, John (1987) "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, Vol. 55, pp. 999–1033.

[50] Ryan, Stephen P. (2012) "The Costs of Environmental Regulation in a Concentrated Industry." *Econometrica*, 80(3): 1019–61.

[51] Salant, Stephen W., Sheldon Switzer, and Robert J. Reynolds (1983) "Losses from Horizontal Merger: The Effects of an Exogenous Change in Industry Structure on Cournot-Nash Equilibrium," *Quarterly Journal of Economics*, Vol. 98, pp. 185-199.

[52] Scherer, Frederic M. (1965) "Firm Size, Market Structure, Opportunity, and the Output of Patented Inventions." *American Economic Review*, 55(5): 1097–1125.

[53] Schumpeter, Joseph A. (1942) *Capitalism, Socialism, and Democracy*, Harper & Brothers, NY.

[54] Scotchmer, Suzanne (2004) *Innovation and Incentives*, MIT Press, Cambridge, MA.

[55] Selten, Richard (1973) "A Simple Model of Imperfect Competition, where 4 Are Few and 6 Are Many," *International Journal of Game Theory*, 2(1): 141–201.

[56] Stahl, C. Jessica (2011) "A Dynamic Analysis of Consolidation in the Broadcast Television Industry," mimeo.

[57] Stigler, George J. (1950) "Monopoly and Oligopoly by Merger," *American Economic Review*, Papers and Proceedings, 40 (2): 23–34.

[58] Sutton, John (1998) *Technology and Market Structure: Theory and History*. Cambridge, MA: The MIT Press.

[59] Takahashi, Yuya (2015) "Estimating a War of Attrition: The Case of the US Movie Theater Industry," *American Economic Review*, 105 (7): 2204–41.

[60] Tirole, Jean (1988) *The Theory of Industrial Organization*, MIT Press, Cambridge, MA.

[61] Werden, Gregory J. and Luke M. Froeb (1994) "The Effects of Mergers in Differentiated Products Industries: Logit Demand and Merger Policy," *Journal of Law, Economics, & Organization*, Vol. 10, No. 2, pp. 407-426.

[62] Williamson, Oliver E. (1968) "Economies as an Antitrust Defense: The Welfare Tradeoffs," *American Economic Review*, Vol. 58, No. 1, pp. 18-36.