UNIL | Université de Lausanne
Département de finance

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

FINANCE RESEARCH SEMINAR SUPPORTED BY UNIGESTION

"Mergers, Innovation, and Entry-Exit Dynamics: Consolidation of the Hard Disk Drive Industry, 1996-2016"

Mitsuru IGAMI,
Yale University, Department of Economics

Friday, June 28, 2019, 10:30-12:00
Room 126, Extranef building at the University of Lausanne

UNIGESTION

Mergers, Innovation, and Entry-Exit Dynamics: Consolidation of the Hard Disk Drive Industry, 1996–2016*

Mitsuru Igami† Kosuke Uetake‡

January 4, 2019

Abstract

How far should an industry be allowed to consolidate when competition and innovation are endogenous? We develop 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 current rule-of-thumb policy, which stops mergers when three or fewer firms exist, strikes approximately the right balance between pro-competitive effects and value-destruction side effects in this dynamic welfare tradeoff.

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

JEL classifications: L13, L41, L63, O31.

*For detailed and insightful suggestions, we thank the editor, Aureo de Paula, and three anonymous referees. For helpful comments, we thank John Rust, Paul Ellickson, April Franco, Joshua Gans, Allan Collard-Wexler, and Philipp Schmidt-Dengler, 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 2016, 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 in a dynamic setting; hence, a given merger increases the likelihood of subsequent mergers.[4] Its impact on subsequent innovation is more complicated because the competition-innovation relationship crucially hinges on demand, supply, and investment.[5] These ex-post changes in competition

[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.

[4] Qiu and Zhou (2007) study a dynamic game with Cournot competition in every period, and find the incremental value from a merger increases as the number of firms decreases as a result of previous mergers.

[5] For example, Marshall and Parra (2018) show that competition *increases* innovation when the leader-follower profit gap is weakly increasing in the number of firms, which holds under some parameterizations of Bertrand and Cournot games with homogeneous goods. They also show that the necessary (but not sufficient) condition for competition to *decrease* innovation is that the gap decreases with the number of firms, which

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 may in turn reduce their ex-ante investments in productivity, survival, 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,[6] 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).[7] Third, workhorse models of dynamic oligopoly games such as Ericson and Pakes (1995) entail multiple equilibria, which makes the application of full-solution estimation methods such as Rust (1987) challenging, because point identification will be difficult 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 2017, 2018). Instead, we propose a random-mover dynamic game in which the turn-to-move arrives stochastically. Dynamic games with stochastically

holds under (some other parameterizations of) a homogeneous-good Cournot game and a differentiated-product Bertrand game. The key parameters of their theoretical model includes the step size of innovation, the fixed cost of innovation, and the convexity of (the variable part of) the cost of innovation. Certain combinations of the parameter values could lead to nonmonotonic relationships as well.

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

[7]CCP-based methods are proposed by Hotz and Miller (1993) and Hotz, Miller, Sanders, and Smith (1994) to alleviate the computational burden for estimating dynamic structural models. Their first step estimates policy functions as CCPs by using data on actions and states. Their second step estimates the underlying structural parameters by calculating value functions that are implied by the empirical CCPs. These methods require the first step to be nonparametric.

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 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., technological improvements in areal density).[8] Our dataset consists of three elements (Panels A, B, and C). 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 major supply disruption due to flood in Thailand in 2011. 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

[8]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. We endogenize it as well.

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 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.[9] 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) and Reinganum (1983). Second, this competition-innovation relationship becomes heterogeneous and nonmonotonic with more than three firms. Thus, our structural competition-innovation curve exhibits a "plateau"

[9] 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 5 different types of firms and 76,160 industry states in each of the 360 periods. We perform this subroutine in C++, and the estimation procedure takes less than a week on a regular desktop PC with a quad-core 3.60GHz CPU, 32GB RAM, and a 64-bit operating system.

shape instead of the famous "inverted U."[10] Third, mergers become more attractive as the industry matures, and all kinds of pairs can merge.

In section 5, we conduct counterfactual policy simulations to answer our main question: How far should the industry be allowed to consolidate? We find the current rule-of-thumb policy (which blocks mergers if three or fewer firms exist) is reasonably close to maximizing the discounted present value of social welfare. We clarify the underlying mechanism by showing the effects of various policy regimes on the number of firms and technological frontier, as well as the firms' endogenous choices between mergers, innovation, and entry-exit that determine the paths of competition and innovation. These results highlight the dynamic welfare tradeoff between the pro-competitive benefits of blocking mergers and the value-destruction side effects. We conclude in section 6 by discussing other policy implications and limitations.

Literature

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 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 (2018, henceforth MNSW) numerical theory of duopoly with

[10] Marshall and Parra (2018) thoroughly investigate these shapes in their theoretical work.

mergers and investments, which Marshall and Parra (2018) 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.[11] 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.[12] 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.[13]

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 (2017, 2018).[14] 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). Stochastically alternating moves have been applied to bargaining games, including Diermeier, Eraslan, and Merlo (2003) and Merlo and Tang (2012). Iskhakov, Rust, and Schjerning (2014, 2016) numerically study Bertrand duopoly with "leap-frogging" process innovations with a random-mover setup.

Igami (2017, 2018) studied the HDD industry as well, but the similarities end there. Our paper differs from his in three major ways: questions, data, and models. The two existing papers studied (i) the introduction of new products and offshoring, respectively, (ii) using old data from 1976 to 1998, (iii) in a model without mergers or stochastically alternating moves. By contrast, we study merger policy, use a completely new data source on the latest process of consolidation (1996–2016), and endogenize mergers without imposing any deterministic order of moves. We also endogenize the advances of the technological frontier (i.e., Kryder's Law), which were assumed exogenous in the previous papers.

[11] Other canonical references include Wolpin (1984, 1987) and Pakes (1986).

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

[13] Much of the dynamic-programming discrete-choice models in labor economics are nonstationary and finite-horizon as well (e.g., Wolpin 1984, 1987). Egesdal, Lai, and Su (2015) propose MPEC algorithm for the estimation of dynamic games. MPEC 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.

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

2 Model

This section describes our empirical model. Our goal is to incorporate a dynamic oligopoly game of mergers and innovation within a standard dynamic discrete-choice model.

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, $\omega_{it} \in \{\omega^1, \omega^2, ...\}$, which represents the level of tacit knowledge embodied by its team of R&D engineers and manufacturing engineers. Given the productivity profile, $\omega_t \equiv \{\omega_{it}\}_{i=1}^{n_t}$, which contains the information on n_t as well, 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 the time-varying demand situation. 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.[15] Upon entry, it becomes active at the lowest productivity level, $\omega_{i,t+1} = \omega^1$.[16] If it stays out, $\omega_{i,t+1} = \omega^0$. Each of the two actions entails a sunk cost, $\kappa^{a^0} + \varepsilon(a_{it}^0)$, where $a^0 \in A^0 = \{enter, out\}$, κ^{a^0} is deterministic, and $\varepsilon(a_{it}^0)$ is stochastic. An incumbent chooses between exit, innovation, merger, innovation-and-merger, and staying alone without taking any major action (which we call "idling"), when its turn arrives. Each of these dynamic actions, $a \in A = \left\{exit, innovate, \{propose\ merger\ to\ j\}_{j \neq i}, \{innovate\ \&\ propose\ j\}_{j \neq i}, idle\right\}$, entails a sunk cost, $\kappa^a + \varepsilon(a_{it})$, where κ^a is deterministic and $\varepsilon(a_{it})$ is stochastic. We assume $\varepsilon(a_{it}^0)$ and $\varepsilon(a_{it})$ are independently and identically distributed (i.i.d.) type-1 extreme value with CDF $\exp(-\exp(-\varepsilon/\sigma))$, where σ is the scale parameter.[17]

The three actions by incumbents induce the following transitions of ω_{it}. First, all exits

[15] 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.

[16] To be precise, our computational implementation uses $\tilde{\omega}_{i,t+1} = \tilde{\omega}^1$, where $\tilde{\omega}$'s reflect a redefined grid relative to the current frontier level, $\tilde{\omega}^L$ (see Appendix D.1 for details). This specification reflects the actual data pattern in which an actual entrant would start operations from the lowest level within the industry's current technological standards, which advance endogenously when the frontier firms innovate, and not the all-time lowest level in the absolute sense.

[17] Note we assume the same distribution of ε's across different actions, which restricts the ways in which their equilibrium choice probabilities respond to changes in payoffs.

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 productivity, $\omega_{i,t+1} = \omega_{it} + 1$. Third, an incumbent may propose merger to one of the other incumbents by making a take-it-or-leave-it ("TIOLI") offer.[18]

Horizontal mergers in the HDD context are not so much about the reallocation of tangible assets (e.g., physical production capacities), which are "perishable" and tend to become obsolete within a few quarters, as about (i) simply eliminating rivals to soften competition and/or (ii) combining teams of engineers who embody tacit knowledge.[19] 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 they 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.

We model the antitrust authority by making mergers infeasible when the number of firms, n_t, reaches a policy threshold, $\underline{N}$. Hence, the option to propose merger (and its associated cost) is relevant to firms' decision-making only when $n_t > \underline{N}$. We set $\underline{N} = 3$ in our baseline specification, because retrospective surveys by the Federal Trade Commission (2013) and Gilbert and Greene (2015) show this threshold was the de-facto rule of thumb for high-tech industries, and the HDD industry participants seemed to share this view. Section 5.1 provides further details and counterfactual policy simulations.[20]

[18] We also consider Nash bargaining with equal bargaining powers between the acquirer and the target ("NB") as an alternative bargaining protocol for sensitivity analysis.

[19] Industry experts explain two reasons for mergers. First, Reggie Murray's narrative (quoted in the introduction) epitomizes a dominant view in the HDD market that most mergers were to kill competitors. Second, 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). Many interviewees reiterated these views, which are not mutually exclusive. Appendix B.1 explains why (our interviewees believed) poaching top engineers from another firm was not sufficient or cost-effective for the second purpose.

[20] Appendix E.2 reports additional results based on alternative specifications with price-based merger policies.

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 (2017, 2018) 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, at the beginning of each period. We set $\rho = \frac{1}{n_{\max}} = \frac{1}{14}$ in our baseline model (where $n_{\max}$ is the maximum number of active players) to accommodate 13 incumbents in the data at the beginning of our sample period and a potential entrant.

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 components of sunk costs associated with the dynamic actions. If i is an incumbent, $\varepsilon(a_{it})$ includes ε_{it}^{x}, ε_{it}^{c}, ε_{it}^{i}, $\left\{\varepsilon_{ijt}^{m}\right\}_j$, and $\left\{\varepsilon_{ijt}^{i\&m}\right\}_j$, for exit, idling, innovation, merger proposal to rival incumbent j, and innovation-and-merger, respectively. These target-specific ε_{ijt}^{m}'s and $\varepsilon_{ijt}^{i\&m}$'s represent transient and idiosyncratic factors, are also sunk, and do not enter merger negotiation.[21]

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 costs, $\kappa^a + \varepsilon(a_{it})$. 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

[21] 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}.

from i to j upon agreement. Our baseline specification of the bargaining protocol is TIOLI.[22] 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, $\phi_t = \phi_0 + \phi_t(\omega_{it})$.

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. Empirical models of a dynamic game typically assume that only firm i observes the stochastic components of sunk costs, $\varepsilon(a_{it})$, because such private shocks are necessary to guarantee the existence of Markov perfect equilibria (c.f., Doraszelski and Satterthwaite 2010) in a simultaneous-move game with an infinite horizon. By contrast, we use a sequential-move formulation with a finite horizon and do not need to assume private information. Regardless of whether firm $j \neq i$ (non-mover) observes $\varepsilon(a_{it})$, there is nothing j can do about it, because i is the only mover at t. Moreover, these shocks are transient and sunk, and do not enter neither the joint surplus from i's merger with j nor their disagreement payoffs (see below). Thus we may assume $\varepsilon(a_{it})$ to be either public or private in our baseline specification with TIOLI offers.[23]

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 (not to be confused with the logit scaling parameter σ), 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} and consider σ_i as a collection of the ex-ante optimal choice probabilities conditional on $(\omega_{it}, \omega_{-it}, t)$.

[22] 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 simple as possible.

[23] By contrast, our alternative specification with Nash bargaining implicitly assumes $\varepsilon(a_{it})$ to be public information. Binmore, Rubinstein, and Wolinsky (1986) provide a non-cooperative foundation of Nash bargaining by showing that its solution coincides with Rubinstein's (1982) alternating bargaining protocol, which is a complete information game. We thank Allan Collard-Wexler and Aureo de Paula for this advice.

The following Bellman equations characterize an incumbent firm's dynamic optimization problem.[24] Mover i's value *after* drawing ε_{it} is

$$V_{it}(\omega_t, \varepsilon_{it}) = \pi_{it}(\omega_t) - \phi_t(\omega_{it}) + \max \left\{ \begin{array}{c} \bar{V}_{it}^{x}(\omega_t, \varepsilon_{it}^{x}), \bar{V}_{it}^{c}(\omega_t, \varepsilon_{it}^{c}), \bar{V}_{i}^{i}(\omega_t, \varepsilon_{it}^{i}), \\ \left\{\bar{V}_{ijt}^{m}\left(\omega_t, \varepsilon_{ijt}^{m}\right)\right\}_j, \left\{\bar{V}_{ijt}^{i\&m}\left(\omega_t, \varepsilon_{ijt}^{i\&m}\right)\right\}_j \end{array} \right\}, \quad (1)$$

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

$$\begin{aligned}
\bar{V}_{it}^{x}(\omega_t, \varepsilon_{it}^{x}) &= -\kappa^{x} + \varepsilon_{it}^{x} + \beta E\left[\Lambda_{i,t+1}(\omega_{t+1}) | \omega_t, a_{it} = exit\right], && (2)\\
\bar{V}_{it}^{c}(\omega_t, \varepsilon_{it}^{c}) &= \varepsilon_{it}^{c} + \beta E\left[\Lambda_{i,t+1}(\omega_{t+1}) | \omega_t, a_{it} = idle\right], && (3)\\
\bar{V}_{it}^{i}\left(\omega_t, \varepsilon_{it}^{i}\right) &= -\kappa^{i} + \varepsilon_{it}^{i} + \beta E\left[\Lambda_{i,t+1}(\omega_{t+1}) | \omega_t, a_{it} = innovate\right], && (4)\\
\bar{V}_{ijt}^{m}\left(\omega_t, \varepsilon_{ijt}^{m}\right) &= -\kappa^{m} + \varepsilon_{ijt}^{m} - p_{ij}(\omega_t) \\
&\quad + \beta E\left[\Lambda_{i,t+1}(\omega_{t+1}) | \omega_t, a_{it} = merge\ j\right], \text{ and} && (5)\\
\bar{V}_{ijt}^{i\&m}\left(\omega_t, \varepsilon_{ijt}^{i\&m}\right) &= -\kappa^{i} - \kappa^{m} + \varepsilon_{ijt}^{i\&m} - p_{ij}(\omega_t) \\
&\quad + \beta E\left[\Lambda_{i,t+1}(\omega_{t+1}) | \omega_t, a_{it} = innovate\ \&\ merge\ j\right]. && (6)
\end{aligned}$$

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

$$\begin{aligned}
EV_{it}(\omega_t) &= E_{\varepsilon}\left[V_{it}(\omega_t, \varepsilon_{it})\right] \\
&= \pi_i(\omega_t) - \phi_t(\omega_{it}) \\
&\quad + \sigma \left\{ \gamma + \ln \left[\begin{array}{c} \exp\left(\frac{\tilde{V}_{it}^{x}}{\sigma}\right) + \exp\left(\frac{\tilde{V}_{it}^{c}}{\sigma}\right) + \exp\left(\frac{\tilde{V}_{it}^{i}}{\sigma}\right) \\ + \sum_{j \neq i} \exp\left(\frac{\tilde{V}_{ijt}^{m}}{\sigma}\right) + \sum_{j \neq i} \exp\left(\frac{\tilde{V}_{ijt}^{i\&m}}{\sigma}\right) \end{array} \right] \right\}, && (7)
\end{aligned}$$

where γ is Euler's constant, σ is the logit scaling parameter, and $\tilde{V}_{it}^{a}$ is the deterministic part of $\bar{V}_{it}^{a}(\omega_t, \varepsilon_{it}^{a})$, that is, $\tilde{V}_{it}^{a} \equiv \bar{V}_{it}^{a}(\omega_t, \varepsilon_{it}^{a}) - \varepsilon_{it}^{a}$. In equations 2 through 6, $\Lambda_{i,t+1}$ represents i's expected value at $t+1$ *before* nature picks a mover at $t+1$,

$$\Lambda_{i,t+1}(\omega_{t+1}) = \rho EV_{i,t+1}(\omega_{t+1}) + \sum_{j \neq i} \rho W_{i,t+1}^{j}(\omega_{t+1}). \quad (8)$$

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

[24] Appendix B.2 features the corresponding expressions for the potential entrant.

an action) is

$$
\begin{aligned}
W_{it}^{j}(\omega_t) \;=\;& \pi_{it}(\omega_t) - \phi_t(\omega_{it}) + E_{it}\left[\Pr(a_{jt} = exit)\right] \times \beta E\left[\Lambda_{i,t+1}(\omega_{t+1}) \,|\omega_t, a_{jt} = exit\right] \\
& + E_{it}\left[\Pr(a_{jt} = idle)\right] \times \beta E\left[\Lambda_{i,t+1}(\omega_{t+1}) \,|\omega_t, a_{jt} = idle\right] \\
& + E_{it}\left[\Pr(a_{jt} = innovate)\right] \times \beta E\left[\Lambda_{i,t+1}(\omega_{t+1}) \,|\omega_t, a_{jt} = innovate\right] \\
& + \left\{E_{it}\left[\Pr(a_{jt} = merge\ i)\right] + E_{it}\left[\Pr(a_{jt} = innovate\ \&\ merge\ i)\right]\right\} \\
& \quad \times p_{ji}(\omega_t) \\
& + \sum_{k \neq i,j} E_{it}\left[\Pr(a_{jt} = merge\ k)\right] \\
& \quad \times \beta E\left[\Lambda_{i,t+1}(\omega_{t+1}) \,|\omega_t, a_{jt} = merge\ k\right] \\
& + \sum_{k \neq i,j} E_{it}\left[\Pr(a_{jt} = innovate\ \&\ merge\ k)\right] \\
& \quad \times \beta E\left[\Lambda_{i,t+1}(\omega_{t+1}) \,|\omega_t, a_{jt} = innovate\ \&\ merge\ k\right],
\end{aligned} \tag{9}
$$

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

$$
\Pr(a_{it} = action) = \frac{\exp\left(\frac{\tilde{V}_{it}^{action}}{\sigma}\right)}{\exp\left(\frac{\tilde{V}_{it}^{x}}{\sigma}\right) + \exp\left(\frac{\tilde{V}_{it}^{c}}{\sigma}\right) + \exp\left(\frac{\tilde{V}_{it}^{i}}{\sigma}\right) + \sum_{j \neq i} \exp\left(\frac{\tilde{V}_{ijt}^{m}}{\sigma}\right) + \sum_{j \neq i} \exp\left(\frac{\tilde{V}_{ijt}^{i\&m}}{\sigma}\right)}. \tag{10}
$$

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 TIOLI bargaining protocol implies the equilibrium acquisition price equals the target firm's outside option (i.e., staying independent),[25]

$$p_{ij}(\omega_t) = \beta\Lambda_{j,t+1}(\omega_{t+1} = \omega_t). \tag{11}$$

We solve this dynamic game for a unique sequential equilibrium in pure strategies that are type-symmetric.[26] 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 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.[27]

[25] Under NB, the two parties jointly maximize the following expression:

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

where $\zeta \in [0,1]$ represents the bargaining power of the acquirer (i here), which equals .5 with 50-50 split (1 under TIOLI). The last term in each bracket is the disagreement payoff. Note that the target-specific ε_{ijt}^{m} or $\varepsilon_{ijt}^{i\&m}$ represent transient/sunk factors that have led to the beginning of the negotiation and does not enter the above. 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). The specification shares the spirit of Crawford and Yurukoglu (2012) and Ho (2009), among others.

[26] By "type-symmetric" in this context, we mean the firms of the same type (productivity level) use the same mapping from the draws of the εs to the actions, and that we do not treat such firms differently based on their identities or other characteristics.

[27] See Appendix B.3 for further explanations on the key assumptions that guarantee uniqueness and whether our approach is akin to some form of equilibrium selection.

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.

Infinite Horizon 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 in dynamic games would often entail loss of point identification.

Continuous Time 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 unknown. Another problem with a shorter 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.

Recognition Probabilities Third, some firms might be more active in M&A than others, and recognition probabilities can accommodate such heterogeneity. For example, making ρ 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's, we prefer keeping it symmetric and instead focus on the extent of heterogeneity in the equilibrium CCP estimates.

Bargaining Powers Fourth, regarding the specifications TIOLI and NB, 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 TIOLI and NB as alternative models, and implement both as a sensitivity analysis.

Synergy Fifth, regarding the nature of synergies, we assume firms do not know their realizations before mergers, 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 be infeasible without strong additional assumptions. 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 could be regarded as a rich source of such 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.[28] 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 see no obvious alternatives to our specification amid many conceptual and practical challenges, and propose it as a baseline model.

3 Data

3.1 Institutional Background and Product Characteristics

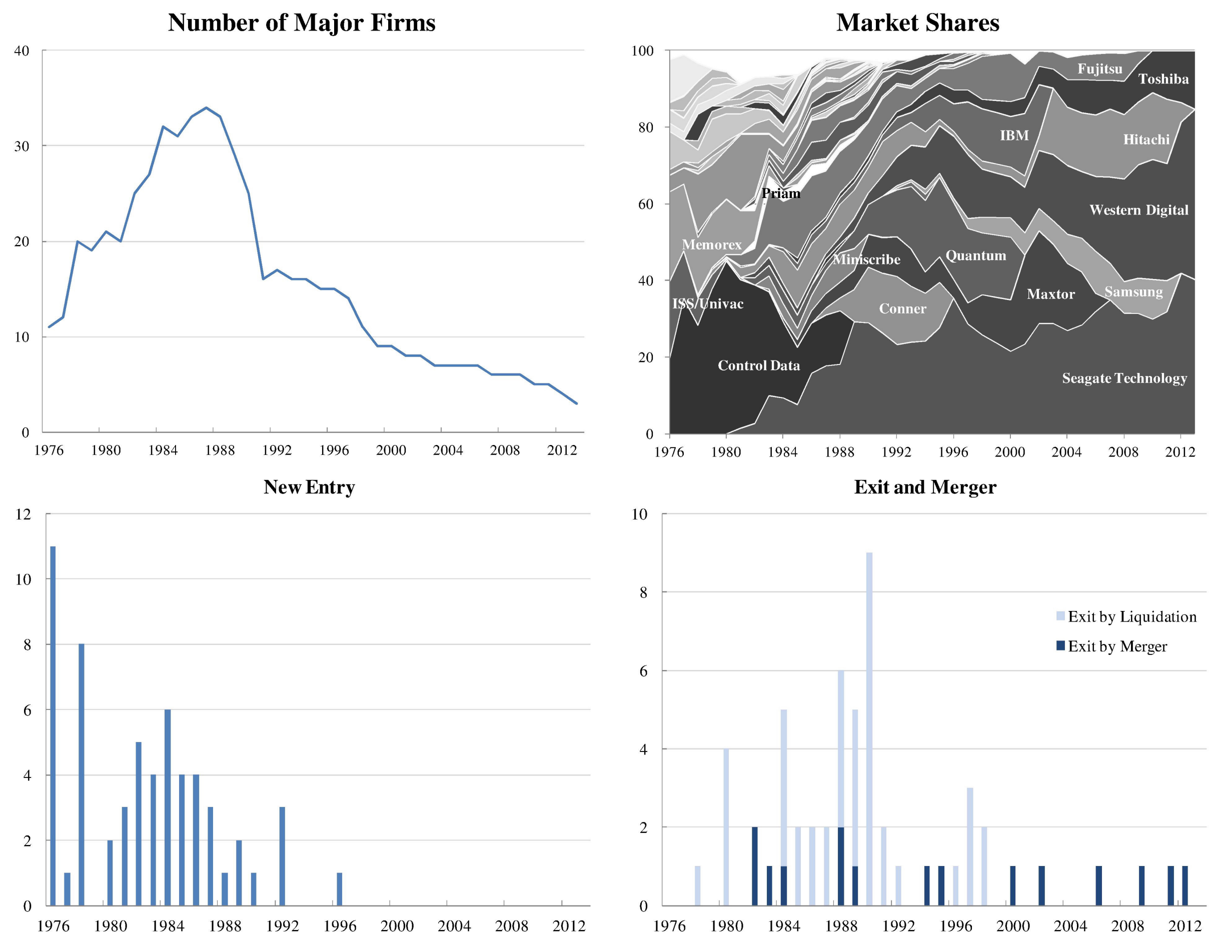
Computers are archetypical high-tech goods that store, process, and transmit data. HDDs, semiconductor chips, and network equipment perform these tasks, respectively. HDDs offer the most relevant empirical context to study mergers and innovation in the process of industry consolidation. The industry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

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 of increasing areal density that constantly improves both quality and efficiency (Kryder's Law).

Despite such complexity, HDDs are also one of the simplest products in terms of economics 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.[29] Consumers typically do not observe or distinguish "brands." Moreover, HDDs are physically durable but do not drive the repurchasing cycle of PCs. Microsoft

[28] From authors' personal interview on April 20, 2015, in Corona del Mar, CA.
[29] From authors' personal interview on June 30, 2015, in Cupertino, CA. See also section 4.1.

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 (2017, 2018) on product and process innovations during the 1980s and 1990s.

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, 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.[30] 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, 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.

[30] PC makers typically do not stockpile HDDs either, because HDDs become cheaper and better over time.

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 average CR4 is 91.6% during the same period.[31]

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.[32]

3.2 Three Data Elements

Our empirical analysis will focus on the period between 1996 and 2016 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 2017), and manufacturing operations in Southeast Asia had achieved the most competitive cost-quality balance (see Igami 2018). Third, our main data source, *TRENDFOCUS*, an industry publication series, started most

[31] 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.

[32] 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.

Table 1: Summary Statistics

Variable	Unit of measurement	Number of observations	Mean	Standard deviation	Minimum	Maximum
Panel A						
HDD shipments, Q_t	Exabytes*	83	16.91	17.34	0.02	53.20
HDD price, P_t	$/Gigabytes*	83	14.09	36.33	0.03	178.62
Disk price, Z_t	$/Gigabytes*	83	1.83	5.11	0.002	23.51
PC shipments, X_t	Million units	83	29.12	7.01	14.47	40.31
Panel B						
Market share, ms_{it}	%	605	13.68	11.43	0.00	45.75
Panel C						
Indicator$\{a_{it} = merge\}$	0 or 1	248	0.0242	0.1540	0	1
Indicator$\{a_{it} = innovate\}$	0 or 1	248	0.2460	0.4315	0	1
Indicator$\{a_{it} = exit\}$	0 or 1	248	0.0202	0.1408	0	1
Indicator$\{a_{it} = enter\}$	0 or 1	248	0.0040	0.0635	0	1
Variable profit, π_{it}	Million $	(see note)	40.80	119.67	0.00	12,040.82

Note: 1 exabytes (EB) = 1 billion gigabytes (GB), and 1 GB = 1 billion bytes. Panel A is recorded in quarterly frequency at the aggregate level, Panel B is quarterly at the firm level (unbalanced panel), and Panel C is a single time series from January 1996 to August 2016 (which summarizes the observable actions of all firms according to the timing convention of our model). $ms_{it} = 0.00$ is recorded for negligible output levels (e.g., the initial periods of a new entrant and the final periods of exiting incumbents). π_{it} is our period-profit estimate and contains 25,285,120 values across 128 productivity levels, 83 quarters, and 76,160 industry states. See sections 4.1 and 4.2.

Source: *TRENDFOCUS* Reports (1996–2016).

of its systematic data collection at the quarterly frequency in 1996.[33]

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,[34] 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. We observe entry, exit, and mergers in the *TRENDFOCUS* reports.[35] Panel C also includes some elements that are derived from the other two panels, such as the indicator of innovation and the equilibrium variable profits.[36]

[33] By contrast, Igami (2017, 2018) 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.

[34] Appendix C.1 features more details on Panel A, including visual plots of these variables.

[35] The antitrust authority has approved all HDD mergers during the sample period. We do *not* observe merger proposals that were rejected in private negotiations. We use a model in which all proposals are accepted in equilibrium.

[36] Appendix D.1 explains the details of this data construction.

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 contain 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.

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.[37] 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,

$$\log Q_t = \alpha_0 + \alpha_1 \log P_t + \alpha_2 \log X_t + \varepsilon_t, \tag{12}$$

[37] From authors' personal interview on June 30, 2015, in Cupertino, CA.

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 i.i.d. 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. We also use as another IV a dummy variable indicating a major supply disruption caused by flood in Thailand in the fourth quarter of 2011.

In Table 3, columns 1 and 2 show OLS estimates, whereas columns 3 and 4 show IV estimates. The estimates for price elasticity, α_1, are similar across specifications and suggest the demand is close to unit-elastic. We use the IV estimates of the full model (4) in the subsequent analysis. Because the data on quantities and prices clearly indicate serial correlations and time trends (see Appendix C.1), we use detrended time series of these variables.

Table 3: Demand Estimates

	(1) OLS	(2) OLS	(3) IV	(4) IV
Log HDD price per GB (α_1)	−1.112	−1.046	−1.054	−1.043
	(0.035)	(0.046)	(0.032)	(0.038)
Log PC shipment (α_2)	–	0.271	–	0.276
	(–)	(0.095)	(–)	(0.086)
Number of observations	83	83	83	83
Adjusted R^2	0.942	0.948	–	–
First-stage regression				
Log disk price per GB	–	–	0.813	0.567
	(–)	(–)	(0.026)	(0.032)
Thai flood dummy	–	–	0.263	0.548
	(–)	(–)	(0.079)	(0.070)
F-value	–	–	585.49	732.12
Adjusted R^2	–	–	0.874	0.946

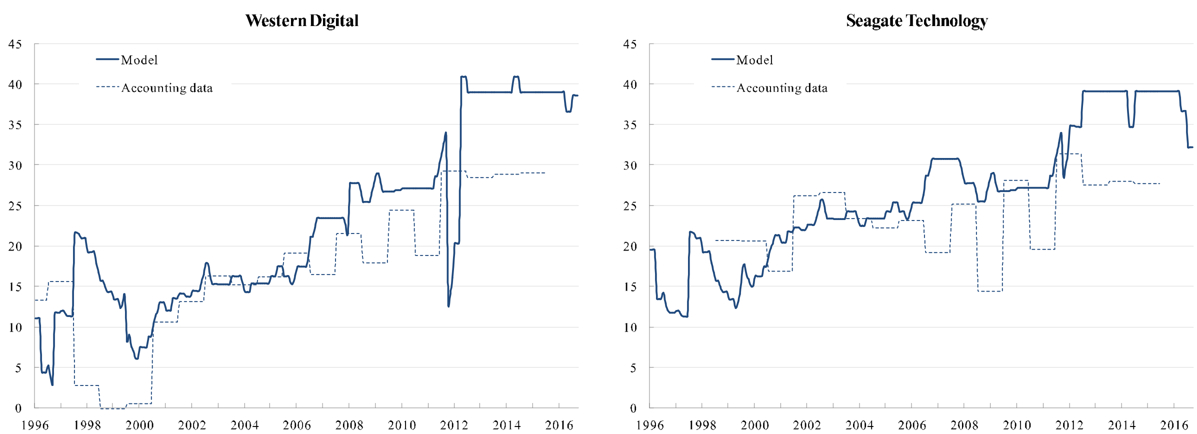
Note: Dependent variable is log total HDD (in EB) shipped. We use detrended quantities and prices of HDD to address nonstationarity in the original time series of these variables. Huber-White heteroskedasticity-robust standard errors are in parentheses.

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 time trend and by analyzing sales at the byte level. 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 2), 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), which brings conceptual clarity and preserves simple economic intuition.

Figure 2: 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 0.75 for Western Digital, and 0.51 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,

$$\pi_{it} = (P_t - mc_{it})\, q_{it}, \tag{13}$$

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}.[38] Firm i's first-order condition is

$$P_t + \frac{dP}{dQ} q_{it} = mc_{it}, \tag{14}$$

which provides one-to-one mapping between q_{it} (observed) and mc_{it} (implied) given P_t in the data and dP/dQ 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).[39]

Figure 2 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,

$$m_{it}(\omega_t) = \frac{P_t(\omega_t) - mc_{it}}{P_t(\omega_t)}, \tag{15}$$

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.

[38] 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 geographical market is global. Thus, one cannot rely on either inter-temporal or cross-sectional variation in data to identify marginal-cost curves nonparametrically.

[39] 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, 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.

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 4.6 and 3.5 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.[40] Thus, correlation is more important than levels, which is 0.75 for Western Digital, and 0.51 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.

We convert these marginal-cost estimates into productivity levels, ω_{it}, for the subsequent dynamic analysis. First, we discretize marginal costs on a 0.1 log-US$ grid. Second, we reverse their rank order, so that higher productivity levels represent lower mc_{it}'s. Third, we keep track of each firm's ω_{it} by looking at its "frontier" (i.e., the highest ω_{it} reached in the industry to date) and how many bins below it a firm is. Appendix D.1 explains further details.

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.

Table 4 lists all the parameters and key specifications of our model. Before engaging in the MLE of the core parameters, $\theta \equiv (\phi_0, \kappa^i, \kappa^m, \kappa^e, \sigma)$, we determine the values of the other

[40] 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.

parameters either as by-products of the previous two steps or directly from auxiliary data.[41]

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 (section 4.3)
Logit scaling parameter	σ	MLE (section 4.3)
Base fixed cost of operation	ϕ_0	MLE (section 4.3)
Time-varying fixed cost of operation	$\phi_t(\omega_{it})$	Accounting data (see Appendix D.2)
Liquidation value	$\kappa^x = 0$	Industry background
3. Dynamics (transitions)		
Annual discount factor	$\beta = 0.9$	Calibrated
Prob. stochastic depreciation	$\delta = 0.04$	Implied by mc_{it}
Average synergy	$\lambda = 1$	Implied by mc_{it} (sensitivity analysis with 0 and 2)
4. Other key specifications		
Terminal period	$T = Dec$-2025	Sensitivity analysis with Dec-2020
Bargaining power	TIOLI: $\zeta = 1$	Sensitivity analysis with NB: $\zeta \in \{0.5, 0.6, 0.7, 0.8, 0.9\}$
Recognition probability	$\rho = \frac{1}{n_{\max}} = \frac{1}{14}$	Sensitivity analysis with $n_{\max} \in \{21, 28\}$

First, we pin down the other two costs as follows. The time-varying (and productivity-dependent) component of the fixed cost of operation, $\phi_t(\omega_{it})$, 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.[42] We set liquidation value, κ^x, to zero because tangible assets quickly become obsolete and have no productive use outside the HDD industry.

Second, three parameters govern transitions. The discount factor is calibrated to $\beta = 0.9$ at an annualized rate. 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 $\delta = 0.04$.[43] Likewise, our mc_{it} estimates suggest the extent of synergy. The average post-merger improvement is a 10% decrease in marginal

[41] Recall we use $\pi_{it}(\omega_t)$, dollar-valued period profits, as data. They help us identify the scale parameter for the ε's, σ. Larger σ would make the model less responsive to these profits, whereas smaller σ would make the predicted CCPs highly sensitive to their changes. The numerical search for the likelihood-maximizing σ tends to be volatile. Hence, we use a grid search with an interval of 0.01, conditional on which we use MATLAB's simplex- and derivative-based search algorithm for the other parameters.

[42] See Appendix D.2 for details.

[43] The "occasional deterioration" in firm productivity is the opposite of the *innovate* action and is recognized in the data as upward changes in the marginal-cost estimates. Such changes occur in 4% of the observations or "time at risk." Because these changes are not desirable for the firms, we model them as exogenous negative shocks to their productivity (i.e., stochastic depreciation that is not controlled by the firms) rather than their active choice.

cost (or a one-level increase in the discretized productivity grid),[44] which constitutes our "estimates" of the Poisson synergy parameter,

$$\hat{\lambda}_{MLE} = \frac{1}{\#_m} \sum_{m=1}^{\#_m} \Delta_m, \tag{16}$$

where $\#_m$ is the number of mergers in the data, and Δ_m is the productivity improvement from merger m.[45] However, mergers in a concentrated industry are rare events ($\#_m = 6$ in our main sample), and antitrust agencies tend to hear merging parties' claim about synergy with skepticism. 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 2016Q3, but the HDD industry does not; hence, we need to assume something about the post-sample end game. Our baseline specification assumes the HDD demand continues to exist until the end of year 2025, with linear interpolation of profit-function estimates between September 2016 and December 2025. Our sensitivity analyses employ a more pessimistic scenario, with $T = Dec$-2020. The second aspect is bargaining protocols. Our baseline specification is TIOLI, $\zeta = 1$, but we also estimate the NB version with $\zeta \in \{0.5, 0.6, 0.7, 0.8, 0.9\}$.

Incorporating a Random-Mover Dynamic Game

Having determined the baseline configuration, we proceed to estimate $\theta = (\phi_0, \kappa^i, \kappa^m, \kappa^e, \sigma)$. The outline of our MLE procedure follows Rust's (1987) NFXP approach, but our model diverges from his 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 point-identification difficult 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.,

[44] See Appendix D.1 for the details of discretization.

[45] The variance of Δ_m in the data is 1 as well. Hence, the Poisson distribution fits our (limited) data well.

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.

The optimal choice probabilities of entry, exit, innovation, and mergers in equation 10 constitute the likelihood function. The contribution of action profile $a_t \equiv (a_{it})_i$ in month t is

$$l_t\left(a_t|\omega_t;\theta\right) = \sum_i \bar{\rho}_{it} \prod_{action \in A_{it}(\omega_t)} \Pr\left(a_{it} = action\right)^{1\{a_{it}=action\}}, \tag{17}$$

where $\bar{\rho}_{it}$ is the realization of turn-to-move in data (see equation 19 below), $A_{it}\left(\omega_t\right)$ is i's choice set in state ω_t, and $1\left\{\cdot\right\}$ is an indicator function. The MLE is

$$\hat{\theta}_{MLE} = \arg\max_{\theta} \frac{1}{T} \sum_t \ln\left[l_t\left(a_t|\omega_t;\theta\right)\right], \tag{18}$$

where T is the number of sample periods.

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 observable action (such as exit, innovation, 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

$$\bar{\rho}_{it} = \begin{cases} 1 & \text{if } a_{it} \in \left\{exit, merge/innovate, enter\right\}, \text{ and} \\ \frac{1}{n_{\max}} & \text{if } a_{it} \in \left\{idle, out\right\} \text{ for all } i \text{ at } t. \end{cases} \tag{19}$$

That is, when exit, merger, innovation, or entry is recorded in the data, we interpret that nature picked the firm that took the action (and set $\bar{\rho}_{jt} = 0$ for all non-mover firms $j \neq i$), 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.[46]

[46] To be precise, when we say we set $\bar{\rho}_{it} = 1$ for some firm i at some t, we are merely registering particular realizations of the random variable (i.e., nature's choice) in the data. This empirical $\bar{\rho}_{it}$ should be distinguished from the generic (ex-ante) recognition probability, $\rho = \frac{1}{n_{\max}} < 1$, because firms do not know in advance who will be chosen. We fix $n_{\max} = 14$, the highest number of firms in the data ($= 13$) plus a potential entrant. Note we use monthly frequency in our empirical model to avoid multiple movers in a period.

Results

In Table 5, column 1 shows our baseline estimates with (i) TIOLI, $\zeta = 1$, (ii) mean synergy from the data, $\lambda = 1$, and (iii) the terminal period, $T = Dec$-2025. As a sensitivity analysis, column 2 alters ζ, columns 3 and 4 alter λ, and column 5 alters T. All the specifications lead to similar estimates that are mostly within the 95% confidence interval of each other. Note we distinguish the innovation costs at frontier firms ($\omega_{it} = 4$) and other firms ($\omega_{it} = 1, 2, 3$), because advancing the industry's technological frontier is fundamentally more difficult and observed less frequently in the data.

Table 5: MLE of Dynamic Parameters and Sensitivity Analysis

Specification	(1)	(2)	(3)	(4)	(5)
Bargaining (ζ):	1 (TIOLI)	0.5 (NB)	1	1	1
Synergy (λ):	1	1	0	2	1
Terminal period (T):	2025	2025	2025	2025	2020
ϕ_0	0.011	0.011	0.012	0.011	0.011
	$[0.001, 0.020]$	$[0.000, 0.021]$	$[0.001, 0.022]$	$[0.001, 0.019]$	$[0.001, 0.020]$
κ^i $(\omega_{it} = 1, 2, 3)$	0.48	0.51	0.52	0.47	0.48
	$[0.26, 0.69]$	$[0.28, 0.75]$	$[0.27, 0.77]$	$[0.26, 0.68]$	$[0.26, 0.70]$
κ^i $(\omega_{it} = 4)$	0.85	0.91	0.97	0.84	0.85
	$[0.39, 1.42]$	$[0.42, 1.54]$	$[0.45, 1.63]$	$[0.26, 0.68]$	$[0.39, 1.43]$
κ^m	1.27	1.21	1.34	1.31	1.27
	$[0.81, 1.86]$	$[0.72, 1.84]$	$[0.81, 2.00]$	$[0.86, 1.88]$	$[0.81, 1.85]$
κ^e	0.17	0.16	0.15	0.18	0.17
	$[-]$	$[-]$	$[-]$	$[-]$	$[-]$
σ	0.55	0.60	0.63	0.54	0.55
	$[0.41, 0.80]$	$[0.45, 0.87]$	$[0.47, 0.91]$	$[0.40, 0.78]$	$[0.41, 0.80]$
Log likelihood	-156.93	-157.23	-157.56	-156.60	-156.96

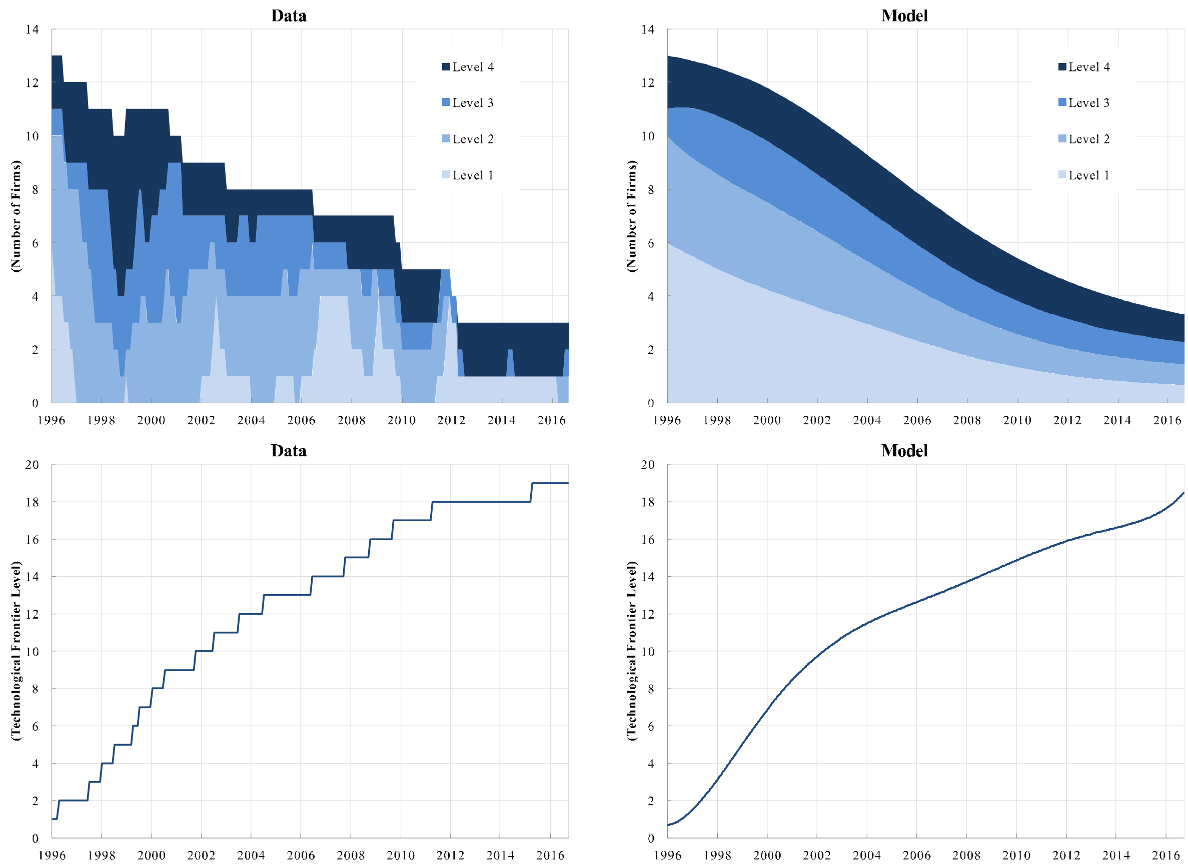
Note: The 95% confidence intervals are constructed from the likelihood-ratio tests. See Table 14 in Appendix D.5 for additional sensitivity analysis.

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.[47] The comparison reveals that at least three out of the six historical transaction values closely match the target firms' predicted values. See Appendix D.4 for further details.

Another way of assessing fit is to compare the actual and predicted trajectories of market structure and technological frontier (Figure 3). Its top panels show 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

[47] 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.

Figure 3: Fit of the Estimated Model



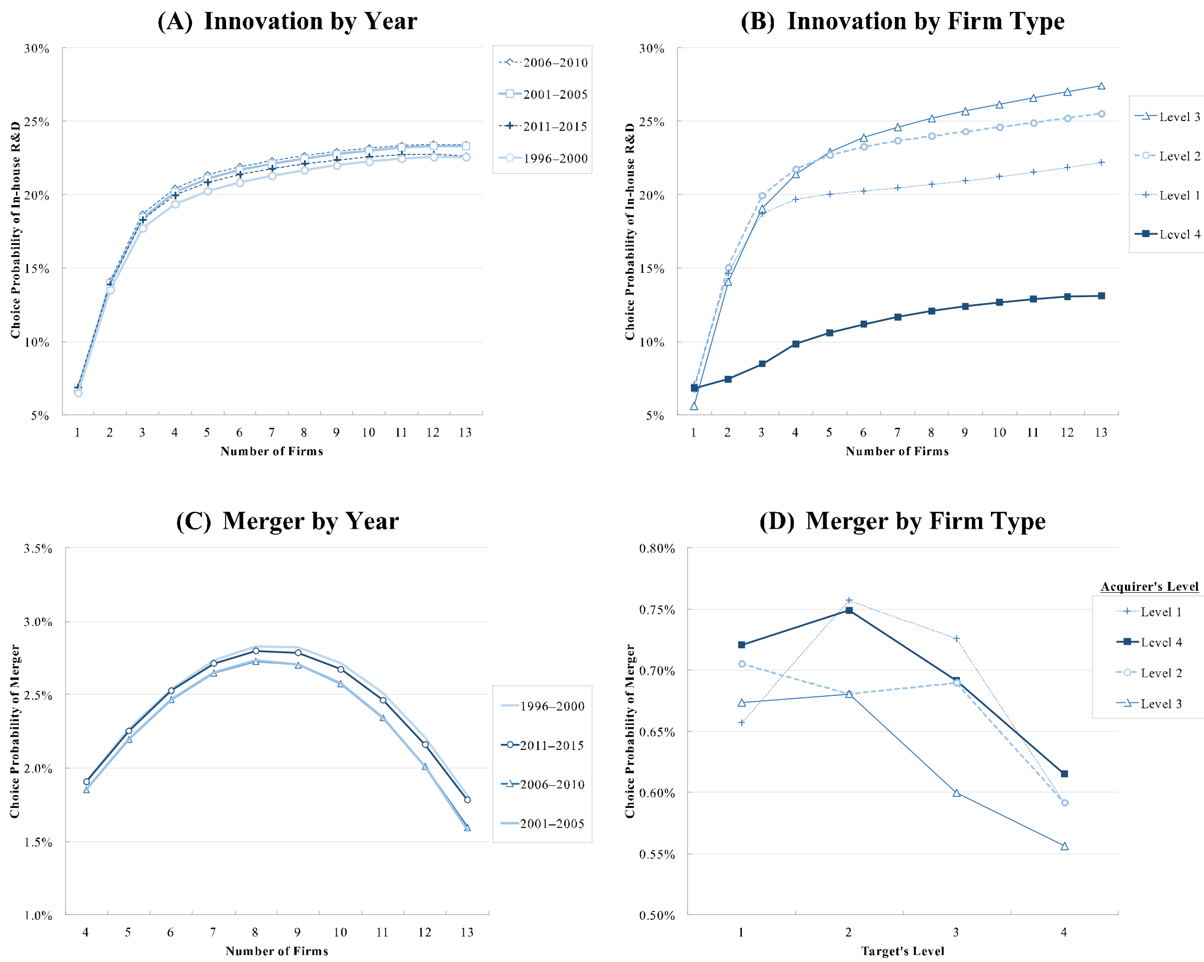
Note: The model outcome is the average of 10,000 simulations based on the estimated model. See Appendix D.1 for the details of discretized productivity levels.

some aspects of their productivity composition (e.g., the survival of a few low-level firms). The bottom panels show the model's average path of the frontier slightly undershoots the data path between 2005 and 2015, but their eventual levels seem reasonably close to each other. Hence, we believe the estimated model provides a reasonable benchmark with which we can compare welfare performances of hypothetical antitrust policies in section 5.

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 structural relationships between competition, innovation, and merger. Figure 4 shows the equilibrium R&D and M&A strategies by year, type, and market structure.

Figure 4: Heterogeneous Plateaus of Equilibrium Strategies



Note: Each graph summarizes the equilibrium strategies for R&D and M&A, by averaging the structural CCP estimates across ω_{it}, n_t, or t. For expositional purposes, the horizontal axis represents the concurrent number of active firms (n_t) as a summary statistic of the underlying state (ω_t, which subsumes both n_t and the productivity profile of all firms) in Panels A, B, and C. In Panel C, the horizontal axis is truncated at 3, because the antitrust authorities do not allow mergers below this point, and our model incorporates this actual policy regime (see section 5 for further details).

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, Reinganum 1983). After four or five firms, however, the slopes become flat. These plateaus exhibit heterogeneity both across time (panel A) and productivity (panel B). Innovation rates are high and increasing with n_t in the peak years of HDD demand (i.e., 2006–2010) and at relatively more productive firms (i.e., levels 2 and 3), because continuation values (and hence the incremental value of investment) are high. By contrast, the incentives are low and often decreases with n_t in later years (i.e., 2011–2015) and at low-productivity firms (i.e., level 1), because the possibility of exit becomes more realistic in such cases. Note the frontier (level 4) firms face a fundamentally more challenging task of advancing the frontier technology; hence their seemingly low CCPs of innovation is not necessarily a sign of reluctance.

These "heterogeneous plateaus" are our structural-empirical finding about the competition-innovation relationship, which have often been theorized or described as an "inverted-U" curve (e.g., Scherer 1965, Aghion et al. 2005). These relationships between competition and innovation are neither accidental findings from particular simulation draws nor mechanical reflections of our modeling choices. In their computational theory paper, Marshall and Parra (2018) show (i) the plateau shapes could arise under fairly general and standard model settings, but (ii) different parameter values could generate either increasing, decreasing, inverse U-shaped, or plateau-shaped patterns.

The incentive for merger is equally intriguing. Panel C plots the inverted-U shaped optimal M&A strategy as a function of time and competition. Mergers are not particularly attractive when a dozen competitors exist, some of which are likely to exit anyway. Incurring the sunk cost of negotiation is not worthwhile when weaker rivals are expected to disappear soon. By contrast, the CCP of merger is the highest when $n_t = 6 \sim 10$. This is the phase of industry consolidation in which many potential merger targets still exist and killing rivals become more profitable (i.e., the incremental profit from reducing n_t increases as n_t decreases). Finally, the CCPs of merger seem to decrease again when market structure becomes more concentrated, but this decline mostly reflects the reduced opportunities (i.e., number of merger targets) and do not necessarily indicate reduced incentives to merge. Once we divide the CCPs by n_t, the merger-competition relationships (per active target) exhibit downward-sloping curves in this region ($n_t = 4 \sim 8$) as well.

Who merges with whom? Panel D plots the CCP of merger (sliced by the *acquiring* firm's level) against the *target* firm's level. In general, all combinations are possible, as is the case in our data.[48] Three underlying forces shape the nonmonotonic patterns. First, firms generally want to merge with higher types than themselves because rationalization

[48] The six mergers between 1996 and 2016 involve the following acquirer-target pairs (with their estimated productivity levels in parentheses): Maxtor (3)-Quantum (3), Hitachi (1)-IBM (2), Seagate (4)-Maxtor (1), Toshiba (1)-Fujitsu (1), Seagate (4)-Samsung (1), and Western Digital (1)-Hitachi (1).

(i.e., the deterministic part of productivity improvement after merger) guarantees that the merged entities are at least as productive as the higher of the merging firms' types. This factor explains, for example, why level-1 firms prefer level-2 or level-3 targets to level-1 targets, as well as why level-2 firms prefer merging with level-3 firms. The second factor is synergy (i.e., the stochastic part of productivity improvement after merger), which we model as purely random because we have neither generally accepted theory nor data on this phenomenon. Level 4 firms may not benefit from rationalization but can expect synergy to help them push the technological frontier. The third factor is the acquisition price, which reflects the continuation value of the target firm and hence is increasing in its productivity level. The balance between the latter two forces explains why some firms prefer lower-level targets to higher-level targets. By contrast, level-4 firms are too expensive to acquire despite (and because of) their high return. Our results incorporate all of these economic forces in equilibrium, and broadly agree with the data pattern in which lower-level firms are acquired relatively more frequently.

5 Optimal Policy and Dynamic Welfare Tradeoff

5.1 Welfare under Counterfactual Policy Regimes

How far should an industry be allowed to consolidate? We are now ready to simulate welfare outcomes under hypothetical merger policies. Table 6 compares welfare outcomes under different policy regimes in which antitrust authorities block mergers if n_t is at or below certain thresholds. Each column reports the discounted sums of consumer surplus (CS), producer surplus (PS), and social welfare (SW) under a hypothetical regime with $\underline{N} \in \{1, 2, ..., 6\}$ and their percentage changes from 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.[49] Thus, $\underline{N} = 3$ is a reasonable 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 Sili-

[49] 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 price changes or HHI instead of $\underline{N}$, but we found no clear price/HHI threshold in the report.

Table 6: Welfare Performance of Counterfactual Policies

Threshold number of firms ($\underline{N}$)	1	2	3 (Baseline)	4	5	6
(A) All periods: 1996–2025						
Consumer surplus	721.93	762.32	765.05	766.01	766.64	767.20
	(−5.64%)	(−0.36%)	(±0%)	(+0.13%)	(+0.21%)	(+0.28%)
Producer surplus	21.94	18.88	18.29	18.01	17.80	17.61
	(+19.94%)	(+3.17%)	(±0%)	(−1.58%)	(−2.71%)	(−3.73%)
Social Welfare	743.88	781.20	783.35	784.02	784.44	784.81
	(−5.04%)	(−0.27%)	(±0%)	(+0.09%)	(+0.14%)	(+0.19%)
(B) First half: 1996–2010						
Consumer surplus	530.75	535.50	535.70	535.77	535.88	535.96
	(−0.92%)	(−0.04%)	(±0%)	(+0.01%)	(+0.03%)	(+0.05%)
Producer surplus	13.64	13.23	13.15	13.09	13.03	12.96
	(+3.73%)	(+0.60%)	(±0%)	(−0.47%)	(−0.95%)	(−1.45%)
Social welfare	544.39	548.73	548.85	548.86	548.90	548.92
	(−0.81%)	(−0.02%)	(±0%)	(+0.00%)	(+0.01%)	(+0.01%)
(C) Second half: 2011–2025						
Consumer surplus	191.18	226.82	229.35	230.24	230.77	231.24
	(−16.64%)	(−1.10%)	(±0%)	(+0.39%)	(+0.62%)	(+0.82%)
Producer surplus	8.30	5.65	5.14	4.92	4.77	4.65
	(+61.38%)	(+9.75%)	(±0%)	(−4.40%)	(−7.22%)	(−9.54%)
Social welfare	199.48	232.47	234.49	235.16	235.54	235.89
	(−14.93%)	(−0.86%)	(±0%)	(+0.28%)	(+0.45%)	(+0.60%)

Note: All welfare numbers are present values as of January 1996 in billion US dollars and are the averages of 10,000 simulations under each policy regime. Their percentage changes from the baseline outcomes under $\underline{N} = 3$ are in parentheses. Note the welfare numbers in Panel (C) are considerably smaller than those in Panels (A) and (B), because the annual discount factor of $\beta = 0.9$ means they are discounted at $\beta^{15} = 0.2059$. See Appendix E.1 for the same analysis under the 50-50 Nash bargaining setup.

con 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.

Computational implementation is straightforward. We estimated the baseline model by searching over the parameter space of θ 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 $n_t \leqslant 3$.[50] Simulating welfare outcomes under an alternative regime is simpler than estimation. First, solve the counterfactual game with the same parameter estimates $\hat{\theta}$ but in a different policy environment ($\underline{N} \neq 3$) just once, and obtain the optimal choice probabilities in the coun-

[50] The latter, extremely high "cost" of mergers is merely a computational implementation of "making mergers infeasible" when the antitrust regulation is binding, and should not be interpreted as part of the economic sunk cost represented by κ^m. In other words, κ^m is the cost of decision-making and entering negotiations *conditional on mergers being feasible* (and therefore relevant from the firms' perspectives).

terfactual 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 its time profile in terms of time-0 discounted present value as of January 1996.

Table 7: Competition and Innovation Outcomes of Counterfactual Policies

Threshold number of firms ($\underline{N}$)	1	2	3 (Baseline)	4	5	6
(A) Number of active firms						
1996–2025 average	5.80	6.12	6.24	6.32	6.39	6.46
As of December 2010	4.70	4.91	4.98	5.09	5.21	5.36
As of December 2025	0.80	1.12	1.23	1.28	1.31	1.33
(B) Frontier technology level						
1996–2025 average	13.62	13.71	13.73	13.74	13.74	13.75
As of December 2010	14.86	14.87	14.87	14.87	14.87	14.86
As of December 2025	18.36	18.69	18.76	18.79	18.79	18.81
(C) Number of mergers						
1996–2025 total	6.08	4.87	4.15	3.60	3.12	2.66
Of which 1996–2010	3.62	3.33	3.21	3.04	2.79	2.48
Of which 2011–2025	2.47	1.54	0.94	0.56	0.33	0.18
(D) Number of innovations						
1996–2025 total	45.45	47.84	48.79	49.41	49.94	50.48
Of which 1996–2010	37.39	37.44	37.49	37.54	37.63	37.75
Of which 2011–2025	8.06	10.40	11.30	11.87	12.31	12.73
(E) Number of entries						
1996–2025 total	0.0999	0.0547	0.0328	0.0236	0.0224	0.0217
Of which 1996–2010	0.0028	0.0022	0.0014	0.0010	0.0010	0.0002
Of which 2011–2025	0.0971	0.0525	0.0314	0.0226	0.0214	0.0215
(F) Number of exits						
1996–2025 total	6.22	7.06	7.65	8.14	8.60	9.03
Of which 1996–2010	4.73	4.80	4.84	4.91	5.03	5.19
Of which 2011–2025	1.49	2.26	2.81	3.24	3.57	3.84

Note: All numbers are the averages of 10,000 simulations under each policy regime.

The first column shows the most permissive policy ($\underline{N} = 1$) is unambiguously bad for both CS and SW, reducing them by more than 5% relative to the $\underline{N} = 3$ baseline. The magnitude of these changes might appear small at first glance, but the all-period sum in Panel A masks the policy's true impact. Panel C shows CS and SW in the second half (2011–2025) decrease by almost 15%, because this is the time period in which the final mergers take place (or are blocked). Likewise, another permissive policy ($\underline{N} = 2$) leads to net welfare losses. Thus, allowing mergers to monopoly or duopoly seems a bad idea, even if we account for potentially positive "side effects" on ex-ante incentives to enter and innovate.

By contrast, stricter policies ($\underline{N} = 4, 5, 6$) lead to better outcomes than in the baseline case. However, their rates of improvement for both CS and SW are limited to less than

1% even in the second half (Panel C). The same analysis under the Nash bargaining setup shows similar patterns, with $\underline{N} = 5$ achieving the highest discounted SW (see Table 15 in Appendix E.1).

This result is surprising, because our policy-function estimates in section 4 (Figure 4) suggested positive relationships between the number of firms and the optimal choice probability of innovation (in-house R&D).[51] One would expect the combination of "more competition" and "more innovation" would lead to substantial improvements in CS and SW.

We investigate the mechanism behind these curious findings in Table 7, which summarizes the main drivers of welfare: the number of firms (Panel A), the frontier technology level (Panel B), and the aggregate count of firms' main actions (Panels C, D, E, and F). These key objects relate to each other as follows.

- Market structure (including, but not limited to, n_t) and the level of technological frontier (as well as each firm's distance from it) collectively determine the firms' markups and costs, respectively, and hence the market price of HDDs as well.[52]
- The state of competition and innovation in each period is the consequence of the firms' mergers, in-house R&D, and entry-exit decisions in the past. More specifically, mergers directly change both the number of firms and their productivity profile; in-house R&D changes only the latter; and entry-exit changes only the former.

Panel A shows permissive policies ($\underline{N} = 1, 2$) reduce the number of firms substantially, whereas stricter policies ($\underline{N} = 4, 5, 6$) do not necessarily lead to comparable changes in the opposite direction. The reason becomes clear in Panels C, E, and F, which suggest stricter policies reduce mergers but increase exits. The pro-competitive effect of blocking mergers is mostly offset by this increase in exits. The change in new entry hardly matters, because it is a rare event in this mature industry with high technological barriers, but its direction of change suggests entry becomes more profitable under permissive policies.[53] The mechanism underlying this increase in (net) exits is the value-destruction side effects of stricter policies. Table 6 shows stricter policies reduce PS (i.e., the industry-wide profits), thereby making firms more likely to exit.

[51] Strictly speaking, Figure 4 showed the equilibrium strategies under $\underline{N} = 3$, which are not policy-invariant, but their counterparts under $\underline{N} \neq 3$ show the same qualitative pattern.

[52] Note the HDD price (P_t) summarizes the welfare-relevant outcomes of each state (s_t), because we analyze HDDs at the level of efficiency unit (i.e., GB of information storage capacity).

[53] Entry is more profitable under permissive policies, because market structure tends to be more concentrated (see Panel A) and the entrants may expect to be acquired by the incumbents. The latter mechanism resonates with MNSW's (2018) theoretical results, which emphasize the role of entrants' incentive to be acquired.

Panel B of Table 7 shows permissive policies ($\underline{N} = 1, 2$) do not help advance the technological frontier despite frequent claims by casual proponents of the "Schumpeterian" hypotheses that monopoly is conducive to innovation. Note our model allows synergies from mergers and fully internalizes the positive effects of future monopoly profits on the ex-ante incentive to innovate. However, synergies and the positive incentive effects of higher expected profits are dominated by the countervailing negative effect of reduced competitive pressure. Recall from Figure 4 that the optimal choice probabilities of in-house R&D decrease precipitously as n_t falls below three. Arrow's (1962) replacement effect dominates, and preemptive motives of Gilbert and Newbery (1982) and Reinganum (1983) evaporate under monopoly when the prospect of new entry is remote (as in our industry context). Thus, even if policy-makers exclusively focus on "promoting innovation," allowing mergers to duopoly or monopoly turns out to be counter-productive.

At the same time, Panel B shows stricter policies are not particularly helpful in promoting innovation either. The slightly higher count of in-house R&D (in Panel D) does not translate into corresponding advances of the frontier, because only R&D by frontier firms ($\omega_{it} = 4$) can press it forward. Blocking mergers does not necessarily help create frontier firms or encourage their investments.

5.2 Merger Policy toward 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.[54] 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 many 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.[55]

Should the authority relax its merger policy in declining industries? We will answer this

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

[55] "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.

Table 8: If the HDD Industry Were Disappearing Faster

Threshold number of firms ($\underline{N}$)	1	2	3 (Baseline)	4	5	6
(A) If T = December 2016						
Consumer surplus	670.01	695.06	696.89	697.79	697.98	697.88
	(−3.86%)	(−0.26%)	(±0%)	(+0.13%)	(+0.16%)	(+0.14%)
Producer surplus	19.02	17.06	16.67	16.45	16.28	16.13
	(+14.11%)	(+2.36%)	(±0%)	(−1.30%)	(−2.34%)	(−3.73%)
Social Welfare	696.89	712.12	713.55	714.24	714.26	714.02
	(−3.44%)	(−0.20%)	(±0%)	(+0.10%)	(+0.10%)	(+0.06%)
(B) If T = December 2020						
Consumer surplus	698.38	730.66	733.02	733.69	733.95	733.74
	(−4.73%)	(−0.32%)	(±0%)	(+0.09%)	(+0.13%)	(+0.10%)
Producer surplus	20.45	17.95	17.46	17.21	17.01	16.86
	(+17.12%)	(+2.78%)	(±0%)	(−1.45%)	(−2.57%)	(−3.45%)
Social welfare	718.84	748.61	750.49	750.90	750.97	750.60
	(−4.22%)	(−0.25%)	(±0%)	(+0.05%)	(+0.06%)	(+0.02%)

Note: All welfare numbers are present values as of January 1996 in billion US dollars and are the averages of 10,000 simulations under each policy regime. Their percentage changes from the baseline outcomes under $\underline{N} = 3$ are in parentheses.

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 September 2016). Our baseline model assumes the demand will linearly decline to zero between September 2016 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 2016 or December 2020 (i.e., $T = Dec$-2016 or Dec-2020), years earlier than our baseline scenario. We solve these new games for equilibrium CCPs, simulate 10,000 histories, and calculate their average welfare performances.

Table 8 shows the welfare performances of different policy regimes under hypothetical scenarios in which the HDD industry disappears earlier than in the baseline model. The terminal period, T, is December 2016 and December 2020 in Panels A and B, respectively. The overall patterns are similar to the baseline model, that is, $\underline{N} = 1$ or 2 reduces CS and SW, whereas $\underline{N} > 3$ marginally increase them. A subtle but important change is that $\underline{N} = 5$ (instead of $\underline{N} \geqslant 6$) maximizes CS and SW in both scenarios. Thus, our results suggest the optimal merger policy becomes (slightly) less stringent in industries that are disappearing faster.[56]

[56] Note $\underline{N} = 5$ is still more stringent than the current de-facto threshold, $\underline{N} = 3$.

5.3 Optimal Ex-Post ("Surprise") 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 might wonder if the authority could craft a smarter policy than simply committing to constant $\underline{N}$. Our short answer is "yes" in the short run and "no" in the long run.

Table 9 considers "smart" policies in which the authority acts opportunistically and alters the merger threshold ex post.[57] 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 should attract innovation and reduce exits early on (i.e., no value-destruction side effects). However, when the industry reaches the true threshold $\underline{N}^{post} > \underline{N}^{pre}$, the planner should start blocking mergers, so that firms have to compete to death. This surprise ban on mergers will ensure sufficient pro-competitive outcomes ex post.[58]

Table 9: Performance of Opportunistic Policies

Promised threshold ($\underline{N}^{pre}$)	1	1	1	1	1	1
Actual threshold ($\underline{N}^{post}$)	1	2	3	4	5	6
Consumer surplus	721.93	761.33	765.90	767.49	767.98	768.86
	(−5.64%)	(−0.49%)	(+0.11%)	(+0.32%)	(+0.38%)	(+0.50%)
Producer surplus	21.94	19.25	18.36	17.99	17.76	17.56
	(+19.94%)	(+5.23%)	(+0.12%)	(−1.66%)	(−2.93%)	(−4.02%)
Social welfare	743.88	780.58	784.26	785.48	785.74	786.42
	(−5.04%)	(−0.35%)	(+0.12%)	(+0.27%)	(+0.31%)	(+0.39%)

Note: All welfare numbers are present values as of January 1996 in billion US dollars and are the averages of 10,000 simulations under each policy regime. Their percentage changes from the baseline outcomes under $\underline{N} = 3$ (both promised and actual) in Table 6 are in parentheses. The most permissive policy, $\underline{N}^{pre} = \underline{N}^{post} = 1$, is the same as $\underline{N} = 1$ in Table 6.

To some readers, this simulation experiment might appear too complicated and unrealistic

[57] 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. We refer the reader to MNSW (2018) and Jeziorski (2014) for such investigations.

[58] 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 $\left(\underline{N}^{pre}, \underline{N}^{post}\right)$ that represents a particular ex-post policy.

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 found the current de-facto merger threshold ($\underline{N} = 3$) is reasonably close to maximizing social welfare, although it could be tightened for small improvements (e.g., $\underline{N} = 4, 5, 6$). By contrast, permitting mergers to duopoly or monopoly ($\underline{N} = 1, 2$) would lead to negative welfare impacts that are larger by an order of magnitude.

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. Nevertheless, our basic findings seem robust to many different parameterizations that we have tried in the course of writing and revising this paper. We may also investigate the implications of our model in different industry environments by performing simulations under different parameter values.

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), Miller and Weinberg (2017), or Igami and Sugaya (2018). 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.

Appendices: Table of Contents

Appendix A lists our interviews. Appendices B, C, D, and E supplement the details of sections 2 (Model), 3 (Data), 4 (Empirical Analysis), and 5 (Policy and Welfare) 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 10: 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

B.1 Why Poaching Top Engineers Might Not Be Sufficient

Section 2.1 explained two motivations for mergers in the HDD industry: (i) killing rival firms to soften competition and (ii) pooling teams of engineers for the next technological breakthrough. Regarding the second motivation, one might wonder why hiring top engineers from another firm is not sufficient. Our interviews and the industry context suggest three reasons.

First, the nature of innovation in the HDD engineering seems to rely not so much on "bright new concepts" that come to the minds of few star scientists. Most of the "new" ideas have been around for decades, but typically no firms could implement them with sufficient precision or reliability until recently. For example, a paradigm called "perpendicular recording" has existed as a concept since 1976 but became commercially viable only in 2005.

Second, the implementation of "new" ideas seems to involve a long tatonnement process to configure the product design, its pilot production line, and its volume production lines (the preparation of which requires careful re-tooling of specialized manufacturing equipment, which are precision instruments themselves). Moreover, the HDD technology spans many different fields, including aero/fluid dynamics, materials science, semiconductor (for read-write heads), signal processing, and other electronic and mechanical engineering expertise. The complex nature of HDD engineering seems to suggest the tacit knowledge is embodied in large teams of engineers rather than a few key individuals.

Third, even if a firm successfully persuaded an entire team of engineers from another firm to move, intellectual property (IP) rights would not automatically follow, because patents are usually "assigned" to their previous employer and not individual inventors. Several of our interviewees told us that there had been occasional lawsuits between major HDD firms concerning the poaching of specific engineers and related IP-rights violations. They said the firms learned that such lawsuits would only make IP lawyers richer and subsequently became less aggressive in pursuing IP-sensitive employee-poaching.

B.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:

$$\bar{V}_{it}^{e}\left(\omega_t, \varepsilon_{it}^{e}\right) = -\kappa^{e} + \varepsilon_{it}^{e} + \beta E\left[\Lambda_{i,t+1}\left(\omega_{t+1}\right)|\omega_t, a_{it} = enter\right], \text{ and} \tag{20}$$

$$\bar{V}_{it}^{o}\left(\omega_t, \varepsilon_{it}^{o}\right) = \varepsilon_{it}^{o} + \beta E\left[\Lambda_{i,t+1}\left(\omega_{t+1}\right)|\omega_t, a_{it} = out\right], \tag{21}$$

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

$$V_{it}^{0}\left(\omega_t, \varepsilon_{it}^{0}\right) = \max\left\{\bar{V}_{it}^{e}\left(\omega_t, \varepsilon_{it}^{e}\right), \bar{V}_{it}^{o}\left(\omega_t, \varepsilon_{it}^{o}\right)\right\}, \tag{22}$$

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

$$EV_{it}^{0}\left(\omega_t\right) = E_{\varepsilon}\left[V_{it}^{0}\left(\omega_t, \varepsilon_{it}^{0}\right)\right] = \sigma\left\{\gamma + \ln\left[\exp\left(\frac{\tilde{V}_{it}^{e}}{\sigma}\right) + \exp\left(\frac{\tilde{V}_{it}^{o}}{\sigma}\right)\right]\right\}. \tag{23}$$

These expressions correspond to equations 1 through 7 in the main text.

When the potential entrant is a non-mover, its expected value is simpler than the incumbent's version in equation 9,

$$\begin{aligned} W_{it}^{0j}\left(\omega_t\right) = \; & \sigma_{it}\left(a_{jt} = exit\right) \times \beta E\left[\Lambda_{i,t+1}\left(\omega_{t+1}\right)|\omega_t, a_{jt} = exit\right] \\ & + \sigma_{it}\left(a_{jt} = stay\right) \times \beta E\left[\Lambda_{i,t+1}\left(\omega_{t+1}\right)|\omega_t, a_{jt} = idle\right] \\ & + \sigma_{it}\left(a_{jt} = innovate\right) \times \beta E\left[\Lambda_{i,t+1}\left(\omega_{t+1}\right)|\omega_t, a_{jt} = innovate\right] \\ & + \sum_{k \neq i,j} \sigma_{it}\left(a_{jt} = merge\ k\right) \\ & \quad \times \beta E\left[\Lambda_{i,t+1}\left(\omega_{t+1}\right)|\omega_t, a_{jt} = merge\ k\right] \\ & + \sum_{k \neq i,j} \sigma_{it}\left(a_{jt} = innovate\ \&\ merge\ k\right) \\ & \quad \times \beta E\left[\Lambda_{i,t+1}\left(\omega_{t+1}\right)|\omega_t, a_{jt} = innovate\ \&\ merge\ k\right], \end{aligned} \tag{24}$$

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 9 and 24 become

$$\begin{aligned} W_{it}^{j}\left(\omega_t\right) = \; & \pi_{it}\left(\omega_t\right) - \phi_t \\ & + \sigma_{it}\left(a_{jt}^{0} = enter\right) \times \beta E\left[\Lambda_{i,t+1}\left(\omega_{t+1}\right)|\omega_t, a_{jt}^{0} = enter\right] \\ & + \sigma_{it}\left(a_{jt}^{0} = out\right) \times \beta E\left[\Lambda_{i,t+1}\left(\omega_{t+1}\right)|\omega_t, a_{jt}^{0} = out\right], \text{ and} \end{aligned} \tag{25}$$

$$\begin{aligned} W_{it}^{0j}\left(\omega_t\right) = \; & \sigma_{it}\left(a_{jt}^{0} = enter\right) \times \beta E\left[\Lambda_{i,t+1}\left(\omega_{t+1}\right)|\omega_t, a_{jt}^{0} = enter\right] \\ & + \sigma_{it}\left(a_{jt}^{0} = out\right) \times \beta E\left[\Lambda_{i,t+1}\left(\omega_{t+1}\right)|\omega_t, a_{jt}^{0} = out\right] \end{aligned} \tag{26}$$

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},

$$\Pr\left(a_{it}^{0} = action\right) = \frac{\exp\left(\frac{\tilde{V}_{it}^{action}}{\sigma}\right)}{\exp\left(\frac{\tilde{V}_{it}^{e}}{\sigma}\right) + \exp\left(\frac{\tilde{V}_{it}^{o}}{\sigma}\right)}, \tag{27}$$

which corresponds to equation 10.

B.3 Uniqueness of Equilibrium

Section 2.3 explained why the equilibrium of our model is unique. This section provides further discussions regarding (i) which of the assumptions are crucial for uniqueness and (ii) whether our modeling approach is akin to some form of equilibrium selection.

First, the distinction between random and deterministic orders of move is not crucial. The combination of sequential (or alternating) moves and a finite horizon *is* crucial, as well as the use of a discrete-choice model at each move (or any other specification that leads to a unique optimal choice at each move).

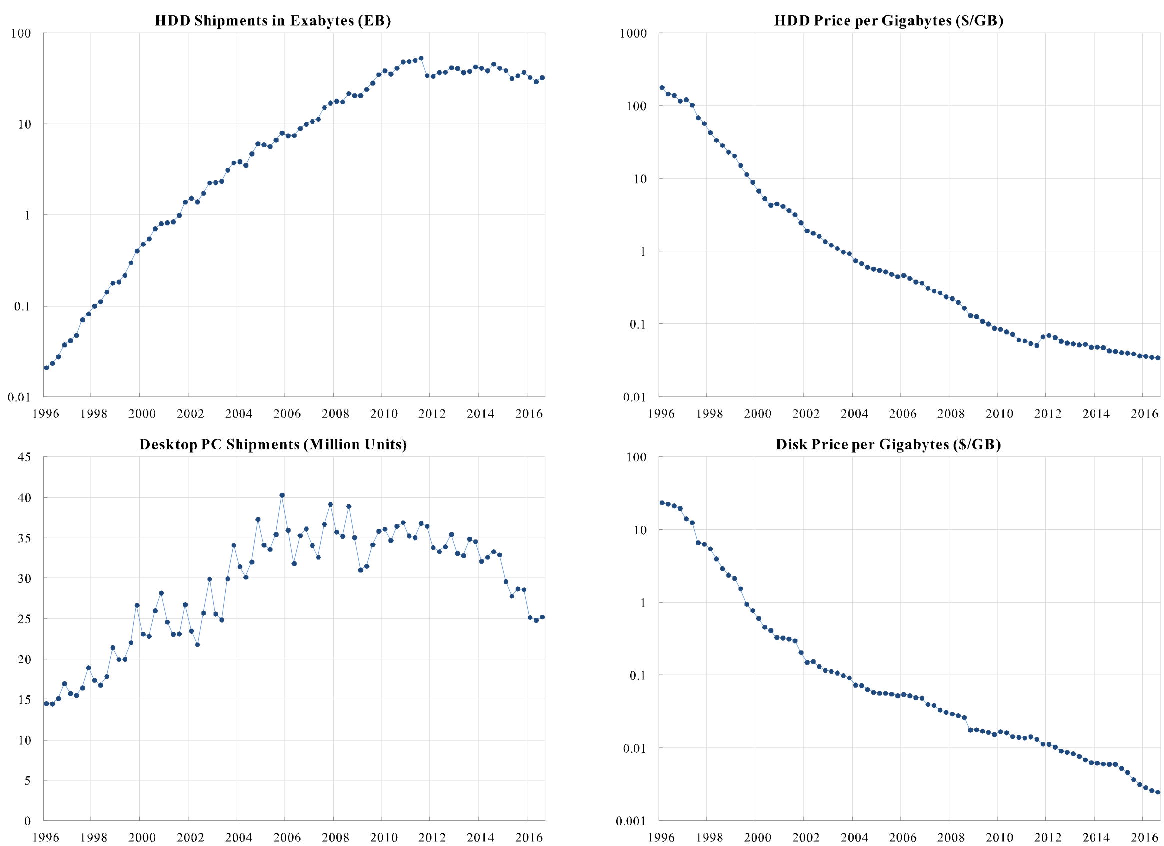
Second, because the equilibrium of our model is unique, there is no "selection" among multiple equilibria in a formal sense. More informally, the question as to whether our model imposes "something akin to equilibrium selection" would be closely related to the choice of $T = December$ 2025 in our baseline specification. This choice is based on the seemingly common view among our interviewees that the HDD industry would remain "relevant" to the broader IT economy for the next 10 years (we conducted our interviews in 2014–2015). Our sensitivity analysis with respect to $T = December$ 2020 (column 5 of Table 5) shows the parameter estimates and the fit change only negligibly from the baseline, which seems to suggest that assumptions about the distant future have limited impacts on our empirical results.

Appendix C Supplementary Materials for Section 3

C.1 Data Patterns Underlying Demand Estimation (Panel A)

Figure 5 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.

Figure 5: 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.

C.2 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).

The two right-most columns of Table 11 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. That is, the Cournot model predicts the *combined* market share of the merging parties should decrease, whereas those of non-merging firms would increase.

Table 11: 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

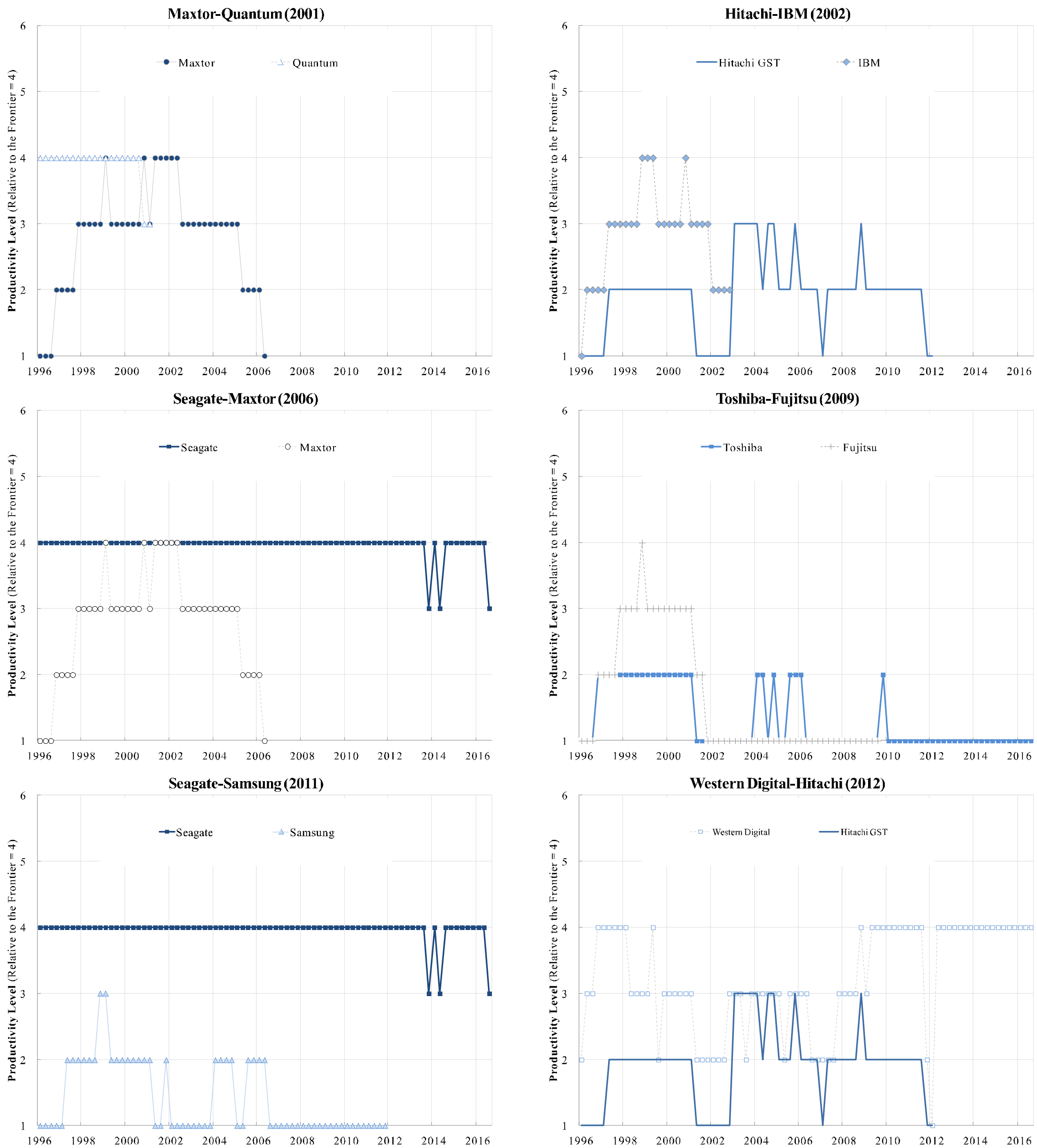
Note: ms^T and ms^A denote the target and the acquiring firms' market shares, respectively. For each merger case, "before" refers to the last calendar quarter in which ms^T was recorded separately from ms^A, and "after" is four quarters after "before." Alternative time windows including 1, 8, and 12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

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

At the same time, the comparison of the right-most column and the column labeled "ms^A Before" suggests the acquiring firms managed to achieve expansions relative to their *individual* pre-merger market shares. This pattern is consistent with our interviews with the industry participants, in which they explained gaining market shares as the primary motivation for mergers. Figure 6 plots the realized productivity paths of the merging parties for the entire sample period, which are calculated based on our marginal-cost estimates and discretization procedures.

Figure 7 overlays the historical HHI on the number of firms, n_t. The HHI correlates

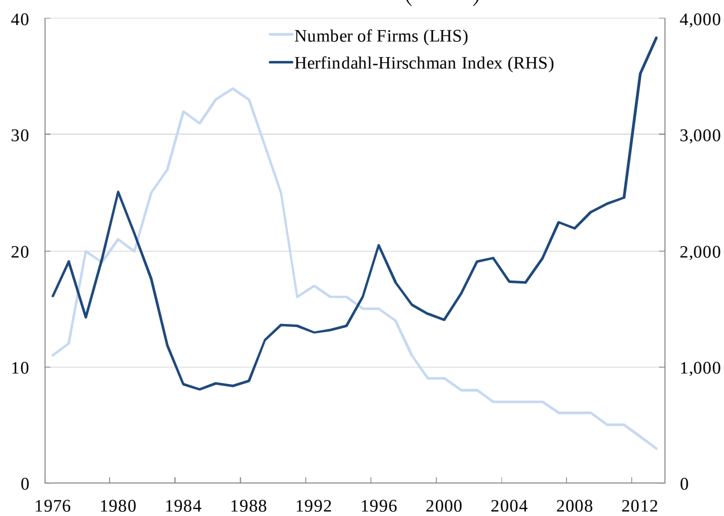
Figure 6: Realized Productivity Paths of the Merging Parties



Note: See sections 4.2 and 4.3 and Appendix D.1 for the details of marginal-cost estimates, productivity, and discretization.

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.

Figure 7: Herfindahl-Hirschman Index (HHI) of the Global HDD Market



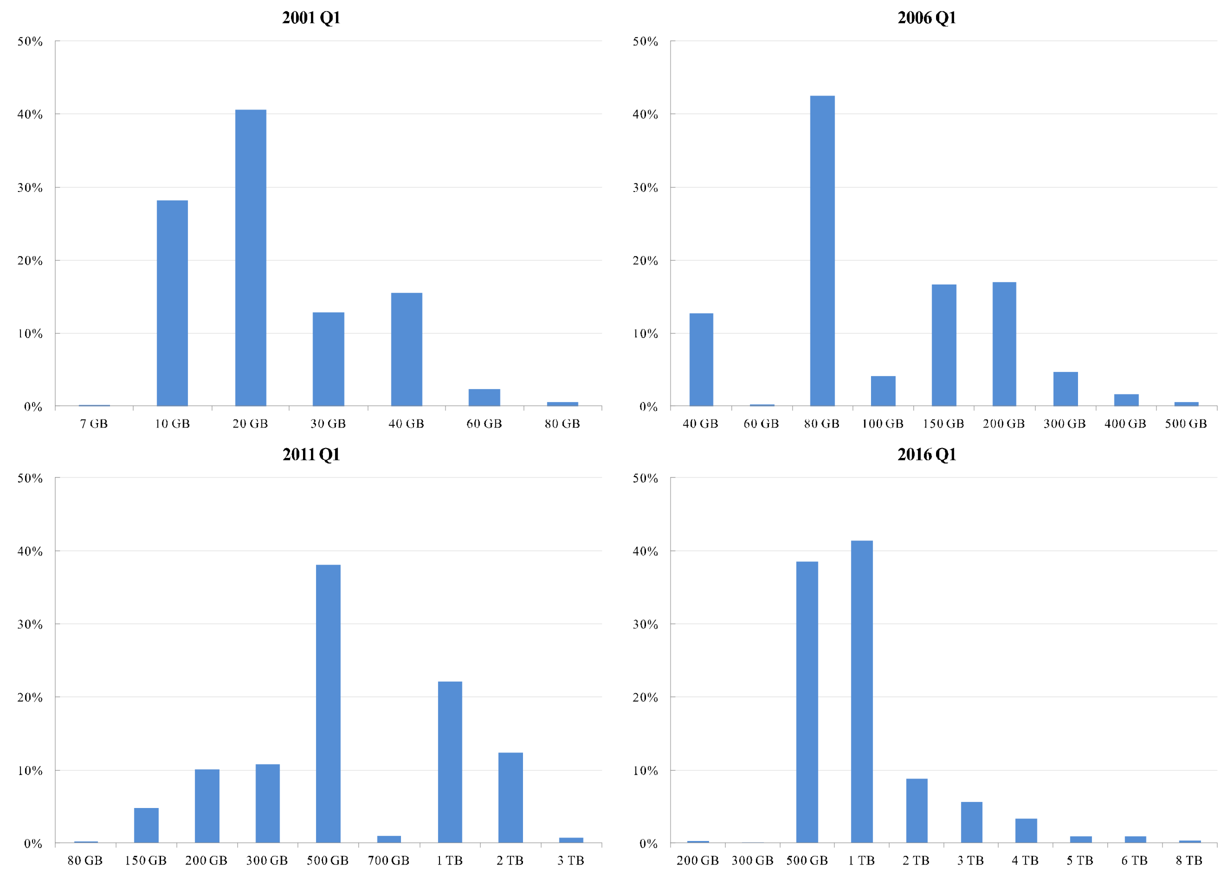
Note: The HHI is the sum of the squares of the firm's market shares.

C.3 Additional Details on Product Characteristics

Section 3 stated all firms carry all capacities at all times, but one might wonder whether "frontier" firms introduce higher-capacity products early and held commercial advantage in terms of quality differentiation. Figure 8 shows four histograms of HDD sales distribution across storage capacities in the first calendar quarter of 2001, 2006, 2011, and 2016, respectively. A few "mainstream" categories of the time tend to account for over 70%–80% of the aggregate sales. The highest capacities (e.g., over 5TB in 2016Q1) typically record only negligible sales. Hence, even though frontier firms of the time tend to lead the product-innovation race, the introduction of highest-capacity HDDs does not necessarily confer competitive advantage in terms of vertical differentiation.

Figure 9 shows the average price of HDDs per quality unit (i.e., per GB of storage capacity) at four different points in time. Within each graph, higher-capacity HDDs tend to sell at lower prices per GB, because magnetic disks are not the only component of HDDs. The price of each HDD also reflects "fixed cost" components, such as electronic and mechanical parts. However, these differences within each graph are dwarfed by the differences across

Figure 8: Distribution of HDD Sales across Storage Capacities

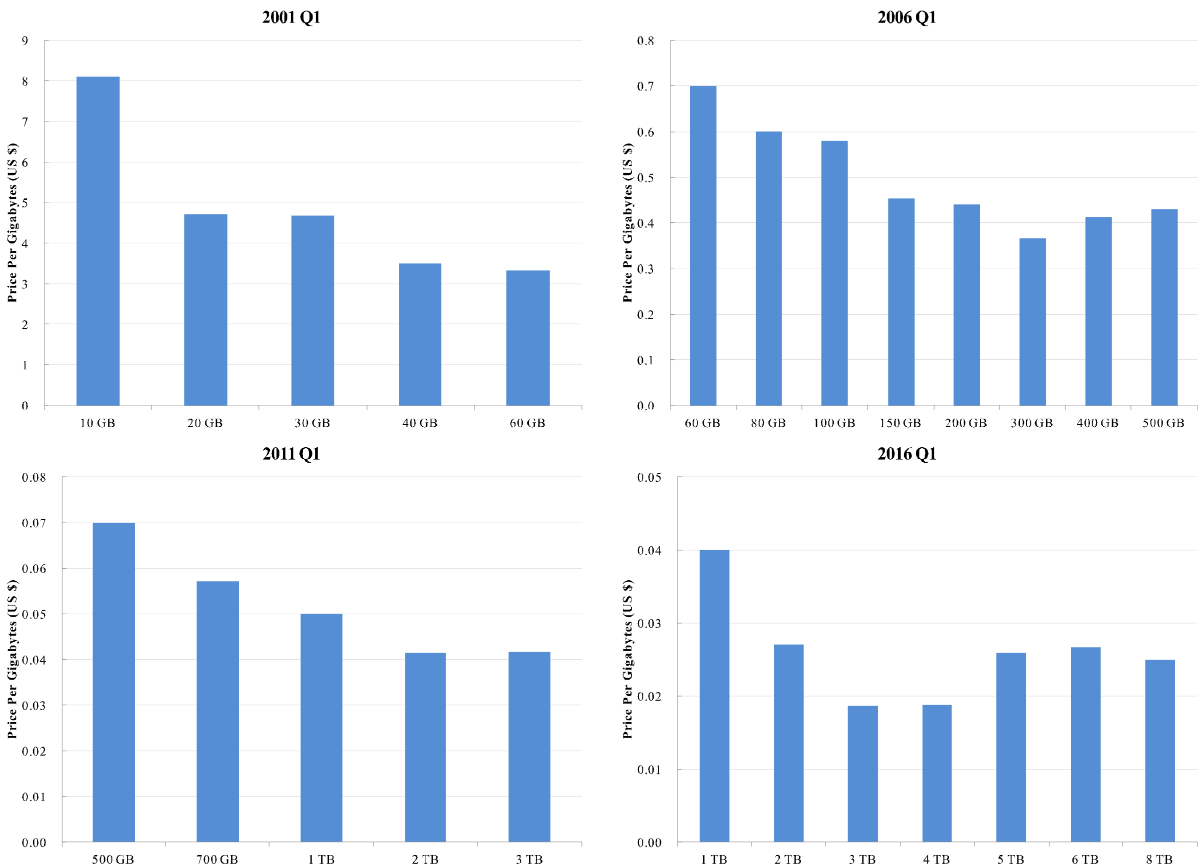


Note: Each panel shows a histogram of sales volume of HDDs for desktop PCs in that calendar quarter. *Source*: *TRENDFOCUS* Reports.

graphs. Note the scale of the vertical axes and how they decrease over time. Therefore, the scale economy in capacity within period seems relatively unimportant given the speed of cost reduction.

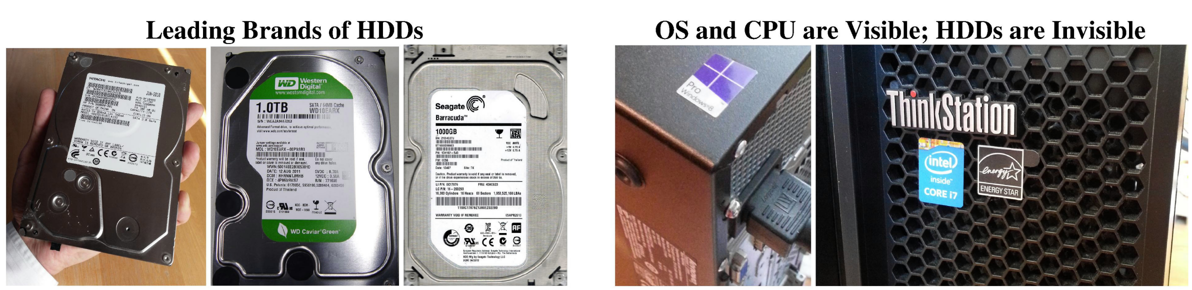
Finally, Figure 10 shows the pictures of HDDs and the exterior of a desktop PC. Most users do not pay attention to the specific "brands" of HDDs that their PCs are using. The demand-side dynamics of PCs might be affected by the generational shifts in OS and CPU but not as much by HDDs.

Figure 9: HDD Price per Gigabytes by Storage Capacity



Note: Each panel shows the average price of desktop HDDs divided by storage capacity in GB. Note the scale differences across the graphs dwarf the within-graph variability.
Source: *TRENDFOCUS* Reports.

Figure 10: Product Characteristics of HDDs



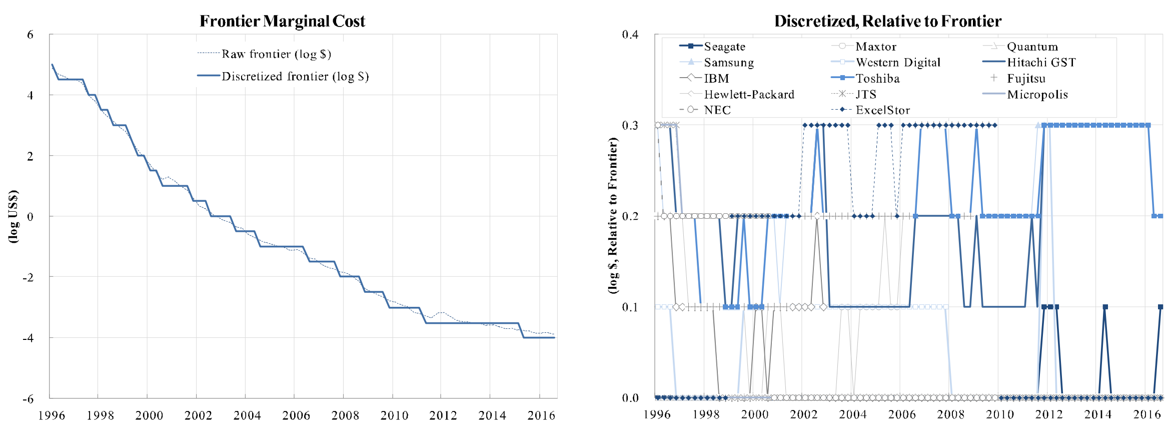
Note: Left panel shows 3.5-inch HDDs of Hitachi GST, Western Digital, and Seagate Technology. Right panel shows evidence of successful marketing efforts by Microsoft and Intel (and lack thereof by HDD makers).

Appendix D Supplementary Materials for Section 4

D.1 Discretization of Productivity Levels

We define the state space for computational implementation by discretizing the levels of firm-specific productivity based on the marginal cost estimates in section 4.2. Figure 11 (left) plots the trajectory of marginal costs at the frontier firms; Figure 11 (right) shows individual firms' positions relative to the frontier.

Figure 11: Marginal Cost Estimates and Their Discretization



Note: The left panel plots the "frontier" marginal cost. The right panel plots each firm's discretized marginal cost on a 0.1 log-dollar grid relative to the frontier. We construct productivity levels by reversing the rank order of discretized marginal costs and keeping track of the frontier as well as each firm's distance to it.

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 as leader-follower differences. 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 is denoted by $\omega_{it} \in \{\omega^1, \omega^2, ...\}$. Note the ranking convention reverses as we employ marginal costs as productivity levels. That is, a lower marginal cost will be referred to as a high-productivity level. We construct the variable indicating *innovate* (one of the dynamic discrete-choice actions) from these discretized productivity estimates.

Our computational implementation allows 128 different absolute levels of ω_{it}. Most of the levels in the top half are never reached in the data. We define such a large state space to avoid imposing a numerical boundary on the realization of simulated industry history.

Literally allowing 128 absolute levels for each firm in each period will exhaust computer memory. We alleviate this burden by exploiting the sparsity of the full state space. As Figure 11 shows, the firms in the data stay close to each other (i.e., within four levels) because of the industry-wide technological spillovers. Denote the highest productivity level ever achieved up to time t by $f_t \equiv \max_{i,\tau \leq t} \{\omega_{i\tau}\}$ (i.e., the "frontier"), and define individual firms' relative productivity levels as,

$$\tilde{\omega}_{it} \equiv \omega_{it} - f_t + L,$$

where L is a constant denoting the frontier firms' relative level. For expositional purposes, we set $L = 4$, so that the frontier firms belong to level 4 and the bottom ones to level 1.

When a frontier firm innovates at t, its productivity increases at $t+1$ by one in absolute levels. At the same time, the relative productivity grid for all firms shifts up at $t+1$ by one in absolute levels as well. Level-1 firms at t continue belonging to level 1 at $t+1$ instead of dropping out of the grid, thanks to the spillovers from the frontier firm's innovation, which advances the state of knowledge for the entire industry as well as the scientific community surrounding the magnetic information-storage technologies. Keeping track of $(f_t, \{\tilde{\omega}_{it}\}_i)$ instead of the original $\{\omega_{it}\}_i$ facilitates computation. This reformulation is similar to Goettler and Gordon (2011) and other "quality ladder" models with moving technological frontiers.

The economic significance of the (quality-adjusted) cost-reducing innovations and "synergies" in our empirical model is as follows. Between 1996 and 2016, the frontier marginal cost (per GB) decreased from $+5$ log US\$ (approximately \$149) to -4 log US\$ (approximately \$0.02). The step size of the frontier marginal cost is 0.5 log US\$, and the step size of each firm's distance to frontier is 0.1 log US\$. Therefore, both an innovation (due to in-house R&D) and the expected efficiency gain (due to synergy) would reduce marginal cost by 10% in the case of a non-frontier (level-1, 2, or 3) firm and by 50% in the case of a frontier (level-4) firm.

In terms of competitive implications, a 10% reduction of marginal cost could increase the firm's market share by 5% $\sim$ 10% in 1996, for example. In terms of antitrust implications, this level of expected synergy could make mergers decrease the equilibrium spot-market prices in some cases (depending on the time period and market structure) but more often lead to price increases.

D.2 Fixed Costs and Accounting Data

We determine the fixed cost of operations and technological catch-up, ϕ_t, 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 2017, 2018; 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 ϕ_t 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 ϕ_t 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, ϕ_t	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 1–4	35	3.521	0.610	2	4
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 ϕ_t 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 13: Fixed-Cost Estimates from Accounting Data

Dependent variable: Fixed cost, ϕ	(1) OLS	(2) OLS	(3) OLS	(4) OLS
Year (t)	89.80^{***}	–	75.75^{***}	47.68^{***}
	(11.16)	(–)	(11.34)	(8.47)
Productivity level (ω_{it})	–	788.48^{***}	510.04^{***}	215.55^{**}
	(–)	(139.63)	(100.66)	(90.17)
$I\{i = Seagate\}$	–	–	–	568.54^{***}
	(–)	(–)	(–)	(117.40)
$I\{(i,t) \in Special\}$	–	–	–	$1,067^{***}$
	(–)	(–)	(–)	(158.3)
Number of observations	35	35	35	35
Adjusted R^2	0.633	0.476	0.774	0.906

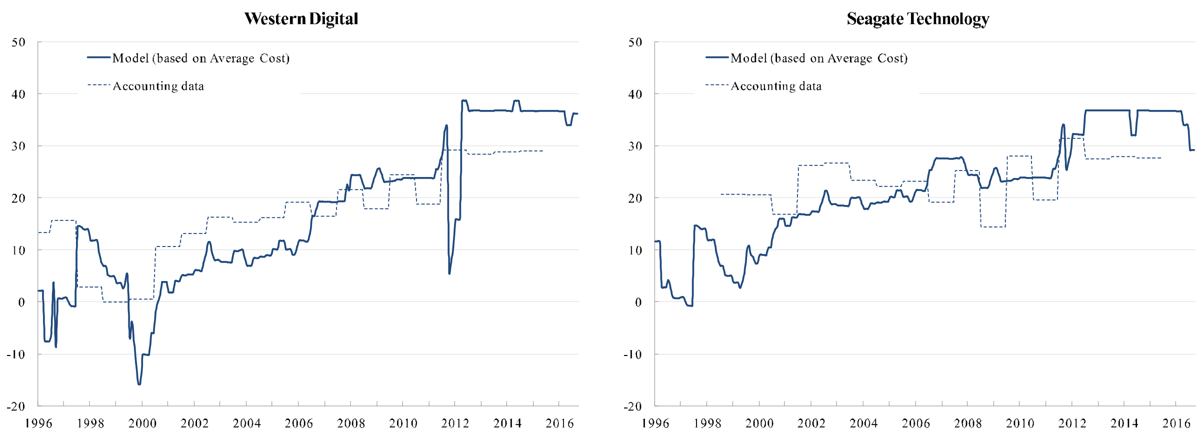
Table 12 shows summary statistics. Sample size is smaller than 40 (i.e., two firms times 20 years) because Seagate became privately owned for financial restructuring in 2000 and its financial statements lost consistency after it went public again. Our main variable is *fixed cost*, which is the sum of 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 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 $\phi_t(\omega_{it})$ in our main estimation task in section 4.3.

D.3 Fit of Profit Margins

In Figure 2 in section 4.2, we showed our estimates of profit margins based on variable-cost estimates. We explained the difference between model and accounting profits is due to fixed costs. Figure 12 shows a different version that incorporates our estimate of the base fixed cost ($\hat{\phi}_0$) as well. The inclusion of $\hat{\phi}_0$ (divided by each firm's concurrent revenue) improves the fit in two respects. First, the differences between economic and accounting margins decrease from 4.6 to -1.6 percentage points for Western Digital (i.e., now our estimate is slightly lower than the accounting gross-profit margin) and from 3.6 to 0.2 percentage points for Seagate, respectively. Second, the correlation coefficients increase from 0.75 to 0.79 for Western Digital, and from 0.51 to 0.83 for Seagate, respectively.

Figure 12: Comparison of Profit Margins (%) in the Model and Financial Statements



D.4 Implied vs. Actual Acquisition Prices

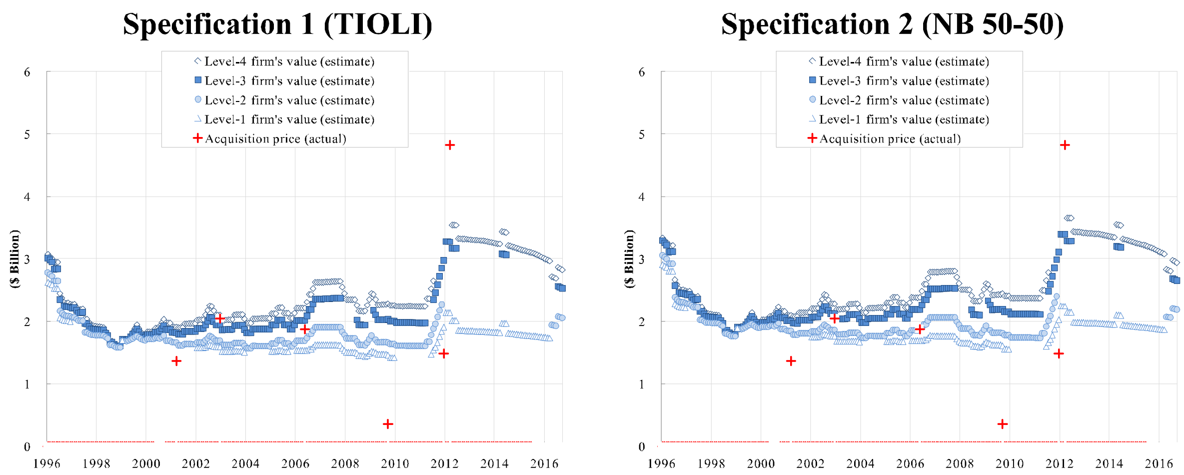
We conduct a sanity check of fit by comparing estimated enterprise values and the actual acquisition prices. Figure 13 plots our firm-value estimates along the historical path of market structure in the data, and overlays the actual transaction prices in the six merger cases from *Thomson*'s financial 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).

The difference between the two graphs may not be highly visible, but the average firm value under the 50-50 Nash bargaining specification ($2.194 billion) is slightly higher than under the TIOLI specification ($2.029 billion).

D.5 Sensitivity Analysis

Table 14 shows additional ML estimates of dynamic parameters. Columns 1 through 4 set the acquirer's bargaining power $\zeta \in \{0.6, 0.7, 0.8, 0.9\}$, respectively, to investigate the intermediate cases between the two protocols in section 4.3 ($\zeta = 0.5$ and 1 in Table 5). The parameter values change only negligibly across these specifications, but their directions

Figure 13: 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 four markers represent our estimates of equilibrium firm values along the historical path of market structure in the data.

Table 14: MLE of Dynamic Parameters (Additional Sensitivity Analysis)

Specification	(1)	(2)	(3)	(4)	(5)	(6)
Bargaining (ζ):	0.6 (NB)	0.7 (NB)	0.8 (NB)	0.9 (NB)	1 (TIOLI)	1
Recog. prob. (ρ):	$\frac{1}{14}$	$\frac{1}{14}$	$\frac{1}{14}$	$\frac{1}{14}$	$\frac{1}{21}$	$\frac{1}{28}$
κ^c	0.010	0.011	0.011	0.011	-0.017	-0.016
	[0.000.0.020]	$[0.001, 0.020]$	$[0.001, 0.020]$	$[0.001, 0.020]$	$[-0.040, 0.006]$	$[-0.040, 0.008]$
κ^i $(\omega_{it} = 1, 2, 3)$	0.50	0.50	0.49	0.48	0.39	0.18
	$[0.28, 0.74]$	$[0.27, 0.73]$	$[0.27, 0.73]$	$[0.27, 0.71]$	$[0.00, 0.78]$	$[-0.26, 0.62]$
κ^i $(\omega_{it} = 4)$	0.90	0.89	0.87	0.86	1.08	0.88
	$[0.41, 1.52]$	$[0.41, 1.49]$	$[0.41, 1.49]$	$[0.40, 1.45]$	$[0.26, 2.08]$	$[-0.05, 1.99]$
κ^m	1.22	1.23	1.25	1.26	1.78	2.06
	$[0.74, 1.84]$	$[0.76, 1.84]$	$[0.78, 1.87]$	$[0.79, 1.85]$	$[0.99, 2.76]$	$[1.19, 3.14]$
κ^e	0.17	0.16	0.17	0.17	0.29	0.22
	[–]	[–]	[–]	[–]	[–]	[–]
σ	0.59	0.58	0.57	0.56	0.93	1.01
	$[0.44, 0.85]$	$[0.43, 0.84]$	$[0.43, 0.82]$	$[0.42, 0.81]$	$[0.66, 1.41]$	$[0.72, 1.54]$
Log likelihood	-157.15	-157.08	-157.02	-156.97	-138.10	-127.73

Note: The synergy parameter $\lambda = 1$ and the terminal period $T =$ December 2025 in all specifications. The 95% confidence intervals are constructed from the likelihood-ratio tests.

are intriguing. The estimated sunk cost of merger negotiation (κ^m) becomes larger with ζ, because: (i) greater bargaining power lowers acquisition prices (p_{ij}) and increases the values of mergers for the acquirers ($\bar{V}_{ij}^m$); (ii) but the count of actual mergers in the data do not change with our calibration of ζ; (iii) hence, the only way for the model to reconcile the model with the data is to increase κ^m, so that the model-generated CCPs of merger goes down to match the CCPs in the data. The estimated costs of in-house R&D (κ^is) seem to move in the opposite direction to balance the choice between innovation and merger.

The scaling parameter of logit errors (σ) also changes systematically across columns 1 through 4. This parameter governs the importance of (logit shocks relative to) the underlying payoffs of different states (recall we estimated $\pi(\omega_{it})$ outside the dynamic framework). If σ were extremely large, period profits and other economic fundamentals would hardly matter to the dynamic discrete choice. By contrast, if σ were extremely small, these underlying payoffs and their differences would have to explain all the patterns of dynamic discrete choice in the data. Thus, the (slightly) lower σs at higher ζs suggest our model of static demand supply plays a (slightly) more important role in explaining mergers, innovation, and entry-exit dynamics when we assume greater benefits for acquirers.

The last two columns of Table 14 investigate the role of recognition probability (ρ). Our baseline model sets $\rho = \frac{1}{n_{\max}} = \frac{1}{14} = 0.071$, because the data start with 13 incumbents (and we endogenize entry by assuming one potential entrant in each month). Column 5 increases the maximum number of players ($n_{\max}$) by 50% and sets $\rho = \frac{1}{21} = 0.048$; column 6 does so by 100% and sets $\rho = \frac{1}{28} = 0.036$. These lower ρs mean a given firm can expect to move only once every two to three years, which seems extreme even for relatively rare events such as mergers. Nevertheless, this sensitivity analysis helps us understand the inner working of the model and certain features of the data. First, the logit scaling parameter (σ) becomes larger, because the model has to rationalize the actual mergers and innovations in the data despite assuming that firms could hardly move. Second, however, the likelihood improves substantially, because lower ρs mechanically fit some part of the data better (i.e., there are many months in which nothing happens). Third, the base fixed cost of operation (ϕ_0) becomes negative, which suggests the model has to rationalize the existence of active firms (i.e., why these firms do not exit despite the unattractive environment in which they can hardly make payoff-maximizing moves) by subsidizing their continued operations.

D.6 Heterogeneous Synergies

We set the mean synergy parameter $\lambda = 1$ in the baseline model, because this is the average improvement in productivity level (above and beyond the benefits of rationalization) that is implied by our marginal-cost estimates before and after the six mergers in the data. Given the limited number of observations, we did not attempt to model or estimate λ as a function of pre-merger productivity, ω_{it}.

Nevertheless, we may investigate how the equilibrium strategies change in response to different parameterizations. We consider two cases. First, one might suspect the acquirer's current productivity is positively correlated with the chance of further improvements due to synergy (i.e., $\frac{d\lambda(\omega)}{d\lambda} > 0$). Second, playing catch-up could be easier than pushing the frontier not only in terms of R&D (i.e., as suggested by our estimates of $\kappa^i(\omega)$) but also in terms of synergies: $\frac{d\lambda(\omega)}{d\lambda} < 0$.

We implement these ideas by parameterizing:

$$\lambda(\omega_{it}) = \frac{\omega_{it}}{c} \tag{28}$$

and

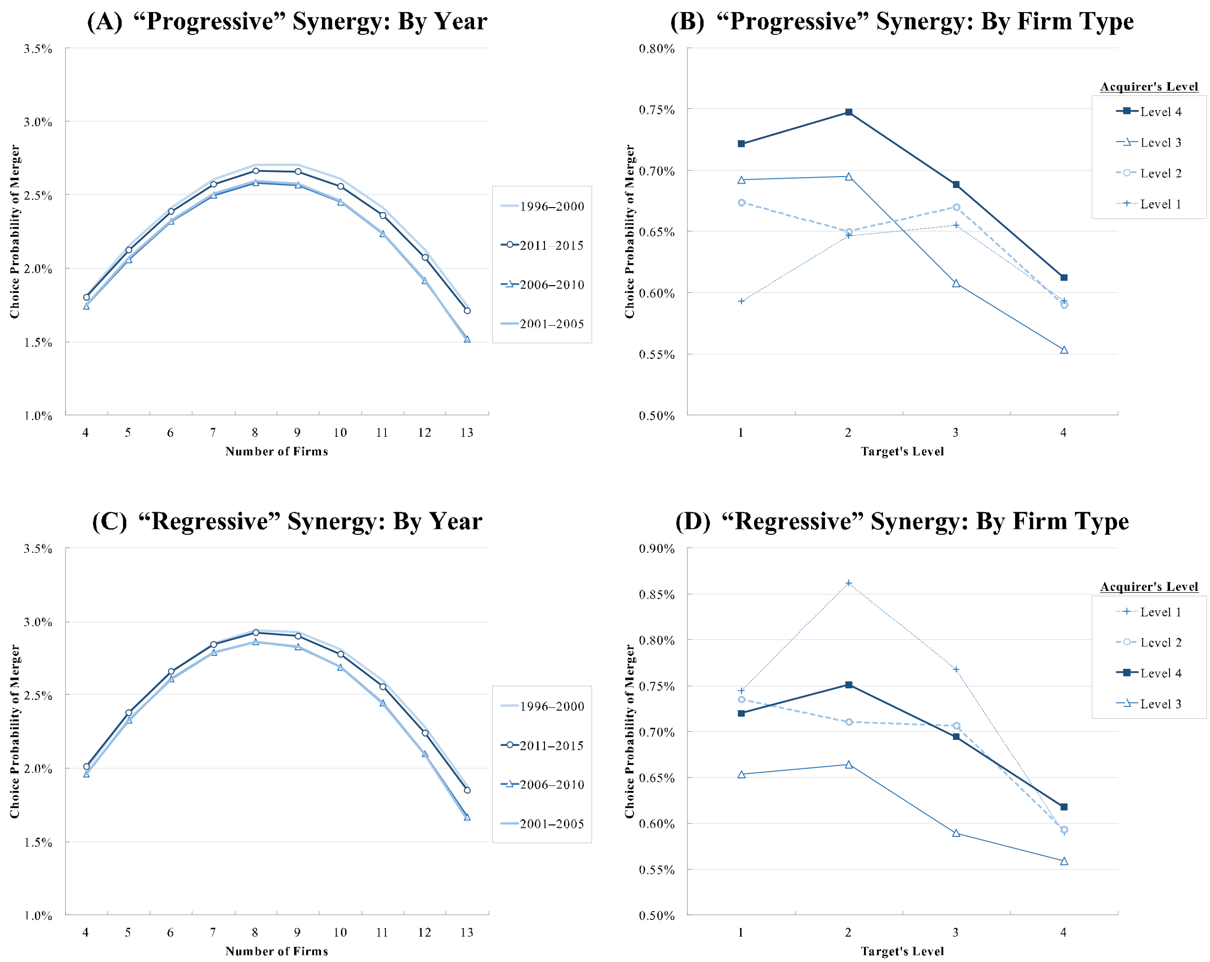
$$\lambda(\omega_{it}) = \frac{5 - \omega_{it}}{c}, \tag{29}$$

where the scalar $c = 2.5$ ensures that the average λ across the acquirer's types $\omega_{it} \in \{1, 2, 3, 4\}$ remains the same as the baseline model's $\lambda = 1$. We recomputed equilibrium under each of the two alternative specifications.

Figure 14 plots the equilibrium CCPs of merger by year and type. Panels A and B show the "progressive" synergy case (equation 28), whereas C and D show the "regressive" synergy case (equation 29). The left panels are almost exactly the same as in the baseline result (Panels C and D in Figure 4). The right panels also share the mostly downward-sloping shapes with the baseline CCPs, but the ordering of types in terms of levels is different. Specifically, the "progressive" specification makes level-4 (the highest productivity) firms the most acquisitive. By contrast, the "regressive" specification makes level-1 firms' CCPs the highest with respect to most target types. These patterns reflect the underlying configurations of $\lambda(\omega_{it})$.

Thus, this sensitivity analysis suggests the equilibrium CCPs are responsive to the way we model heterogeneity of $\lambda(\omega_{it})$ in an intuitive manner. The rank order of likely acquirers mostly follow that of expected synergies. At the same time, the overall level of CCPs ($0.55\% \sim 0.75\%$) remains similar across specifications as long as the average λ remains the same.

Figure 14: Equilibrium Merger Strategies under Heterogeneous Synergies



Note: Each graph summarizes the equilibrium strategies for M&A, by averaging the structural CCP estimates across ω_{it}, n_t, or t. For expositional purposes, the horizontal axis represents the concurrent number of active firms (n_t) as a summary statistic of the underlying state (ω_t, which subsumes both n_t and the productivity profile of all firms).

Appendix E Supplementary Materials for Section 5

E.1 Counterfactual Policies under the 50-50 Nash Bargaining Setup

Section 5.1 showed counterfactual policy simulations under the baseline specification of the bargaining protocol (TIOLI). Table 15 shows an alternative version under the 50-50 Nash bargaining setup. The results are qualitatively similar. An interesting difference is that $\underline{N} = 5$ leads to the highest discounted SW instead of $\underline{N} = 6$ in Table 6. Taken together, these results seem to suggest the limit of welfare improvement by tougher antitrust regimes even when the equilibrium relationship between competition and innovation is positive (c.f., Panel A of Figure 4).

Table 15: Welfare Performance of Counterfactual Policies under 50-50 Nash Bargaining

Threshold number of firms ($\underline{N}$)	1	2	3 (Baseline)	4	5	6
(A) All periods: 1996–2025						
Consumer surplus	755.56	763.49	764.95	765.82	766.11	766.13
	(−1.23%)	(−0.19%)	(±0%)	(+0.11%)	(+0.15%)	(+0.15%)
Producer surplus	19.25	18.68	18.42	18.20	18.00	17.81
	(+4.48%)	(+1.40%)	(±0%)	(−1.22%)	(−2.29%)	(−3.33%)
Social Welfare	774.81	782.17	783.38	784.02	784.11	783.94
	(−1.09%)	(−0.15%)	(±0%)	(+0.08%)	(+0.09%)	(+0.07%)
(B) First half: 1996–2010						
Consumer surplus	535.83	536.37	536.50	536.63	536.59	536.51
	(−0.13%)	(−0.03%)	(±0%)	(+0.02%)	(+0.02%)	(+0.00%)
Producer surplus	13.32	13.28	13.25	13.20	13.14	13.07
	(+0.53%)	(+0.22%)	(±0%)	(−0.38%)	(−0.83%)	(−1.36%)
Social welfare	549.15	549.65	549.75	549.83	549.73	549.58
	(−0.11%)	(−0.02%)	(±0%)	(+0.01%)	(−0.00%)	(−0.03%)
(C) Second half: 2011–2025						
Consumer surplus	219.73	227.12	228.45	229.20	229.52	229.61
	(−3.81%)	(−0.58%)	(±0%)	(+0.33%)	(+0.47%)	(+0.51%)
Producer surplus	5.93	5.40	5.18	5.00	4.86	4.74
	(+14.58%)	(+4.42%)	(±0%)	(−3.36%)	(−6.03%)	(−8.35%)
Social welfare	225.66	232.53	233.62	234.20	234.38	234.36
	(−3.41%)	(−0.47%)	(±0%)	(+0.25%)	(+0.33%)	(+0.31%)

Note: All welfare numbers are present values as of January 1996 in billion US dollars and are the averages of 10,000 simulations under each policy regime. Their percentage changes from the baseline outcomes under $\underline{N} = 3$ are in parentheses. Note the welfare numbers in Panel (C) are considerably smaller than those in Panels (A) and (B), because the annual discount factor of $\beta = 0.9$ means they are discounted at $\beta^{15} = 0.2059$.

E.2 Estimates and Counterfactuals under Price-based Merger Policies

We modeled antitrust policy as a threshold number of firms ($\underline{N} = 3$ in the baseline case). This rule of thumb seemed commonly shared by the HDD industry participants that we interviewed. Several former chief economists at the DOJ and the FTC also informed us that the "burden of proof" tends to shift at this number, at least informally. However, this rule of thumb is not always supported by theory. Even in a simple static model such as a homogeneous-good Cournot game, a merger's impact on welfare depends the cost profile of firms as well as the identity of the merging parties. Hence, even though a rule of thumb might provide a useful reference point, the antitrust agencies would try to evaluate whether particular mergers would raise prices or not.[59]

For these reasons, we designed an alternative model with a different antitrust policy rule that *blocks mergers that would raise prices by more than $X\%$ in the comparative statics of the Cournot model*. Theoretically, this threshold X should be set to zero, so that only welfare-enhancing mergers (in expectations) would be approved. Empirically, however, all of the six approved mergers in the data are expected to increase prices by $0.21\% \sim 7.68\%$ *ceteris paribus*, which suggests $X > 7.68\%$ in our context of the global HDD industry.

Table 16: MLE of Dynamic Parameters under Price-based Policy (10% Threshold)

Specification	(1)	(2)
Merger policy:	Block if $\underline{N} \leq 3$ (Baseline)	Block if $\Delta P > 10\%$
ϕ_0	0.011	0.011
	$[0.001, 0.020]$	$[0.004, 0.016]$
$\kappa^i \quad (\omega_{it} = 1, 2, 3)$	0.48	0.47
	$[0.26, 0.69]$	$[0.33, 0.63]$
$\kappa^i \quad (\omega_{it} = 4)$	0.85	0.85
	$[0.39, 1.42]$	$[0.51, 1.26]$
κ^m	1.27	1.24
	$[0.81, 1.86]$	$[1.00, 1.74]$
κ^e	0.17	0.17
	$[-]$	$[-]$
σ	0.55	0.55
	$[0.41, 0.80]$	$[0.44, 0.68]$
Log likelihood	-156.93	-156.68

Note: For comparison, column 1 shows the baseline estimates (same as column 1 of Table 5). The two specifications are the same except for the modeling of merger policy. The 95% confidence intervals are constructed from the likelihood-ratio tests.

We also reviewed retrospective studies of past mergers and learned "the mean price change

[59] Farrell and Shapiro's (1990) Proposition 1 clarifies a necessary and sufficient condition for prices to fall after two firms merge in the static Cournot model with constant marginal costs.

for the forty-two true mergers is an increase of 7.22 percent," according to a meta-survey of American cases by Kwoka (2015, p. 110). Moreover, he finds "the outcomes all lie between a price increase of 30 percent and a decrease of 10 percent" except for a few outliers (p. 95), which suggests a de-facto threshold policy with $X = 30$. This policy seems extremely lenient as a general guideline and might reflect occasional forecasting errors by the agencies, because the measured price effects exceeding 5% or 10% are "generally interpreted, and viewed in the Merger Guidelines, as indicating significant competitive problems with a merger" (p. 95).

Consequently, we set $X = 10$ in our alternative model and re-estimate the key parameters. Table 16 compares our baseline model (column 1) with the alternative model (column 2). The parameter estimates are almost identical. The only visible differences are that $\hat{\kappa}^i$ (when $\omega_{it} = 1, 2, 3$) and $\hat{\kappa}^m$ are slightly lower in the new model, which suggests innovation and mergers are slightly less attractive under the new specification. But these differences are negligible and well within the 95% confidence intervals.

Table 17: Welfare Performance of Price-based Counterfactual Policies

Threshold price increase (%)	30	20	10 (Baseline)	5	1	0
(A) All periods: 1996–2025						
Consumer surplus	765.33	766.41	767.72	772.59	783.13	786.48
	(−0.31%)	(−0.17%)	(±0%)	(+0.64%)	(+2.01%)	(+2.44%)
Producer surplus	18.56	18.31	18.08	17.64	16.48	16.24
	(+2.63%)	(+1.27%)	(±0%)	(−2.47%)	(−8.89%)	(−10.20%)
Social Welfare	783.89	784.73	785.80	790.23	799.60	802.72
	(−0.24%)	(−0.14%)	(±0%)	(+0.56%)	(+1.76%)	(+2.15%)
(B) First half: 1996–2010						
Consumer surplus	536.23	536.38	536.68	538.71	542.91	544.29
	(−0.08%)	(−0.06%)	(±0%)	(+0.38%)	(+1.16%)	(+1.42%)
Producer surplus	13.18	13.15	13.09	12.88	12.24	12.13
	(+0.70%)	(+0.46%)	(±0%)	(−1.61%)	(−6.48%)	(−7.34%)
Social welfare	549.41	549.53	549.77	551.58	555.14	556.42
	(−0.07%)	(−0.04%)	(±0%)	(+0.33%)	(+0.98%)	(+1.21%)
(C) Second half: 2011–2025						
Consumer surplus	229.10	230.03	231.04	233.88	240.22	242.19
	(−0.84%)	(−0.43%)	(±0%)	(+1.23%)	(+3.98%)	(+4.83%)
Producer surplus	5.38	5.17	5.00	4.76	4.24	4.11
	(+7.70%)	(+3.39%)	(±0%)	(−4.74%)	(−15.18%)	(−17.69%)
Social welfare	234.48	235.20	236.03	238.65	244.46	246.30
	(−0.66%)	(−0.35%)	(±0%)	(+1.11%)	(+3.57%)	(+4.35%)

Note: All welfare numbers are present values as of January 1996 in billion US dollars and are the averages of 10,000 simulations under each policy regime. Their percentage changes from the baseline outcomes under the 10% threshold policy are in parentheses. Note the welfare numbers in Panel (C) are considerably smaller than those in Panels (A) and (B), because the annual discount factor of $\beta = 0.9$ means they are discounted at $\beta^{15} = 0.2059$.

Table 17 shows the results of counterfactual simulations with $X \in \{0, 1, 5, 20, 30\}$, where higher X means more lenient merger enforcement. The overall patterns look similar to the results in Table 6, with higher CS (and lower PS) under more stringent enforcement. The dollar values are also similar, although Table 17 tends to feature slightly higher numbers than Table 6. In terms of percentage change, the biggest difference seems to occur between $X = 5$ and $X = 1$ in Table 17, whereas the largest change in Table 6 is between $\underline{N} = 1$ and $\underline{N} = 2$.

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] Diermeier, Daniel, HülyaEraslan, and Antonio Merlo. (2002) "Coalition governments and comparative constitutional design," *European Economic Review*, 46 (4–5): 893–907.

[16] Doraszelski, Ulrich, and Mark Satterthwaite. (2010) "Computable Markov-perfect industry dynamics." *RAND Journal of Economics*, 41 (2): 215–243.

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

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

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

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

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

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

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

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

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

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

[27] ——— (1999) "A Dynamic Model of Endogenous Horizontal Mergers," *RAND Journal of Economics*, Vol. 30, pp. 56-83.

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

[29] Igami, Mitsuru (2017) "Estimating the Innovator's Dilemma: Structural Analysis of Creative Destruction in the Hard Disk Drive Industry, 1981–1998," *Journal of Political Economy*, 125 (3): 798–847.

[30] ——— (2018) "Industry Dynamics of Offshoring: The Case of Hard Disk Drives," *American Economic Journal: Microeconomics*, 10 (1): 67–101.

[31] ——— and Takuo Sugaya (2018) "Measuring the Incentive to Collude: The Vitamin Cartels, 1990–1999," manuscript, Yale University and Stanford University.

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

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

[34] ——— (2016) "Recursive Lexicographical Search: Finding All Markov Perfect Equilibria of Finite State Directional Dynamic Games." *Review of Economic Studies*, 83 (2): 658-703.

[35] 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.

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

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

[38] 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.

[39] Kwoka, John (2015) *Mergers, Merger Control, and Remedies: A Retrospective Analysis of U.S. Policy*. MIT Press.

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

[41] Marshall, Guillermo, and Álvaro Parra (2018) "Mergers in Innovative Industries," mimeo.

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

[43] Merlo, Antonio, and Xun Tang. (2012) "Identification and Estimation of Stochastic Bargaining Model." *Econometrica*, 80 (4): 1563–1604.

[44] Mermelstein, Ben, Volker Nocke, Mark A. Satterthwaite, and Michael D. Whinston (2018) "Internal Versus External Growth in Industries with Scale Economies: A Computational Model of Optimal Merger Policy," *Journal of Political Economy*, forthcoming.

[45] Miller, Nathan H., and Matthew C. Weinberg (2017) "Understanding the Price Effects of the MillerCoors Joint Venture," *Econometrica*, 85 (6): 1763–1791.

[46] 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.

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

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

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

[50] ———, 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.

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

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

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

[54] Reinganum, Jennifer F. (1983) "Uncertain Innovation and the Persistence of Monopoly." *American Economic Review*, 73 (4): 741–748.

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

[56] Rubinstein, Ariel. (1982) "Perfect Equilibrium in a Bargaining Model." *Econometrica*, 50 (1): 97–109.

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

[58] 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.

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

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

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

[62] 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.

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

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

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

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

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

[68] 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.

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

[70] Wolpin, Kenneth I. (1984) "An Estimable Dynamic Stochastic Model of Fertility and Child Mortality." *Journal of Political Economy*, 92 (5): 852–874.

[71] ——— (1987) "Estimating a Structural Search Model: The Transition from School to Work." *Econometrica*, 55 (4): 801–817.